These notes are for the half of the Electronics course given by Dr Lesurf. There are just eight lectures in this half of the course, but there are also a set of laboratories. These notes should be combined with the labwork booklet to cover the analog portion of the course. There is also additional information, colour versions of the diagrams, etc, on the "Scots Guide to Electronics" website. To view this, direct a web-browser to the URL

http://www.st-and.ac.uk/~www_pa/Scots_Guide/intro/electron.htm

and then click the link for the "analog and audio" section of the guide. Items in other sections of the Scots Guide may also be of interest. The examples given here and on the website tend to focus on applications like audio (hi-fi) as I think this helps to make the topic more understandable and interesting. Hence the website contains extra information on audio-related topics which you can understand on the basis of the explanations provided in this course.

The lectures in this booklet are:

1) Analog and amplifiers	pg 2
2) Limits and class prejudice	pg 10
3) Frequencies and filters	pg 18
4) Feedback	pg 27
5) Power supplies	pg 35
6) Wires and cables	pg 44
7) Transmission and loss	pg 51
8) Op-Amps and their uses	pg 58
Tutorial Questions	pg 66

Please read through the notes for each lecture before the lecture itself is given. Then see of you can answer the relevant tutorial questions after you have attended the lecture. The notes should mean you can then concentrate upon the explanations given in the lecture. If you are unclear as to any point you can ask about this during the lectures, or during one of the whole-class tutorials.

J. C. G. Lesurf
3rd Nov 2010

Lecture 1 – Analog and Amplifiers

This course is intended to cover a range of electronic techniques which are likely to be of some use or interest to physical scientists and engineers. The lectures covered in these notes focus on the 'Analog' side of things. In general, electronic systems tend to be made up by combining a range of simpler elements, circuits, and components of various kinds. We will examine a number of these in detail in later lectures, but for now we'll start with a simple over-view of amplifiers and their uses.

1.1 The uses of amplifiers.

At the most basic level, a signal amplifier does exactly what you expect – it makes a signal bigger! However the way in which this is done does vary with the design of the actual amplifier, the type of signal, and the reason why we're wanting to enlarge the signal. We can illustrate this by considering the common example of a 'Hi-Fi' audio system.

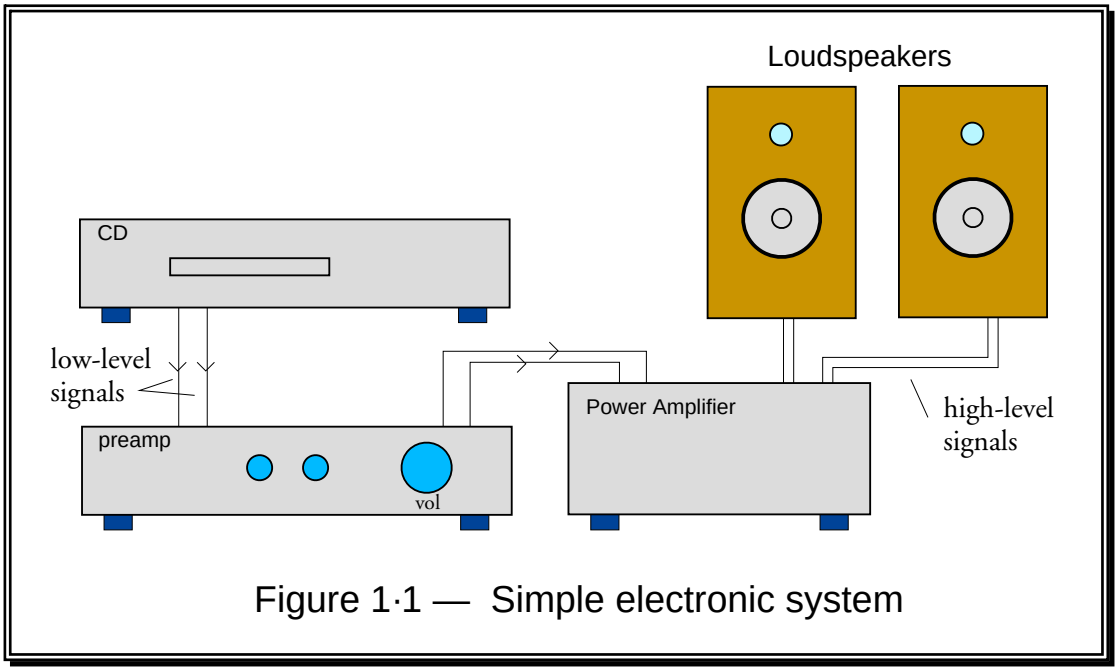


Figure 1·1 — Simple electronic system

In a typical modern hi-fi system, the signals will come from a unit like a CD Player, FM Tuner, or a Tape/MiniDisc unit. The signals these produce have typical levels of the order of 100mV or so when the music is moderately loud. This is a reasonably large voltage, easy to detect with something like an oscilloscope or a voltmeter. However the actual power levels of these signals is quite modest. Typically, these sources can only provide currents of a few milliamps, which by $P = V I$ means powers of just a few milliwatts. A typical loudspeaker will require between a few Watts and perhaps over 100 Watts to produce loud sounds. Hence we will require some form of *Power Amplifier* to 'boost' the signal power level from the source and make it big enough to play the music.

We usually want to be able to control the actual volume – e.g. turn it up to annoy neighbours, or

down to chat to someone. This means we have to have some way of adjusting the overall *Gain* of the system. We also want other functions – the most obvious being to select which signal source we wish to listen to. Sometimes all these functions are built into the same piece of equipment, however is often better to use a separate box which acts as a *Pre-Amp* to select inputs, adjust the volume, etc. In fact, even when built into the same box, most amplifier systems use a series of 'stages' which share the task of boosting and controlling the signals.

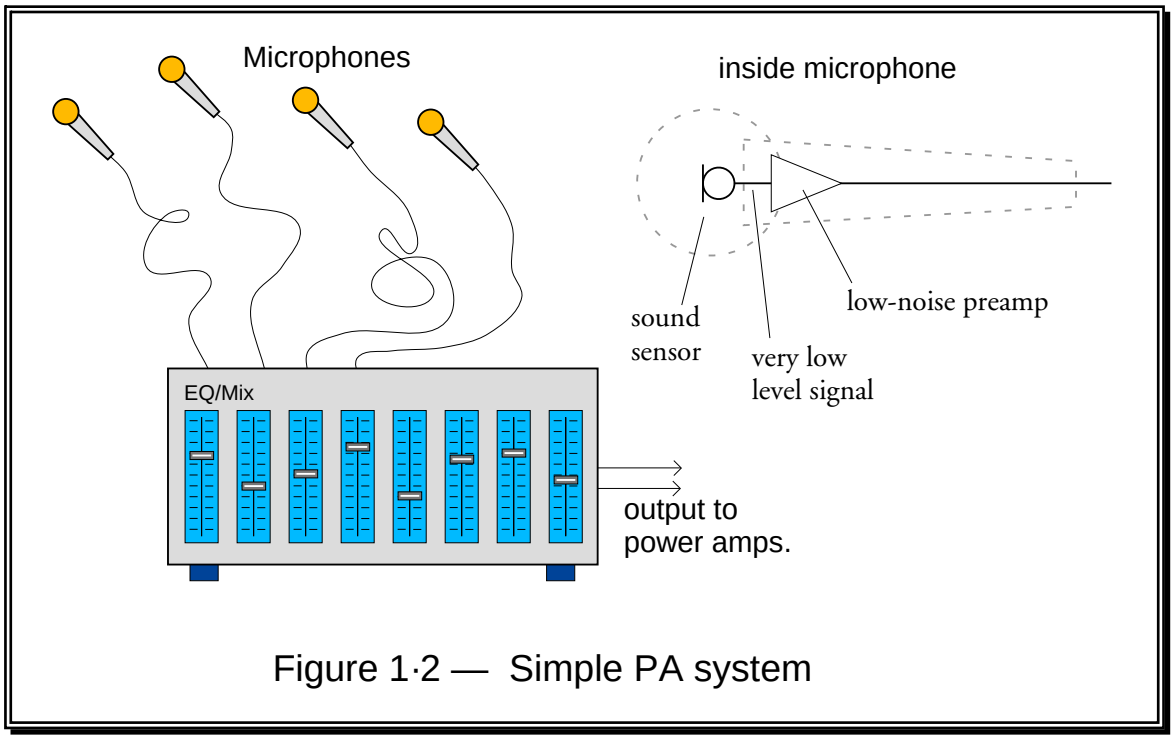


Figure 1·2 — Simple PA system

The system shown in figure 1.2 is similar to the previous arrangement, but in this case is used for taking signals from microphones and 'mixing' them together to produce a combined output for power amplification. This system also include an 'EQ' unit (Equalising) which is used to adjust and control the frequency response.

Microphones tend to produce very small signal levels. typically of the order of 1 mVrms or less, with currents of a few tens of microamps or less – i.e. signal powers of microwatts to nanowatts. This is similar to many other forms of sensor employed by scientists and engineers to measure various quantities. It is usual for sensors to produce low signal power levels. So low, in fact, that detection and measurement can be difficult due to the presence of *Noise* which arises in all electronic system. For that reason it is a good idea to amplify these weak, sensor created, signals, as soon as possible to overcome noise problems and give the signal enough energy to be able to travel along the cables from the source to its destination. So, as shown in figure 1.2, in most studios, the microphone sensor will actually include (or be followed immediately) by a small Low-Noise Amplifier (LNA).

Due to the range of tasks, amplifiers tend to be divided into various types and classes. Frequently a system will contain a combination of these to achieve the overall result. Most practical systems make use of combinations of standard 'building block' elements which we can think of as being the 'words' of the language of electronics. Here we can summarise the main types and their use. Most types come in a variety of forms so here we'll just look at a few simple examples.

1.2 The Voltage Amplifier

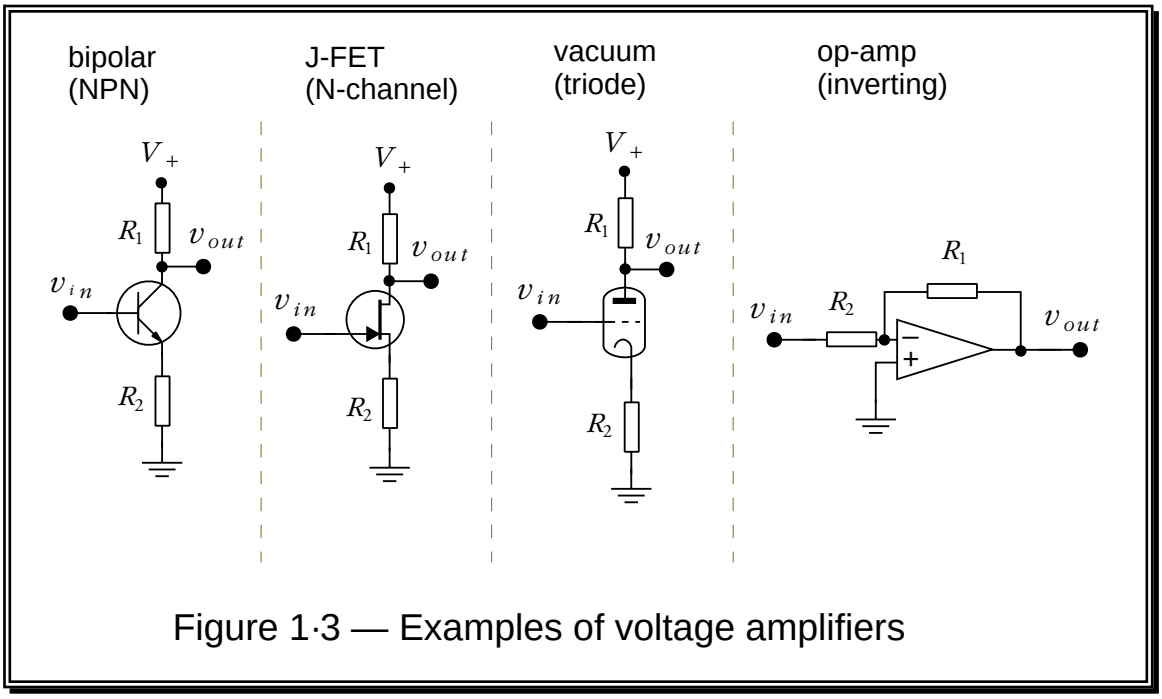


Figure 1·3 — Examples of voltage amplifiers

Figure 1·3 shows four examples of simple voltage amplifier stages using various types of device. In each case the a.c. voltage gain will usually be approximated by

$$A_v \approx -\frac{R_1}{R_2} \qquad \ldots (1.1)$$

provided that the actual device has an inherent gain large enough to be controlled by the resistor values chosen. Note the negative sign in expression 1.1 which indicates that the examples all invert the signal pattern when amplifying. In practice, gains of the order of up to hundred are possible from simple circuits like this, although for reasons we will discuss in later lectures (on feedback and distortion) it is usually a good idea to keep the voltage gain below this. Note also that vacuum state devices tend to be called "valves" in the UK and "tubes" in the USA.

Many practical amplifiers chain together a series of voltage amplifier *stages* to obtain a high overall voltage gain. For example a PA system might start with voltages of the order on 0·1μV from microphones, and boost this to perhaps 10 – 100 V to drive loudspeakers. This requires an overall voltage gain of × 10^9, so a number of voltage gain stages will be required.

1.2 Buffers and Current Amplifiers

In many cases we wish to amplify the signal current level as well as the voltage. The example we can consider here is the signal required to drive the loudspeakers in a Hi-Fi system. These will tend to have a typical input impedance of the order of 8 Ohms. So to drive, say, 100 watts into such a loudspeaker load we have to simultaneously provide a voltage of 28 V_{rms} and 3·5 A_{rms}. Taking the example of a microphone as an initial source again a typical source impedance will be around 100 Ohms. Hence the microphone will provide just 1 nA when producing 0·1μV. This means that to take this and drive 100 W into a loudspeaker the amplifier system must amplify the signal current by a factor of over ×10^9 at the same time as boosting the voltage by a similar amount. This means that the overall power gain required is ×10^{18} – i.e. 180 dB!

This high overall power gain is one reason it is common to spread the amplifying function into separately boxed pre- and power-amplifiers. The signal levels inside power amplifiers are so much larger than these weak inputs that even the slightest 'leakage' from the output back to the input may cause problems. By putting the high-power (high current) and low power sections in different boxes we can help protect the input signals from harm.

In practice, many devices which require high currents and powers tend to work on the basis that it is the signal voltage which determines the level of response, and they then draw the current they need in order to work. For example, it is the convention with loudspeakers that the volume of the sound should be set by the voltage applied to the speaker. Despite this, most loudspeakers have an efficiency (the effectiveness with which electrical power is converted into acoustical power) which is highly frequency dependent. To a large extent this arises as a natural consequence of the physical properties of loudspeakers. We won't worry about the details here, but as a result a loudspeaker's input impedance usually varies in quite a complicated manner with the frequency. (Sometimes also with the input level.)

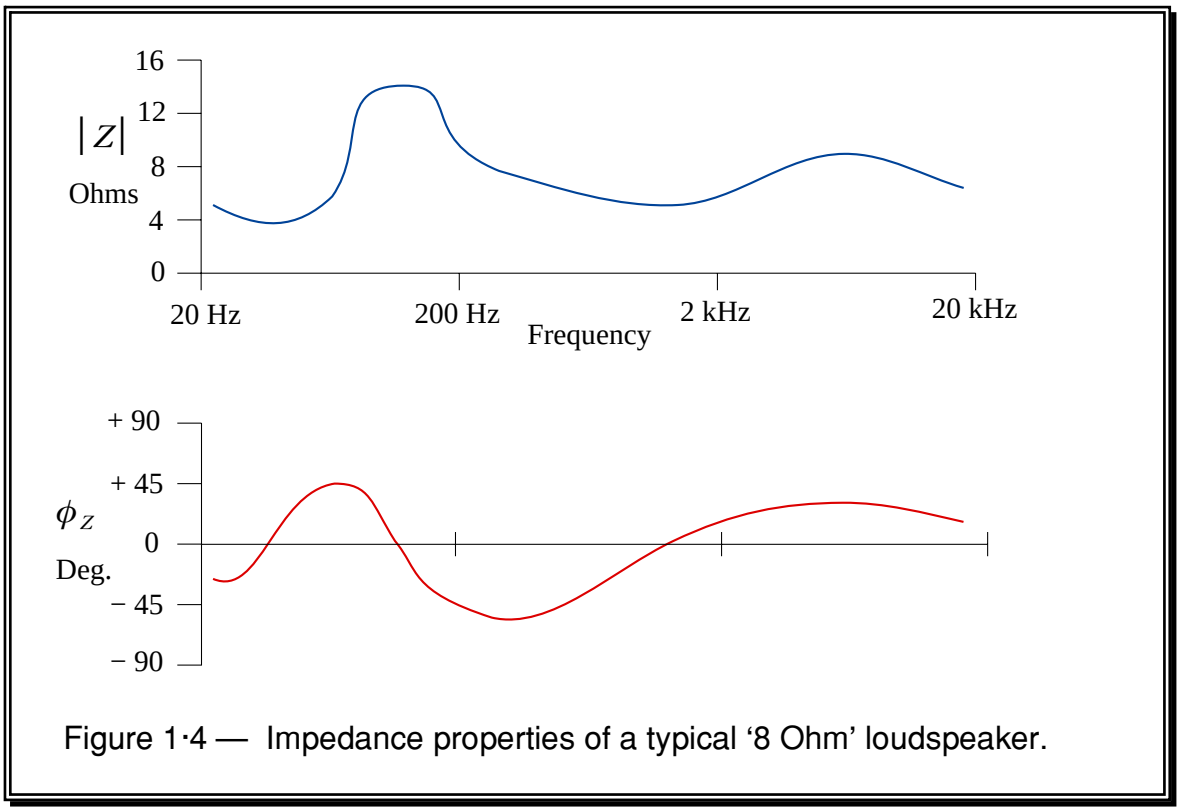


Figure 1·4 — Impedance properties of a typical '8 Ohm' loudspeaker.

Figure 1·4 shows a typical example. In this case the loudspeaker has an impedance of around 12 Ohms at 150 Hz and 5 Ohms at 1 kHz. So over twice the current will be required to play the same output level at 1 kHz than is required at 150 Hz. The power amplifier has no way to "know in advance" what kind of loudspeaker you will use so simply adopts the convention of asserting a voltage level to indicate the required signal level at each frequency in the signal and supplying whatever current the loudspeaker then requires.

This kind of behaviour is quite common in electronic systems. It means that, in information terms, the signal pattern is determined by the way the voltage varies with time, and – ideally – the current required is then drawn. Although the above is based on a high-power example, a similar

situation can arise when a sensor is able to generate a voltage in response to an input stimulus but can only supply a very limited current. In these situations we require either a *current amplifier* or a *buffer*. These devices are quite similar, in each case we are using some form of gain device and circuit to increase the signal current level. However a current amplifier always tries to multiply the current by a set amount. Hence is similar in action to a voltage amplifier which always tries to multiply the signal current by a set amount. The buffer differs from the current amplifier as it sets out to provide whatever current level is demanded from it in order to maintain the signal voltage it is told to *assert*. Hence it will have a higher current gain when connected to a more demanding load.

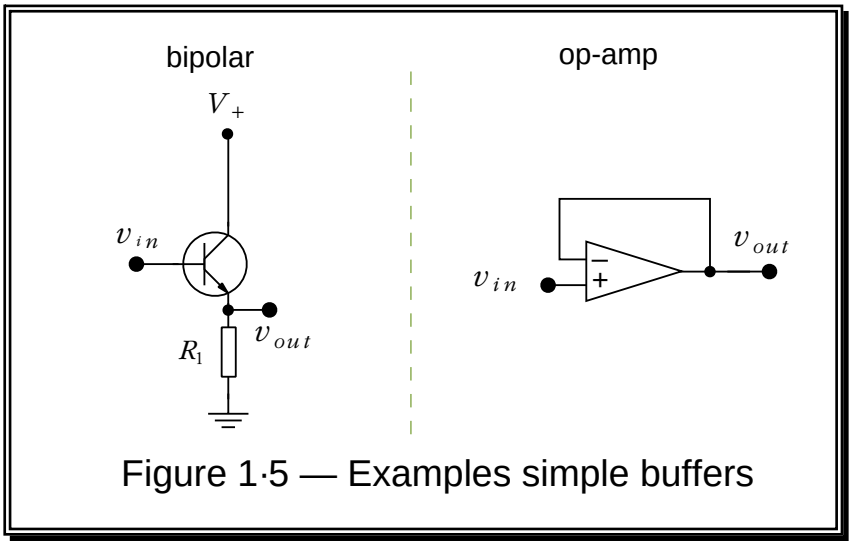


Figure 1·5 — Examples simple buffers

Figure 1·5 shows two simple examples of buffers. In each case the gain device (an NPN transistor or an op-amp in these examples) is used to lighten the current load on the initial signal source that supplies v_{in}. An alternative way to view the buffer is to see it as making the load impedance seem larger as it now only has to supply a small current to ensure that a given voltage is output.

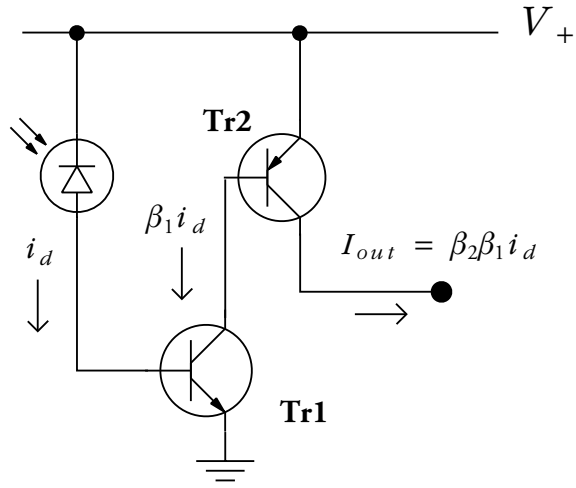


Figure 1·6 — photodiode and 2-stage current amplifier

Figure 1·6 shows an example of a 2-stage current amplifier in a typical application. The

photodiodes often used to detect light tend to work by absorbing photons and releasing free charge carriers. We can then apply a potential difference across the photodiode's junction and 'sweep out' these charges to obtain a current proportional to the number of photons per second striking the photodiode. Since we would require around 10^{16} electrons per second to obtain an amp of current, and the efficiency of the a typical photodiode is much less than 'one electron per photon' this means the output current from such a photodector is often quite small. By using a current amplifier we can boost this output to a more convenient level. In the example, two bipolar transistors are used, one an NPN-type, the other PNP-type, with current gains of β_1 and β_2. For typical small-signal transistors $\beta \approx 50 - 500$ so a pair of stages like this can be expected to amplify the current by between × 250 and × 250,000 depending upon the transistors chosen. In fact, if we wish we can turn this into a voltage by applying the resulting output current to a resistor. The result would be to make the circuit behave as a high-gain voltage amplifier.

1.3 The diff-amp

There are many situations where we want to amplify a small difference between two signal levels and ignore any 'common' level both inputs may share. This can be a achieved by some form of *Differential Amplifier*. Figure 1·7 shows an example of a differential amplifier stage that uses a pair of bipolar transistors.

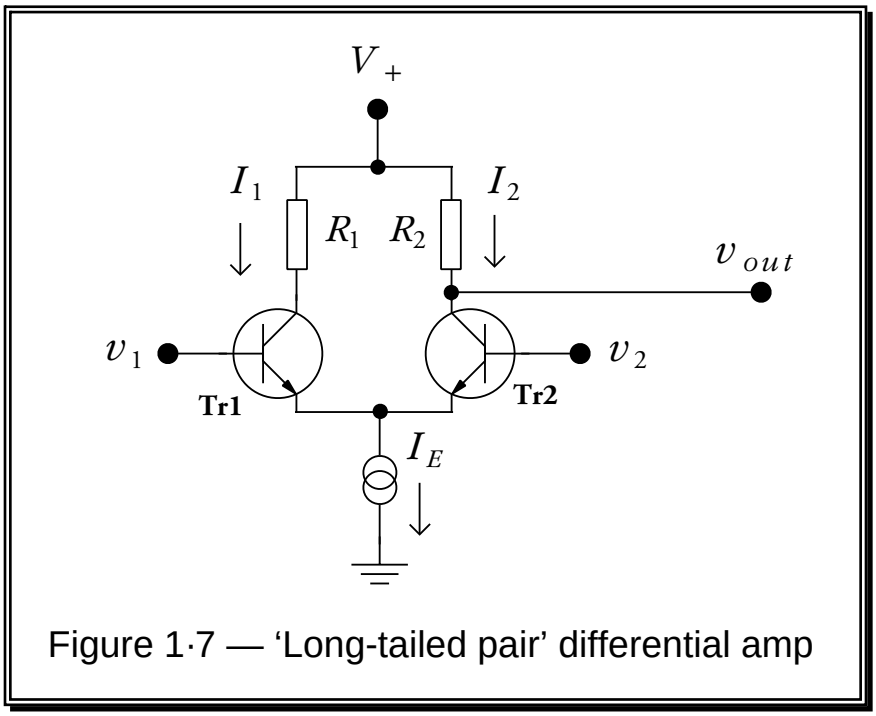


Figure 1·7 — 'Long-tailed pair' differential amp

This particular example uses a *Long-Tailed Pair* connected to a *Current Source*. We can understand how the circuit works simply from an argument based upon symmetry. For simplicity in this diagram we represent the Current Source by the standard symbol of a pair of linked circles. We will look at how a current source actually works later on. For now we can just assume it is an arrangement that insists upon always drawing a fixed current which in this case has the value I_E.

Assume that $R_1 = R_2$ and that the two transistors are identical. Assume that we start with the two input voltages also being identical. The circuit is arranged so that $I_E = I_1 + I_2$. By symmetry, when $v_1 = v_2$ it follows that $I_1 = I_2$, hence $I_1 = I_2 = I_E / 2$.

Now assume we increase, say, v_1 by a small amount. This will mean that the transistor, **Tr1**, will try to increase its current level and hence lift the voltage present at its emitter. However as it does this the base-emitter voltage of **Tr2** will fall. Since the current in a bipolar transistor depends upon its base-emitter voltage the result is that the current, I_1, rises, and I_2 falls. The sum of these currents, I_E, does not alter very much, but the balance between the two transistor currents/ voltages changes. The result is that the rise in v_1, keeping v_2 fixed, causes more current to flow through R_1 and less through R_2. The reduction in I_2 means the voltage drop across R_2 will reduce. hence the voltage at its lower end moves up towards V_+.

To evaluate this more precisely, we can assume that for each transistor in the pair the collector-emitter current is related to the base-emitter voltage via an expression

$$HI_{CE} = V_{BE} \qquad \ldots (1.2)$$

i.e. we can say that

$$HI_1 = (V_{B1} - V_E) \quad ; \quad HI_2 = (V_{B2} - V_E) \qquad \ldots (1.3 \,\&\, 1.4)$$

where V_E is the emitter voltage which the two transistors have in common and V_{B1} and V_{B2} are the base voltages. The value H represents the voltage-current gain of each transistor. The sum of these two currents always has a fixed value imposed by the current source, but any difference between them will be such that

$$H(I_1 - I_2) = v_1 - v_2 \qquad \ldots (1.5)$$

where V_E vanishes as it is common to both of the initial expressions.

Since $I_1 + I_2 = I_E$ we can say that the above is equivalent to saying that

$$H(I_E - 2I_2) = v_1 - v_2 \qquad \ldots (1.6)$$

When only concerned about a.c. signals we can ignore the constant I_E and say that this becomes

$$i_2 = -\frac{1}{2h_{ib}}(v_1 - v_2) \qquad \ldots (1.7)$$

where h_{ib} is one of the small-signal h-parameter values for a bipolar transistor, i_2 represents the change in the current in R_2 produced when we change the input voltage so that their imbalance from being equal is $(v_1 - v_2)$. Since this change in current appears across R_2, and the potential at the top of this is held fixed at V_+ it follows that the lower end of R_2 which we are using as the output will change in voltage by

$$v_{out} = \frac{R_2}{2h_{ib}}(v_1 - v_2) \qquad \ldots (1.8)$$

i.e. the stage has an effective voltage gain of $R_2 / (2h_{ib})$.

