EEE3068F

2013 Rev 1

Course Notes

Samuel Ginsberg

Table of Contents

EEE3068F........1
Introduction........5
Chapter 1: Resistors........6
Power Dissipation........6
Construction of resistors........6
Specifications 6
Parasitics........8
Potentiometers........8
Potentiometer Usage Pointers........8
Chapter 2: Capacitors........10
Capacitance Marking Codes........11
Parasitics........12
Trimming Capacitors........12
Chapter 3: Inductors........13
Parasitics........13
Specifications........14
Trimmable Inductors........14
Chapter 4: Diodes........15
Rectifier Diodes and Small Signal Diodes........15
Schottky Diodes........16
Zener Diodes........16
Thyristors........17
Triacs........18
Chapter 5: Transistors........19
Bipolar Junction Transistors........19
PNP Bipolar Junction Transistors........21
MOSFETS........22
Review of Operational Amplifiers........24
Typical Opamp Internal Circuit........25
Opamp parameters and departure from ideal........27
Types of Opamps, single supply, rail to rail, decompensated etc........30
Differential Amplifiers and Instrumentation Amplifiers........33
Power Amplifiers........36
Bridge Amplifiers........43
Chapter 7: Noise and Frequency Response........44
Types of Noise........44
Noise Colour........46
Signal to Noise Ratio........46
Frequency Response........47
Bode Plots........47
Chapter 8: Filters........51
The purpose of a filter........51
Implementation Techniques........51
Filter Types........52
Transfer Functions and Complex Impedance........54
Quality Factor........56

Common Active Filter Configurations....57
Normalisation....60
Approximations to Brick Walls....60
Table Based Filter Design....62
Chapter 9: Oscillators....64
The purpose of an Oscillator....64
Types of Oscillator....64
Sine Wave Oscillators....64
Non- Sine Wave Oscillators....68
Opamp Stability – Amplifiers Sometimes Oscillate....71
Chapter 10: Linear Power Supplies....73
Raw DC Supplies....73
Voltage Regulator Circuits....75
Power Supply Protection....77
Power Dissipation....80
Negative regulators....80
Voltage Regulator Chips....80
Heatsink Design....82
Chapter 11: Switched Mode Circuits....85
Introduction....85
Step Down Regulators....86
Step Up Regulators....87
Inverting Regulator....88
Isolated Regulators....89
Integrated Switched Mode Regulator Devices....90
Charge Pump Voltage Multipliers....91
Class D Amplifiers....92
MOSFET Drivers....95
Safe Operating Area of Power Devices....96
Chapter 12: Unintentional Circuits....99
Wiring....99
Power Supply Decoupling....100
Physical Layout....101
Transmission Lines....104
Appendix 1 – Open Loop versus Closed Loop Gain....107
Appendix 2 – Gain Bandwidth Product....109
Appendix 3 – Stability of Opamp Circuits....112
Appendix 4 – Output Impedance of Opamp Circuits....118
References....119
Chapter 3....119
Chapter 4....119
Chapter 5....119
Chapter 6....119
Chapter 7....119
Chapter 8....119
Chapter 9....120
Chapter 10....120
Chapter 11....120
Chapter 12....120
Appendix 1....120

Appendix 2..120
Appendix 3..121

Introduction

This set of notes is aimed at EEE3068F, a one semester, third year electronics course for electrical engineering and mechatronics students.
The notes consist of three parts.
The first part consists of the opening chapters which contain material that is considered prerequisite for the course, as well as useful supplementary information. These chapters will not be examined explicitly, however, material in later chapters will rely on an understanding of these concepts.

The second part is the examinable section of the course and starts with the sixth chapter entitled "Amplifiers".

The appendices, which are the third part, are intended for students who want a deeper understanding of various issues and phenomena that pertain to material covered in the notes. On occasion the full details behind several "rules of thumb" are not given in the main body of the notes because of the need to constrain the scope of the course. Where it is felt that these will contribute to the understanding of more advanced students, appendices have been added to give further details for their reference. Appendices are not examinable.
Throughout the notes one will find exercises in italics which are to be answered by students.

Thanks go to Mr James Gowans for proofreading and adding quality to these sections. Thanks go to Ms Catherine Dollman for her editing work on these notes.
Please report any errors, omissions or comments directly to Samuel Ginsberg, samuel.ginsberg@uct.ac.za.

Chapter 1: Resistors

Resistors are devices that obey Ohm's Law. This means that for any resistor, in any application V = IR, where V is the voltage across the terminals of the resistor, I is the current through the resistor and R is the resistance (also referred to as the "value") of the resistor. This applies in *ANY* situation for an ideal resistor. In many cases our simple, cheap resistors are close enough to ideal to consider them perfect.

Power Dissipation

The power dissipated by a resistor is P = VI, where P is the dissipated power, V is the voltage across the resistor and I is the current through it. Ohm's law allows a few alternate forms of this equation to be derived. The power dissipated by a resistor is emitted as heat, and thus resistors will warm up to a greater or lesser extent depending on their operating condition and ability to "dispose" of that heat.

Construction of resistors

There are many possible ways of making a resistor but despite the numerous processes of producing resistors, the completed resistors are divided into two categories (types): "wirewound" resistors and "film" resistors.
Wirewound resistors are made by winding a high resistance wire (such as Nichrome) around a bobbin. Wirewound resistors are mostly used for low resistance high power dissipation applications. Film resistors are made by depositing a resistive film on a surface and etching the film into patterns to get a resistive path. With both types the longer and thinner the resistive path the higher the resistance obtained. There are limits on the range of resistances achievable with both technologies.

Specifications

When buying resistors a few key specifications need to be given.
These are:
a) Resistance

Resistors come in standard values. These are called the E12 or E24 series.
The E12 series consists of multiples of the following values:
1, 1.2, 1.5, 1.8, 2.2, 2.7, 3.3, 3.9, 4.7, 5.6, 6.8, 8.2.
The E24 series consists of the following 24 base values and multiples thereof:
1, 1.1, 1.2, 1.3, 1.5, 1.6, 1.8, 2, 2.2, 2.4, 2.7, 3, 3.3, 3.6, 3.9, 4.3, 4.7, 5.1, 5.6, 6.2, 6.8, 7.5, 8.2, 9.1.
Resistors are commonly available in these series from 1 ohm to 10 megaohms, with values down to 0.1 ohm available for high power resistors. Lower value resistors are available for measuring current. These are called "shunts".

b) Tolerance

The maximum deviation of the value of the resistor from its specified value. Resistors are commonly available in 1% and 5% tolerances but tighter tolerances are available. A 1 kilo ohm resistor with a tolerance of 5% could have a measured value of anywhere between 950 - 1050 ohms.

c) Power dissipation/package

Through-hole resistors are commonly available in power ratings of 0.25W, 0.5W, 1W, 2W and 5W. The physical sizes of these packages differ. The power rating of a resistor is specified so that at an ambient temperature of 25 degrees Celsius the resistor will be at its maximum operating temperature. Most resistors can operate at 120 degrees and above, so if a resistor dissipates its maximum power it will reach well over 100 degrees, enough to burn a person and discolor the circuit board, melt breadboards etc. As a rule of thumb try to limit the dissipation to no more than about a third of the rated maximum. Also remember that resistors operating at elevated temperatures can dissipate less power. Consult a “Derating Curve” for this information.