Differential amplifiers are particularly useful in three applications:

- When we have an input which has come from some distance and may have had some added interference. Using a pair of wires to send the signal we can then take the difference in potential between them as the signal and reject any 'common mode' voltages on both wires as being induced by interference.
- In feedback arrangements we can use the second input to control the behaviour of the amplifier.
- When we wish to combine two signals we can feed one into one transistor, and the second signal into the other.

Most *Operational Amplifier* integrated circuits have differential amplifier input stages and hence amplify the difference between two given input levels. Many use bipolar pairs of the kind shown in figure 1·7, but similar arrangements using Field Effect Transistors are also often used.

The Current Source

Having made use of one, we should finish with an explanation of the *Current Source*. Figure 1·8 shows a simple example.

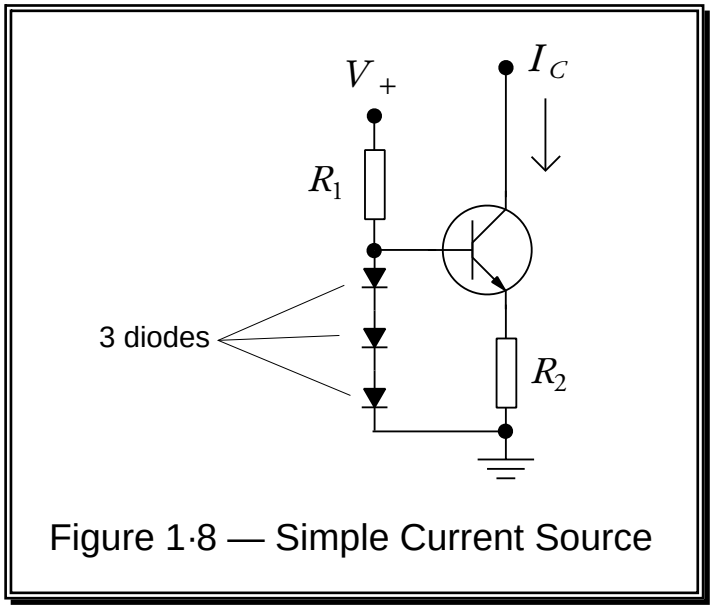


Figure 1·8 — Simple Current Source

A positive voltage is applied to the base of a bipolar transistor, but this base is also connected to a lower potential via a series of three diodes. The potential across such a diode when conducting current will tend to be around 0·6 V. Since there are three of these, the base of the transistor will sit at a potential about 1·8 V above the lower end of R_2. The base-emitter junction of a bipolar transistor is also a diode, hence it will also drop about 0·6 V when conducting. As a result, there will be 1·8 - 0·6 = 1·2 V across R_2. This means the current through this resistor will be $1{\cdot}2 / R_2$. since the transistor will probably have a current gain of over × 100 this means that more than 99% of this current will be drawn down from the transistor's collector.

Hence approximately, we can say that

$$I_c = \frac{1{\cdot}2 \text{ Volts}}{R_2} \qquad \ldots (1.9)$$

in practice the current will vary slightly with the chosen values of R_1 and V_+. However provided the current through the diodes is a few mA the above result is likely to be reasonably accurate. Provided that V_+ is much larger than 1·8 V the exact value does not matter a great deal. The circuit therefore tends to draw down much the same current whatever voltage appears at the transistor's collector – assuming that it, too, is more than a couple of volts. So the circuits acts to give a 'fixed' current and behaves as a steady current source. Although this example uses a bipolar transistor, as with many other forms of circuit we can make similar arrangements using other types of device.

Summary

You should now know the difference between some of the basic types of amplifier and the kinds of electronic 'building blocks' that can be used as amplifier stages. It should also be clear that the range of tasks and signal details/levels varies enormously over a wide range of applications. You should now also know how the basic amplifier arrangements such as Long-Tailed Pairs, etc, work.

Lecture 2 – Limits and Class Prejudice

Every real amplifier has some unavoidable limitations on its performance. The main ones that we tend to have to worry about when choosing and using them are:

- Limited bandwidth. In particular, for each amplifier there will be an upper frequency beyond which it finds it difficult/impossible to amplify signals.
- Noise. All electronic devices tend to add some random noise to the signals passing through them, hence degrading the SNR (signal to noise ratio). This, in turn, limits the accuracy of any measurement or communication.
- Limited output voltage, current, and power levels. This will mean that a given amplifier can't output signals above a particular level. So there is always a finite limit to the output signal size.
- Distortion. The actual signal pattern will be altered due non-linearities in the amplifier. This also reduces the accuracy of measurements and communications.
- Finite gain. A given amplifier may have a high gain, but this gain can't normally be infinite so may not be large enough for a given purpose. This is why we often use multiple amplifiers or stages to achieve a desired overall gain.

Let's start by looking at the limits to signal size.

2.1 Power Limitations

Figure 2·1 shows a simple amplifier being used to drive output signals into a resistive load.

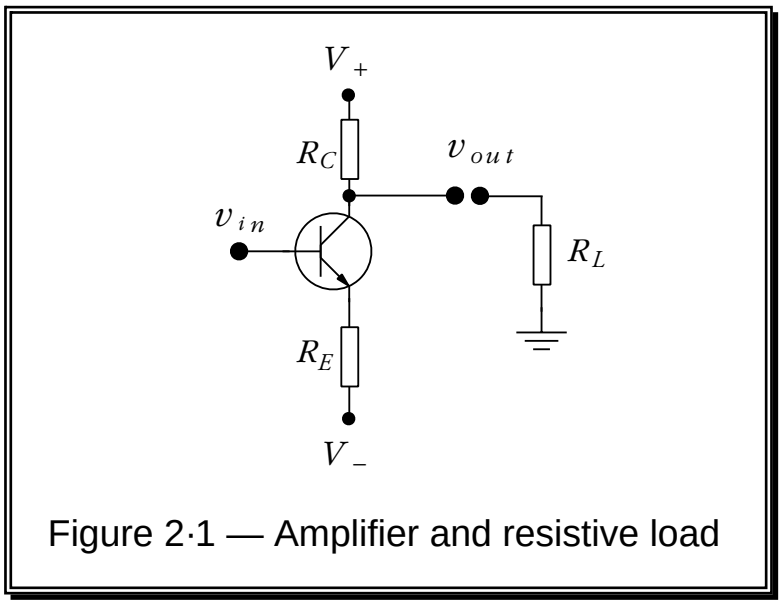


Figure 2·1 — Amplifier and resistive load

The amplifier is supplied via power lines at V_+ and V_- so we can immediately expect that this will limit the possible range of the output to $V_+ > v_{out} > V_-$ as the circuit has no way to

provide voltages outside this range. In fact, the output may well be rather more restricted than this as we can see using the following argument.

In effect, we can regard the transistor as a device which passes a current whose level is primarily determined by its base-emitter voltage, and hence by the input, v_{in}. However any current it draws from its collector must pass through R_C or R_L, and any current emerging through its emitter must then flow through R_E. In practice it really acts a sort of 'variable resistor' placed in between these other resistors. We can therefore represent it as being equivalent to the circuit shown in Figure 2·2a where the transistor is replaced by a variable resistance, *VR*.

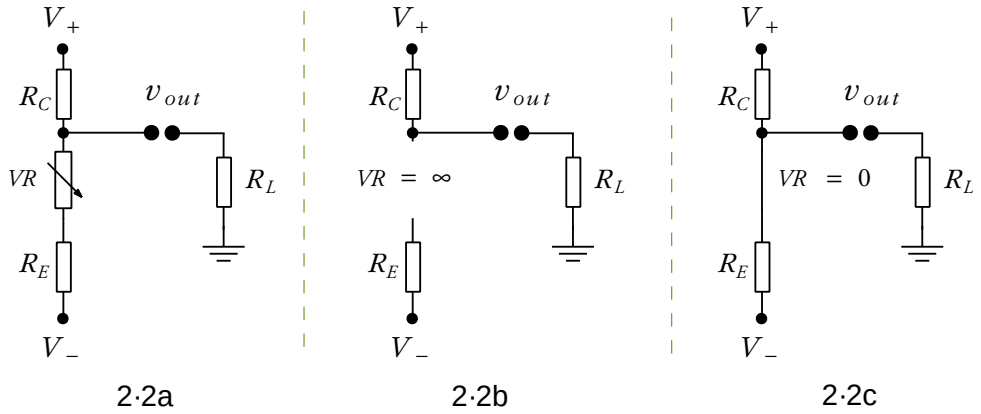


Figure 2·2 — Transistor as variable resistor

When we lower the input voltage, v_{in}, we reduce the transistor's willingness to conduct, which is equivalent to increasing its effective collector-emitter resistance. Taking this to the most extreme case we would reduce the current it passes to zero, which is equivalent to giving it an effective collector-emitter resistance of infinity. This situation is represented in figure 2·2b. Alternatively, when we increase v_{in} we encourage the transistor to pass more current from collector to emitter. This is equivalent to reducing its effective collector-emitter resistance. Again, taking this to the extreme we would reduce its resistance to zero as shown in 2·2c. We can now use the extreme cases shown in 2·2b and 2·2c to assess the maximum output the circuit can provide, and the implications for the design.

From 2·2b we can see that the highest (i.e. most positive) current and voltage we can provide into a load, R_L, will be

$$v_{out} = V_{max} = \frac{V_+ R_L}{R_L + R_C} \quad : \quad I_{max} = \frac{V_+}{R_L + R_C} \qquad \ldots (2.1 \,\&\, 2.2)$$

as the collector resistor acts a a Potential Divider with the load to limit the fraction of the available positive power line voltage we can apply to the output. Looking at expressions 2.1 and 2.2 we can see that we'd like $R_C \lll R_L$ to enable the maximum output to be able to be as big as possible – i.e. to approach V_+. For the sake of example, lets assume that we have therefore chosen a value for R_C equal to $R_L / 10$.

Now consider what happens when we lower the transistor's resistance (i.e. increase the current it will conduct) to the maximum case shown in 2·2c. We now find that the minimum (most negative) output voltage and current will be

$$v_{out} = V_{min} = \frac{R_L V_a}{R_a + R_L} \quad : \quad I_{min} = \frac{V_a}{R_a + R_L} \qquad \ldots (2.3 \& 2.4)$$

where we can define

$$V_a \equiv \frac{R_E V_+ + R_C V_-}{R_E + R_C} \quad : \quad R_a = \frac{R_C R_E}{R_E + R_C} \qquad \ldots (2.5 \& 2.6)$$

Either by considering the diagram or examining these equations we can say that in order for V_{min} to be able to approach V_- we require $R_E \lll R_C \,//\, R_L$ (where // represents the value of the parallel combination). Since we have already chosen to set $R_C \lll R_L$ this is equivalent to choosing a value of $R_E \lll R_C$. Again, we can assume for the sake of example that R_E is chosen to be $(R_C \,//\, R_L)/10$ which means it will also apprimate to $R_L/100$.

To consider the implications of the above lets take a practical example based on a power amplifier for use as part of an audio system. This means we can expect R_L to be of the order of 8 Ohms. For simplicity, lets assume $R_L = 10\Omega$. Following the assumptions above this means we would choose $R_C = 1\Omega$ and $R_E = 0{\cdot}1\Omega$. Again, for the sake of example, lets assume that the voltage rails available are ±25 V. This would mean that for an ideal amplifier we could expect to output peak levels of about 90% of ±25 Vwhich for a sinewave would correspond to about 30 Watts rms maximum power delivered into the 10Ω load.

Now consider what happens when the amplifier is on, but isn't trying to deliver a signal to the load – i.e. when $v_{out} = 0$. This means that there will be 25 Volts across R_C. Since this has a resistance of 1Ω it follows that $I_C = 25$ Amps when $v_{out} = 0$. i.e. the amplifier will demand 25 Amps from its positive power supply. Since none of this is being sent to the load this entire current will be passed through the transistor and its emitter resistor and away into the negative power rail supply. The result is that 25 × 50 = 1250 Watts will have to be dissipated in the combination of R_C, the transistor, and R_E in order to supply no actual output!

It would be fair to describe this kind of design as being "quite inefficient" in power terms. In effect it will tend to require and dissipate over 1 kW per channel if we wanted to use it as a 30 Watt power amplifier. This is likely to require large resistors, a big transistor, and lots of heatsinks. The electricity bill is also likely to be quite high and in the summer we may have to keep the windows open to avoid the room becoming uncomfortably warm. Now we could increase the resistor values and this would reduce the "standing current" or *Quiescent Current* the amplifier demands. However it would also reduce the maximum available output power, so this form of amplifier would never be very efficient in power terms. To avoid this problem we need to change the design. This leads us to the concept of amplifier *Classes*.

2.2 Class A

The circuit illustrated in Figure 2·1 and discussed in the previous section is an example of a *Class A* amplifier stage. Class A amplifiers have the general property that the output device(s) always carry a significant current level, and hence have a large *Quiescent Current*. The Quiescent Current is defined as the current level in the amplifier when it is producing an output of zero. Class A amplifiers vary the large Quiescent Current in order to generate a varying current in the load, hence they are always inefficient in power terms. In fact, the example shown in Figure 2·1 is particularly inefficient as it is *Single Ended*. If you look at Figure 2·1 again you will see that the amplifier only has direct control over the current between **one** of the two power rails and the load. The other rail is connected to the output load through a plain resistor, so its current isn't really under control.

We can make a more efficient amplifier by employing a *Double Ended* or *Push-Pull* arrangement. Figure 2·3 is an example of one type of output stage that works like this. You can see that this new arrangement employs a pair of transistors. One is an NPN bipolar transistor, the other is a PNP bipolar transistor. Ideally, these two transistors have equivalent properties – e.g. the same current gains, etc – except for the difference in the signs of their voltages and currents. (In effect, a PNP transistor is a sort of electronic "mirror image" of an NPN one.)

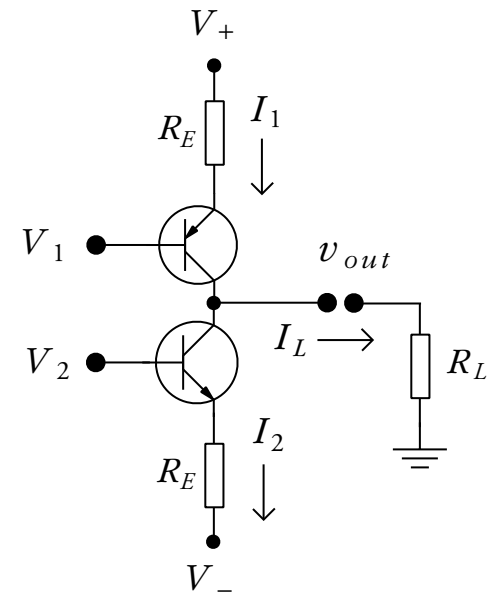


Figure 2·3 — Push-Pull Output

The circuit shown on the left now has two transistors which we can control using a pair of input voltages, V_1 and V_2. We can therefore alter the two currents, I_1 and I_2 independently if we wish.

In practice, the easiest way to use the circuit is to set the Quiescent Current, I_Q to half the maximum level we expect to require for the load. Then adjust the two transistor currents 'in opposition'.

It is the imbalance between the two transistor currents that will pass through the load so this means the transistors 'share' the burden of driving the output load.

When supplying a current, I_L to the load this means that we will have

$$I_1 = I_Q + \frac{I_L}{2} \quad : \quad I_2 = I_Q - \frac{I_L}{2} \qquad \ldots (2.7 \,\&\, 2.8)$$

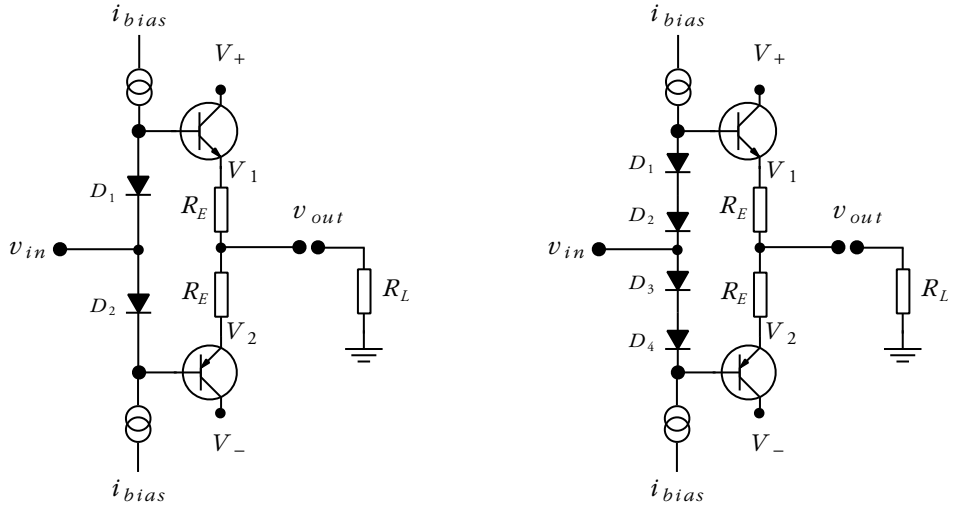
In linear operation this limits us to a current range $0 \rightarrow 2I_Q$ in each transistor, and load currents in the range $-2I_Q \leqslant I_L \leqslant 2I_Q$. Hence this arrangement is limited in a similar way to before. However the important difference to before becomes clear when we take the same case as was used for the previous example, with ±25 V power lines and a 10Ω load.

Since we wish to apply up to 25 V to a 10Ω load we require a peak output load current of 2·5 Amps. This means we can choose $I_Q = 1{\cdot}25$ A. As a result, each transistor now only has to dissipate 1·25 × 25 = 31 Watts – i.e. the output stage dissipates 62 Watts in total when providing zero output. Although this still isn't very efficient, it is rather better than the value of well above 1 kW required for our earlier, single ended, circuit!

2·3 Class B

Note that as is usual in electronics there are a number of arrangements which act as Class A amplifiers so the circuit shown in Figure 2·3 is far from being the only choice. In addition we can make other forms of amplifier simply by changing the Quiescent Current or *Bias Level* and operating the system slightly differently. The simplest alternative is the *Class B* arrangement. To illustrate how this works, consider the circuit shown in Figure 2·4a. Once again this shows a pair of devices. However their bases (inputs) are now linked by a pair of diodes.

The current in the diodes is mainly set by a couple of constant current stages which run a bias current, i_{bias}, through them. If we represent the forward voltage drop across each diode to be V_d we can say that the input to the base of the upper transistor sits at a voltage of $v_{in} + V_d$, and the input to the base of the lower transistor sits at $v_{in} - V_d$. Now since the base-emitter junction of a bipolar transistor is essentially a diode we can say that the drop between the base and emitter of the upper transistor will also be V_d. Similarly, we can say that the emitter of the lower (PNP) transistor will be V_d above its base voltage.



2·4a Class B 2·4b Class AB

Figure 2·4 — Class B and AB Power stages

This leads to the very interesting result where the emitter voltages in 2·4a will be

$$V_1 = V_2 = v_{in} \qquad \ldots (2.9)$$

This has two implications. Firstly, when $v_{in} = 0$ the output voltages will be zero. Since this means that the voltages above and below the emitter resistors (R_E) will both be zero it follows that there will be no current at all in the output transistors. The Quiescent Current level is zero and the power dissipated when there is no output is also zero! Perfect efficiency!

The second implication is that as we adjust v_{in} to produce a signal, the emitter voltages will both tend to follow it around. When we have a load connected this will then draw current from one transistor or the other – but not both. In effect, when we wish to produce an positive voltage the upper transistor conducts and provided just enough current to supply the load. None is wasted in the lower transistor. When we wish to produce a negative voltage the lower transistor conducts and draws the required current through the load with none coming from the upper transistor. Again this means the system is highly efficient in power terms.

This all seems to be too good to be true, there must be a snag... and yes, there are indeed some serious drawbacks. When power efficiency is the main requirement then Class B is very useful. However to work as advertised it requires the voltage drops across the diodes and the base-emitter junctions of the transistors to all be **exactly** the same. In practice this is impossible for various reasons. The main reasons are as follows:

- No two physical devices are ever **absolutely** identical. There are always small differences

between them.

• The diodes and the transistors used will have differently doped and manufactured junctions, designed for different purposes. Hence although their general behaviours have similarities, the detailed IV curves of the transistor base-emitter junctions won't be the same as the IV curves of the diodes.

• The current through the transistors is far higher than through the diodes, and comes mostly from the collector-base junctions which are intimately linked to the base-emitter. Hence the behaviour is altered by the way carries cross the device.

• The transistors will be hotter than the diodes due to the high powers they dissipate. Since a PN junction's IV curve is temperature sensitive this means the transistor junctions won't behave exactly like the diodes, and may change with time as the transistor temperatures change.

• When we change the applied voltage it takes time for a PN junction to 'react' and for the current to change. In particular, transistors tend to store a small amount of charge in their base region which affects their behaviour. As a result, they tend to 'lag behind' any swift changes. This effect becomes particularly pronounced when the current is low or when we try to quickly turn the transistor "off" and stop it conducting.

The overall result of the above effects is that the Class B arrangement tends to have difficulty whenever the signal waveform changes polarity and we have to switch off one transistor and turn on the other. The result is what is called *Crossover Distortion* and this can have a very bad effect on small level or high speed waveforms. This problem is enhanced due to *Non-Linearities* in the transistors – i.e. that the output current and voltage don't vary linearly with the input level – which are worst at low currents. (We will be looking at signal distortions in a later lecture.)

This places the amplifier designer/user in a dilemma. The Class A amplifier uses devices that always pass high currents, and small signals only modulate these by a modest amount, avoiding the above problems. Alas, Class A is very power inefficient. Class B is far more efficient, but can lead to signal distortions. The solution is to find a 'half-way house' that tries to take the good points of each and minimise the problems. The most common solution is *Class AB* amplification, and an example of this is shown in Figure 2·4b.

2·4 Class AB

The Class AB arrangement can be seen to be very similar to the Class B circuit. In the example shown it just has an extra pair of diodes. The change these make is, however, quite marked. We now find that – when there is no output – we have a potential difference of about $2 \times V_d$ between the emitters of the two transistors. As a consequence there will be a Quiescent Current of

$$I_Q \approx \frac{V_d}{R_E} \qquad \ldots (2.10)$$

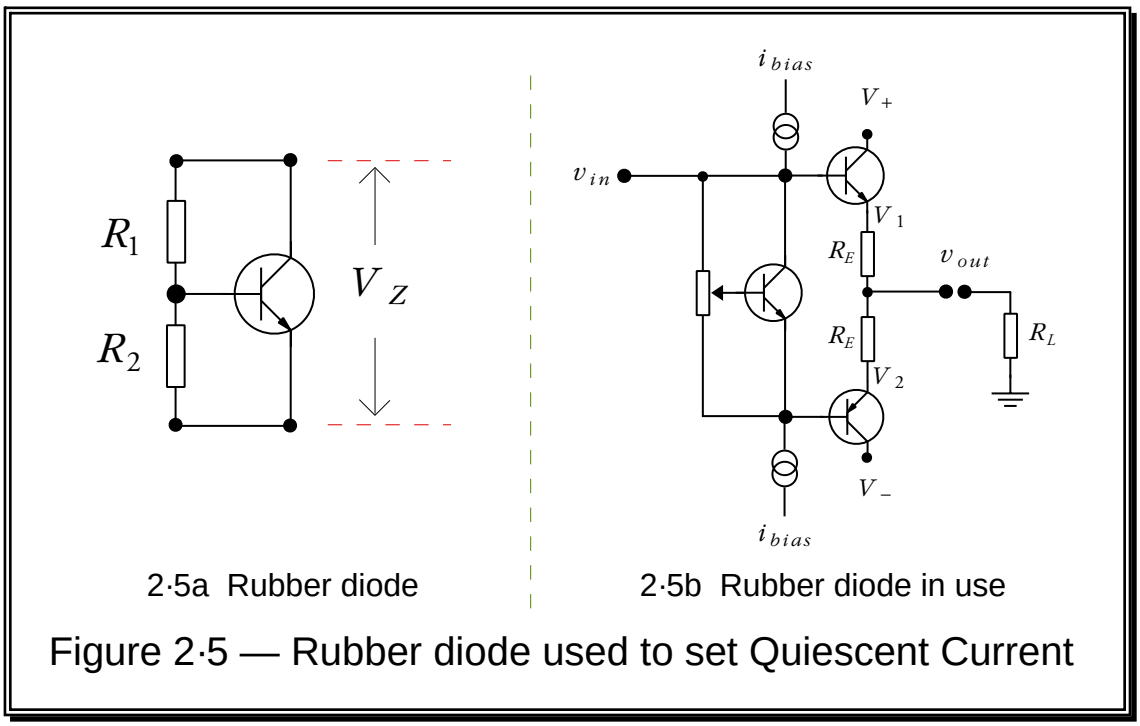
flowing through both transistors when the output is zero. For small output signals that require output currents in the range $-2I_Q < I_L < 2I_Q$ **both** transistors will conduct and act as a double ended Class A arrangement. For larger signals, one transistor will become non-conducting and the other will supply the current required by the load. Hence for large signals the circuit behaves like a Class B amplifier. This mixed behaviour has caused this approach to be called Class AB.

When driving sinewaves into a load, R_L, this means we have a quasi-Class A system up to an rms voltage and output power of

$$V'_{rms} \approx \frac{\sqrt{2}\, V_d R_L}{R_E} \quad : \quad P_L' \approx \frac{2 V_d^2 R_L}{R_E^2} \qquad \text{... (2.11 \& 2.12)}$$

If we take the same audio example as before ($R_L = 10\Omega$), assume a typical diode/base-emitter drop of 0·5V, and chose small value emitter resistors of $R_E = 0{\cdot}5\Omega$ this leads to a maximum 'Class A' power level of 20 Watts. The Quiescent Current will be around 1 Amp, so with ±25V power rails this system would still dissipate around 50 Watts in total. In this case the system is more like Class A than Class B. We do, however have some control over the circuit and can adjust the bias current level.

The obvious way to change the bias current is to choose a different value for R_E. Choosing a larger value will reduce I_Q and P_L' whilst still preserving a region of Class A behaviour which can be large enough to avoid the Crossover and imperfection problems mentioned earlier. However if we make R_E too large they will 'get in the way' as all the current destined for the load must pass though these emitter resistors. Most practical Class AB systems therefore use a different approach. The most common tends to be to replace the diodes with an arrangement often called a 'rubber diode' or 'rubber Zener'.



2·5a Rubber diode

2·5b Rubber diode in use

Figure 2·5 — Rubber diode used to set Quiescent Current

Figure 2·5a shows the basic 'rubber zener' arrangement. Looking at this we can see that it consists of a pair of resistors and a transistor. The resistors are connected together in series to make a potential divider which feeds a voltage to the base of the transistor. However the transistor can 'shunt' away any current applied and control the potential difference appearing across the pair of resistors. The arrangement therefore settles into a state controlled by the fact that the base-emitter junction of the transistor is like a diode and the transistor conducts when the base-emitter voltage, V_d, is approximately 0·5 V. Since V_d is also the voltage across R_2 we can expect that

$$V_d \approx \frac{V_Z R_2}{R_1 + R_2} \qquad \text{... (2.13)}$$

provided that the base current is small enough compared to the currents in the two resistors to be ignored. The above just comes from the normal action of a potential divider. However since in this case the transistor is involved it will act to adjust V_Z to keep V_d at about 0·5V. The result is therefore that we can set the voltage

$$V_Z \approx \frac{V_d\,(R_1 + R_2)}{R_2} \qquad \ldots (2.14)$$

This means we can choose whatever voltage value, V_Z, we require by selecting an appropriate pair of resistors. In a real power amplifier the pair of fixed resistors is often replaced by a potentiometer and the circuit in use looks as shown in Figure 2·5b. The Quiescent current can now be adjusted and set to whatever value gives the best result since we can now alter V_Z and hence the potential difference between the emitters of the output transistors. In practice the optimum value will depend upon the amplifier and its use. Values below 1mA are common in op-amp IC's, whereas high-power audio amps may have I_Q values up to 100mA or more.

Class AB arrangements are probably the most commonly employed system for power amplifier output sections, although 'Pure' Class A is often used for low current/voltage applications where the poor power efficiency isn't a problem. Although we won't consider there here, it is worth noting that there are actually a number of other arrangements and Classes of amplifier which are used. For example, Class 'C' and Class 'E' which tend to use clipped or pulsed versions of the signal.