Surface mount resistors are generally specified by their package name. Common packages are 1206, 0805, 0603 and 0204. These names specify the length and width of the resistor in imperial units. A 1206 resistor is 0.12 inches long and 0.06 inches wide. A 1206 resistor can typically dissipate 0.2W maximum.

d) Temperature Coefficient

Any conductor's resistance will change as its temperature changes. This is called the Temperature Coefficient of Resistance (TCR) and is different for different types of resistors. Typical low cost resistors have a TCR of around 400 parts per million per degree Celsius. Some resistor types have a negative TCR, indicating that their resistance decreases with increasing temperature.

e) Maximum Current

A resistor can carry a maximum current. Power dissipation is one factor which limits the maximum current, but not the only factor. Short term pulses of current might not cause the body of the resistor to overheat because of their duration, however, they might damage internal connections, blowing the resistor open-circuit like a fuse. For that reason resistors specify a maximum peak current in addition to a maximum power dissipation.

f) Maximum Voltage

If the voltage placed across a resistor is too high there is a possibility that it will spark across the resistor, or across one of the etched gaps in the resistive film. For that reason a maximum voltage is specified for any resistor. This might be lower than expected from the resistor's power dissipation specifications. Typically, a 1/4W resistor is rated for 250VDC from end to end. A 230VAC power supply, such as domestic mains reaches a peak of 230Vrms*1.414 = 325V. This means that a single resistor with mains across it is likely to fail (usually in a disastrous way!). It might not fail immediately, but is likely to do so after time. The solution is to use series resistors of equal value. This will split the voltage across each resistor in half, and this is comfortably within rated values.

Parasitics

Thus far we have considered our resistors to be perfect resistances. In reality there are parasitic effects which hinder the performance of practical resistors.
All resistors suffer from parasitic capacitance between parallel conductors and parasitic inductance. Some resistors are better than others. Normal through hole resistors are spirals of film around a ceramic rod, and this gives relatively high inductance. There are lower inductance resistors available, in which the first half of the spiral is clockwise and the second half is anti-clockwise. In general, surface mount resistors have lower parasitic inductance than through-hole types because of their planar construction. At radio frequencies the physical size of the resistor becomes important, with physically smaller resistors typically giving more ideal performance.

Potentiometers

Potentiometers are adjustable resistor dividers. Internally they have a resistive track and a “wiper” which moves along the track as the potentiometer is adjusted. The track is normally made of a resistive composite material, known as “Cermet”. Low value high power potentiometers are sometimes wire wound.

Potentiometers come in a variety of types. Some common ones are:

Trim pots: These small pots are designed for adjusting (trimming) values in a circuit. These pots are compact and generally cheap. They have a limited rotational life, often as few as 100 cycles.
Multi-turn pots: These pots are used for applications which require very fine adjustment of resistance. By having many turns of the shaft to cover the adjustment range they make fine-tuning easier. They are available in circuit board mounted styles for trimming and in panel mount for use in instruments.
Servo pots: These pots have no end stops, so they can turn indefinitely. After 360 degrees of rotation their end-to-wiper resistance drops to zero.
Multi-gang pots: These are two (or more) potentiometers coupled mechanically to a single shaft.

Potentiometers are usually available with values in multiples of 1, 2 and 5 from 100 ohms to 10 megaohms.
In addition, there are also pots which offer non-linear relationships between shaft angle and wiper-to-end resistance. The most common non-linear taper is the logarithmic taper used for volume controls on audio equipment.

Potentiometer Usage Pointers

Limit the range of adjustment. One should insert series resistors to limit the range of output from the potentiometer to that which is actually needed. If a 10% adjustment range is needed then turning the potentiometer from one end to the other should produce a 10% adjustment range, and not much more. A particularly bad design mistake is to design circuits so that some setting of the potentiometer can cause damage. Keeping the adjustment range small will make it easier to trim the circuit and will reduce long term drift caused by the wiper “wandering” because of vibration and thermal effects.

Keep wiper current low. This is important for two reasons. Firstly, any load on the wiper of the potentiometer will affect the output of the resistor divider. This must be taken into account when designing the circuit. A second, more subtle reason is that the resistance between the resistive track and the wiper is not negligible, and increases as the potentiometer ages. If the wiper is loaded then there will be a voltage drop across this resistance, affecting the accuracy of the output. For these reasons it is quite common to feed the wipers of potentiometers into a high impedance voltage follower stage, often an opamp based unity gain stage.

Chapter 2: Capacitors

Capacitors are devices characterised by the following equation: Q = CV, where Q is charge, C is capacitance and V is voltage across the capacitor. From this the following useful forms may be derived:

$$I=C\left(\frac{dV}{dt}\right)$$

and

$$X_c=\frac{1}{2j\pi f C}$$

where I is current through the capacitor and f is frequency.

The simplest form of capacitor is made by placing two conductive plates parallel to each other with an insulating layer between them. This leads to the device being physically large for there to be any useful amount of capacitance. For this reason two things are done to reduce the size of the capacitor. Firstly, the parallel plates are rolled up or stacked into many layers. Secondly, a specialised insulating layer (dielectric) is used. Dielectrics chosen for capacitors have as high a "dielectric constant" (a misnomer) as possible. When using a dielectric material the capacitance of the capacitor is multiplied by the dielectric constant. Both of these optimisations have complex effects on the behaviour of the capacitor, causing it to deviate from the ideal, as described by the equations above.

When researching capacitors one will come across terms such as "ceramic Y5V" and "electrolytic". These terms refer to the dielectric material. Common dielectrics for electronics include: Ceramic (actually a family of materials), Tantalum, Electrolytic, Polyester, Polypropylene, and a variety of other "poly" materials. Tantalum and Electrolytic capacitors are polarised, meaning they have a distinct positive and negative lead. Operating them the wrong way around will lead to early failure, and sometimes to spectacular failure. Electrolytic capacitors usually have their negative lead marked, while tantalum capacitors usually have their positive lead marked. Some of the properties of different dielectrics will be discussed in the final chapter of these notes. "The Art of Electronics" has a useful table of capacitor types and their properties.