In linear audio amplifiers the most well known alternative is the '*Current Dumping*' arrangement developed by QUAD (Acoustical Manufacturing) during the 1970's. This combines a small Class A amplifier and a pair of 'Dumping' devices. The Dumping devices just supply wadges of current and power when required, but carry no Quiescent Current so they are highly power efficient, but prone to distortion. The accompanying Class A amplifier is arranged to 'correct' the errors and distortions produced by the Dumping devices. In effect this means it just has to 'fill in the gaps' so only has to provide a relatively small current. Hence the Quiescent level required by the Class A 'fill in' amplifier can be small, thus avoiding a poor overall efficiency.

Summary

You should now be aware of of the basic building blocks that appear in Power Amplifiers. That *Class A* amplifiers employ a high *Quiescent Current* (sometimes called bias current or standing current) large enough that the transistor currents are large even when the output signal level is small. It should be clear that the power efficiency of Class A amplifiers is therefore poor, but they can offer good signal performance due to avoiding problems with effects due to low-current level nonlinearities causing *Distortion*. You should now also see why a *Double Ended* output is more efficient than a *Single Ended* design. You should also know that *Class B* has a very low (perhaps zero) Quiescent Current, and hence low standing power dissipation and optimum power efficiency. However it should be clear that in practice Class B may suffer from problems when handling low-level signals and hence *Class AB*, with its small but controlled Quiescent Current is often the preferred solution in practice. It should also be clear how we can employ something like a *Rubber Zener* to set the optimum Quiescent Current and power dissipation for a given task.

Lecture 3 – Frequencies and filters

One of the limitations which affect all real amplifiers is that they all have a finite *Signal Bandwidth*. This means that whatever amplifier we build or use, we find that there is always an upper limit to the range of sinewave frequencies it can amplify. This usually means that high frequencies signals won't be boosted, and may attenuated instead. It also often means that high frequency signals may be *Distorted* – i.e. their waveshape may be changed. Some amplifiers also have a low-frequency limitation as well, and are unable to amplify frequencies below some lower limit. We therefore need to be aware of, and be able to assess, these effects to decide if a specific amplifier is suitable for a given task.

In addition to the above, there are many situations where we deliberately want to alter the signal frequencies a system will pass on or amplify. There are many situations where deliberate filtering is needed. Some common examples are:

- Low Pass Filters used to limit the bandwidth of a signal before Analog to Digital Conversion (digital sampling) to obey the Sampling Theorem
- Bandpass filters to select a chosen 'narrowband' signal and reject unwanted noise or interference at other frequencies.
- Bandreject filters used to 'block' interference at specific frequencies. e.g. to remove 50/100 Hz 'hum' (or 60/120 Hz if you happen to be in the USA!)
- 'Tone Controls' or 'EQ' (Equalisation) controls. These tend to be used to alter the overall tonal balance of sound signals, or correct for other unwanted frequency-dependent effects.
- 'A.C.' amplifiers that don't amplify unwanted low frequencies or d.c. levels.

In this lecture we will focus on some examples of filters, but bear in mind that similar effects can often arise in amplifiers.

3.1 Ways to characterise filters

The most common and most obvious way to describe the effect of a filter is in terms of is *Frequency Response*. This defines the filter's behaviour in terms of its Gain as a function of frequency. figure 3·1 shows a typical example.

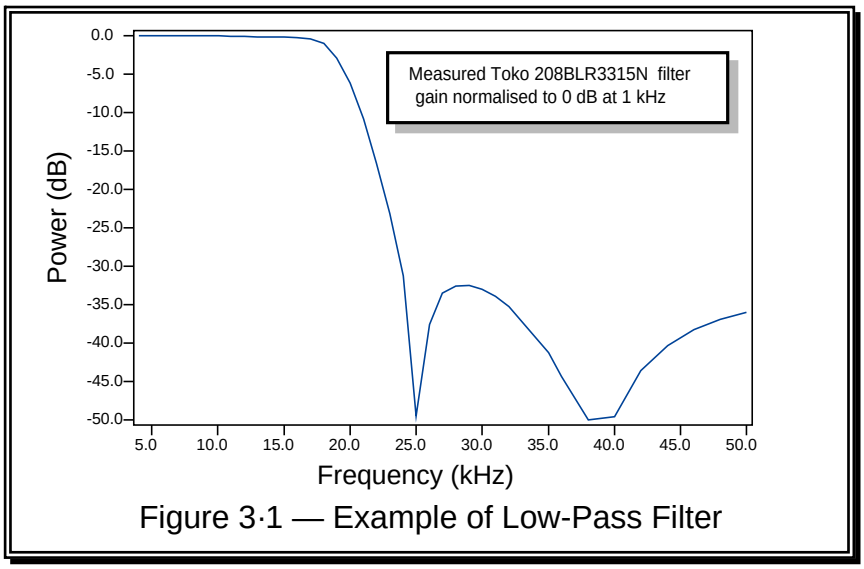


Figure 3·1 — Example of Low-Pass Filter

In this case Figure 3·1 shows the measured power gain (or loss!) as a function of frequency for a specific low-pass filter. This example is based upon a commercial audio filter sold under the 'Toko' brand name for use in hi-fi equipment. This filter is Passive – i.e. includes no gain devices, just passive components like resistors, capacitors, or inductors. As a result we need to counteract any unwanted losses in the filter. The overall circuit measured in this case is shown in figure 3·2.

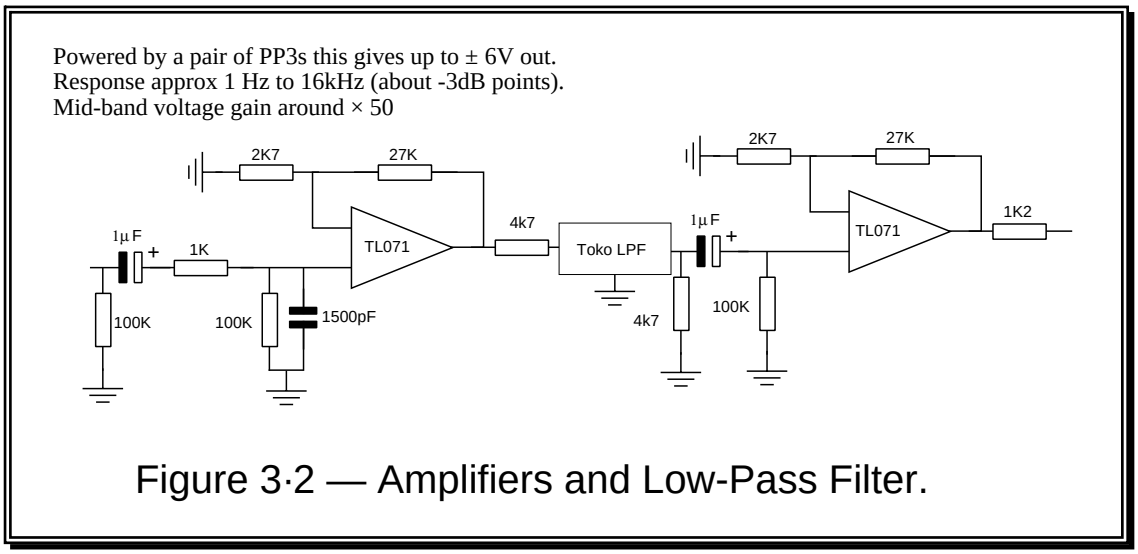


Figure 3·2 — Amplifiers and Low-Pass Filter.

Here the filter actually sits in between two amplifiers which raise the signal level. The filter's job in the system this example is taken from is to stop high frequencies from reaching the output which is to be sampled by an ADC. The overall voltage gain is around ×50 at 1 kHz. The Toko filter passes signals from d.c. up to around 18 kHz but attenuates higher frequencies. Note also the pair of 1µF capacitors in the signal path. These, taken in conjunction with the following 100kΩ resistors, act as *High Pass* filters to remove any d.c. as it this was unwanted in this case. The combination of the 1kΩ resistor and 1500pF capacitor at the input also act as an additional low-pass filter.

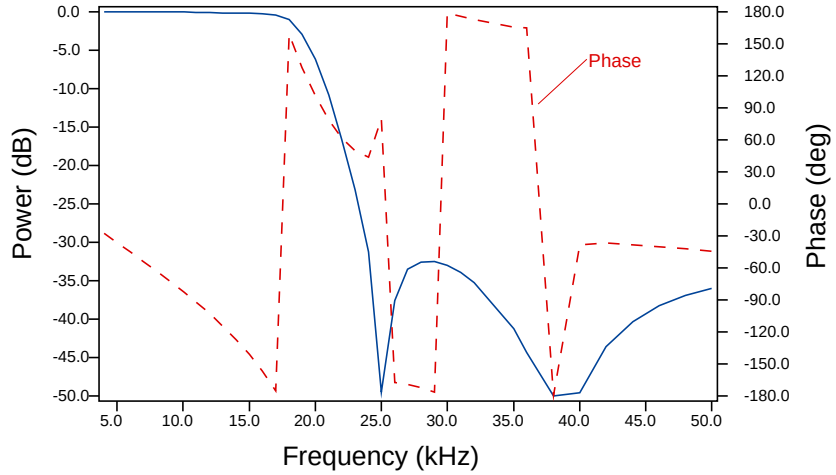


Figure 3·3 — As fig 3·1 but including phase plot

Although a frequency response of the kind shown in figure 3·1 is the most common way to display the behaviour of a filter, it doesn't tell the whole story. When sinewave components pass

through the filter they may also be delayed by an amount that depends upon their frequency. This effect is often represented in terms of a plot of the Phase Delay as a function of signal frequency. For the same filter as before the measured Phase Delay as a function of frequency is shown in figure 3·3.

In figure 3·3 the phase seems to 'jump around' in quite a complex way. However the underlying behaviour is simpler than it appears. There are two reasons for this. Firstly, the phase **measurements** are always in the range $-180° \rightarrow 180°$, but the actual signal delay may be more than a half-cycle. When making phase measurements we can't normally distinguish a delay of, say, 50 degrees, from 360 + 50 = 410 degrees, or any other multiple number of cycles plus 50 degrees. If we look at the phase plot shown in figure 3·3 the phase seems to 'jump' from −180° to about +160° at around 16 kHz. In fact, what has happened is that the real signal delay has increased from −180° to around −200°. However, since the measurement system only gives results in the limited range it assumes that this is the same as +160°. Despite the apparent hopping around, the **actual** signal phase delay tends to increase fairly steadily with frequency. The second reason for the complex plot is that at some points the signal level becomes very small. e.g. in figure 3·3 at around 25kHz. When we try to make a phase measurement on a tiny signal the result becomes unreliable due to measurement errors and noise. Hence the value at such a frequency can't be taken as being as reliable as at other frequencies.

Measured plots of phase delay versus frequency therefore have to be interpreted with some care. That said, the phase behaviour is important as we require it along with the amplitude behaviour to determine how complex signals may be altered by passing through the filter. In practice, though, it is often more convenient to measure or specify the filter's time-effects in terms of a time delay rather than a phase. We can say that a phase delay of $\phi\{f\}$ (radians) at a frequency f is equivalent to a time delay of

$$\Delta t = \frac{\phi\{f\}}{2\pi f} \qquad \ldots (3.1)$$

Figure 3·4 shows the time delay as a function of frequency for the same filter as used for the earlier figures.

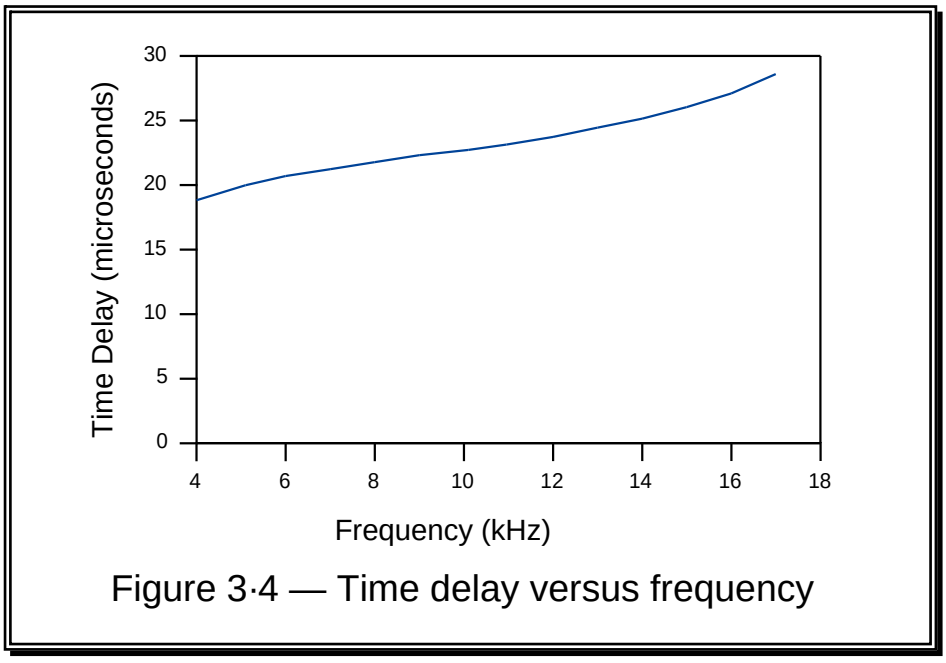


Figure 3·4 — Time delay versus frequency

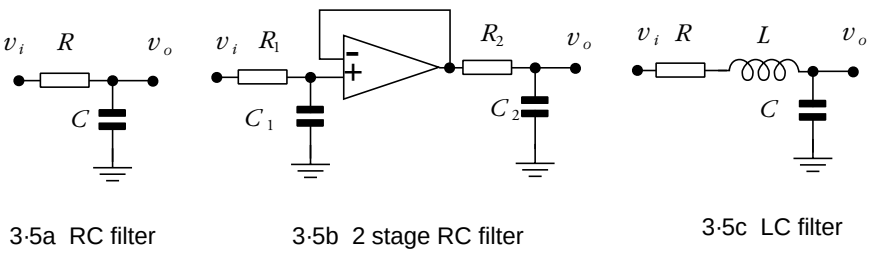
Looking at figure 3·4, we can see that when expressed in terms of a time, the delay seems rather more uniform and does not vary as quickly with frequency as does the phase. This behaviour is quite common in filters. In effect, the filter tends to delay signals in its *Passband* by a fairly uniform amount. However since a given time delay looks like a phase delay that increases linearly with frequency, this behaviour isn't obvious from phase/frequency plots. The term *Group Delay* is used to refer to the average time delay imposed over the range of frequencies the filter is designed to pass through. In this case we can see that for our example filter the Group Delay value is around 22 microseconds. The term comes from considering a 'group' of signal frequencies.

In fact, we can always expect signals to take a finite time to propagate through **any** circuit or element in a system. Hence we can also quote Group Delay values for amplifiers, transistors, or even cables. Unless the delay is so long as to make us wait an annoyingly long time for a signal, a uniform Group Delay is normally fine as it just means we have to allow a short time for the signal to arrive but its waveform shape is unaffected. A non-uniform delay will mean that – even if the amplitudes of frequency components remain relatively unchanged – the signal waveform will be distorted. Hence in most cases we try to obtain filters and other elements that have a uniform delay. That said, there are situations where we require a *Dispersive* element – i.e. one which delays differing frequencies by differing times. Some forms of filter are actually designed to be *All Pass*. i.e. they don't set out to alter the relative amplitudes of the frequency components in the signals, but are designed to alter the relative times/phases of the components.

Although we won't consider it in detail here, it should be clear from the above that we can model the behaviour of filters just as well by using their *Temporal Response* to waveforms like short pulses or 'steps' (abrupt changes in level) as by using sinewave methods. In fact, the two ways of understanding the signal processing behaviour are linked via Fourier Transformation. We can therefore characterise amplifiers, filters, etc, in either the *Time Domain* or the *Frequency Domain as we wish*. When dealing with continuous waveforms the frequency domain is usually more convenient, but when dealing with pulses the time domain is more useful.

3·2 Order order!

Having established some of the general properties of filters, lets now look at some specific types and applications. Broadly speaking we can separate filters into *Active* versus *Passive*, As the names imply, an Active filter makes use of some gain device(s) as an integral part of the operation of the filter. Passive filters don't actually require a gain device to work, but are usually accompanied by some form of amplifiers or buffers for convenience. Filters are also often referred to as 'First Order', 'Second Order', etc. This refers to the number of components (capacitors and inductors, not resistors or transistors) that affect the 'steepness' or 'shape' of the filter's frequency response. To see what this means we can use a few examples.



3·5a RC filter 3·5b 2 stage RC filter 3·5c LC filter

Figure 3·5 — Some Low-Pass filters

Figure 3·5 shows three simple types of low-pass filter. In each case we can use the standard methods of a.c. circuit analysis to work out the circuit's voltage gain/loss, $A_v \equiv v_o / v_i$, as a function of frequency and obtain the following results

$$\text{for 3·5a} \qquad A_v = \frac{1}{1 + j\omega RC} \qquad \ldots (3.2)$$

$$\text{for 3·5b} \qquad A_v = \left[\frac{1}{1 + j\omega R_1 C_1}\right] \times \left[\frac{1}{1 + j\omega R_2 C_2}\right] \qquad \ldots (3.3)$$

$$\text{for 3·5c} \qquad A_v = \frac{1}{1 + j\omega RC - \omega^2 LC} \qquad \ldots (3.4)$$

In practice the above results would be modified by the impedances connected to the input and output of each filter, but the results shown are good enough our purposes here. We can in fact now re-write all of the above expressions in the general form

$$A_v = \frac{1}{\sum_{i=0}^{N} a_i \omega^i} \qquad \ldots (3.5)$$

i.e. in each case the frequency response is given by an expression in the form of the inverse of a polynomial of some *order*, N. The relevant values for the three examples are as follows

Filter	order N	a_0	a_1	a_2
3·5a (simple RC)	1	1	jRC	-
3·5b (two-stage RC)	2	1	$j(R_1C_1 + R_2C_2)$	$-R_1C_1R_2C_2$
3·5c (LC)	2	1	jRC	$-LC$

So 3·5a shows a 'first order' low-pass filter, but 3·5b and 3·5c are 'second order'. In general, we can expect a filter of order N to cause the value of A_v to fall as ω^N when we are well into the frequency region where the filter is attenuating any signals (i.e. when ω is large for the low-pass examples). We can apply a similar general rule to other types of filter – high pass, etc – although very complex filters may require a more general expression of the form

$$A_v = \frac{\sum_{j=0}^{M} b_j \omega^j}{\sum_{i=0}^{N} a_i \omega^i} \qquad \ldots (3.6)$$

3.3 Active filters, Normalisation, and Scaling

The above examples are all 'passive' filters. OK, the circuit shown in Figure 3·5b does use an amplifier (probably an IC op-amp) as a buffer between the two RC stages, but this is just present to ensure that the two stages don't *interact* in an unwanted manner. Active filters generally employ some form of *feedback* to alter or control the filter's behaviour. This makes the designs more flexible in terms of allowing for a wider choice of shapes of frequency response, time-domain behaviour, etc. In theory we can make passive equivalents to most active filter so it may seem as the need for an amplifier is an unnecessary extra complexity. However the active filer offers some useful advantages. In particular, they help us avoid the need for 'awkward' components. For example, it is often good practice to avoid using inductors as these can pick up unwanted interference due to stray magnetic fields. Also, some passive filters may require large-value capacitors or inductors, which would be bulky and expensive. By using an active equivalent we can often 'scale' all the component values in order to be able to use smaller and cheaper components.

Most of the analysis and design of active filters is done in terms of *Normalised* circuits. To understand what this means, look again at the circuit shown in figure 3·5a. The way in which the filter's voltage gain, A_v, varies with frequency is set by the value of RC. The important thing to note is that the behaviour isn't determined by either R or C taken independently. So, for example, if we were to make R ten times bigger, and C ten times smaller, the behaviour would be unchanged. The RC value sets a 'turn-over frequency' we can define as

$$\omega_0 \equiv \frac{1}{RC} \qquad \ldots (3.7)$$

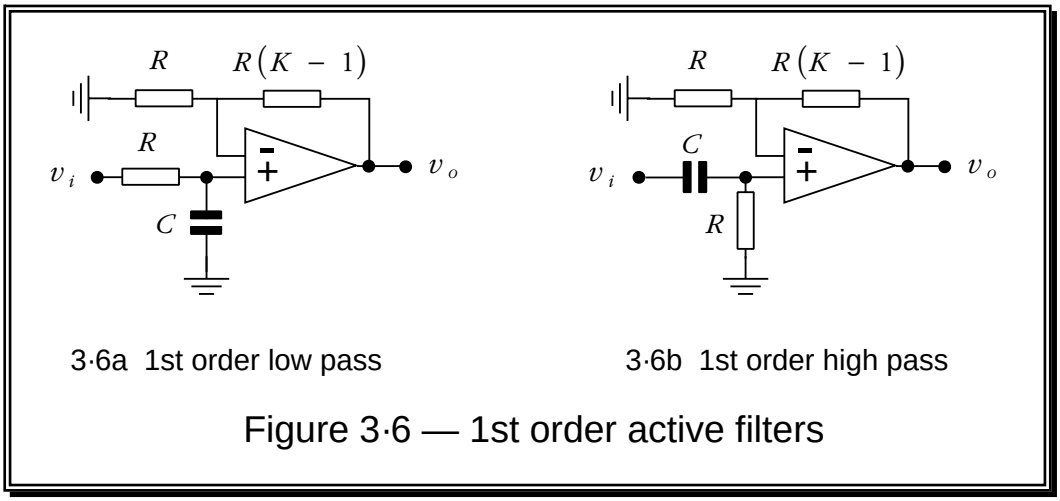
A *Normalised* version of the circuit would assume that R was one Ohm, and C was one Farad. This would set the turn-over frequency of $\omega_0 = 1$ Radians/sec (i.e. $f_0 = 1/2\pi$ Hz).

In reality, normalised values will be impractical, and we will usually want a specific turn-over frequency very different to the above value. However we can scale the normalised version of each circuit as follows:

- To shift the turn-over frequency to ω_0 we choose the R and C values so that expressions 3.7 is correct
- To obtain convenient component values we then 'trade off' the values of R and C against each other until we get something suitable. e.g. we may increase R by some factor, α, and then reduce C by the same factor.

The choice of sensible values is a matter of practical experience so is hard to define in a lecture course. This is one of the areas of electronics where design is a skill, not a science which provides a uniquely correct answer. When using op-amps for the amplifiers in the active filer, we generally try to end up with resistor values in the range 1kΩ to 100kΩ, and capacitors in the range from 1μF down to 10pF.

Lets look at some examples of designs, then try scaling one for a specific result.



3·6a 1st order low pass 3·6b 1st order high pass

Figure 3·6 — 1st order active filters

Figure 3·6 shows 1st order low and high pass active filters. For a Normalised version we would assume that $R = 1\Omega$ and $C = 1$F. More generally we can say that the frequency response of A_v for each of these will have the form

$$\text{for 3·6a} \qquad A_v = \frac{K\omega_0}{S + \omega_0} \qquad \ldots (3.8)$$

for 3·6b $$A_v = \frac{KS}{S + \omega_0} \quad \ldots (3.9)$$

where we use a commonly used standard definition

$$S \equiv j\omega \quad \ldots (3.10)$$

Let's now take as an example a real filter which we want to have a turn-over frequency of, say, $f_0 = 5\text{kHz}$. This means that in this case $\omega_0 = 1/RC = 2\pi \times 5000 = 31{,}415$. If we try choosing a resistance value of $R = 10\text{k}\Omega$ this means we would need to choose $C = 3{\cdot}18\text{nF}$ to set the required turn-over frequency.

In the case of the above circuits the K value just sets the gain for frequencies well within the *Passband* of the filter. So we can choose K to amplify the output by whatever amount is suitable for our purpose. (Although in practice it is usually advisable to choose a gain below ×100.) However for more complex higher-order filters the gain often alters the details of the frequency response, particularly around the chosen turn-over frequency. These effects can be seen in the following examples of second order filters.

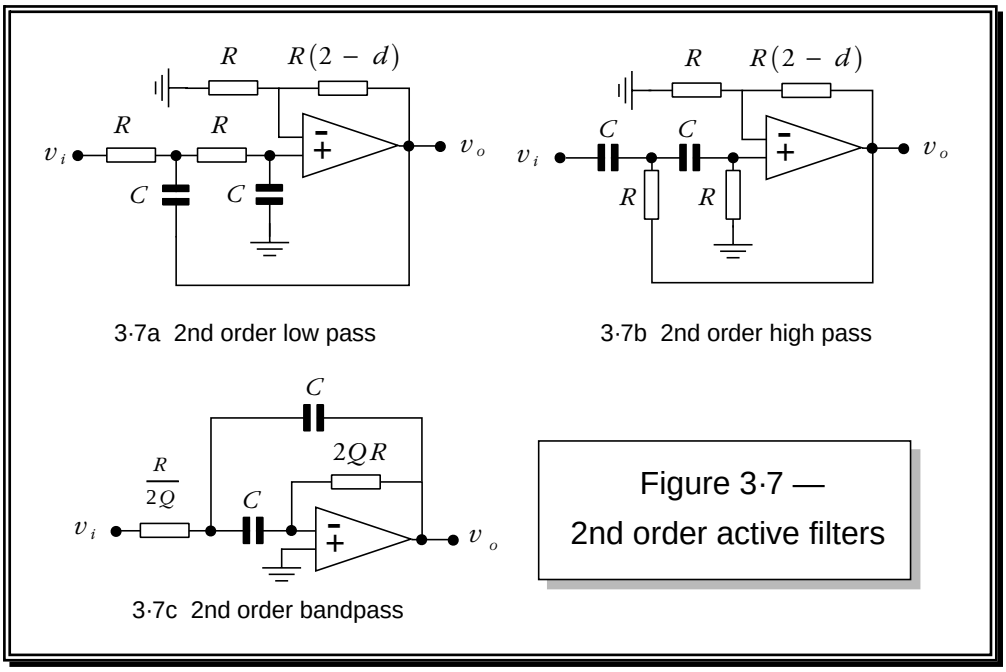


Figure 3·7 — 2nd order active filters

In this case an example of a band-pass filter is also included. The examples shown introduce two new quantities that have standard definitions. The *Damping* value, d, and the *Quality* factor, Q The gain value, K is related to these via the expressions

$$K = 3 - d \quad : \quad K = -2Q^2 \quad \ldots (3.11 \text{ \& } 3.12)$$

The frequency response of each of the above is as follows:

3·7a 2nd order lowpass $$A_v = \frac{K\omega_0^2}{S^2 + d\omega_0 S + \omega_0^2} \quad \ldots (3.13)$$

3·7b 2nd order highpass $$A_v = \frac{KS^2}{S^2 + d\omega_0 S + \omega_0^2} \quad \ldots (3.14)$$

3·7c 2nd order bandpass $$A_v = \frac{K\omega_0 S}{S^2 + \omega_0 S/Q + \omega_0^2} \quad \ldots (3.15)$$

The damping and quality factors have a very important effect upon the filter response. To see this we can start by take the damping factor's effect on the high/low pass filters.

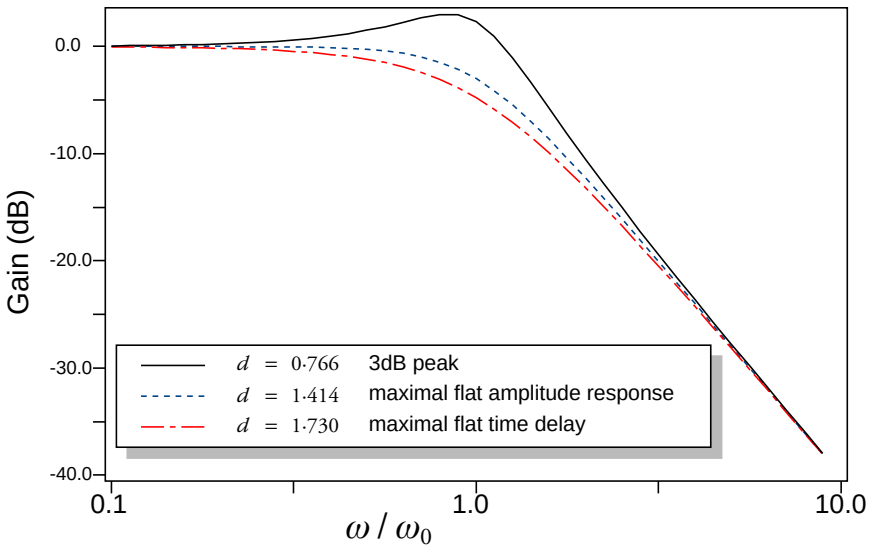


Figure 3·8 — Effect of damping upon low-pass response

Figure 3·8 shows the relative frequency/amplitude response shapes for three possible choices for d. It can be seen that d values below unity produces a 'peak' in the response at frequencies around ω_0. (In fact, if we were to reduce $d \rightarrow 0$ the circuit becomes an oscillator and will oscillate at ω_0.) Values of d greater than unity tend to 'damp down the response and smooth out the corner at around ω_0. A choice of $d = 1{\cdot}414$ (i.e. the square root of 2) produces a *Maximally Flat Amplitude Response*. i.e. the amplitude response covers the smallest possible range in the passband up to ω_0. A choice of $d = 1{\cdot}730$ produces a *Maximally Flat Time Delay*, i.e. the most uniform Group Delay in the passband up to ω_0. We can therefore in practice select a d value which suits best our purpose. Note that the gain in the passband also depends upon d. To make the changes in shape clearer the above graphs are normalised to a 0dB level of K^2. Note also that the above example uses the low-pass filter. The high-pass filter behaves in a similar manner but we would be referring to a passband that extends downward from high frequencies to ω_0 instead of up to ω_0.

The bandpass filter has a frequency response which has a width (to the -3dB points) of

$$\pm\Delta f = f_0 / Q \qquad \ldots (3.16)$$

and has a phase response that changes from $+\pi$ radians at $f_0 - \Delta f$ to $-\pi$ at $f_0 + \Delta f$. In effect, 'Q' sets the narrowness or sharpness of the bandpass filter. A sharper filter will allow us to pick out a wanted frequency from noise, etc, at other frequencies so a narrow filter is often desirable. This is why this variable is therefore often called the *Quality Factor*, and given the letter 'Q'.

The filters used above as examples are just a few of the many types of filter that are available. Higher orders of filter can be made with many slight variations in the relative component values, and with various arrangements. The relative choices of component values (which then determine the a_i and b_j coefficients in equation 3.6) determine what is called the *Alignment* of the filter and set its detailed response. The most common forms of alignment have names like *Butterworth* filters, *Elliptic* filters, etc. Each has its good and bad points but can be designed and analysed using the same methods as used above.

In practice it is common to build high order filters by cascading two or more lower order (1st or 2nd) order filters. This makes the system easier to design and means it can be less demanding to build. High order filters often require close-tolerance components with strange values which then have to be put together from series/parallel combinations of more usual values. There are also a number of specialist filters that use quite different techniques. The most common analog types are those based upon 'biquad' and 'NIC' (Negative Impedance Convertor) arrangements. The NIC is particularly useful for special applications as it permits us to make a circuit which provides otherwise impossible values – e.g. a negative resistance or inductance.

At radio frequencies there are a number of special filters based upon the properties of some materials – Crystal Filters, Surface Acoustic Wave Filters, etc. And of course, these days, it is common to perform signal filtering in the digital domain, so there are a variety of digital filtering systems available. Since here we are concentrating on basic analog methods we won't here examine how these work, but it is worth bearing them in mind for situations where they may prove more convenient than the analog filters described in detail in this lecture.

Summary

You should now understand how the behaviour a filter may be characterised by its *Frequency Response* (both in terms of amplitude gain/loss, and the way in which phase varies with the signal frequency). That this *Frequency Domain* behaviour also relates to a *Time Domain* behaviour. That it is often useful to define or use the Time Delay as a function of frequency to assess the filter's effect upon signals. You should now also know that the *Group Delay* indicates the average time delay that signals encounter across the filter's passband, and that a uniform Group Delay will avoid *Dispersive* distortions of the filtered signal.

You should now understand the concept of the *Order* of a filter, and know how to design simple 1st and 2nd order filters. That the *Quality Factor* or *Damping* of a filter affects the detailed response, and that these can be set by the component values selected. You should also now be aware that the *Alignment* of a filter can be chosen for optimal amplitude or phase or time properties in a given application, and that some 'standard' alignments with names like Butterworth Filters, etc, exist.

Lecture 4 – Feedback

In lecture 2 we looked at some of the basic limitations of signal amplifiers, in particular power and signal size limitations. Now we will look at some of some of the other problems and see how a specific technique called *Feedback* can be used to deal with them. However, before doing so it is worth remembering that all 'active' systems use amplifiers. For example, the active filters we looked at in the last lecture require op-amps or similar amplifiers. What is less obvious is that amplifiers, in themselves, tend to act as filters...

4·1 The Problems

There are two basic problems we want to examine in this lecture. The first is a consequence of the fact that all real components – resistors, capacitors, transistors, etc – suffer from what is know as 'stray' or 'parasitic' effects. For example, every real capacitor will have leads that have a non-zero resistance and inductance. Similarly, every transistor will have some capacitance between its leads, inductance in the leads, etc. As a result, every circuit we ever make will include lots of unintended, invisible 'components' which will affect the performance.

In addition, all real gain devices take a finite time to respond to a change in their input. For example, the ability of a bipolar transistor to pass current from its emitter to collector will depend upon the density of charges in its base region. When we change the base current it will take a finite time for the new current level to alter the charge density right across the base. It then takes another finite time before the emitter-collector current can react to this change. In bipolar transistors these effects are described in terms of an 'f_T' value (*transit frequency*). This indicates the highest input frequency which will be able to change the output current. Higher frequencies tend to be ignored by the transistor as they change too swiftly for the transistor to be able to react.

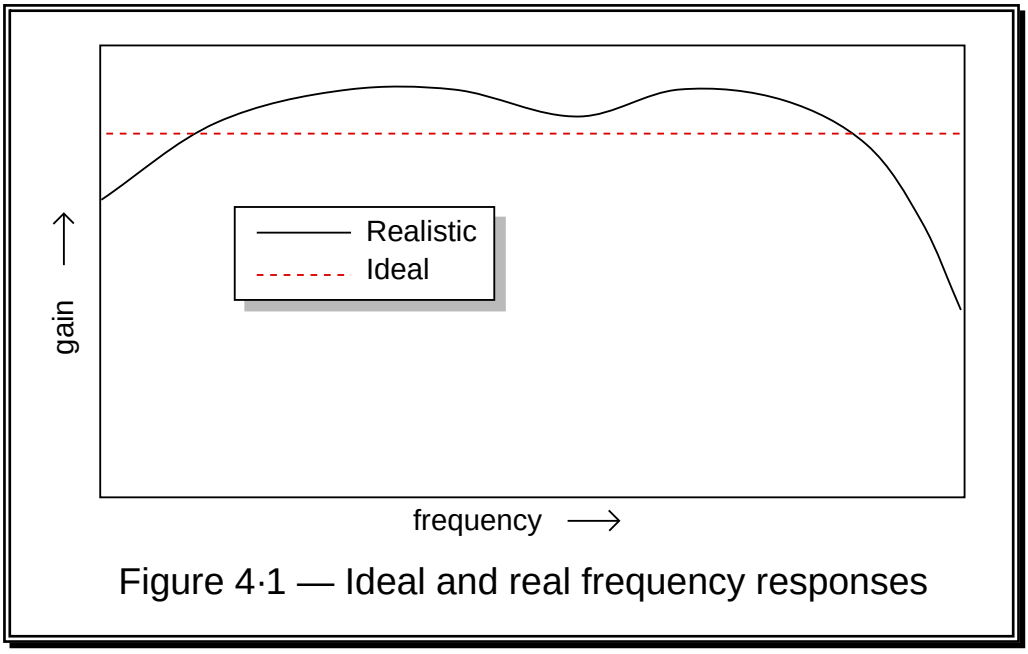


Figure 4·1 — Ideal and real frequency responses

We don't really need to worry here about all the detailed reasons for these effects. We can start from accepting that the result will always be an amplifier that will have a frequency response that

isn't just a uniform gain value at all frequencies. A real amplifier will be more likely to exhibit the complex kind of shape shown by the solid line in figure 4·1 than the ideal 'flat' response. In particular there will always be a tendency for the gain to fall at high frequencies. The phase/time delays may also vary with frequency. In addition, the complexity of the circuit (with all the stray effects) may mean that its frequency/time behaviour is far from uniform even at frequencies it can amplify reasonably well. The result of these effects is twofold. Firstly, the band of frequencies the amplifier can amplify may not be wide enough for our purposes. Secondly, the non-uniform gain/phase/time behaviour may alter the signal waveforms in a way we do not want. The signals can be distorted by non-uniform time/amplitude behaviour, although this term isn't usually used for this effect.

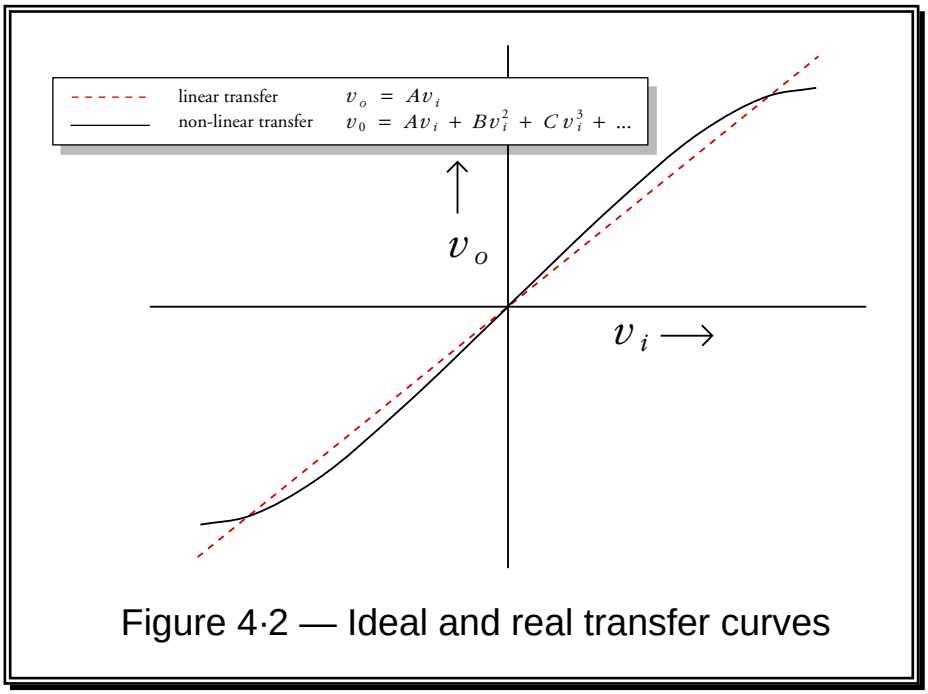


Figure 4·2 — Ideal and real transfer curves

The term *Distortion* is usually used to mean that the amplifier has a gain which depends upon the signal amplitude (or sometimes also the frequency). The effect is described as an effect which arises due to *non-linearity*. Figure 4·2 illustrates the kind of non-linearity we are concerned with here. We can distinguish an ideal amplifier from a realistic one in terms of the *Transfer Curve* – the plot of v_o versus v_i – and the expressions that relate the input to the output.

For an ideal, linear amplifier, the transfer curve will actually be a straight line, and we find that

$$v_o = Av_i \qquad \ldots (4.1)$$

i.e. the output level is simply proportional to the input and we can say that the amplifier's voltage gain will be $A_v = A$, a set value.

For a real amplifier, the transfer curve will not be a perfectly straight line. Now we have to represent the relationship between input and output using a more complicated equation. The most convenient form for our purpose is a polynomial, so we would now write something like

$$v_o = Av_i + Bv_i^2 + Cv_i^3 + \ldots \qquad \ldots (4.2)$$

where the individual coefficients are constants for a specific amplifier. The result of non-zero values for B, C, etc means that the effective gain now becomes

$$A_v = A_0\left[1 + \alpha_1 v_i + \alpha_2 v_i^2 + \ldots\right] \qquad \ldots (4.3)$$

where since we can define $A_v \equiv v_o / v_i$ we can say that

$$\alpha_1 = \frac{B}{A_0} \quad : \quad \alpha_2 = \frac{C}{A_0} \quad \text{etc}\ldots \qquad \ldots (4.4\ \&\ 4.5)$$

For a well-designed amplifier we can hope that the $|\alpha_i| \ll 1$ for all i. This will mean the the amount of non-linearity, and hence the amount by which a signal is distorted during amplification, will be small.

4·2 Negative Feedback and performance improvements

Having outlined the problems we can now examine the application of feedback and see how this may help us reduce their magnitude.

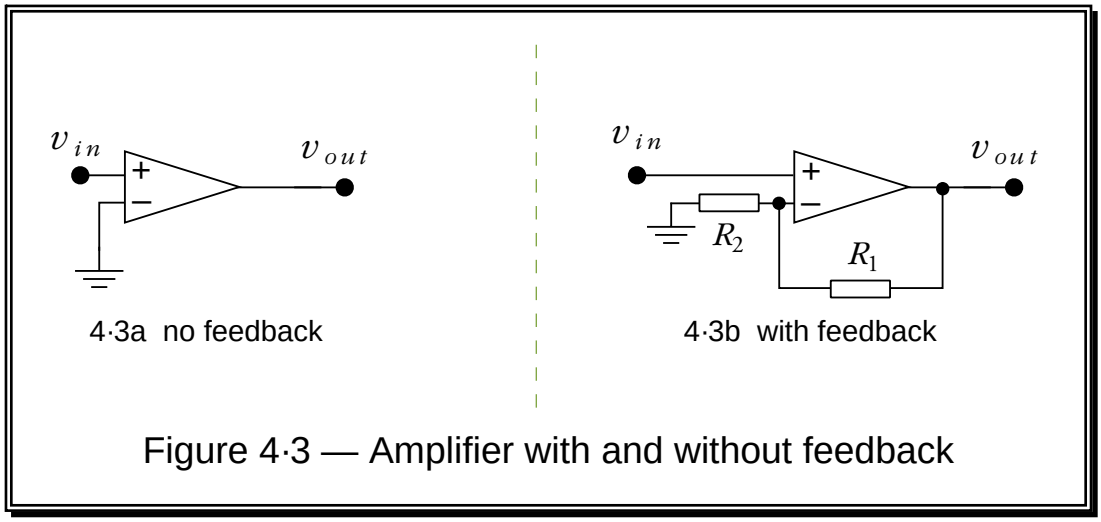


Figure 4·3 — Amplifier with and without feedback

Figure 4·3 shows the amplifier in two situations. 4·3a shows the amplifier with no applied feedback. 4·3b shows a modified circuit where a pair of resistors applies feedback to the non-inverting input of the amplifier. This feedback is subtracted from the actual input and tends to reduce the signal levels, hence it is conventionally called *Negative Feedback*.

The feedback applied to the amplifier in 4·3b means that the voltage applied to the inverting input of the amplifier will be

$$v' = v_o \frac{R_2}{R_1 + R_2} \qquad \ldots (4.6)$$

we can now define the *Feedback Factor*

$$\beta \equiv \frac{R_1 + R_2}{R_2} \qquad \ldots (4.7)$$

to simplify the following expressions.

From the explanations given earlier, we can expect the effective gain A_v to depend to some extent upon the signal level and the frequency. When the apply feedback the output will become

$$v_0 = (v_i - v')A_v = \left(v_i - \frac{v_0}{\beta}\right)A_v \qquad \ldots (4.8)$$

We can now rearrange this to define the overall voltage gain, A_v', of the system with feedback

applied as

$$A_v' \equiv \frac{v_0}{v_i} = \frac{A_v}{1 + A_v/\beta} = \frac{\beta}{1 + \beta/A_v} \quad \ldots (4.9)$$

It is usual to call A_v the amplifier's *Open Loop Gain* as it is the gain we obtain if we were to 'break' the feedback connection. In a similar way, it is usual to call A_v' the system's *Closed Loop Gain*. In practice we can usually arrange for $A_v \gg \beta$. Hence can say that for a suitably small β/A_v we can approximate the above to

$$A_v' = \beta\left(1 - \frac{\beta}{A_v}\right) = \beta - \frac{\beta^2}{A_v} \quad \ldots (4.10)$$

This result is an important one for two reasons. Firstly it tells us that the gain of the system with feedback applied is largely determined by the choice of the feedback factor, β, not the inherent amplifier gain, A_v. This is because it follows from the above that $A_v' \to \beta$ when $A_v \to \infty$.

The second reason can be seen to follow from the first. When $A_v \gg \beta$ the gain when feedback is applied only has a weak dependence up A_v. This implies that changes in A_v will have little effect. To see the effect of the feedback upon distortion let us assume that a change in signal level or frequency has caused the open loop gain to change from $A_v \to A_v(1 + e)$, where we can assume that $|e| \ll 1$ is a small fractional *Error* in the expected gain. This means that the output level is in error by a fraction, e, so we can use this as a simple measure of the amount of signal distortion/alteration produced by the unwanted change in gain.

The change in open loop gain will, in the feedback system, cause a related change, e', in the closed loop gain such that

$$A_v'(1 + e') = \beta - \frac{\beta^2}{A_v(1 + e)} \approx \beta - \frac{\beta^2(1 - e)}{A_v} \quad \ldots (4.11)$$

Combining expressions 4.10 and 4.11 we obtain

$$A_v' e' = -\frac{e\beta^2}{A_v} \quad \ldots (4.12)$$

which since we can expect $A_v' \approx \beta$ is similar to

$$e' \approx -\frac{e\beta}{A_v} \quad \ldots (4.13)$$

i.e. the fractional magnitude of the amount of error (and hence the degree of distortion) is reduced by the factor β/A_v. Now it is quite common for operational amplifier ICs to have low-frequency open loop gain values of the order of 10^6. Used with a feedback factor of 10, the result is a system whose voltage gain is almost exactly 10 where any distortions due to variations in gain with signal level or frequency are reduced by a factor of 10^5. The result is an amplifier system whose behaviour is largely determined by the choice of the resistors used for the feedback loop. Given its simplicity and the dramatic effect it can have, it is perhaps no surprise that Negative Feedback is often regarded as a 'cure all' for correcting amplifier imperfections. In reality, though, feedback can itself lead to problems...

4.3 Limitations of Feedback

The most obvious 'limitation' imposed by negative feedback is that the overall gain is reduced. As a result, when we want a lot of gain we may have to use a chain or *Cascade* of amplifiers. This will be more expensive, and require more space. At first glance it also looks as it will undo the benefit

of low distortion since each amplifier in the chain will add its own distortions to the signal. Fortunately, this isn't usually a real problem as we can see from the following example. Let's assume that we have a basic amplifier design which – open loop – has a voltage gain of ×1000, and we want to amplify an actual input signal voltage by ×1000.

The simplest thing to do would be to use the amplifier, open loop, to provide all the gain. However in order to reduce the distortion we apply a feedback factor of $\beta = 10$, and then use a chain of similar stages, each with this feedback factor. The result still have an overall gain of 10× 10×10 = 1000, but the feedback reduces the error level (and hence the distortion level) by a factor of 10/1000 = 0·01 in each stage.

The problem is that each stage adds its own distortion. We can therefore say that the total gain error of the chain of three amplifiers will be given by

$$1 + e_t = (1 + e_1')(1 + e_2')(1 + e_3') \qquad \ldots (4.14)$$

(where we are assuming that the signal passes through the amplifiers in the order, '1, 2, 3'). Now when we obtain the same output level from the end of the chain as we would using a single stage with no feedback we can expect that

$$e_3' = -\frac{e\beta}{A_v} \qquad \ldots (4.15)$$

However the distortion levels produced by the other amplifiers are **not** equal to e_3'. The reason for this is that the error (distortion) level varies with the size of the signal. In the middle amplifier. The signal level at the input/output of the middle amplifier ('2') in the chain will be β times smaller than seen by '3' at its input/output. Similarly, the input/output level seen by the first amplifier will be β^2 times smaller than seen by '3'.

For the sake of simplicity we can assume that the error level varies in proportion with the signal voltage, hence we can say that

$$e_2' = \frac{e_3'}{\beta} \qquad : \qquad e_1' = \frac{e_3'}{\beta^2} \qquad \ldots (4.16 \text{ \& } 4.17)$$

We can now approximate expression 4.14 to

$$1 + e_t \approx 1 + e_1' + e_2' + e_3' = 1 + e_3'\left(1 + \frac{1}{\beta} + \frac{1}{\beta^2}\right) \qquad \ldots (4.18)$$

i.e. for the values used in this example

$$e_t \approx -\frac{e\beta}{A_v} \times \left(1 + \frac{1}{\beta} + \frac{1}{\beta^2}\right) = 0{\cdot}0111 \times e \qquad \ldots (4.19)$$

i.e. we find that in practice $e_t \approx e_3'$ and the overall distortion level tends to nearly all come from the distortions produced by the final stage.

In practice, the behaviour of a real system will be more complex but the explanation given above remains generally valid in most cases. As a result, a chain of feedback-controlled amplifiers tends to provide lower distortion than using a single stage without feedback despite the sacrifice in gain per stage. The only significant case where this result is not correct is when the error or distortion level does **not** fall with the signal size – i.e. if there is some kind of 'kink' or discontinuity in the amplifier transfer curve in the region around zero volts. In practice the situation where this arises is when we have used a pure class 'B' amplifier which will exhibit cross-over distortion. In these cases the feedback will reduce the distortion level for reasonably powerful signals, but small signals will experience high levels of gain error, and hence high distortion even when feedback is used. For this reason class 'B' should be avoided – especially when we are dealing with small signals – if

we wish to avoid distortion.

From the simple explanation of feedback given in section 4·2 we might think that "the more, the better". In practice, however, it is usually a good idea to only use a modest amount of feedback as excessive use can lead to some serious problems. The reason for this can be seen by looking again at expression 4·9 are remembering that in general A_v, β, and hence A_v' will all be *complex* values. i.e. the gain and the feedback normally include some frequency-dependent phase/time effects as well as changes in signal amplitude. To see the consequences of this lets take another example.

Consider a two-stage amplifier chain. Each stage has an open loop voltage gain of ×100 so the total low-frequency gain is $\times 10^4$. However one of the amplifiers has internal stray capacitances that mean its gain tends to 'roll off' and behave as if it includes a first order low-pass filter with an f_0 of 200 Hz. The other amplifier has a similar limitation which gives it a turn-over frequency of 5kHz. We then measure the overall frequency response both with and without some feedback. The results will be as shown in figure 4·4.

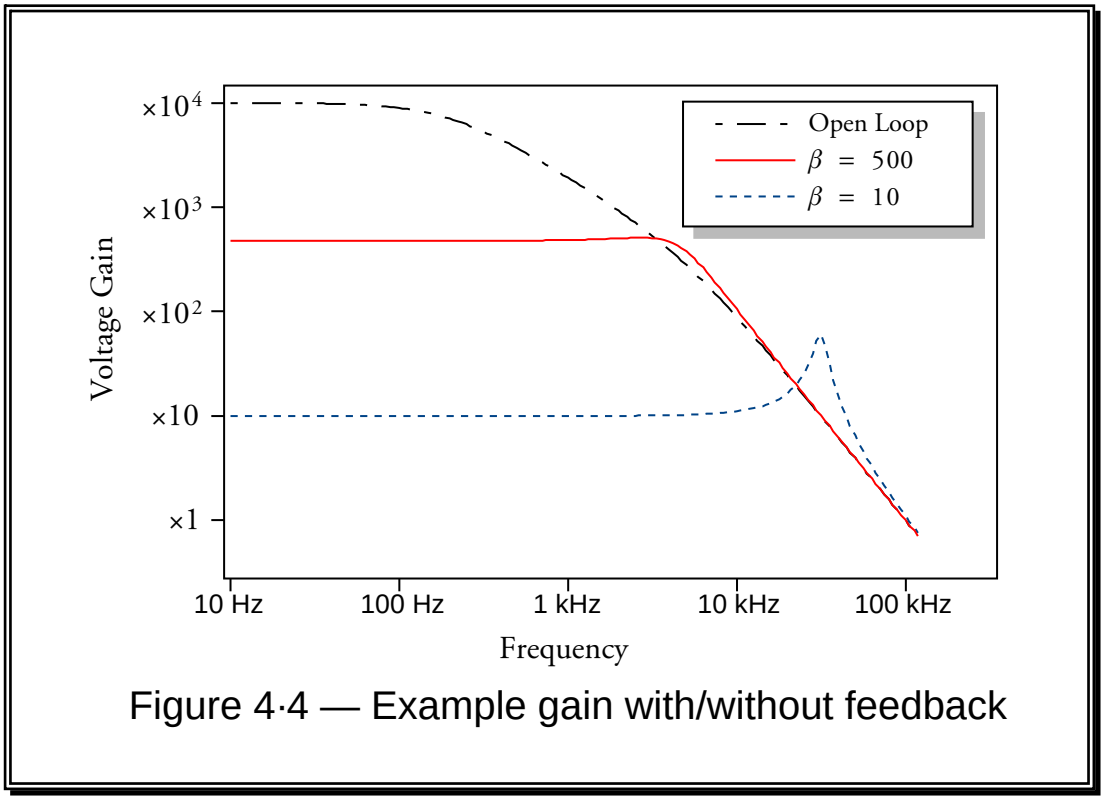


Figure 4·4 — Example gain with/without feedback

The broken line with alternating long/short dashes shows the open loop gain of the system as a function of frequency. The solid (red if you see this in colour) line shows the frequency response with a feedback factor of $\beta = 500$, and the broken (blue) line shows the response with a feedback factor of $\beta = 10$. Looking at the graphs we can see that the $\beta = 500$ feedback has the beneficial effect of 'flattening' the response and ensuring that the gain remains reasonably uniform over a wider range than is the case for the system without feedback. i.e. without feedback the response is only fairly flat up to a few hundred Hz, with with a feedback of $\beta = 500$ applied the response is flat to nearly 10 kHz. Thus the bandwidth over which the amplifier can be used is increased by the application of feedback. In effect we have 'flattened down' the response, but it still almost fits inside the open loop response. Note, however that it doesn't quite fit and that it slightly exceeds the open loop gain at high frequency.

It is customary to describe the amount of feedback applied to a system in terms of the ratio A_v / β. When this value is around unity the feedback does almost nothing. However as we reduce β in comparison with A_v the gain reduces, the distortion level tends to fall, and the flat bandwidth tends to increase. However if A_v / β becomes too large things can go wrong as indicated by the $\beta = 10$ line in figure 4·4. Here we can see that the frequency response has developed an odd 'peak' at around 30 kHz. The reason this happens can be understood by looking at figure 4·5 which displays the phase change versus frequency for each of the situations shown in figure 4·4.

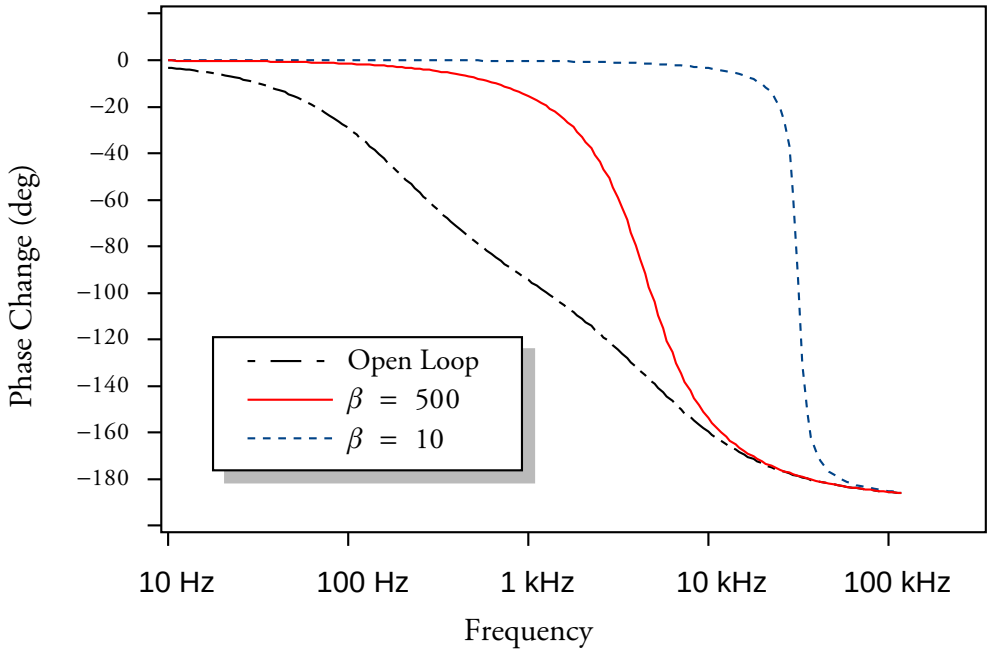


Figure 4·5 — Phase shifts with/without feedback

Figure 4·5 shows two interesting points. Firstly, we can see that the feedback tends to suppress variations of the phase as a function of frequency just as it tends to reduce variations in the amplitude of the gain. However the second point is that this is only the case when the open loop phase change is less than about 90 degrees. In particular, looking at the graphs we can see that when the open loop phase change approaches −180 degrees, the closed loop systems also move rapidly towards −180 degrees.