When purchasing capacitors there are a few key things that need to be specified. The dielectric material is one of these, the capacitance is the another. This is specified in Picofarads, Nanofarads, Microfarads and, recently, in Farads. Remember that 1000pF = 1nF, 1000nF = 1uF and 1000 000uF = 1F. The correct prefix symbol for microfarads is the Greek letter mu, but a 'u' is also commonly used. Because of confusion surrounding the prefix "milli" and its prefix symbol "m" being confused with "micro", I suggest that "milli" be spelled out in the rare cases that you wish to denote millifarads.
Capacitors are available in the E12 series, although not all members of the series are easily available in all types of capacitors. The price of capacitors depends to some extent on their availability and popularity. Thus, a 1000uF electrolytic capacitor is likely to cost less than a 680uF electrolytic capacitor. One will often have flexibility when choosing component values so this can be used to optimise cost.

As with all components there is a tolerance associated with the specified capacitance. Cheap electrolytic capacitors sometimes have tolerances of -20% and +50%. Other types are usually better, but in general capacitors have poor tolerances compared to resistors. When designing RC circuits it is pointless to design with a high precision resistor and a low precision capacitor. Dielectric materials also have a temperature coefficient associated with them, and thus some capacitors have more stable values than others over variation in temperature. The operating voltage is also a critical specification. Capacitors operated beyond their voltage limits will suffer a shortened lifespan. This can range from microseconds for serious violations to months for marginal cases. The marginal cases are really problematic because the problem may only surface once the product has been installed. Remember that the capacitor must be rated for the peak voltage that it will experience, rather than the average. Remember that capacitors running off mains supplies (with or without a transformer) will experience variations in peak voltage from time to time and must have a safety factor to accommodate these.

There is a variety of packages available. Capacitors are available in through-hole types, screw terminal types, types with flexible wires for use in appliances, and in a range of surface mount types. Surface mount capacitors are generally available in similar packages as resistors. With capacitors the trade-off is between physical size, capacitance and operating voltage. For example, for a chosen physical size of the capacitor, the operating voltage will be lower for a larger capacitance, and vice-verse. If you have a specified capacitance and operating voltage then one is restricted in physical size. For this reason it is important to consult the data from capacitor manufacturers in order to select the correct physical package.

Capacitors carrying AC will sometimes have substantial current flowing through them. In such cases the capacitor must be rated to carry the required current. Capacitors used in applications that need to supply high charge and discharge currents need high “ripple current” ratings. Capacitor data sheets often specify a “maximum rate of voltage rise” across them. Remembering that $I = C(dv/dt)$ will allow you to translate that into a maximum current into the capacitor. This is important in “snubber” applications as described later.

As with all components, capacitors have operating temperature limits. Some electrolytic capacitors only have an operating temperature rating of 85 degrees Celsius. This is quite low for some types of electronic equipment and if the application demands higher temperature operation then high temperature types are available.

Some applications have higher safety requirements. Capacitors used in applications such as mains filters and capacitive drop power supplies could result in lethal conditions if the dielectric fails and the plates become short circuited. In those cases the capacitors must be chosen with the correct safety rating depending on the situation. Follow the manufacturers' guidelines carefully and do not attempt to cut any corners.

Capacitance Marking Codes

Sometimes capacitors are marked with a value in pF, nF or uF and sometimes a three digit code is used to save space. The first two digits are then the value, followed by the number of zeros specified by the third digit. Generally the capacitance given is then in picofarads. For example, a 100nF capacitor is 100 000pF. This would have a code of 104, indicating that it would be 10 followed by four zeros in pF.
Similarly a “475” capacitor would be 47 00000pF = 4.7uF. The code is sometimes ambiguous because the coding scheme used is not marked on the capacitor. Consequently, a capacitor with a

marking of “100” might be 100pF or 10pF.
Most surface mount capacitors are simply not marked at all. This is very problematic. One should then use an LCR bridge to measure their capacitance.

Parasitics

Capacitors suffer from a variety of parasitic effects. There is an inductive effect, caused mainly by the leads and internal connections inside the capacitor, and two main resistive effects.

The first of the two resistive effects causes the capacitor to discharge through the dielectric. This is known as “leakage”. Leakage is an important parameter in some cases. Examples of such cases are RC timers (in which the capacitor must charge very slowly), memory backup capacitors (in which the leakage can discharge the capacitor faster than the backup load) and sample-and-hold circuits (in which the capacitor must “remember” a voltage for as long as possible).

The second resistive effect is called Equivalent Series Resistance (ESR) and limits the rate of charge and discharge. This causes the capacitor to dissipate power and therefore heat up. Low ESR capacitors are available and are used in applications such as power supplies. An ideal capacitor dissipates zero power because it is purely reactive, so the current through it and the voltage across it are at 90 degrees to each other. In a real capacitor the waveforms are not quite at 90 degree angles. “Loss Angle” is a useful specification of how the parasitics affect the AC performance of a capacitor.

Trimming Capacitors

Variable capacitors are available. They generally work by varying the overlapping area between parallel plates. Generally their range of capacitances is very limited, with most trim capacitors having a maximum value of a few tens of picofarads and a minimum value of a few picofarads.

When adjusting trimmer capacitors it is important to use a non-metallic screwdriver. These are often called “trim tools”. If a normal screwdriver is used the metal tip will affect the trimmer's capacitance. As the screwdriver is removed the capacitance changes slightly and the carefully trimmed capacitance will no longer be as desired.

Chapter 3: Inductors

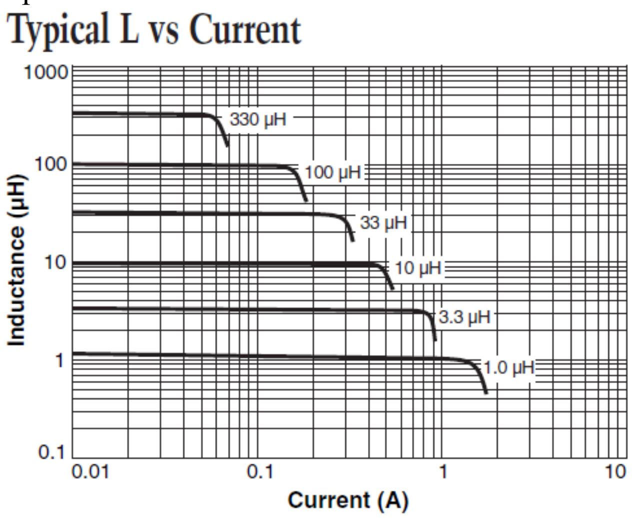
Inductors are devices which implement the equation v = L(di/dt) where v is the voltage across the inductor's leads, L is the inductance of the inductor and I is the current through the inductor. As a complex impedance the inductor gives X = j*2*pi*f*L.

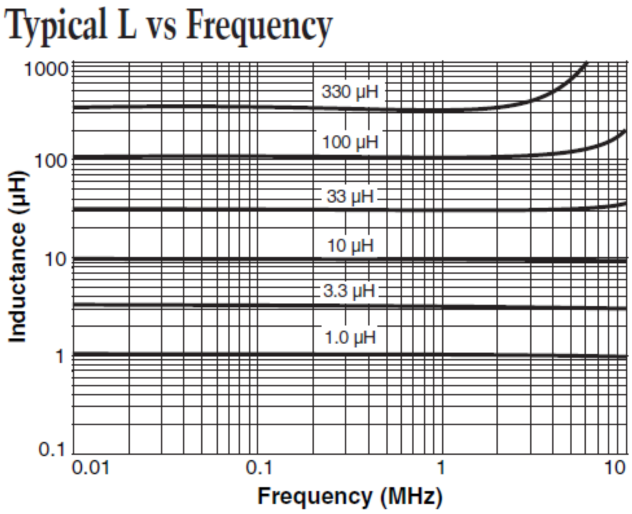
In practice, inductors are coils of wire, very often wound around a ferrite or steel core. The purpose of the core is to give more inductance with fewer turns. This introduces some advantages to the inductor such as a reduced physical size, as well as some disadvantages such as a slightly varying inductance as the current in the inductor changes.

Parasitics

Inductors suffer from two main parasitic effects. The first is the resistance of the wire used to make them. Substantial lengths of wire are used to wind high value inductors and the resistance of this wire is a parasitic element in series with the inductor. The parasitic resistance will limit the amount of current that the inductor can handle because of the heating effect of the resistance and it will also limit the Q (quality factor) of resonant circuits in which the inductor is placed.
There is also a parasitic capacitance between turns of the inductor. This limits the operating frequency of the inductor. At some (usually very high) frequency resonance between the inductor and parasitic capacitance will occur.
In addition, the magnetic core of the inductor will saturate if too much current flows through the inductor. This imposes a further limit on the operating current of the inductor. The core will also have a maximum operating frequency.

The effect of the core on the inductance as the current and frequency are changed is shown in these two graphs:





Source: *Shielded SMT Power Inductors LPS3008 Series* 2009, CoilCraft Document 438-2, p.2.

Specifications

The first important parameter to specify is the desired inductance. This is specified in Henrys, with values from 1nH to several H being available.
The next important parameter is the current that the inductor needs to handle. The peak current must not be high enough to saturate the core. The RMS current must not overheat the inductor. Because of these different requirements inductors often specify two current figures, one related to the core and the other related to the resistive losses.
The DC resistance specifies the wire resistance. This is important for loss calculations and for determining the Q of resonant circuits in which the inductor might be used.
The Self Resonant Frequency (SFR) is also normally specified in MHz.

Inductors are available in a wide range of packages from very tiny surface mount 0603 packages up to large assemblies which bolt down to the circuit board. As inductance and operating current increase, the physical size of the inductor will increase.

If there is no suitable inductor for a particular application it is possible to custom wind them. A range of ferrite cores is available for this purpose.

Trimmable Inductors

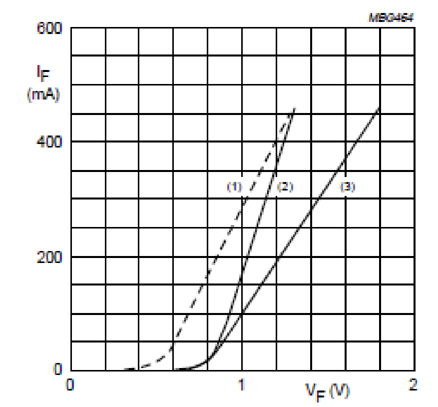
Adjustable inductors are available, albeit not popular, in modern electronics. They have a threaded ferrite slug with a slot for a trimming tool (not a screwdriver!). As the slug is turned the amount of ferrite in the coil changes, thus changing the inductance of the inductor.

Chapter 4: Diodes

Diodes are devices which control the direction of current flow and have other useful properties. They come in a large variety of types, sizes and shapes; some common ones are discussed here.

Rectifier Diodes and Small Signal Diodes

These diodes are generally used to pass signals in one direction. Their characteristics are shown on a graph which plots the voltage across the diode versus the current flowing through it. A typical graph is shown here, of the 1N4148 diode:



(1) T_j = 175 °C; typical values.
(2) T_j = 25 °C; typical values.
(3) T_j = 25 °C; maximum values.

Fig.3 Forward current as a function of forward voltage.

Source: *High-Speed Diodes (1N4148; 1N4448)* 2004*,* Koninklijke Philips Electronics, Philips Semiconductors Product Specification Data Sheet, p.4.

These diodes are generally only used in forward conduction or reverse blocking modes and so the negative part of the Vf curve is not shown in many datasheets. This graph shows that a diode's forward voltage drop is not constant, as is often assumed. This diode will enter conduction at around 0.6V, but the forward drop will be well over 1V at a forward current of 450mA.

For these diodes the key specifications are the currents that they can handle and the reverse voltage that they can block. This is known as “Peak Inverse Voltage” (PIV). There are two very important current specifications usually given: the first is the average forward current the second current specification is the peak forward current.

Signal diodes are diodes made to carry small currents (usually several hundred milliamps) and tolerate fairly low PIV's (generally around 100V). They are designed to operate fast, with many being able to change from reverse blocking to forward conduction in several nanoseconds. They typically have parallel parasitic capacitances of several picofarads.

Rectifier diodes are physically large diodes which are made to carry large currents and typically have large PIV's. They tend to be quite slow, with parasitic parallel capacitances of tens of picofarads and reverse recovery times of tens of microseconds. These diodes are primarily used for rectification of low frequency AC.