Now a −180 degree change at a given frequency means that the complex gain has a **negative** sign. From before, we can say that the closed loop gain will be

$$A_v' = \frac{\beta}{1 + \beta / A_v} \qquad \ldots (4.20)$$

however we may now find that β / A_v is negative. Worst still, we may find that $\beta / A_v \rightarrow -1$. If this happens it means that $(1 + \beta / A_v) \rightarrow 0$, which in turn means that $A_v' \rightarrow \infty$. i.e. unless we take care we find that the feedback we intend to use to reduce the gain and make it more uniform may produce an unexpectedly high gain at a specific frequency (or set of frequencies) where we happen to have arranged for the above condition to arise! Since most of the effects which cause the open loop response to vary will affect both the amplitude and phase this problem

is always potentially present. The more feedback we apply the wider the range of frequencies we bring into the intended 'flat' region, and the more chance there is for this problem to arise.

In fact we have encountered a consequence of what is called the *Barkhausen Criterion*. This says that we can make an oscillator by arranging for $\beta / A_v = -1$. The peak at the end of the flat portion of the $\beta = 10$ response indicates that the system is near to becoming a 30 kHz oscillator. In practice, the load we may attach to the output of a real amplifier will tend to alter its gain behaviour, so if we allow the system to be 'near' oscillation we may find that it **does** oscillate under some circumstances when certain loads are attached. Avoiding this is particularly important in situations like hi-fi power amplifiers where the amplifier manufacturer can only be confident that a wide – even alarming! - range of loudspeakers with weird impedance behaviour will be used by various customers! An amplifier system which has a flat, unpeaked, response, with no risk of unwanted oscillations irrespective of the output load is said to be *Unconditionally Stable*. For obvious reasons this is highly desirable if it can be achieved.

To avoid these problems and ensure stability we have to adopt two approaches. One is to build an amplifier that works as well as possible without feedback, and then only apply a moderate amount. (Note that the terms "well as possible" and "moderate" here are matters of personal judgement not strict engineering.)

The second approach is to *compensate* the feedback or the amplifier to try and remove the problem. This can be done in various ways but two examples are as follows.

Firstly, we can adjust the amplifiers so that the open loop value of A_v has a magnitude of much less than unity at any frequency where the phase shift approaches 180 degrees. This make it very difficult for β / A_v to be able to approach -1.

Secondly we can apply *lag/lead* compensation. Up until now we have assumed the feedback circuit is just a pair of resistors. However we can add other components (capacitors or inductors) to the *Feedback Network* so that either the magnitude of β or its phase change with frequency in a way that ensures a more uniform behaviour of β / A_v as a function of frequency and keeps its value well clear of -1 at any frequency. The details of this process are outwith the scope of this lecture, but you should see that these methods can help the situation, although the best policy is usually to avoid the problem by applying feedback with care.

Summary

You should now understand how *Negative Feedback* can be used to improve the performance of an amplifier system to give it a wider 'flat response' bandwidth and reduce the level of signal distortion. However it should also be clear that this feedback tends to reduce the available gain and has some associated problems. That the main limitations are of two kinds

- The risk of oscillations or instability
- Class B type distortions

We can use feedback with care to avoid these problems but it should be clear that although some feedback helps it is no substitute for designing or using an amplifier whose open loop behaviour is as good as possible.

Lecture 5 – Power Supplies

Virtually every electronic circuit requires some form of power supply. This normally means that it requires one or more *Power Rails* – lines held at specific steady d.c. voltages – from which the required current and hence power may be drawn. There is also usually a need for an *Earth* connection, both as a reference level and for safety purposes. In some cases the power can come from a battery or some other kind of chemical cell. However most systems require a Power Supply Unit (PSU) that can take a.c. power from the mains and convert this into the required d.c. levels. This lecture will examine the basic types of mains-to-d.c. analog PSUs that are used and consider their main properties.

5·1 The basic PSU

Figure 5·1 illustrates the traditional basic form of a simple mains PSU. It consist of a *Transformer*, a diode, and a smoothing or *Reservoir* capacitor.

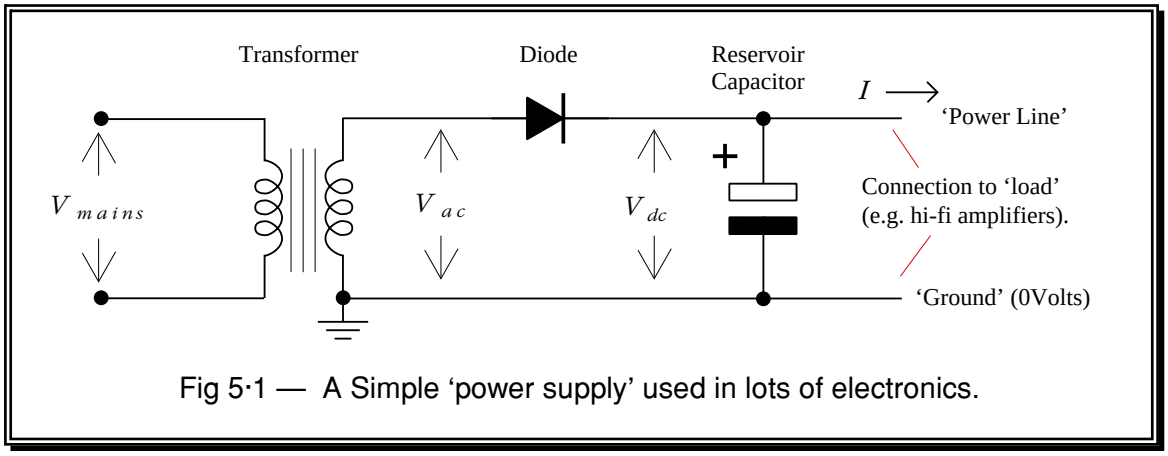


Fig 5·1 — A Simple 'power supply' used in lots of electronics.

We will assume that we don't need to consider the details of how diodes and capacitors work at this point. Instead we can just make a few relevant comments. The most important property of a diode in the context of a PSU is that it will conduct current when a *Forward Bias* is applied, but refuses to conduct when a *Reverse Bias* is applied. Hence when the output from the transformer means that $V_{ac} > V_{dc}$ the diode will pass current and the capacitor will be charged up by the applied voltage from the transformer. However when $V_{ac} < V_{dc}$ the diode will refuse to conduct. Hence none of the charge in the capacitor will be removed again via the diode. Instead it will be available to the output load and the capacitor can supply current to the amplifiers, etc, which use the PSU output. in practice, there will always be a modest voltage drop across the diode when it conducts, so V_{dc} will never be as large as the peak positive value of V_{ac}.

Note that the capacitor shown in figure 6·1 is an *Electrolytic* type. In theory this isn't essential. However we often find that we want very large capacitance values (perhaps tens or hundreds of thousands of μF) and these would become physically very large and expensive unless we choose electrolytics. Electrolytic capacitors have a number of properties which make them undesirable in high-performance circuits, and can explode if used 'the wrong way around' or at too high a voltage. However they offer large amounts of capacitance in a given volume for a given cost so are widely used in PSUs.

Figure 5·2 illustrates how the voltages and current in the circuit behave in a typical situation.

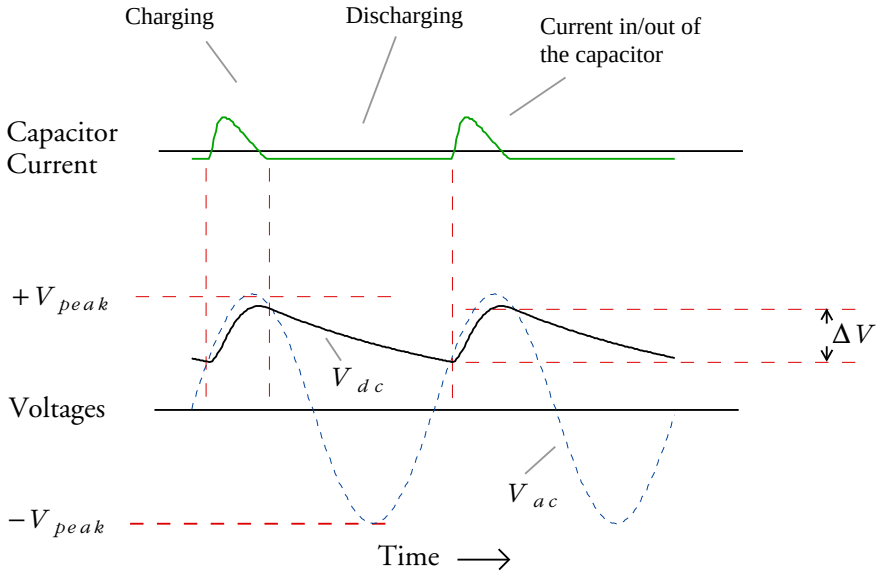


Fig 5·2 — Voltages & currents in power supply.

The input transformer 'steps down' the mains input to provide a 50Hz sinewave voltage[1], V_{ac}, which swings up & down over the range $+V_{peak}$ to $-V_{peak}$ as shown in figure 5·2. The value of V_{peak} depends on the turns ratio of the transformer. The voltage on the capacitor, V_{dc}, at any moment depends on how much charge it holds.

When $V_{ac} > V_{dc}$ the diode is forward biassed. Current then flows through the diode, charging the capacitor. As V_{ac} rises to $+V_{peak}$ during each cycle it pumps a pulse of charge into the capacitor & lifts V_{dc}. Using a silicon diode we can expect the peak value of V_{dc} to approximately be $V_{peak} - 0{\cdot}5$ Volts since the diode always 'drops' around half a volt when it's conducting. When V_{ac} swings below V_{dc} the diode stops conducting & blocks any attempt by charge in the capacitor to get out again via the diode. Since the mains frequency is 50Hz these recharging pulses come 50 times a second.

In the gaps between recharging pulses the dc current drawn out of the supply will cause V_{dc} to fall. How much it falls during these periods depends upon the amount of current drawn & the size of the capacitor. The time between pulses will be

$$\Delta t \approx \frac{1}{50} = 20 \text{ milliseconds} \qquad \ldots (5.1)$$

During this time a current, I_{dc}, would remove a charge

$$\Delta q = I_{dc}\Delta t \qquad \ldots (5.2)$$

from the capacitor, reducing the d.c. voltage by an amount

[1] I am assuming we are in the United Kingdom, hence have 50Hz mains power.

$$\Delta V \approx \frac{I_{dc}\Delta t}{C} \qquad \ldots (5.3)$$

before it can be lifted again by the next recharging pulse.

As a result of this repetitive charging/discharging process during each mains cycle the output voltage *ripples* up & down by an amount ΔV. Hence the output isn't a perfectly smooth d.c. voltage. However, from equation 5.3 we can see that it's possible to reduce the amount of ripple by choosing a larger value capacitor for the reservoir, *C*. This is why PSUs tend to use large capacitances – it minimises the amount of ripple. Ideally, a PSU will provide a fixed d.c. level, but in reality it will have ripple, and may alter when the current demanded by the load changes. In general a 'better' supply will provide a more stable, constant output d.c. level.

5·2 The Transformer

From the explanation given above we can expect the simple PSU to provide an output d.c. level of around V_{peak} − 0·5 Volts. To set a given voltage we must therefore be able to choose the amplitude of the input a.c. waveform delivered to the diode. This is the task of the transformer. To explain the detailed behaviour of a transformer is quite complicated. Here, for the sake of clarity, we can use a very basic model to explain its basic operation. Please note, however, that what follows is a rather simplified explanation and doesn't really give the full picture. It is good enough for our current purposes, though.

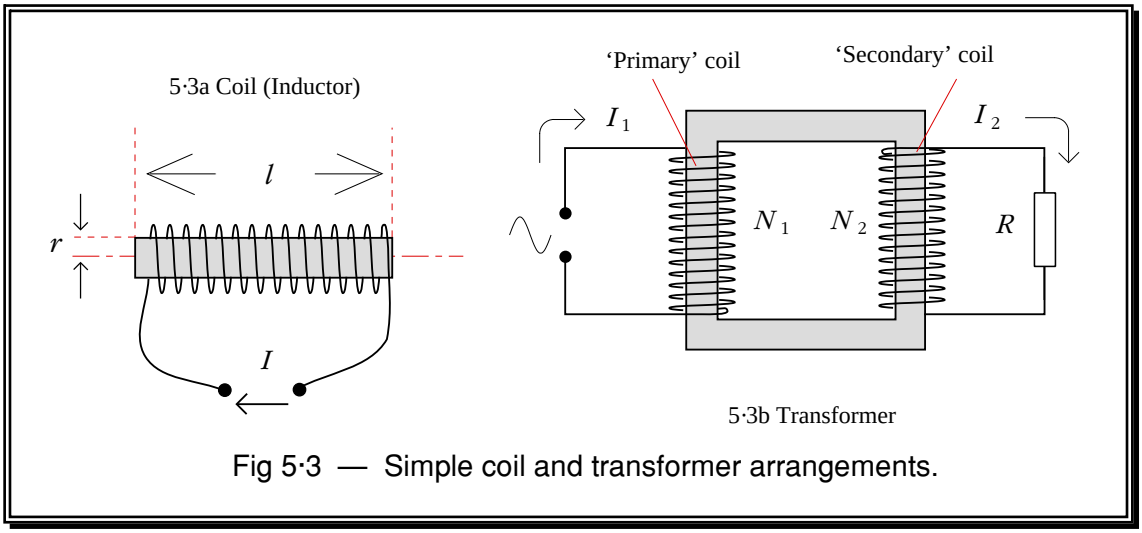


Fig 5·3 — Simple coil and transformer arrangements.

The basic properties of a *transformer* can be explained in terms of the behaviour of *inductors*. Figure 5·3a shows a coil of wire wound around a rod. Passing a current along the wire will produce a magnetic field through & around the coil. This current-field relationship works both ways — i.e. applying a magnetic field through the coil will produce a current. By looking at a good book on electromagnetics we can find the relationship

$$B = \frac{\mu N I}{\sqrt{4r^2 + l^2}} \qquad \ldots (5.4)$$

where

- B = field level at the centre of the coil (Teslas);
- I = current in coil (Amps);
- μ = permeability of the rod (Henries/metre);

- N = number of turns of wire;
- r = radius of coil (metres);
- l = length of coil (metres)

The important feature of a *transformer* is that two (or more) coil *windings* are placed so that any magnetic field in one has to pass through the other. The simplest way to do this is to wind two coils onto one rod. However, this method isn't used much because it produces a lot of field outside the rod. This 'stray' field can cause problems to other parts of a circuit and may waste energy. Most modern transformers are wound on some kind of 'ring' of material as illustrated in figure 5·3b. This works as if we'd 'bent around' the rod & brought its ends together. The magnetic field is now *coupled* from coil to coil by the ring of magnetic material. It's conventional to call the coils on a transformer the *primary* (input) & *secondary* (output) windings. Consider now what happens if we apply an an a.c. current, $I_1\{t\}$, to the primary of the transformer shown in figure 5·3b. This winding has N_1 windings, so it produces a magnetic field

$$B\{t\} = kN_1I_1\{t\} \quad \ldots (5.5)$$

where k is a constant whose value depends upon the size/shape/material of the transformer's ring. Since this field also passes through the secondary winding it must produce a current, $I_2\{t\}$, in the secondary, such that

$$B\{t\} = kN_2I_2\{t\} \quad \ldots (5.6)$$

putting these equations together we get

$$I_2\{t\}\,N_2 = I_1\{t\}\,N_1 \quad \ldots (5.7)$$

The transformer's output is connected to a resistive *load*. Since the current, $I_2\{t\}$, passes through this, the voltage across the resistor will be

$$V_r\{t\} = I_2\{t\}\,R \quad \ldots (5.8)$$

The power dissipation in the resistor will therefore be

$$P\{t\} = V_r\{t\}\,I_2\{t\} \quad \ldots (5.9)$$

Now, from the principle of energy conservation, we can say that this power must be coming from somewhere! The only place it can come from in this system is the signal source which is driving the input current through the primary. Since we need to provide both volts & amps to transfer power this means that a voltage , $V_{in}\{t\}$, must be applied across the transformer's input (primary) terminals to set up the input current. The size of this voltage must be such that

$$V_{in}\{t\}\,I_1\{t\} = P\{t\} = V_r\{t\}\,I_2\{t\} \quad \ldots (5.10)$$

We can now use equation 9 to replace the currents in this expression & obtain the result

$$V_r\{t\} = V_{in}\{t\} \times \frac{N_2}{N_1} \quad \ldots (5.11)$$

This result is a very useful one. It means we can 'step up' or 'step down' the size of an alternating voltage by passing it through a transformer with an appropriate *turns ratio*, N_2/N_1. Note that when this is done the ratio of the input/output currents also changes by the 1/(same ratio). This is necessary as the transformer can't, itself, generate any electrical power. All it does is *transform* the power provided from the source. By choosing a suitable ratio, however, we can arrange for the required input a.c. voltage to drive the diode which then rectifies this to obtain the desired d.c. level.

5.3 A better PSU

The simple arrangement shown in figure 5·1 isn't particularly effective. There are a range of other, more complex arrangements which provide a more stable d.c. level or more power. Here

we will just look briefly at a few of the many improvements that can be made. The first, and most common of these is the use of a *Full-Wave Rectifier Bridge*. This is shown in figure 5·4. Comparing figure 5·4 with 5·1 we can see that the change is that we now use four diodes instead of one. These diodes act to 'steer' the input as follows:

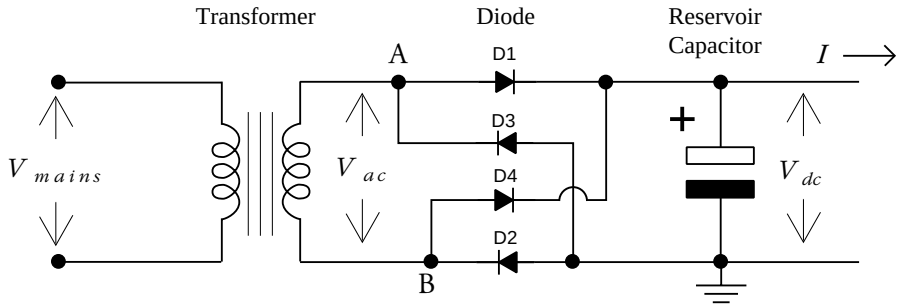
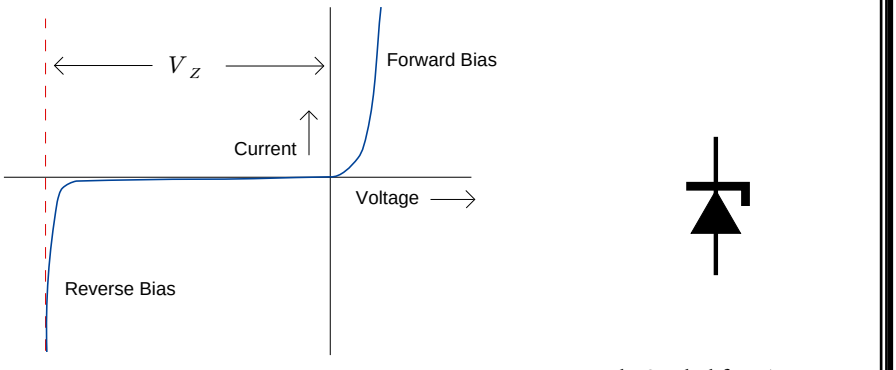


Fig 5·4 — Improved PSU with Full-Wave Rectifier bridge

As with the simple arrangement, the reservoir capacitor is charged whenever the input a.c. is large enough. However this can now happen in two ways which we can understand by considering the relative voltages, V_A and V_B at the two points, 'A' and 'B' shown on figure 5·4

- When $V_{ac} \rightarrow V_{peak}$ we will find that $V_A - V_B > V_{dc}$. This permits the pair of diodes, D1 & D2, to conduct, thus recharging the capacitor.
- When $V_{ac} \rightarrow -V_{peak}$ we will find that $V_B - V_A > V_{dc}$. This permits the pair of diodes, D3 & D4, to conduct, recharging the capacitor.

The result is the capacitor can now be recharged **twice** during every input a.c. cycle. So when the input is a 50Hz sinewave we recharge and 'top up' the capacitor's voltage 100 times per second (i.e. every 10 milliseconds). This is better than the simple one diode supply which can only top up the reservoir every 20 milliseconds. Since the time between recharges is halved we can expect the amplitude of the ripple to be halved. Hence the result is a smoother d.c. output. Since diodes are quite cheap, it is usual to use a full-wave rectifier. The only drawback is that the charging current flowing from the transformer to the capacitor now has to pass through two diodes in series. Hence the 'drop' in voltage lost across the diodes is doubled from around 0·5 V to about 1 V.



5·5a Zener IV curve

5·5b Symbol for a Zener

Fig 5·5 — The Zener Diode

The second commonly employed improvement is the use of a *Zener Diode*, or some similar device. The Zener has a typical IV curve whose shape is shown in figure 5·5a. The device is engineered to suffer a 'breakdown' when a large enough reverse bias voltage is applied. By suitable semiconductor design and manufacturing the voltage at which this occurs can be set for a specific diode to be at almost any level, V_Z, from around a couple of volts to many tens of volts. The breakdown can be quite 'sharp'. As a result we can approximate the Zener's properties in terms of three rules:

1 – In forwards bias (not normally used for a Zener) the behaviour is like an ordinary semiconductor diode.

2 – In reverse bias at voltages below V_Z the device essentially refuses to conduct, just like an ordinary diode.

3 – In reverse bias, any attempt to apply a voltage greater than V_Z causes the device to be prepared to conduct a very large current. This has the effect of limiting the voltage we can apply to around V_Z .

We can therefore build a circuit of the kind shown in figure 5·6 and use it to suppress power line voltage variations. Note that in this circuit I have used the standard symbol for a full-wave diode rectifier bridge. This arrangement of four diodes is so commonly used that it can be bought with the four diodes in a single package and has its own symbol as shown.

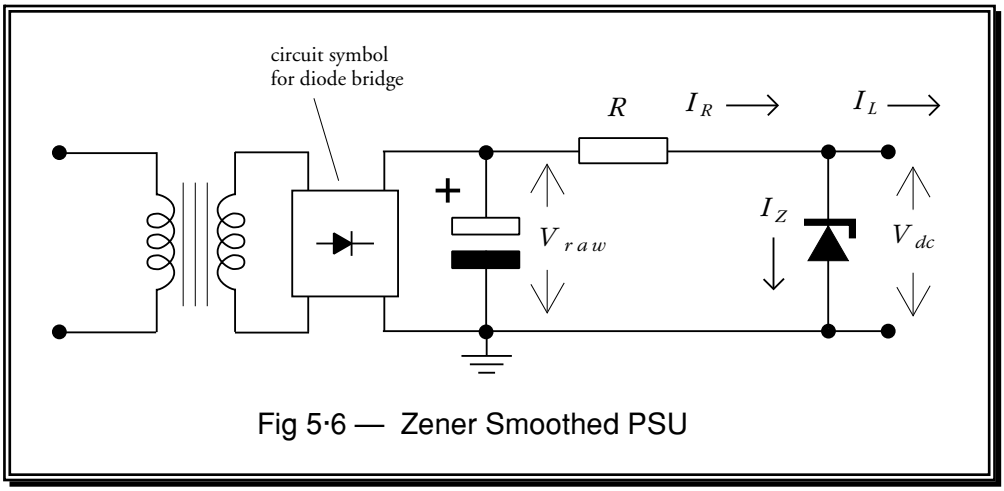


Fig 5·6 — Zener Smoothed PSU

The effect of the Zener in this circuit can be explained as follows. Let us assume that the d.c. level, V_{raw}, provided by the bridge and reservoir capacitor is greater than the chosen Zener's breakdown voltage, V_Z. This means that the series resistor will see a potential difference between its ends of $V_{raw} - V_Z$ and we find that the output voltage, $V_{dc} = V_Z$ almost irrespective of the choice of the resistor value, R, and of V_{raw}. When no load current, I_L, is being drawn by anything connected to the supply, we find that the current through the resistor and hence the Zener will therefore be

$$I_R = I_Z = \frac{V_{raw} - V_Z}{R} \quad \text{when} \quad I_L = 0 \qquad \ldots (5.12)$$

Now consider what happens when the attached load draws a current in the range $0 < I_L < (V_{raw} - V_Z)/R$. In effect, the current arriving via the resistor, R, is now 'shared' by the Zener and the load. This means that there will still be a non-zero current for I_Z, hence it maintains the output level at about $V_{dc} = V_Z$. As a result, provided we don't attach a load that tries to draw a current larger than I_R, the voltage across the Zener will be 'held' at the same value, V_Z. In practice, the action of the Zener is just to draw whatever current is required in order to

ensure that the voltage drop across R will be whatever is needed to ensure that the voltage across the Zener is V_Z.

Looking back at Figure 5.5 can see that the Zener does not actually maintain exactly the same voltage when we allow the current it takes to vary, but the variation is relatively small provided we ensure the Zener current does not fall too close to zero. When designing a Zener controlled PSU in more detail we could use a value for the *dynamic resistance* of the chosen Zener in order to take this into account. This dynamic resistance value is a measure of the slope of the reverse bias breakdown part of the curve shown in figure 5.5. Lists of components will generally give a value for this resistance as well as the nominal Zener voltage, and power handling, for each Zener diode. We would then use this data to select the Zener required for a given purpose.

If we allow the current drawn by the load to exceed the current level set by expression 5.12 the situation will change. Once $I_L > I_R$ there will be "no current left over" for the Zener to draw. In this situation the zener has no effect. The current demanded by the load must come via the resistor, R. Hence we now have a situation where the current through the resistor will be I_L and have a value greater than the value of I_R we would calculate from expression 5.12. From Ohm's Law. this increase in resistor current means that the voltage across the resistor must now exceed $V_{raw} - V_Z$. Unless we have arranged for the input, V_{raw}, to rise, this means that V_{dc} must now be less than V_Z. Any further increase in the current demanded by the load causes a further fall in the output voltage. As a result, at these high load currents the zener has no effect and we just see the system behaving in a way set by the presence of the resistor, R, in series with the raw d.c. level. The overall behaviour can be represented by the graph shown in figure 5·7.