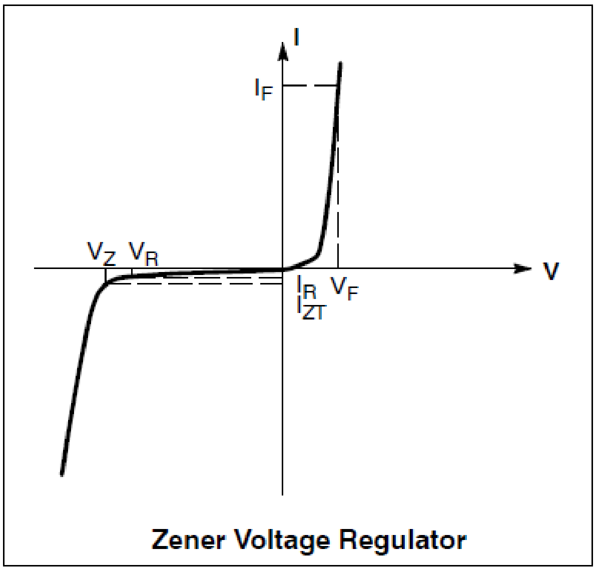
Fast Recovery diodes are power diodes with enhanced reverse recovery times, down to tens of nanoseconds. They are primarily used in high frequency power electronics and will be discussed in a later chapter.

Schottky Diodes

Schottky diodes (also called hot carrier diodes) exhibit a lower forward voltage drop than normal junction diodes, thus approximating the ideal diode more closely. They will typically enter conduction with 0.3V across them. As with normal diodes their voltage drop increases with forward current. In any case, they offer lower forward drop than standard diodes. They also offer fast reverse recovery, often in the tens of nanoseconds.
The disadvantage of schottky diodes is that they have much lower PIV handling than standard rectifier diodes. Most schottky diodes have PIV ratings of less than 100V. This limits their usefulness, although they are widely used in low voltage high current high frequency applications. They also see wide usage in reverse polarity protection circuits, where their low forward drop is a distinct advantage.

Zener Diodes

Zener diodes are designed to operate in the reverse breakdown region and have reasonably well-defined reverse breakdown voltages. Their characteristic curve is the same shape as a standard rectifier diode, however, normal diodes may suffer damage in reverse breakdown. The curve is shown here:



Source: *Zener Voltage Regulator* 2009*,* BZX84B4V7LT1, BZX84C2V4LT1 Series: Zener Voltage Regulators, On Semiconductor Data Sheet for BZX84 Diodes, p. 2

In forward conduction the zener behaves much like a standard diode.
The point Vz is the “Zener Voltage”. When current is fed backwards through the diode the voltage across the diode will be as indicated on the graph. Notice from the graph that the reverse voltage varies a bit with changes in reverse current. Furthermore, the actual zener voltage is subject to a tolerance. Zener diodes are not very accurate voltage references.

Because the diode has current through it and voltage across it there will be power dissipation. Zener diodes are available in a variety of packages, with a range of dissipation capabilities. These typically range from 0.2W up to about 5W. Note that zener diodes are not designed to absorb large transients, so be wary when using them for surge suppression applications.

As shown above, zeners should not be operated at too low a reverse current or else they will not give the correct zener voltage. In addition, the voltage across them will become noisy. The datasheets for the zener will give a minimum operating current, in this case Izt.

Thyristors

Thyristors (also called Silicon Controlled Rectifiers, SCR's) are diodes, but with an important difference. They have three terminals. As with diodes there is an Anode and a Cathode but they have a third terminal which is called the “Gate”. Initially, when a voltage is placed in the forward direction across the anode and cathode no current flows. This state is maintained until a current is made to flow through the gate and out of the cathode. This is done by raising the gate voltage with respect to the cathode. Once this gate current has been made to flow, the thyristor will allow current to flow through the anode to the cathode. This state persists even if the gate current is removed. The device stays in conduction until the anode/cathode current drops to below some threshold, called the holding current.

The basic specifications for a thyristor are the blocking voltage and the average forward current that can be tolerated. There is also a maximum dissipation specification. When an SCR is turned on it will have a voltage drop across it (typically slightly higher than a diode of similar ratings) and this causes power dissipation. In the blocking state an SCR will not dissipate power because there is negligible current flow.

There is also a limit on the rate of rise of voltage across the thyristor as it turns off. For a resistive load this is not an issue as the thyristor only turns off when the current through it is minimal. For inductive loads this an important thing to be aware of as the inductor will attempt to force current through the thyristor as it turns off (called the inductive kick), and this will result in a rising voltage as the thyristor turns off. Common solutions to this problem include an anti-kickback diode across the inductive load or an RC circuit across the anode and cathode of the inductor. This is called a “snubber” circuit.

Because the gate is current triggered there is also a specification for minimum and maximum gate currents. The minimum gate current is the current that is guaranteed to trigger conduction. In general, the bigger the thyristor the higher the trigger current, but there are “sensitive gate” types which need less trigger current.

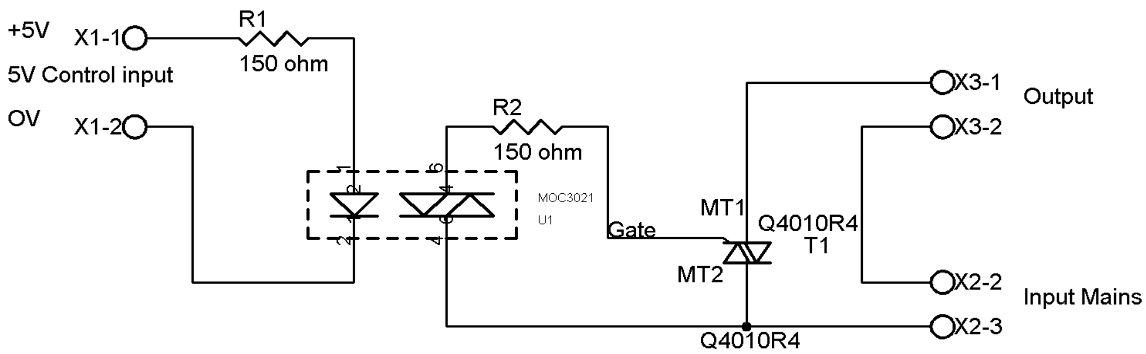
Triacs

Triacs can be crudely described as bidirectional thyristors. They are similar in that they block current until triggered by a gate pulse, and they are different in that current can flow in either direction through them and the trigger pulse can also flow in either direction.
Triacs have three terminals called the Main Terminal 1 (MT1), Main Terminal 2 (MT2) and Gate. In order to trigger the triac the trigger pulse must be applied from gate to MT1. This gives rise to four possibilities:

Quadrant 2 • Conventional Current flow from MT2 to MT1 • Gate trigger pulse current flows out of gate	Quadrant 1 • Conventional Current flow from MT2 to MT1 • Gate trigger pulse current flows into gate
Quadrant 3 • Conventional Current flow from MT1 to MT2 • Gate trigger pulse current flows out of gate	Quadrant 4 • Conventional Current flow from MT1 to MT2 • Gate trigger pulse current flows into gate

Most triacs can be triggered in any of the four quadrants, but on most triacs the quadrants are not equally sensitive, so the amount of gate current will depend on which quadrant is used.