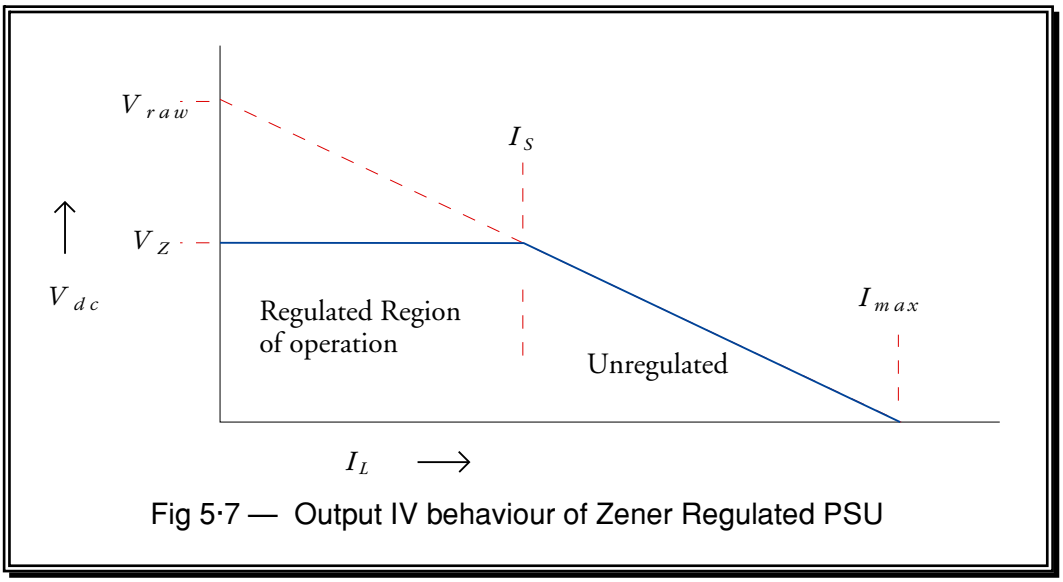


Fig 5·7 — Output IV behaviour of Zener Regulated PSU

We can divide the operation into two areas – regulated and unregulated. In the regulated area the output voltage is controlled by the action of the Zener, and in the unregulated region it is not. In the regulated region we have

$$V_{dc} = V_Z \quad : \quad 0 < I_L < I_S \qquad \text{... (5.13)}$$

$$V_{dc} = V_{raw} - I_L R \quad : \quad I_S \leqslant I_L < I_{max} \qquad \text{... (5.14)}$$

where I_S represents the maximum regulated current value

$$I_S \equiv \frac{V_{raw} - V_Z}{R} \qquad \text{... (5.15)}$$

and the maximum possible available unregulated current is

$$I_{max} \equiv \frac{V_{raw}}{R} \qquad \ldots (5.16)$$

In principle, a Zener regulated supply works well provided that we do not attempt to demand a load current greater than I_S. Within limits, it permits us to hold a steady output level even when the demanded current varies. It also tends to 'reject' any ripple. A small change in the voltage, V_{raw}, will alter the value of I_s by a corresponding amount. However provided that I_s always remains greater than the current demanded by the load, the Zener will still take some current and effectively hold the output voltage at a steady level. Hence the Zener helps to suppress unwanted ripple. Although very useful, the circuit shown in Figure 5.6 does have one serious drawback in situations where we require a significant output voltage and current. As an example to illustrate this, consider a situation where we require an output V_{dc} of 15 Volts and have to supply load currents at this voltage up to 2 Amps.

In order to ensure we can bias the system we can choose a value of V_{raw} a few volts greater than 15 V – lets take V_{raw} of 20 Volts for the sake of example. Since we require currents up to 2 Amps we have I_s = 2A. Using expression 5.15 this means we must choose a resistor, R = 2·5Ω. The first consequence of the way the circuit works is that the resistor always has to pass a current of at least 2 Amps, so must dissipate 5 Watts even when no output is required. The second consequence is that when no output (load) current is demanded all of this current then passes though the Zener, hence meaning it has to dissipate 15 × 2 = 30 Watts! This indicates that the arrangement shown in figure 5·6 suffers from the 'Class A problem'. i.e. It has a relatively large quiescent current and power dissipation level. The cure for this is actually quite simple. We can add a *Series Pass* transistor as shown in figure 5·8.

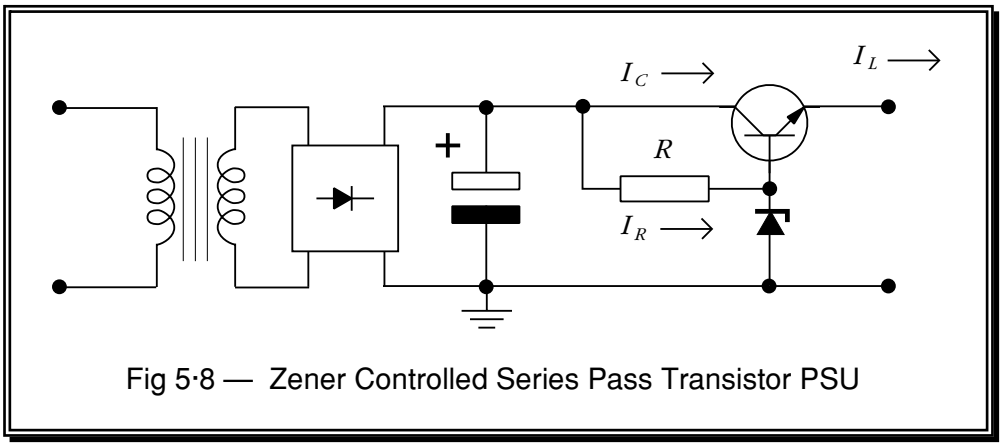


Fig 5·8 — Zener Controlled Series Pass Transistor PSU

Here the output from the resistor-Zener combination isn't used to directly supply the load current. Instead, it sets the voltage on the base of an output transistor. This means that when a given maximum current, I_S, is required by the load, the resistor-Zener only have to supply a maximum current of I_S / β where β is the current gain of the transistor. Given, say, a transistor with a current gain value of ×100 this means the quiescent current required through the resistor and Zener fall by a factor of 100. i.e. We now only need to pass 20 mA through the resistor and Zener to have a power supply capable of providing up to 2 Amps at a regulated voltage. The result is that the quiescent power dissipation in the PSU is 100 times lower than previously. Now, the transistor only passes the required current (and dissipates power) as and when the attached load demands it.

The above design is considerably more power efficient and practical than the simple Zener system. However it does suffer from the drawback that there is an extra 'diode drop' in the way

due to the base-emitter junction between the Zener and the output. This needs to be taken into account when designing a PSU for a specific output voltage. Most real PSU circuits are more complex than is shown in figure 5·8 but the extra complications are usually designed to improve specific aspects of the performance and the operation is basically as explained above. There are in fact a wide range of types of PSU including 'Switched Mode' types. Each have their particular good and bad points. However circuits of the type outlined above generally work well in simple cases where a reliable PSU is required.

Summary

You should now understand how the combination of a *rectifier* diode and a *reservoir* capacitor can be used to obtain a d.c. level from and a.c. input. How a transformer can be used to step up or down the a.c. level, and hence choose the desired size of input to the diode bridge. You should also know how a *Full-Wave Rectifier Bridge* improves performance by doubling the number of recharging opportunities per input a.c. cycle. You should also now understand how a Zener Diode can be used to Regulate the d.c. level and reduce any ripple and that the use of a Series Pass Transistor can improve the power efficiency and capability of the PSU.

Lecture 6 – Wires and Cables

6.1 Which way?

Plain old metal wires turn up a lot in electrical systems, electronics, and even in computing and optoelectronics. The metal patterns used for microwave 'striplines', and even the connections inside integrated circuits are essentially 'cables' or 'wires'. There is also a serious sub-industry as part of the Hi-Fi equipment business that sells fancy connecting cables, often at high prices. Given how many cables there are, and the wide range of tasks they perform, it is seems a good idea to look at their properties with some care.

In general, metal cables and wires serve two purposes.

- To carry electrical power from place to place.
- To carry information-bearing signals from place to place.

Now in order to communicate a signal we have to transfer some energy from place to place. This is because a signal without energy cannot have an effect upon a receiver. Hence it doesn't matter whether the wires are for simple power or signal transfer, power or energy has to be transferred in both cases. We therefore need to establish how this can take place.

To get an initial idea of what is going on, consider the situation shown in figure 6·1.

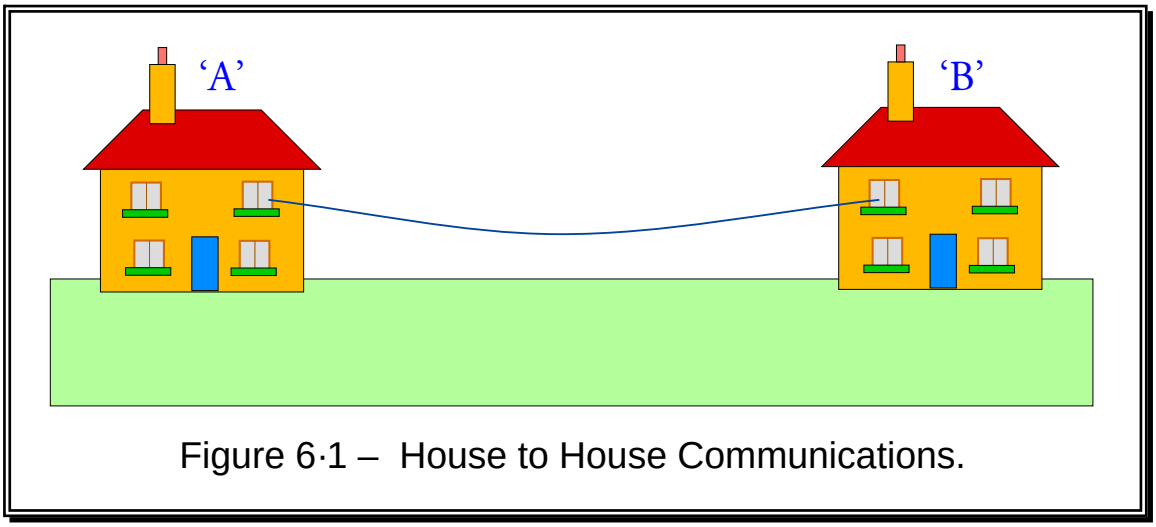


Figure 6·1 – House to House Communications.

This shows a wire hung between two houses. It is being used to send signals from one house to the other. The transmitter is a switch (Morse Key) and a battery, the receiver is a light bulb. (We are deliberately making this as simple as possible to avoid all the details of fancy signal communications equipment!) With stunning originality and imagination, we can call the houses 'A' and 'B'. (We could call them 'Dunpayin' or whatever, but that would just make the following equations longer to type!)

An eavesdropper wants to find out, "Which house are the signals being sent from?" Can he tell this by examining the signals on the wire? For his eavesdropping he just has a voltmeter and a current meter which he can attach. By using the voltmeter he can determine the potential

difference between the wire and the ground. Using the current meter he can determine which way any currents flow in the wire. Note that, as is usual, EM signals on wires actually require a conducting 'loop'. This provides a complete circuit for the current to flow around. In between the signal source and its destination this also means there are **two** conductors, and any applied voltage will appear as a potential difference between them. In this case one of the conductors is actually the ground (earth) upon which both houses sit. So the eavesdropper also measures voltages with reference to the ground. In the situation described he only notices the current in the wire and ignores the current flow in the ground.

Now $P = IV$ so for power to flow from house to house (i.e. from transmitter to receiver – or from battery to light bulb in this case) we require the product of the observed voltage and the observed current to be non zero. The voltmeter and ammeter used by the eavesdropper are 'center zero' types that show both a magnitude and a sign. The transmitter (battery) can be two possible locations, and can be arranged to apply either a positive or a negative potential to the wire, There are therefore four possible situations which may arise whenever the switch is closed and energy flows from battery to bulb. These are illustrated in figure 6·2.

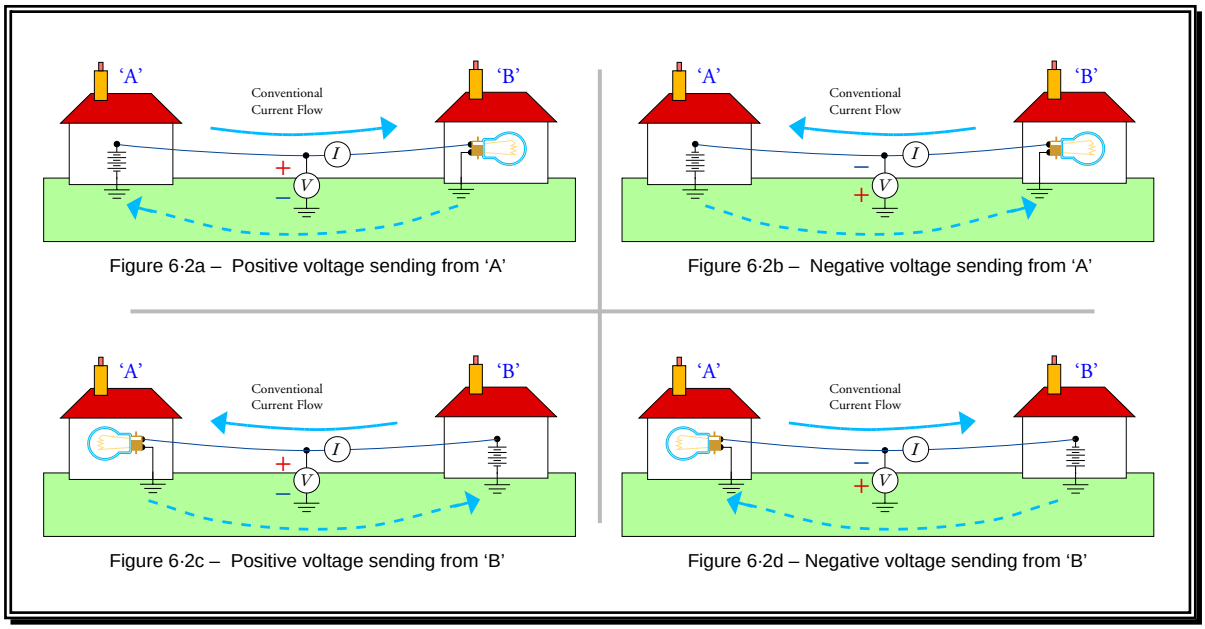


Figure 6·2a – Positive voltage sending from 'A'

Figure 6·2b – Negative voltage sending from 'A'

Figure 6·2c – Positive voltage sending from 'B'

Figure 6·2d – Negative voltage sending from 'B'

The results obtained by observing the meters in each case can be listed as follows if we define a flow of conventional current from A → B as having a positive sign we can draw up the following table of results.

Figure	Observed voltage	Conventional Current Flow	Sign of $P = VI$
6·2a	+ve	A → B (+ve)	+ve
6·2b	–ve	B → A (–ve)	+ve
6·2c	+ve	B → A (-ve)	–ve
6·2d	–ve	A → B (+ve)	–ve

By looking at this table we can see that we can't decide where the signal source is located simply by examining the current or voltage alone. However by considering their product (i.e. the power) we obtain a value whose sign tells us where the signal is coming from. This is because $V\,I$ tells us the rate and **direction** of energy flow along the wires. Although in this case we aren't noticing the current flow in the ground we would find that it would tell us the same thing if we could find the very low current density and voltage in the ground itself since both of these always have to opposite sign to the wire's.

6.2 Energy and Moving Electrons

In the above we only took an interest in the voltage (potential difference) and the current. The simple 'school textbook' model of electricity tends to describe electricity as being similar to water flowing in a pipe. In fact this isn't a very good way of looking at what is happening as it tends to hide some important features. To appreciate this, we can picture the electrons flowing along the wire as illustrated in figure 6·3.

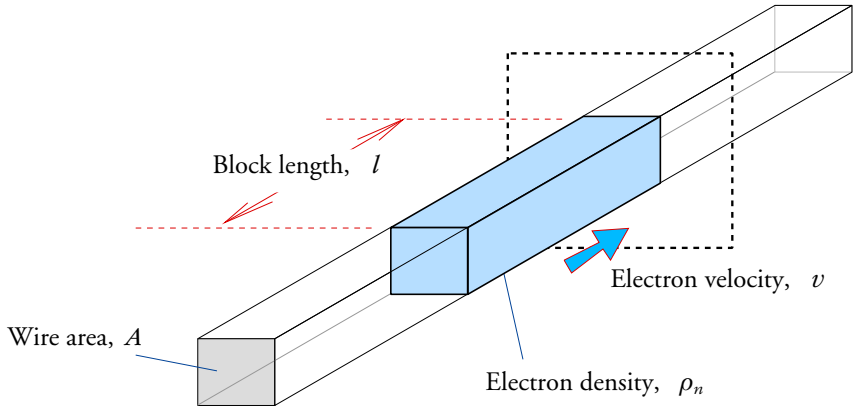


Figure 6·3 – Simple Model of Electron flow along a wire

Here we imagine the 'free' electrons that can provide the moving charges that constitute the current as having a uniform number density inside the material of ρ_n electrons per unit volume. The current is the result of these moving along the wire at an average velocity, v. We can now define the current to be

$$I \equiv \frac{n\,q}{t} \qquad \ldots (6.1)$$

where n is the number of electrons that cross a boundary cutting across the wire in a time, t, and q is the charge per electron. By measuring the current in Amps, and the charge in Coulombs, we can say that I Amps corresponds to $I\,/\,q$ electrons per second.

Since we know the density of the electrons in the wire we can turn this number into an effective volume of the electron distribution that will cross the boundary per second. This volume can be taken as a 'block' of the free charges which extends along a length, l, of the wire and covers its entire cross sectional area. Hence we can determine the length

$$l = \frac{I}{Aq\rho_n} \qquad \ldots (6.2)$$

Which implies that the free electrons are all tending on average to move this distance per second.

Hence we find that the mean velocity of the moving electrons is

$$v = \frac{I}{Aq\rho_n} \quad \ldots (6.3)$$

Lets now take an example of a copper wire with a square cross section of 1mm by 1mm.

Avogadro's Number tells us the number of atoms (or molecules) per gram-mol of the materal. The value of this number is $6{\cdot}0225 \times 10^{23}$. A 'mol' can be regarded as the atomic weight divided by the valancy value of the material. copper has an atomic weight of 63·54, and a nominal valancy of one. We can therefore say that 63·54 grams of copper will contain $6{\cdot}0225 \times 10^{23}$ atoms, and each on contributes one free electron. Thus each gram of copper will contain $6{\cdot}0225 \times 10^{23}/63{\cdot}54 = 9{\cdot}4783 \times 10^{21}$ free electrons. The density of copper is 8·95 grams/cc. So the above is equivalent to saying that each cubic centimetre of copper will contain $8{\cdot}4830 \times 10^{22}$ free electrons per cc – i.e. in S.I. units we can say that $\rho_n = 8{\cdot}4830 \times 10^{28}$ per cubic metre. The wire's area is 10^{-6} m^2 so taking a current of one Amp for the sake of example, we can use expression 6.2 to calculate a mean electron velocity of just under 0·075 mm/second.

This result is an interesting one for two related reasons. Firstly, it is many orders of magnitude less than the speed of light in vacuum. Hence we can immediately see that it doesn't correspond to the velocity of signals, or energy transfer, along metal cables. If it were simply the movement of the electrons that carried the signal we might have to wait a very long time for a reply when speaking over a transatlantic telephone cable! A less obvious reason becomes apparent if we now work out the kinetic energy of the above 'block' of electrons. Each has a mass, m_e, of the order of 9×10^{-31} Kg. So the kinetic energy passing the boundary per second will be

$$P_K = \frac{I m_e v^2}{2q} \quad \ldots (6.4)$$

which, using the above values come to $1{\cdot}5 \times 10^{-20}$ Watts.

Now it must be remembered that the above calculations are only rough, hand-waving, estimates of the correct values. (For example, the effective mass of an electron will change when it is in a material.) However, this result shows that the rate at which the electrons move, and the amount of kinetic energy they carry, can be much smaller than we might expect. For example, it is common in domestic house wiring for a current of an Amp or so to support power levels of a few hundred Watts – i.e. many orders of magnitude greater than the value calculated above. We therefore have to conclude that the real burden of the signal and energy transfer is being carried elsewhere, and **not** simply by the electrons in the wires.

6.3 EH Fields and Domestic Waveguides.

In fact, we can now reveal that it is the electromagnetic fields that surround metal wires that actually carry the signal energy. Here we can examine three standard cases, starting with one that looks like the 'house to house' system we looked at earlier. In each case it turns out to be the product of $E \times H$ that carries the power and the electrons are almost irrelevant except as a convenient place to 'pin' or 'control' the fields. The wires (more precisely, the electrons inside the wires) act to guide the fields, but it is the fields that do the real work!

In terms of electromagnetism, we can define the power flow in terms of the *Poynting Vector*

$$S \equiv E \times H \quad \ldots (6.5)$$

As a vector, this quantity has both a magnitude and a direction, so it in indicates both the rate of

energy flow, and the direction in which the energy travels. To determine the total flow from one place to another we would need to integrate the value of *S* over a suitable surface which is placed so that any path from place to place must pass through the surface. For wires, waveguides, etc, this usually means a plane surface perpendicular to the wires, located between the signal source and destination. We also should average the value over a suitable time (often one cycle of a periodic signal) to obtain a mean or average power value.

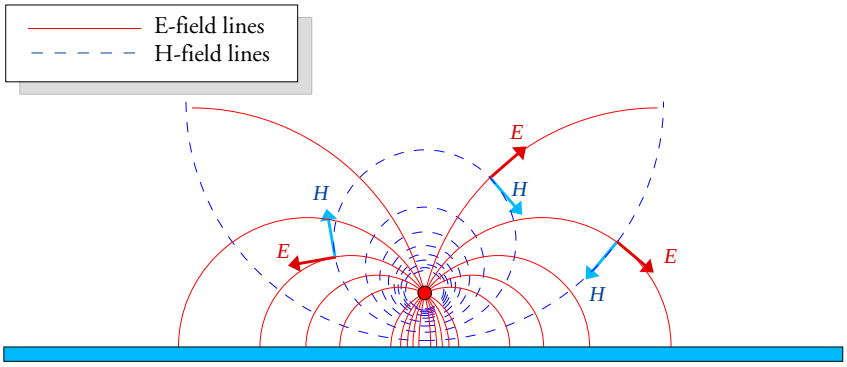


Figure 6·4 EH field around a wire near to a ground plane.

Figure 6·4 shows the E-field (solid lines) and H-field (broken lines) that are set up in the kind of arrangement considered in figures 6·1 and 6·2. Perhaps surprisingly, the ground tends to act as a fairly good conductor so far as the EH fields are concerned. Hence we usually obtain a pattern similar to that between a small wire and a flat conducting metal plane. (Which is often called a 'ground plane' for reasons that should now be obvious!) Since the metal is a good conductor, the E-field lines strike it normal to the surface, and the H-field at the surfaces is parallel to the surface. If we analyse the shapes in detail we discover the the field lines are actually all circles or arcs of circles.

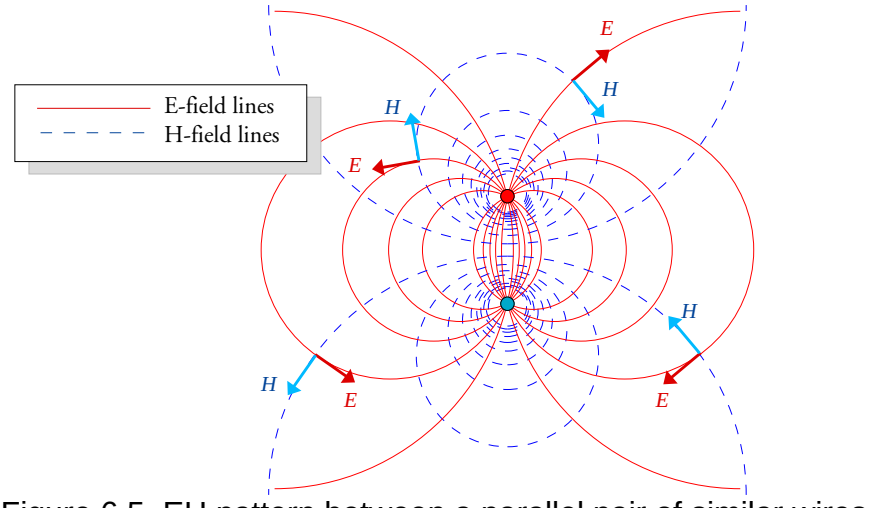


Figure 6·5 EH pattern between a parallel pair of similar wires

A good conductor acts like a 'mirror'. We can therefore say that the fields we see above the ground behave just as if there was an image of this pattern below ground level. Of course, it only looks like this when we are above ground. However it leads to the related result shown in figure

6·5 which shows the field between a parallel pair of wires used as a 'twin feeder' to convey signal energy.

For the illustrations the top wire in each case is assumed to be given a positive potential w.r.t. the ground plane or lower wire. The current flow is assumed to be away from the observer in the top wire, and towards the observer in the ground plane or lower wire. This then gives the E-field and H-field directions indicated by the arrows on the diagrams. By the right-hand rule we can see that everywhere the Poynting Vector – and hence the energy flow – is away from the observer. If we reverse both the potentials and the currents (as would happen if the signal source reversed the polarity of the signal they are applying) both E and H would change signs. Hence the Poynting Vector's sign would remain unchanged. The behaviour of the fields is therefore consistent with the currents and voltages we considered earlier. We can therefore use the Poynting Vector to show us the direction energy (and hence signal) flow.

If we ignore the 'Earth' wire, which is present (in the UK at least!) purely for safety reasons, normal house mains wiring is a form of twin feeder and acts essentially as a form of 'waveguide' to direct the electrical power from generating station, via National Grid, to the lights, TVs, etc, in our house. (In fact, at the risk of complicating things, long distance power transmission often uses three wires as this is more efficient, but we can ignore that here as being a detail.)

A pair of wires acting as a twin feeder is fairly cheap to make, and easy to use. It does, however suffer from two disadvantages. Firstly, the EH fields spread for some distance around the wires. As a consequence, any pieces of metal or dielectric near the wires will be in the field. Hence signal power may be lost by being coupled onto these objects, or the signal propagation behaviour altered by their presence. Secondly, at high frequencies the wires will tend to act like as an 'antenna' and may radiate some of the power rather than guiding it to the intended destination. To minimise these effects it is advisable to keep the spacing between the wires, and their diameters, as small as we can. It is also common to wind the wires around each other the wires together and make a 'twisted pair'. Ideally we want to keep other objects much further away than the wire spacing, and keep the wire spacing much smaller than the free-space wavelength of the highest frequencies that we wish to carry along the wires.

Unfortunately, meeting the above requirements can be difficult at times. Also, thinner wires are likely to have a larger resistance, so will waste signal power heating up the wires. To overcome these problems it is common to use alternative forms of wiring or guiding. The topic of signals guides and fibres is a complex one, so here we will only briefly mention one widely used solution to the above problems – the *Coaxial Cable*. Often called 'Co-ax' .

Coaxial Cable consists of two conductors, one of which surrounds the other. We can think of this as a variation on the wire-over-a-groundplane shown in figure 6·4. However we now bend the ground plane and 'wrap it around' the wire. The result is a wire inside a 'tube' of conductor with an insulating space in between. The two conductors usually have circular symmetry, and share the same long axis (hence the name). The space between them is also usually filled with a suitable dielectric material to keep the conductors apart and help the cable maintain its overall size and shape. The result has a typical form shown in figure 6·6. In this case the E-field is radial and the h-field is circumferential. As before, the power is carried by the EH fields. However unlike the previous examples, this field is now all 'inside' the cable. This means that the signal energy guided along the co-ax is unlikely to be affected by objects which come near to the outside of the cable.

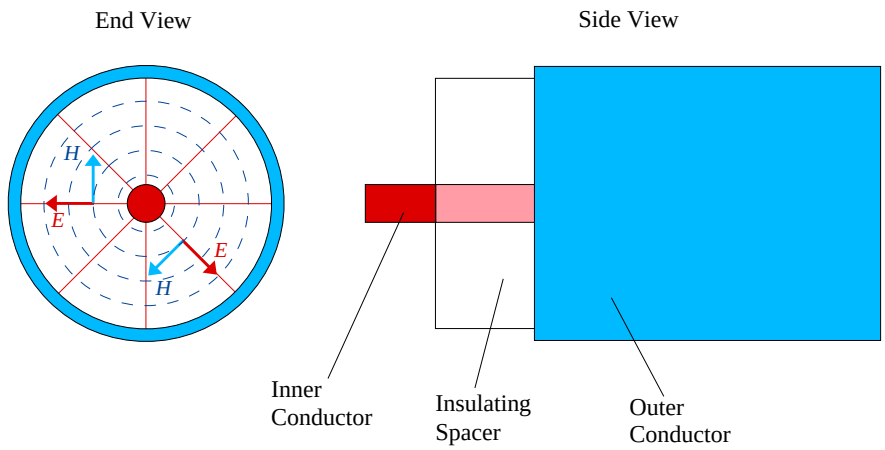


Figure 6·6 – Co-axial cable

The corollary of the above is that any unwanted signals (interference or noise) from elsewhere will now also be unlikely to inject itself onto the cables. Co-ax therefore falls into a general class of cables and guides which is said to be 'shielded'. It efficiently guides the power from place to place, with little chance of any of the signal being affected by the cable's surroundings. For this reason co-ax is particularly useful when we are dealing with low-power and/or high frequency signals. Hence it is used a great deal in radio systems and in sensitive measurement equipment. For similar reasons, many signal and radio cable connectors also have a co-axial form.