The most common use of triacs is switching large AC currents and voltages. A typical circuit is this:



Source: Author's own diagram, (2013).

In this circuit the component in the dashed box is an opto-triac. When the control input is energised the opto-triac conducts bidirectionally between its output terminals. This allows current to flow from the input through the opto-triac, through the gate and via MT1. This turns the triac on. R2 is for limiting the current into the gate. Once the triac is in conduction it reduces its own gate drive signal beause of the low voltage between MT1 and MT2, but, since it is latching, the triac will remain in conduction.

Question: Which quadrants will the triac operate in for this circuit?

The specifications for triacs are very similar to those for thyristors. Diode based anti-kickback circuits are usually not practical for AC circuits and so RC snubbers are generally used.

Chapter 5: Transistors

There is a wide range of different types of transistors available. We will only discuss the two which feature heavily in the remainder of these notes. They are Bipolar Junction Transistors (BJT) and Metal Oxide Semiconductor Field Effect Transistor (MOSFET) devices.

Bipolar Junction Transistors

The BJT is a current-controlled amplifier. A small current flowing into the base terminal will allow a larger current to flow from the collector to the emitter. The maximum collector current is related to the base current by a ratio called Beta, also known as Hfe and commonly called "current gain". Current gain is a fairly imprecise specification. As an example of how Hfe can change for a given transistor, here is a graph of current gain versus collector current:

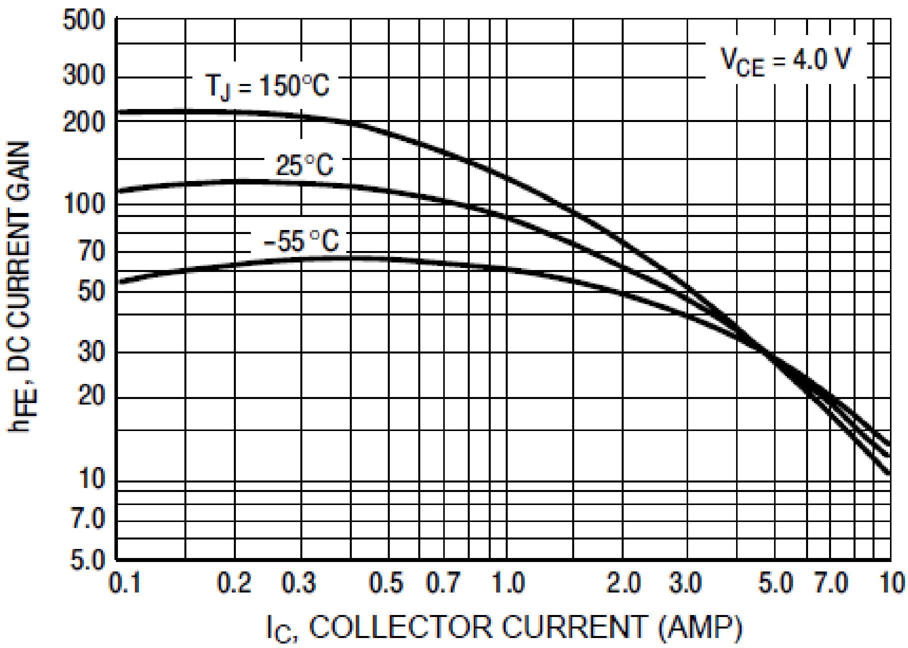


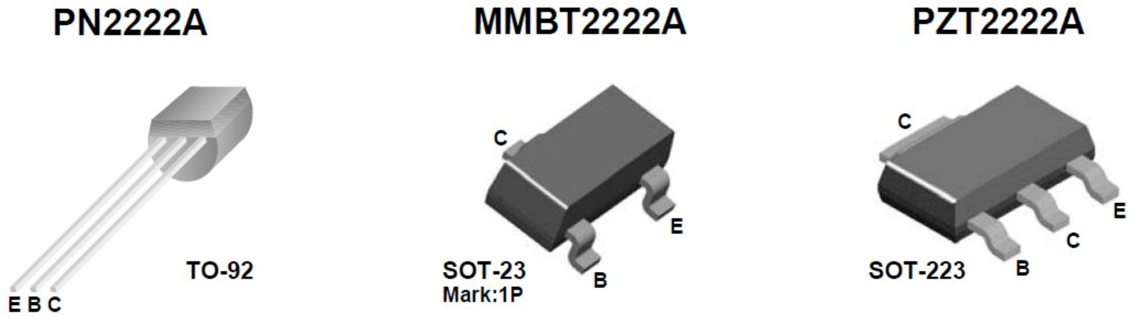
Figure 3. DC Current Gain, 2N3055 (NPN)

Source: *DC Current Gain* 2005*,* 2N3055(NPN), MJ2955(PNP): Complementary Silicon Power Transistors, On Semiconductor Data Sheet, p. 3

In addition to current gain there are a number of other specifications that are of importance when choosing BJT's. The maximum voltage that they can block from collector to emitter is often vital and the amount of current that they can handle must also be considered. These parameters will be discussed at a later stage in these notes.
The frequency at which a transistor can operate is specified by means of a parameter called Ft, the unity gain frequency. This is the frequency at which the gain of the transistor is reduced to unity. Normally transistors are operated at frequencies much lower then their unity gain frequency.