Summary

You should now understand the way in which wires and cables carry signals and signal energy in the form a an EH field pattern. You should also now know that although the charge carriers (free electrons) move in the conductors they only carry a tiny amount of energy, and move much slower than the actual signals. You should now be able to see the underlaying similarity of a wire over a return/ground plane, a pair of wires acting as a 'twin feeder', and a co-axial cable. The useful property of co-ax in keeping signals 'shielded' and preventing them from being lost should also be clear.

Lecture 7 – Transmission and Loss

7.1 Transmission Lines

Cables are generally analysed as a form of *Transmission Line*. The E-fields between the conductors mean that each length of the pair of conductors has a capacitance. The H-fields surrounding them mean they also have inductance. The longer the cable, the larger the resulting values of these might be. To allow for this variation with the length, and also take into account the fact that signals need a finite time to pass along a cable, they are modelled in terms of a given value of capacitance and inductance 'per unit length'. A run of cable is then treated as being a series of incremental (i.e. small) LC elements chained together as shown in figure 7·1. (Note that I have identified these as L' and C' and use the primes to indicate that they are values-per-unit-length.)

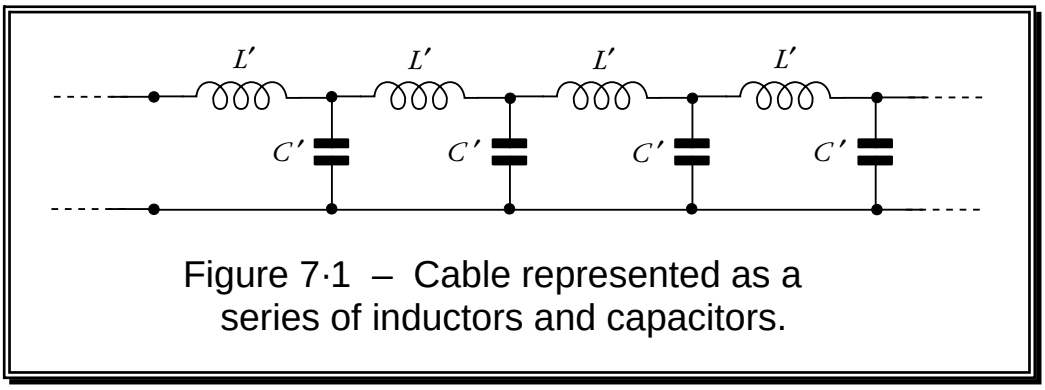


Figure 7·1 – Cable represented as a series of inductors and capacitors.

The amount of capacitance/metre and inductance/metre depends mainly upon the size and shape of the conductors. The inductance relates the amount of energy stored in the magnetic field around the cable to the current level. The capacitance relates the amount of energy stored in the electric field to the potential difference between the conductors. For co-ax and twin feeders the values are given by the expressions in the table below, where the relevant dimensions are as indicated in figure 7·2.

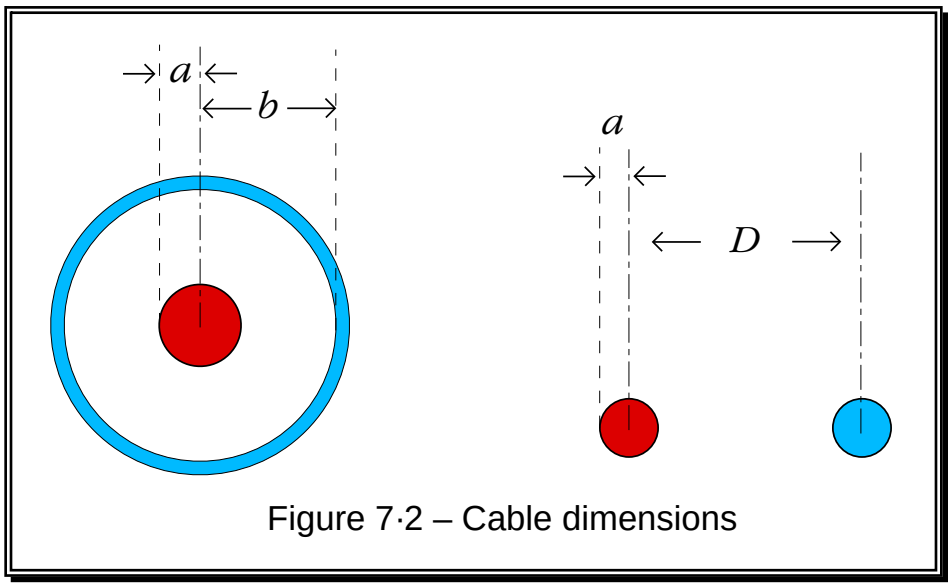


Figure 7·2 – Cable dimensions

	Capacitance F/m	Inductance H/m	Impedance Ω
Co-axial cable	$\frac{2\pi\varepsilon_r\varepsilon_0}{\ln\{b/a\}}$	$\frac{\mu_r\mu_0 \ln\{b/a\}}{2\pi}$	$\frac{138}{\sqrt{\varepsilon_r}} \ln\{b/a\}$
Twin Feeder	$\frac{\pi\varepsilon_r\varepsilon_0}{\text{Arccosh}\{\frac{D}{2a}\}}$	$\frac{\mu_r\mu_0 \text{Arccosh}\{\frac{D}{2a}\}}{\pi}$	$\frac{276}{\sqrt{\varepsilon_r}} \text{Arccosh}\left\{\frac{D}{2a}\right\}$

Note that when $x \geqslant 1$ we can say that $\text{Arccosh}\{x\} = \ln\{x + \sqrt{x^2 - 1}\}$. Since we require the wires not to touch, this condition will alway be true for a twin feeder, so we can use an alternative form for the above expressions. Some textbooks assume that $D \gg a$. Since when this is the case, $\text{Arccosh}\{x\} \rightarrow \ln\{2x\}$, they then give a simplifier set of expressions for the properties of a twin feed where the above $\text{Arccosh}\{D/2a\}$ terms are replaced with $\ln\{D/a\}$.

For cables of the kinds considered here, we can make a few general comments which apply whatever their detailed shapes, sizes, etc.

The *Characteristic Impedance* depends upon the ratio of the values of the capacitance per metre and inductance per metre. To understand its meaning, consider a very long run of cable that stretches away towards infinity from a signal source. The source transmits signals down the cable which vanish off into the distance. In order to carry energy, the signal must have both a non-zero current, and a non-zero potential. (i.e. both the E-field and the H-field must exist and propagate along, guided by the cable.) Since the far end is a long way away, the signals transmitted from the source can't initially be influence by the properties of any destination before they finally arrive. Hence the E/H ratio of the field carried along the cable (and hence the current/voltage ratio) are determined solely by properties of the cable. The result, when the signal power vanishes, never to be seen again, is that the cable behaves like a resistive load of an effective resistance set by the cable itself. This value is called the Characteristic Impedance, Z_0, of the cable. In general, for a loss-free cable, its value is given by

$$Z_0 = \sqrt{\frac{L'}{C'}} \qquad \ldots (7.1)$$

and the velocity with which the signal (and hence its energy) propagate along the cable is given by

$$v = \frac{1}{\sqrt{L'C'}} \qquad \ldots (7.2)$$

which, if we assume no magnetic materials is involved, will be equivalent to a velocity of

$$v = \frac{c}{\sqrt{\varepsilon_r}} \qquad \ldots (7.3)$$

where ε_r represents the relative permittivity of the insulating (dielectric) materials in which the conductors are nominally embedded.

As a matter of convention, there is a tendency for many co-axial cables to be designed and manufactured to have an impedance of either 75Ω (used by TV and VHF radio manufacturers) or 50Ω (used by scientists and engineers for instrumentation and communications). Engineers use a variety of type and impedances of cables. At RF, the use of 300Ω twin feed is fairly common, and 600Ω is often used at audio frequencies. Standard values tend to be adopted for convenience in a given application area as this makes it easier for people to build systems from compatible elements.

7.2 Dispersion and Loss

If you look in books on microwaves, RF, or optics, you'll see that there are many kinds of transmission line. Although they all share the same basic properties as the kinds of wires and cables we are considering here their detailed behaviour is often more complex. Here we'll ignore forms of waveguide that only work at very high frequency, but have a look at the problems that can arise when using 'conventional' types of cables and wires to carry low frequency signals. This topic is of interest to scientists and engineers who need to be alert for situations where the signals may be altered (or lost!) by the cables they are using as this will compromise our ability to make measurements and carry out tasks. Hi-Fi Audio fans also take a keen interest in the effects cables and wires might have upon signals as they fear this will alter the sounds they wish to enjoy!

The main imperfections of conventional wires and cables arise from two general problems. Firstly, the materials used may not behave in the 'ideal' way assumed so far. Secondly, the cables may be used incorrectly as part of an overall system and hence fail to carry the signals from source to destination as we intend. Here we can focus on materials/manufacturing problems of cables. The most obvious of these is that real metals aren't perfect conductors. They have a non-zero resistivity. The signal currents which must flow in the wires when signal energy is transmitted must therefore pass though this material resistance. As a result, some signal power will tend to be dissipated and lost, warming up the metal.

A less obvious problem is that the dielectric material in between the conductors will also tend to 'leak' slightly. Dry air has quite a high resistivity (unless we apply many kV/metre when it will break down and cause a spark), but high isn't the same as infinite. In addition, all real dielectrics consist of atoms with electrons. The electrons are tightly bound to the nuclei, but they still move a little when we apply an electric field. Hence dissipation will also occur when we apply a varying field to a dielectric as work has to be done moving these bound electrons. This can be regarded as a non-zero effective a.c. conductivity of the dielectric.

As a consequence of the above effects a better model of a cable is a series of elements of the form shown in figure 7·3.

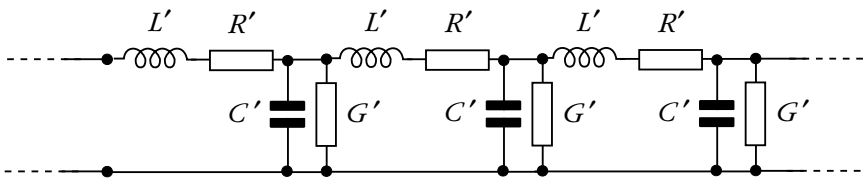


Figure 7·3 – More realistic cable model.

Here the dissipative loss due to conductor resistance is indicated by a resistance-per-length value of R' and loss due to the dielectric is represented by a conductance-per-length value of G'. As before, here I have used primes to emphasise that the values are per unit length and avoid confusion with total values or those of specific components.

The presence of the losses alters the propagation behaviour of the cables. The characteristic impedance and signal velocity now become

$$Z_0 = \sqrt{\frac{j\omega L' + R'}{j\omega C' + G'}} \qquad \ldots (7.4)$$

$$v = \frac{\omega}{\mathrm{Im}\left[\sqrt{(R' + j\omega L')(G' + j\omega C')}\right]} \qquad \ldots (7.5)$$

where ω is the (angular) signal frequency and Im [] means the imaginary part. Note that, unlike the case for a 'perfect' cable, the impedance and velocity will now usually be frequency dependent. The frequency dependent cable velocity may be particularly significant when we are transmitting broadband signals as the frequency components get 'out of step' and will arrive a differing times. This behaviour is called *Dispersion* and is usually unwanted as it distorts the signal patterns as they propagate along the cables. Alas, all real cables will have some resistance/ conductance so this effect will usually tend to arise. However, the good news is that **if** we can arrange that

$$\frac{G'}{C'} = \frac{R'}{L'} \qquad \ldots (7.6)$$

then the two forms of loss have balancing effects upon the velocity and the dispersion vanishes. Such a cable is said to be 'distortionless' and the above requirement which has to be satisfied for this to be true is called *Heaviside's Condition*, named after the first person to point it out.

The rate of energy or power loss as a signal propagates down a real cable is in the form of an exponential. For a given input power, P_0, the power that arrives at the end of a cable of length, Z, will be

$$P = P_0 \,\mathrm{Exp}\{-\alpha Z\} \qquad \ldots (7.7)$$

where α is defined to be the *Attenuation Coefficient* whose value is given by

$$\alpha = \mathrm{Re}\left[\sqrt{(R' + j\omega L')(G' + j\omega C')}\right] \qquad \ldots (7.8)$$

where Re []indicates the real part. For frequency independent cable values the above indicates that the value of α tends to rise with frequency. This behaviour, plus some dispersion, are usual for most real cables. The question in practice is, "How big are these effects, and do they matter for a particular application?" Not, "Do these effects occur?"

The actual values of R' and G' depend both upon the materials chosen for the cables, and the sizes and shapes of the conductors. In general, larger cables tend to have a greater cross section of metal, and a wider dielectric gap, and hence can have lower loss, but this rule can easily be broken if one cable is manufactured with more care than another. The most common metal for cables tends to be copper, but aluminium and steel are also used quite often. Although aluminium and steel have a lower conductivity than copper they are cheaper, aluminium is lighter, and may resist tarnishing ('rusting') better. Silver, graphite (carbon), and even gold are used in special cases. Typical values for the resistivity of some of these materials are shown below.

	Copper	Aluminium	Silver	Graphite	Gold
nanoΩm	16·9	26·7	16·3	13,750·0	22·0

(In practice, most cable materials are not 'pure' and are alloys *with resistivity values which may differ from the above.)*

From the table, silver would seem to be the best choice as it has a very low resistivity and hence would be expected to minimise signal losses. Alas silver suffers from two drawbacks. It is very expensive, and tarnishes very easily. Since a surface tarnish nearly always degrades electrical performance this means it should only be used in situations where the surface can be protected from the atmosphere – e.g. by a coating. Copper is much cheaper, and has almost as low a resistivity as silver, hence its popularity, although it, too, will tarnish.

The most commonly used dielectrics are: air (which has the obvious advantage of being free!), polyethylene, ptfe, and pvc. Their main properties are summarised in the table below. Note that is customary to specify the a.c. losses of a dielectric in terms of a value which is called the *Loss Tangent* ($\tan\delta$) or *Power Factor* (F_p). This is a dimensionless number based upon the ratio of the real and imaginary parts of the material's dielectric constant.

	Dry Air	Polyethylene	PTFE	PVC
ε_r	1·0006	2·2	2·1	3·2
$\tan\delta$	low	0·0002	0.0002	0.001
Ωm (d.c.)	high	10^{15}	10^{15}	10^{15}
Breakdown (MV/m)	3	47	59	34

(*The compositions of all these materials varies from sample to sample, so the above values should be only taken as typical.)*

To take these values into account we can define an effective shunt conductivity per unit length value of

$$G' \equiv G'_0 + \omega C' \tan\delta \qquad \ldots (7.9)$$

and substitute this as required into the earlier expressions for the cable impedance, velocity, and attenuation coefficient.

Looking back to the table of resistivities the inclusion of graphite seems curious. It has a resistivity about 1000 times greater than the other metals that are often used as conductors, hence seems a rather poor choice. In fact, there is a specific reason why a higher conductivity can sometimes be very useful. This is linked to what is referred to the as the *Skin Effect.*

The Skin Effect arises when EM waves are incident upon, or are guided by, conducting surfaces. The E-fields set up currents in the surface and hence the field only penetrates for a finite distance. This in turn means that the currents only exist near the surface. In practice, in a 'thick' conductor the current level falls exponentially with the depth below the metal surface. The result is that the currents on conductors associated with a guided field only make use of a finite metal thickness. Hence the resistance experienced by the currents (which leads to dissipation losses) is influenced by this thickness as well as the material's resistivity. The magnitude of the current falls exponentially with a 1/e scale depth given – for good conductors – by the approximate expression

$$d \approx \frac{1}{2 \times 10^{-3} \times \sqrt{f\sigma}} \qquad \ldots (7.10)$$

where f is the signal frequency, and σ is the conductor's conductivity. (This expression assumes the conductor is 'non-magnetic' – i.e. it has a $\mu_r = 1$.)

The significant point is that this thickness depends upon the signal frequency as well as the conductivity. The table below gives the skin depth values for copper and graphite at a series of frequencies.

	100Hz	10kHz	1MHz	100MHz	10GHz
d copper	6·5 mm	0.65 mm	65 µm	6·5 µm	0·65 µm
d graphite	185 mm	18·5 mm	1·85 mm	0·185 mm	18·5 µm

In general, the wires in many cables tend to have thicknesses or diameters of the the order of 0·1mm to 1mm. Hence when made of copper, we find that the full bulk of the material is only used to support signal conduction for frequencies up to around 100 kHz. At higher frequencies, the cross-sectional area of the material where conduction occurs tends to fall with increasing frequency. The result is that the dissipation resistance per length, R', tends to rise with frequency. This frequency dependence affects both the signal velocity – causing dispersion distortions – and preferentially attenuates high frequencies – causing a change of the amplitude spectrum. Hence it is to be avoided when possible. There are two main ways to avoid these problems.

The first is to use a material of higher resistivity, such as graphite. This means that a conductor of the same size as before will tend to keep a fairly uniform R' value to a frequency around 1000 × higher than when using copper. The drawback being that the losses will be higher (but more uniform) at lower frequencies. The second method (which is sometimes combined with a change in material) is to use many fine, individually insulated wires in a 'woven' arrangement. This means that each individual wire has a small thickness, hence its entire bulk will tend to be used to a higher frequency than for a fatter wire. The snag is that, by having a smaller cross-section, it may have a high resistance. However, by using many such wires all used 'in parallel' we can reduce the overall resistance to a low level. For this kind of reason, many co-axial cables use 'braided' or woven meshes of many thin wires. Similarly, wound components and wires employed for 'rf' purposes sometimes use a woven construction called 'Litz wire' after its developer.

The woven arrangement for coaxial cables also has the effect of making them more flexible. This is convenient when we want to bend them to fit around corners. However it also increases their sensitivity to *Microphony*. This is an effect where any changes in external force (or atmospheric pressure) applied to the cable tends to alter its outer diameter, or shape. This will change its capacitance per length. Now the cable's conductors must be charged when there is a potential difference between them. Altering the capacitance can't immediately create or destroy charges, so the potential difference changes. The result is that the voltage between the inner and outer conductor changes in response to pressure.

In some cases a coaxial cable may be deliberately given a large steady charge, and then the voltage between its inner and outer conductors measured. The result is a linear 'capacitor microphone' which can be used to detect soundwaves, bending of the wire, or even footfalls, anywhere along the cable. A particular use is in the security industry where such a system can be used as an intruder detector in any location where the wires are laid. In other situations this microphonic behaviour is a potential problem. For example, it would be undesirable in cables used as part of a home hi-fi as it would mean that delayed sound signals might be picked up radiating from loudspeakers when music is being played at high level. This could then be re-amplified and appear as a form of 'echo' or acoustical feedback, altering the total sound pattern being produced.

7.3 A short run home

Although much of what has been discussed applies equally to both very long and very short lengths of cable, there are a number of useful simplifying approximations we can make when dealing with the most usual cases where cable lengths are short. In electronic terms we have to define 'short' and 'long' here relative to the wavelength (along the cable) of the highest frequency in use in the signals of interest. To some extent this is a matter of judgement but a useful rule of thumb is to check if the cable length is at least a couple of orders of magnitude less than a quarter-wavelength. If it is, we can often treat the cable as 'short'. If not, it may be best to regard it as 'long' and employ the full metal jacket of transmission line theory.

As an example, consider a 1 metre long co-axial cable for use as part of an audio system. Is this cable 'short'? Well, depending on how good your ears are, you can probably hear sinewaves up to a frequency between 15 kHz and 25 kHz. If we take the upper frequency and assume a velocity along the cable similar to that of EM waves in free space we obtain a length for a quarter-wavelength at 25 kHz of approx 3 km. Hence a 1 m 'audio interconnect' is more than three orders of magnitude shorter, so we can treat it as being 'short' for most purposes. This means that instead of regarding it as a series of incremental lengths which pass the signal along we can treat it as one 'lumped' section placed between the signal source and destination as shown in figure 7·4.

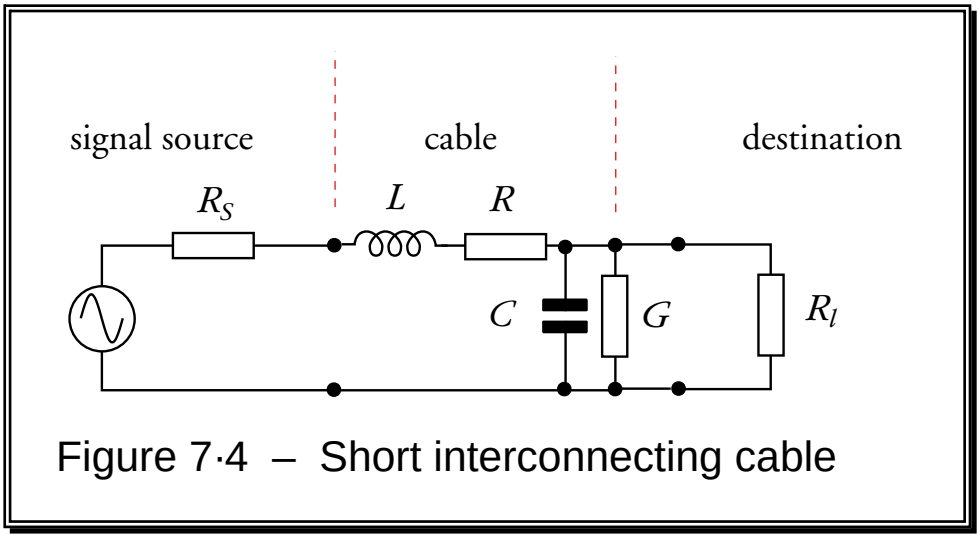


Figure 7·4 – Short interconnecting cable

Where for a cable of length, z, we can simply say that

$$L = zL' \quad : \quad R = zR' \quad : \quad C = zC' \quad : \quad G = zG' \qquad \ldots (7.11)$$

and analyse the effect of the cable as if it were a form of 'filter' placed between the signal source and destination. To do this we can use simple a.c. circuit theory although we may need to take into account the fact that R' and G' may be frequency dependent to some extent.

Summary

You should now understand how a cable can be modelled and analysed as a series of incremental lengths in the form of a Transmission Line. That this line may have losses due to physical imperfections of both the conducting and dielectric elements. You should know how the cable's *Characteristic Impedance* and the velocity of signal transfer depend upon the cable's electrical properties. That a cable may be *Dispersive*, and that dispersion may, in principle, be prevented by

arranging for the cable to satisfy *Heaveyside's Condition*. You should also understand how the *Skin Effect* can cause problems with wideband signals, but that this may be combatted using either woven/braided wiring and/or a change of conductor. Finally, you should now be able to see when a cable can be treated as a 'short' section and its effect assessed without the need for the full details of transmission line theory.

Lecture 8 – Op-Amps & their uses

Operational Amplifier ICs are very widely used in analog signal processing systems. Their obvious use is as amplifier, however they can be used for many other purposes – for example in the active filters we considered in lecture 3. This lecture looks at the basic way Op-Amps work, considers the common types, and some typical applications.

8·1 Op Amp Circuits

The details of the circuitry inside an op-amp integrated circuit can be very complex, and can vary from type to type, and even from manufacturer to manufacturer. They can also make use of semiconductor effects and constructions that you don't normally encounter outside of an IC. For simplicity, here we will just look at a simplified version of the design. The basic circuit arrangement is shown in figure 8·1.

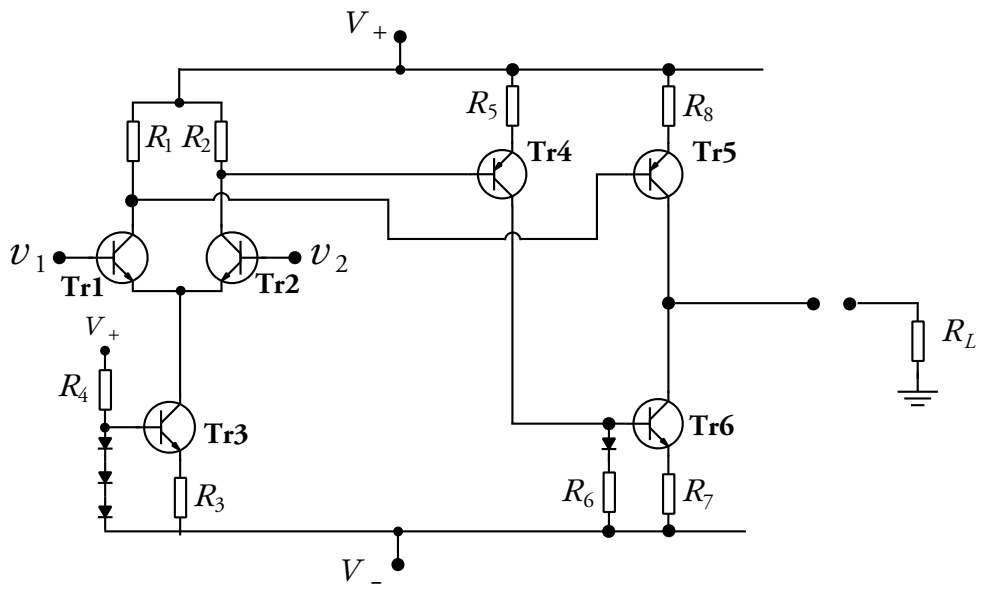


Figure 8·1 — Basic Op-amp Circuit

Tr1/2 form a *Long-Tailed Pair* differential amplifier for which **Tr3** provides a *Constant Current* source for their common emitters. The input pair drive one of the output transistors (**Tr5**) directly. The other output transistor (**Tr6**) is driven via **Tr4**. Taken together **Tr5/6** form a *Class A* output stage. In this kind of circuit it us usual to have the values chosen so that $R_1 = R_2$ and $R_5 = R_6 = R_7 = R_8$, and to ensure that the transistors have similar gain values. For simplicity we can assume the transistors all have a very high gain so we can treat the base currents

as being so small that we can ignore them and assume they are essentially zero. We can therefore understand how the system works as follows:

When the two input are the same (i.e. when $v_1 = v_2$) the currents through R_1 and R_2 will be the same – each will pass an emitter and collector current of approx $1{\cdot}2 / (2R_3)$. This means that the current levels on the two transistors, **Tr4** and **Tr5**, will be almost identical. Since the same current flows though R_5 and R_6 it follows that the value of the potential difference across R_5 is the same as that across R_6. Now the diode in series with R_6 means that the voltage applied to the base of **Tr6** is one junction-drop higher than the potential at that end of the resistor. This counters the voltage drop between the base and emitter of **Tr6**. Hence the potential across R_7 is the same as that across R_6. i.e. we find that the potential across R_7 equals that across R_5.

In effect, the result of the above is that the potential across R_7 (and hence the current that **Tr6** draws) is set by the current through R_2. Similarly, the potential across R_8 (and hence the current provided by **Tr5**) is set by the current through R_1. When $v_1 = v_2$ the currents through **Tr5** and **Tr6** are the same. As a result when we connect a load we find that no current is available for the load, hence the amplifier applies no output voltage to the load, R_L.

However, if we now alter the input voltage so that $v_1 \neq v_2$ we find that the currents through the output devices **Tr5** and **Tr6** will now try to differ as the system is no longer 'balanced'. The difference between the currents through **Tr5** and **Tr6** now flows through the load. Hence an imbalance in the input voltages causes a voltage an current to be applied to the load. The magnitude of this current and voltage will depend upon the gains of the transistors in the circuit and the chosen resistor values. The system acts as a *Differential Amplifier*. When $v_1 > v_2$ the output will be positive and proportional to $v_1 - v_2$. When $v_1 < v_2$ the output will be negative but still proportional to $v_1 - v_2$ (which will now be negative).

This form of circuit is quite a neat one in many ways. It acts in a 'balanced' manner so that any imbalance in the inputs creates an equivalent imbalance in the output device currents, thus producing an output. Since the operation is largely in terms of the internal currents the precise choice of the power line voltages, V_+ and V_- , doesn't have a large effect on the circuit's operation. We just need the power lines levels to be 'large enough' for the circuit to be able to work. For this reason typical Op-Amps work when powered with a wide range of power line voltages – typically from a minimum of ±5 to ±15 Volts. In practice this usually means that the output is limited to a range of voltages a few volts less than the range set by the power lines. In most cases ±15 Volt lines are used, but this isn't essential in every case.