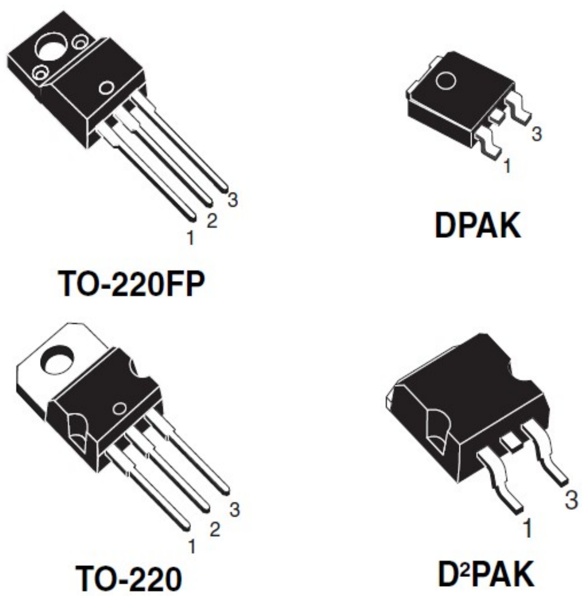
Transistors come in a variety of physical sizes and shapes. Bigger packages allow greater power dissipation and smaller packages allow more compact electronic devices. In addition, there are

special packages for high frequencies. Transistors are freely available in surface mount formats as well.
Some common packages are shown here. Low power transistors are often in the TO-92, SOT-23 or SOT-223 packages:



Source: *NPN General Purpose Amplifier* 2010*, PN2222A/MMBT222A/ PZT2222A NPN General Purpose Amplifier*, Fairchild Semiconductor Data Sheet for PN222A/MMBT2222A/PZT2222A p. 1

High power transistors ore often in the T0-220, DPAK or D^{2}PAK packages. Sometimes the TO-220 package has a metal mounting tab and sometimes an insulating plastic tab.



Source: N-channel 525 V, 1 Ω, 5 A, D^{2}PAK, DPAK, TO-220FP, TO-220
SuperMESH3™ Power MOSFET 2011, STB6N52K3,STD6N52K3 STF6N52K3, STP6N52K3: Supermesh Power MOSFET, ST Microelectronics Data Sheet P. 1

PNP Bipolar Junction Transistors

PNP transistors perform the same function as NPN transistors but with their polarities reversed.

In a NPN transistor conventional current flows from collector to emitter.
In a PNP it flows from emitter to collector.

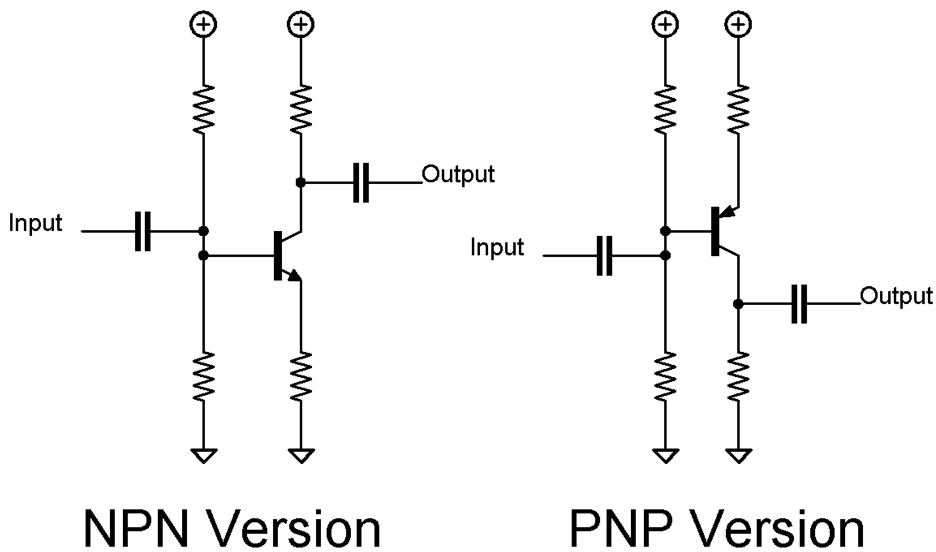
In a NPN transistor conventional current flows into the base.
In a PNP it flows out of the base.

As one would expect, in order to obtain the current flows in the correct direction, the potentials on the leads of the transistor must also be reversed. Thus:

NPN	PNP
Emitter is most negative voltage	Emitter is most positive voltage
Conventional current flows out of emitter	Conventional current flows into emitter
Base must be about 0.6V above emitter to turn on	Base must be about 0.6V below emitter to turn on
Base current flows in to base lead	Base current flows out of base lead

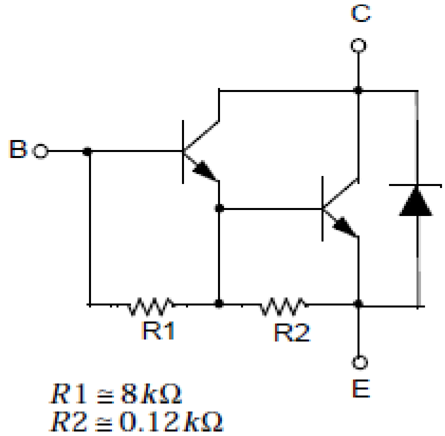
When the symbols for semiconductor devices are designed the arrow always points from P type silicon to N type silicon. For the BJT and diode symbols the arrow indicates the direction of current flow through the device.

The common emitter amplifier may be built up with either NPN or PNP transistors. The two circuits below are shown to illustrate the above points:



Source: Author's own diagram, (2013).

If high current gains are needed it is possible to buy Darlington transistors. These are three lead devices in standard transistor packages. Internally they consist of two transistors. Normally they also include a resistor to speed up the switching operation of the Darlington pair. The Fairchild Semiconductor datasheet on the popular TIP122 NPN Darlington transistor gives the following internal diagram:



Source: *NPN Epitaxial Darlington Transistor Equivalent Circuit* 2001*,* Fairchild Semiconductor Data Sheet, TIP 120/121/122: Medium Power Linear Switching Applications, p.1.

R2 prevents leakage current from the incoming transistor turning on the output transistor and the diode is reverse polarity protection.

MOSFETS

Metal Oxide Semiconductor Field Effect Transistors (MOSFET's) have a controllable conductive channel. The control is affected by varying the electric field strength in the channel. This is done by varying the voltage on the gate terminal with respect to the channel. MOSFET's thus have at least three terminals: The Drain (in some sense equivalent to the collector on a BJT), the Source (similar to the BJT's emitter) and the Gate (the control terminal).
In order to turn a MOSFET on, the gate terminal must have a potential applied with respect to the source terminal.
MOSFET's are available in N channel and P channel types. These are analogous to the BJT's NPN and PNP variants. In order to turn a N channel MOSFET on the gate must also be made more positive than the source. For a P channel MOSFET the gate must be more negative than the source.

Because of the construction of MOSFETS the gate has a fairly low allowable voltage. This is seldom above 20V and often far below that. Higher voltages between gate and source will cause destructive breakdown of the gate insulator and destroy the device. This, combined with ultra high gate impedance, make MOSFET's sensitive to static electricity. Caution should be exercised when handling them.