Due to the way the circuit operates it has a good *Common Mode Rejection Ratio* (CMRR) – i.e. any common (shared) change in both input levels is largely ignored or 'rejected'. Most real Op-Amps use many more transistors than the simple example shown in figure 8·1 and hence can have a very high gain. Their operation is otherwise almost the same as you'd expect from the circuit shown. This means they tend to share its limitations. For example, the pure Class A means that there is usually limit to the available output current (double the quiescent level) and that this can be quite small if we wish to avoid having high power dissipation in the Op-Amp. For a typical 'small signal' Op-Amp the maximum available output current is no more than a few milliamps, although higher power Op-Amps may include Class AB stages to boost the available current and power.

There are an enormous variety of detailed types of Op-Amp. Many include features like Compensation where an internal capacitance is used to control the open-loop gain as a function

of frequency to help ensure stability when feedback is applied between the output and the input(s). This is why common Op-Amp types such as the **741** family have a very high gain at low frequencies (below a few hundred Hertz) which falls away steadily at higher frequencies. Many Op-Amps such as the **TL071** family use FET input devices to minimise the required input current level. Although the details of performance vary, they all are conceptually equivalent to the arrangement shown in figure 8·1.

8·2 Op-Amps as Amplifier Stages

The obvious use of Op-Amp ICs is as signal amplifiers. In general terms these are of four main types. Inverting, Non-Inverting, Differential, and Buffers. We have already discussed some of these in previous lectures but figure 8·2 shows a comparison of the required circuits, all using the same basic Op-Amp. It is conventional to call the two inputs *Inverting* and *Non-Inverting* depending upon the sign of the resulting output and the gain when a signal is fed to the relevant input whilst the other is connected to zero volts (shown as an Earth symbol). In figure 8·1 v_1 was shown entering the non-inverting input and v_2 entering the inverting input. It is also conventional to label these inputs with a plus sign for the non-inverting input and a minus sign for the inverting input to indicate the sign of the gain that will be applied to the signal entering via that connection.

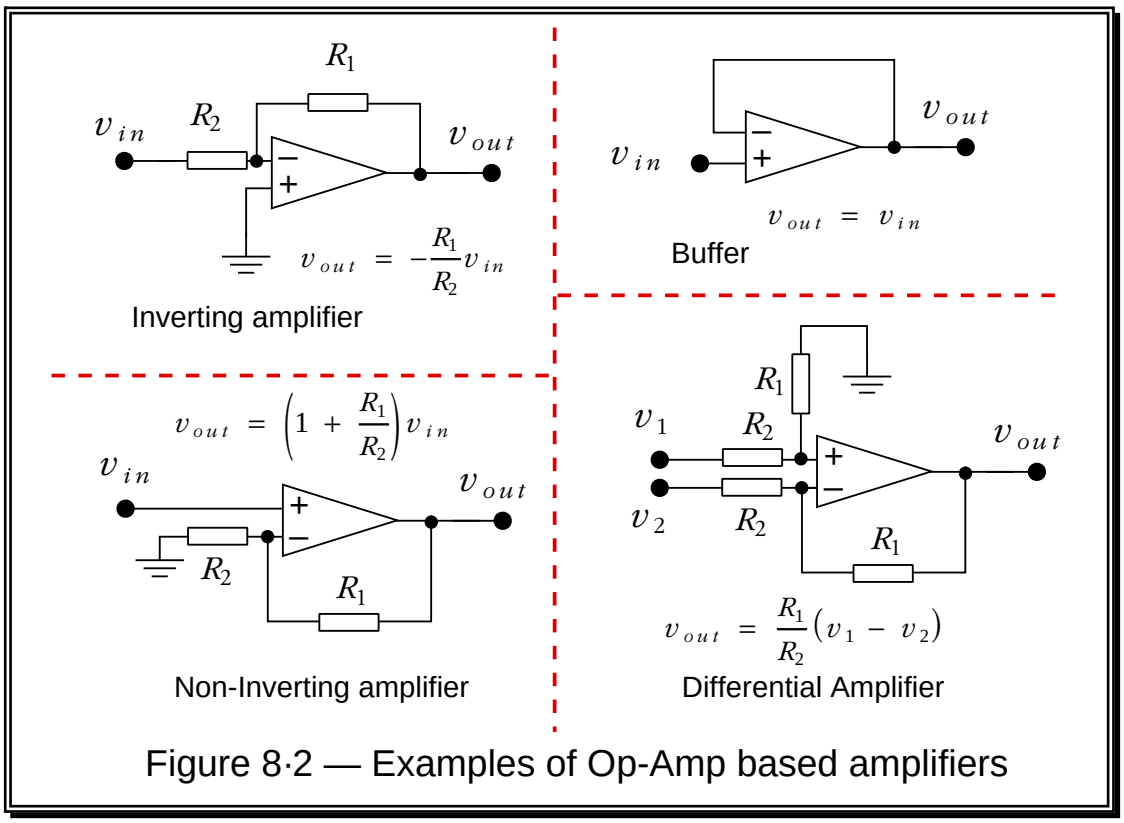


Figure 8·2 — Examples of Op-Amp based amplifiers

In each case the behaviour of the amplifier is controlled by the feedback from the output to the inverting input (i.e. the input where v_2 is shown to be applied in figure 8·1). Given an Op-Amp with a very high open-loop gain we can expect the voltages at the two inputs to always be very similar when the output is a modest voltage. e.g. If the amplifier has an open loop differential voltage gain of $A_v = 10^8$ (quite a common value) then when the output is, say, 10 Volts, the two inputs will only differ in voltage by 0·1microvolts.

In a circuit like the Inverting arrangement the non-inverting input is connected to Earth (nominally zero volts) directly. Hence the output always adjusts to keep the inverting input to the Op-Amp at almost zero volts (give or take a few microvolts). Hence the input to an inverting input connection sees a resistance (R_1 in these examples) whose other end is connected to a *Virtual Earth*, so sees an effective input resistance of R_1. Signals presented to a non-inverting input in the other arrangements see an input which the Op-Amp tends to adjust to almost equal the input. Hence the non-inverting arrangement has a very high input resistance. The output impedance of all the arrangements is very low provided we don't ask for more current than the Op-Amp can supply since the feedback tries always to assert the output voltage required.

8.3 Filters, Tone Controls, and EQ

Although often used to amplify signals, Op-Amps have many other uses. We have already seen in lecture 3 how an amplifier with a differential input can be used as part of a active filter. In principle, these active filters are just feedback amplifiers with highly frequency-specific feedback networks that manipulate the closed-loop gain as a controlled function of the signal frequency. in addition to the high/low/bandpass (or band reject) filters outlined in lecture 3 there are a number of other frequency-dependent functions which Op-Amps can be used to perform. Here we can take the example of *Tone Controls* circuits sometimes used in audio systems.

Although Tone Controls are now rarely provided in domestic Hi-Fi equipment, they still appear in some professional items and can be very useful in improving less-than-perfect source material. Their main task is to adjust the frequency response to obtain a more natural result. However they can also be useful for special purposes such as deliberately manipulating the sound. In scientific areas beyond Hi-Fi some form of adjustment of the frequency response can be very useful in '*pre whitening*' the spectrum of a signal. This means boosting some frequencies and attenuating others in order to obtain a spectrum which has a more uniform power spectral density. It allows recording or transmission systems to be used more efficiently and provide an optimum signal to noise performance.

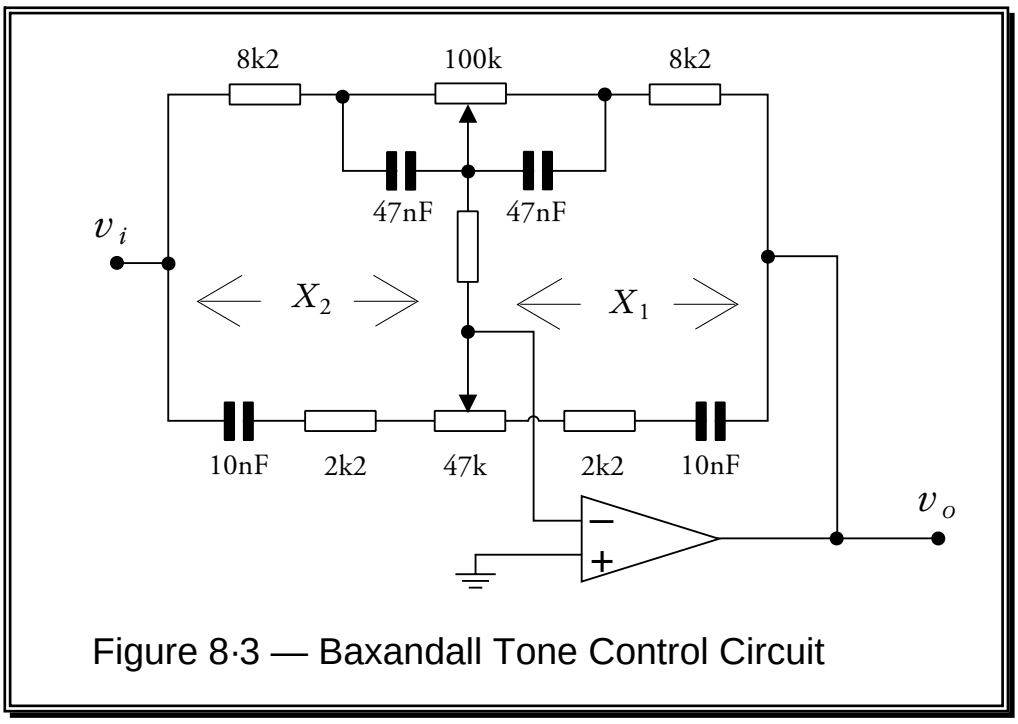


Figure 8·3 — Baxandall Tone Control Circuit

The classic form of tone control in Hi-Fi is the Baxandall arrangement shown in figure 8·3. This arrangement is called a Baxandall tone control, named after its inventor. We can understand how it works by noticing that it is actually a development of the non-inverting amplifier arrangement shown in figure 8·2. However the normal pair of feedback resistors have been replaced by quite complicated arrangements of resistance and capacitance. The circuit is laid out in a symmetric manner. In this case the impedance between the signal input and the inverting input of the amplifier is X_2 and the impedance between the output and the inverting input is X_1. The voltage gain is therefore now

$$A_v = -\frac{X_1}{X_2} \qquad \ldots (8.1)$$

where both X_1 and X_2 may be complex and have frequency-dependent values. However if we set both potentiometers to their central positions we find that despite being individually frequency dependent we get $X_1 = X_2$ at all frequencies. Hence when the pots are centered the frequency response is nominally flat and has a gain of −1. However if we move either potentiometer setting away from its central position we imbalance the system and produce a value of A_v which varies with frequency.

Consider first the effect of the upper pot (the 100kΩ one). At high frequencies the pair of 47 nF capacitors act as a short circuit and clamp the three wires of this 100k pot together. Hence adjusting the 100k pot does not alter the high frequency behaviour of the circuit. However at lower frequencies the impedance of the capacitors rises and the pot has some effect. Hence the 100k pot acts as a *Bass Control* and allows us to boost or cut the relative gain for low frequency signals. Now consider the lower pot (the 47kΩ one). Here the effect of the associated capacitors is reversed. The 10 nF capacitors mean that the arm of the circuit which contains the 47k pot essentially loses contact with the input and output at low frequencies. Hence the 47k pot has not effect upon low frequency signals, but it does upon high frequencies. It therefore acts as a *Treble Control* and can be used to boost or cut the relative gain at high frequencies.

Various forms of Tone adjustment can be applied. For example, the *Graphic Equaliser* circuits which are popular in studios and PA systems use a bank of bandpass filters to break the signal's frequency range into chunks. Each frequency section is then amplified and the results added (or subtracted) back together with various controlled gains to rebuild the overall signal. By altering the relative gains of the bandpass filters specific tonal bands can be boosted or cut to alter the sound. These complex circuits do reveal one of the main potential problems of Tonal adjustments. Any slight unwanted imbalances mean that it can be almost impossible to get a flat response should it be desired! For this reason, professional or high quality system use close tolerance components and usually have a 'defeat' switch that allows the signal to bypass the entire system when tonal adjustments are not required. Given the good quality of signals that are often available these days, tonal adjustment is usually only of value for special purposes or for reducing the severity of problems with historic recordings, or ones made incorrectly. There is therefore something of a 'theological' debate in Hi-Fi as to whether people should have, or use, such systems at all. Purists say not. Realists find them useful. As with most engineering this is a case of "Yer pays yer money and yer takes yer choice"!

8.4 Special Purposes

In the lectures so far we have seen circuits which show how an op-Amp could be used as part of a feedback amplifier or filter. In fact, Op-Amps have many other uses and we can give a few examples here just to illustrate the range of possibilities.

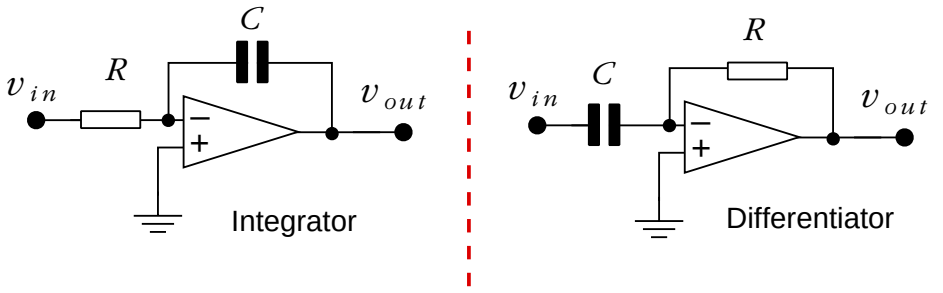


Figure 8·4 — Signal integration and differentiation

Figure 8·4 shows a pair of circuits which we can use to integrate or differentiate a signal value with respect to time. The *Integrator* acts to provide an output level proportional to the time-integral of the input level. It provides an output voltage at a time, t_1 given by

$$v_{out}\{t_1\} = v_{out}\{t_0\} - \int_{t_0}^{t_1} \frac{v_{in}\{t\}}{RC}\, dt \qquad \ldots (8.2)$$

where $v_{out}\{t_0\}$ represents the output voltage at some initial time, $t = t_0$, and $v_{out}\{t_1\}$ represents the output voltage at a later moment, $t = t_1$. Note that the output is actually proportional to **minus** the integral of the input value $v_{in}\{t\}$, and that there is a constant of proportionality, $1/RC$. This arrangement is useful whenever we want to integrate or sum over a series of values of a signal over some period of time. In practice we often arrange to add a switch connected across the capacitor and close this, to set $v_{out}\{t_0\} = 0$ and the instant we start the summing or integration. Signal integration is a very useful function in signal and data collection as we frequently wish to sum signal levels to improve a measurement by performing an average over many readings.

The *Differentiator* performs the opposite function. Here the output at some time, t, will be

$$v_{out}\{t\} = \frac{1}{RC} \times \frac{dv_{in}\{t\}}{dt} \qquad \ldots (8.3)$$

This allows us to observe the rate of change of a signal level. (Note that the Differentiator here is a completely different function to the Differential amplifier we considered in earlier lectures.) in both of the above circuits the scaling factor RC is called the *Time Constant* as it has the dimensions of time and is often represented by the symbol, τ.

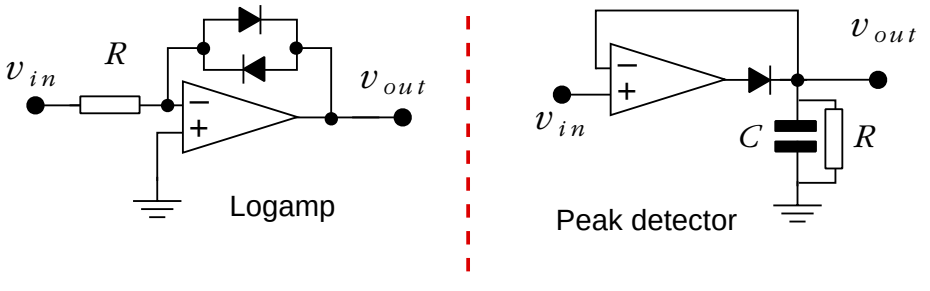


Figure 8·5 — Logamp and peak detector

Figure 8·5 shows two examples of circuits which combine an Op-Amp with diodes to perform some useful non-linear function. The *logamp* exploits the fact that the effective resistance V / I of a diode varies with the applied voltage. For a 'textbook' diode we can say that the current and voltage will be linked via an expression of the form

$$I = I_0\left(\mathrm{Exp}\left\{\frac{qV}{kT}\right\} - 1\right) \qquad \ldots (8.4)$$

where k is Boltzmann's constant, T is the absolute temperature of the diode, q is the charge on an electron, and I_0 is the saturation current value of the diodes chosen. In the logamp circuit shown in figure 8·5 a pair of diodes replace the usual feedback resistor.

Now the Op-Amp has a very high differential gain. This means that when it can it tries to adjust its output to keep its two input voltages very similar, and it also only draws a small input current. This means that the point where the resistor, R, and the diodes meet will be held almost precisely at zero volts to ensure that the voltage at the inverting input almost exactly matches the zero potential applied to the non-inverting input. The low current requirement means that almost none of any current passing through the other components will flow in or out of the Op-Amp's inputs. Hence we can say that the potential across the diodes will equal v_{out} and the potential across R will equal v_{in}. We can also expect the current in the pair of diodes to equal that through .

Since the diodes are connected in parallel, but facing opposing ways, the total current they pass when the output voltage is v_{out} will be

$$I = I_0\left(\mathrm{Exp}\left\{\frac{qv_{out}}{kT}\right\} + \mathrm{Exp}\left\{-\frac{qv_{out}}{kT}\right\} - 2\right) \qquad \ldots (8.5)$$

Whereas the current through R will be v_{in} / R. Putting these to be equal we therefore find that

$$v_{in} = RI_0\left(\mathrm{Exp}\left\{\frac{qv_{out}}{kT}\right\} + \mathrm{Exp}\left\{-\frac{qv_{out}}{kT}\right\} - 2\right) \qquad \ldots (8.6)$$

Clearly under most circumstances this would be a truly awful choice for an amplifier as it will distort the signal's time-varying voltage pattern quite severely. However its usefulness becomes apparent when we consider what happens when the voltages are large enough to ensure that $|v_{out}| \gg q / kT$. We can then approximate expression 8·6 to

$$|v_{in}| \approx RI_0\,\mathrm{Exp}\left\{\frac{q}{kT} \times |v_{out}|\right\} \qquad \ldots (8.7)$$

which we can rearrange into

$$|v_{out}| \approx \frac{kT}{q}\,\ln\left\{\frac{|v_{in}|}{RI_0}\right\} \qquad \ldots (8.8)$$

i.e. we find that when the signal levels are large enough the output voltage varies approximately as the natural log of the input voltage. Hence the circuit's name *Logamp*. The circuit is very useful for *Compressing* the range of voltage levels. This means that amplitude measurements become easier to make, and signals easier to observe without becoming too small to notice or so large as to become overpowering. It can permit a system to work over a wider Dynamic Range. The sacrifice is that the actual signal pattern will be deformed in the process. By the way, note that the amplifier is still voltage inverting so v_{out} always has the opposite sign to v_{in}. Also note that 8.8 is just an approximation and 'blows up' if you make the error of assuming it is correct when v_{in} approaches zero!

The second circuit shown in figure 8·5 acts as a positive peak detector. In this case the diode is being used as a 'switch' that can only pass current in one direction. The precise form of its

nonlinearity doesn't matter. The Op-Amp tries to behave like a voltage buffer and assert a voltage on the capacitor which equals its input. However it can only do this whilst the diode is prepared to conduct. Hence when the input is greater to or equal to the voltage on the capacitor the circuit behaves in this way.

However whenever the input voltage falls below the voltage on the capacitor the diode becomes reverse biassed. The Op-Amp cannot then force the capacitor to discharge.The charge stored can only leave via the resistor, *R*. Hence the circuit tends to 'remember' most recent peak positive input voltage level, but slowly forgets unless the voltage rises to a new peak. If we wish we can remove the resistor and replace it with a switch. The output then always remains at whatever the peak positive value of the input has been since the last time the switch was closed. The circuit is therefore useful for holding peak values which only occur briefly. This function is useful in 'peak hold' meters and displays.

Summary

You should now understand the basic circuit arrangement used in most Op Amps and why it acts as a differential amplifier. The Class A output with small transistors also explains why most Op Amps can only output small current and power levels. You should now know how Op Amps can be used to perform various *linear* and *nonlinear* functions. These include the linear examples of tone controls, differentiators, integrators as well as the nonlinear ones of Logamps and peak detectors.

Tutorial Questions

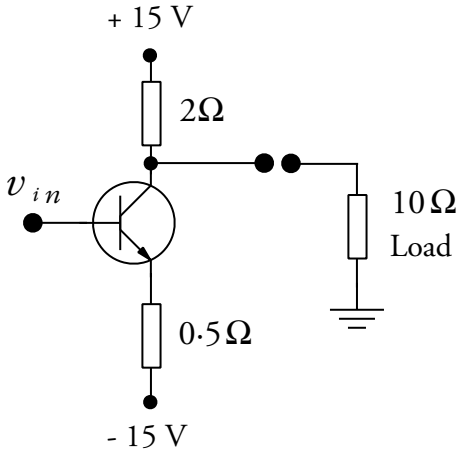
Lecture 1

Q1) Draw a diagram of a Constant Current Source that uses a Bipolar transistor. Say what emitter resistor value in your circuit would give a constant current value of 200mA. **[6 Ohms if you have three diodes in your circuit.]**

Q2) Draw a diagram of a *Long Tailed Pair* differential amplifier stage and give a brief explanation of how it works.

Lecture 2

Q1) Explain the difference between a *Class A* and a *Class B* amplifier circuit and say why in many cases where high powers and low distortion are required, *Class AB* is chosen.



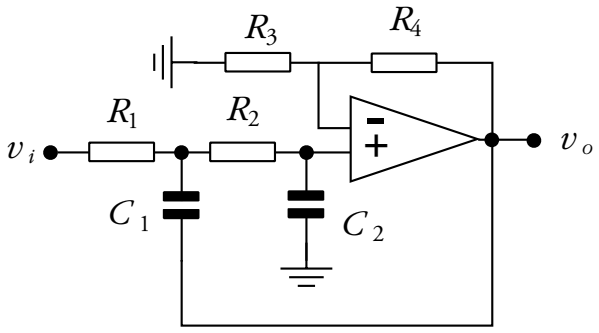
Q2) The diagram on the left shows a single ended Class A amplifier output stage. What are the values of the maximum positive and negative voltage the amplifier can apply to the load resistor? How much power does the amplifier dissipate when providing no output? What is the mean value of the maximum undistorted sinewave power level the amplifier can supply to the load? [**max +ve +12·5V. most -ve -8·65 V, quiescent dissipation 225 W, 3·75 W**]

Lecture 3

Q1) Sketch a graph showing how the frequency response (voltage gain versus frequency) of a 2nd order active low pass filter varies with the choice of the value of the damping value, *d*.

Q2) Explain what we mean by the *Quality Factor* of a resonant filter. Explain briefly how the width of the passband of a resonant pass filter varies with the value of its quality factor.

Q3) A 2nd order lowpass filter is required with a turn-over frequency value of 5kHz and a maximally flat response (i.e. d = 1·414). The circuit is required to have an input impedance of at least 10kΩ and the arrangement shown below is chosen. Work out and write down suitable values for the resistors and capacitors in the circuit. What will be the value of the circuit's low frequency voltage gain? **[Example results: All R's except R_4 choose 10kΩ, then R_4 = 5·8 kΩ. Both C's then 3·18 nF. LF voltage gain 1·58]**



Lecture 4

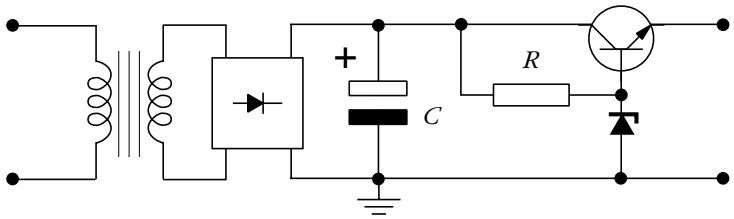
Q1) An amplifier has a low-frequency open loop voltage gain of 1,000,000. A feedback factor of β = 1,000 is applied. Say why this can be expected extend the range of frequencies over which the amplifer's response is fairly flat. By what factor would you expect the feedback to reduce any nonlinearity distortions? Say why it can be unwise to apply too much feedback and reduce the closed-loop gain too much. **[Distortion should fall by a factor of 1,000]**

Q2) With the aid of a suitable diagram of an amplifier, explain what is meant by the term *Negative Feedback*, and show how it is applied. Define the terms *Open Loop Gain* and *Closed Loop Gain*.

Q3) A differential-input amplifier has low-frequency open loop voltage gain value of $\times 10^5$. The amplifier gain is uniform at this value up to 5kHz, but above this frequency falls by 10dB each time the signal frequency is doubled. Feedback with a feedback factor (β) value of $\times 10^3$ is applied to the amplifier. What will be the value of the amplifier's voltage gain at low frequency once the feedback is applied? Sketch a graph showing the gain/frequency behaviour with and without the applied feedback, including the frequency region where the gain of the system with feedback applied begins to fall away. **[With feedback, gain flat up to around 80kHz. LF gain essentially 10^3 in voltage terms.]**

Lecture 5

Q1) A power supply is built according to the circuit diagram shown below.



The transformer provides an an output a.c. voltage to the full-wave diode bridge of 20 Volts rms at a frequency of 50Hz. What will be the value of the peak voltage supplied to the reservoir capacitor, C? **[27 Volts]**

Q2) Given a reservoir capacitor of C = 10,000 μF, what will be the magnitude of the ripple at

the capacitor when a load draws 500 mA from the power supply? **[0·5V]**

Q3) The Zener used in the circuit has a V_Z value of 15·5 Volts. What will be the stablised output voltage level supplied to loads that do not demand excessive currents? Give a brief explanation saying why this output doesn't equal the Zener voltage. **[14·9V]**

Q4) The resistor R = 100 Ω and the transistor has a current gain value of ×100. Give a brief explanation of what sets the maximum current the supply can provide whilst maintaining a stabilised output voltage. What is the highest value of load current the supply can be expected to provide to a load whilst maintaining its stablised output voltage? **[11·5 A ignoring ripple. With ripple this drops by a factor** $\left(1 + \frac{\Delta t \beta}{C R}\right)$ **to 5·75 A.]**

Q5) A transformer has 800 turns on its primary (input) and 250 turns on its secondary (output). The output is connected to a 470Ω resistor as a load, and the input is connected to mains (230V rms). How much power will be dissipated in the resistor? What will be the rms value of the current in the primary? **[10·9 W, 47mA]**

Lecture 6

Q1) Sketch diagrams of *Co-axial Cable* and *Twin Feeder*. Which type of cable would you expect to offer better protection from external interference, and why?

Lecture 7

Q1) Why do cable losses often tend to also make the cable dispersive? Why is it sometimes advisible to use Braided/Woven cables or conductors with a high resistivity?

Q2) A co-axial cable is 20 metres long. It has a capacitance per unit length of 200pF/m, and the internal dielectric has a relative permittivity of ε_r = 5·0. No magnetically active materials are used in the cable. At what velocity would you expect electromagnetic signals to propagate along the cable? The cable is used to convey signals from a source which has an output impedance of 1kΩ to a destination which has a very high input impedance. At what frequency would you expect the system to show a 3dB loss in signal power compared to its behaviour at low frequency? Suggest some steps you might take if you wished to use the cable over a wider range of frequencies than this value. **[1·34 × 10^8 m/s. 79·5 kHz]**

Q3) A coaxial cable has a Characteristic impedance of 75Ω and a capacitance per metre of 50 pF/m. What will be the value of the cable's inductance per metre? (You may ignore the effects of any losses in the cable when calculating this value.) The outer conductor of the cable has a diameter of 10 mm and the gap between the inner and outer is filled with a material whose ε_r = 4. What will be the diameter of the inner conductor? **[L' = 0·28 μH/m. 3·3mm]**

Lecture 8

Q1) Draw a diagram of a *Peak Detector* circuit and give a brief explanation of how it works.

Q2) Draw a diagram of an I*ntegrator* and give a brief explanation of how it works.