MOSFET's can be used for linear control. They can produce a linear channel response to an analogue gate drive. There are instances (such as in some large audio amplifiers) in which this is used, but the vast majority of MOSFET applications use them as simple switches. The channel is either fully blocking or fully conducting. This is achieved by driving the gate either to the same potential as the source (blocking channel) or at a potential above some threshold. This threshold is known as the “Gate Threshold Voltage”. When the MOSFET is in conduction the channel behaves like a resistor. The resistance of the channel is a key specification in determining how close to ideal the device is. This figure is called “R_{ds}” (Resistance from Drain to Source) and the most modern low voltage MOSFETS achieve milliohm figures.

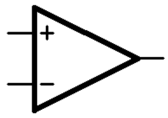
Because the gate is insulated from the rest of the MOSFET structure there is negligible leakage current into the gate. The resistance of the gate with respect to the source is many gigaohms and is usually ignored in design. The capacitance between gate and source is often more of an issue. This parasitic limits the rate of change of the gate and thus the rate at which the device can change from blocking to conduction and vice versa. Gate capacitance is often related to the power handling capabilities of the device. The effects of this on switching circuits will be discussed at a later stage.

MOSFET's are available in the same physical packages as BJT's.

Chapter 6: Amplifiers

Review of Operational Amplifiers

Ideally, the opamp is a three terminal device as shown here:



Source: Author's own diagram, (2013).

This three terminal device implements the following equation:
$V_{out} = A_v(V_+ - V_-)$

- The terminal with the + symbol is called the non-inverting input because it has the same polarity as the output.
- The terminal with the – symbol is called the inverting input because it has the opposite polarity to the output, because of the subtraction in the equation.
- The A_v parameter is called the Open Loop Gain of the device.
- Ideally A_v is infinite.
- Ideally no current flows into either of the inputs.
- The output impedance is infinitely low.

In reality A_v is not infinite, but at DC it will typically be in the range of 100 000 to 1000 000. A small current does flow into the inputs. This is often negligible, but will be discussed at a later stage.
The opamp does not generate any voltage, it merely channels power from the power supply to the output, in order to fulfill the equation. It follows therefore that the voltage on the output pin can never exceed the supply rail's voltage.

From the above equation we can see that if $V_+ > V_-$ then the output is going to be as positive as it can be (within the supply rail range).
From the above equation we can also see that if $V_- > V_+$ then the output is going to be as negative as it can be (within the supply rail range).
This is called comparator action. It is useful in many applications for comparing two voltages.

Suppose that we arrange our circuit so that V_{out} affects V_-. In this case, if V_{out} increases, so does V_-. What will happen? We can analyse this intuitively with two scenarios:

- If the voltage on V_+ increases relative to V_- then the V_{out} will also increase.
- If V_{out} increases then V_- also increases.
- If V_- increases then V_{out} decreases.

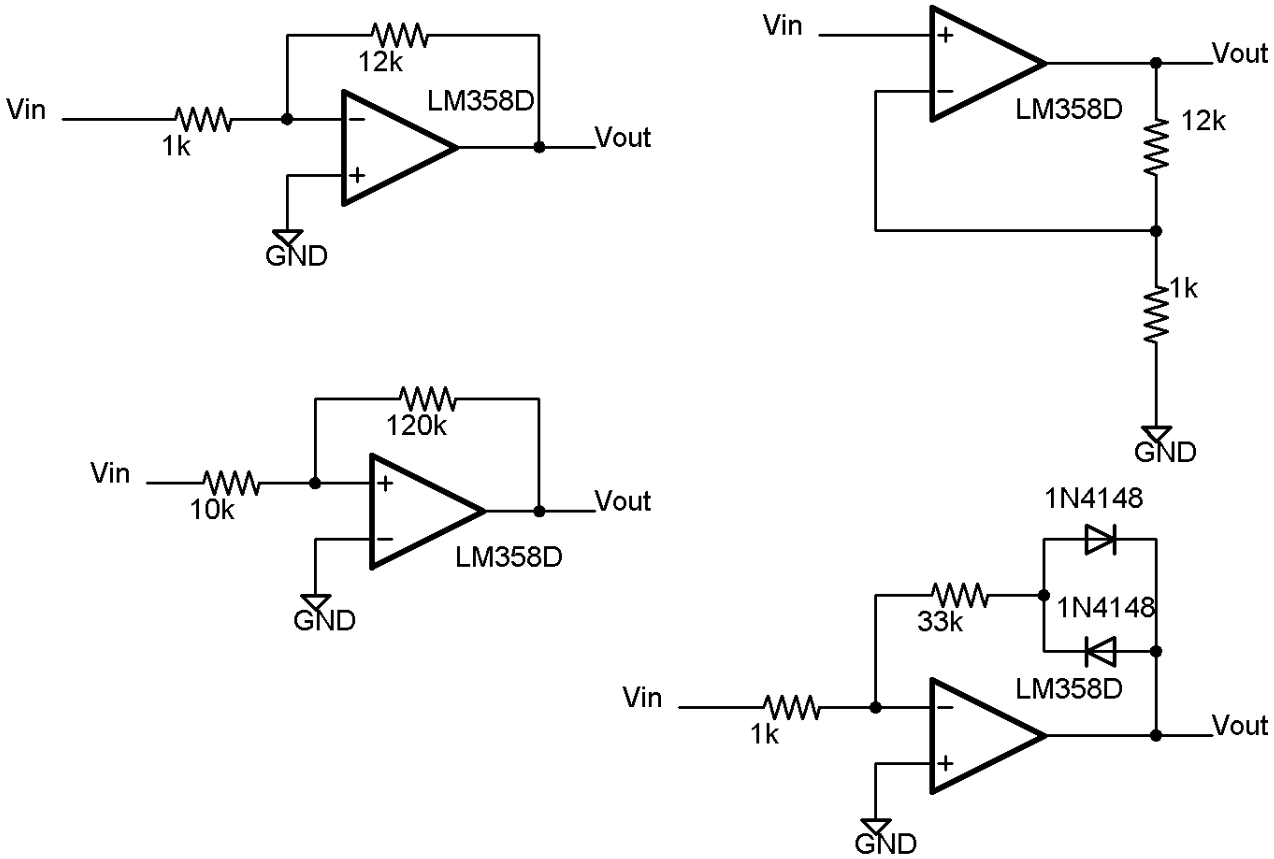
Furthermore,

- If the voltage on V_+ goes down relative to V_- then the V_{out} will go down.
- If V_{out} goes down then V_- also goes down. If V_- goes down then V_{out} goes up.

The result of this is that if the opamp is arranged in a negative feedback configuration the inputs will be driven extremely close to equality so that the opamp equation is satisfied.

Exercise:

Derive the equations for the following circuits:



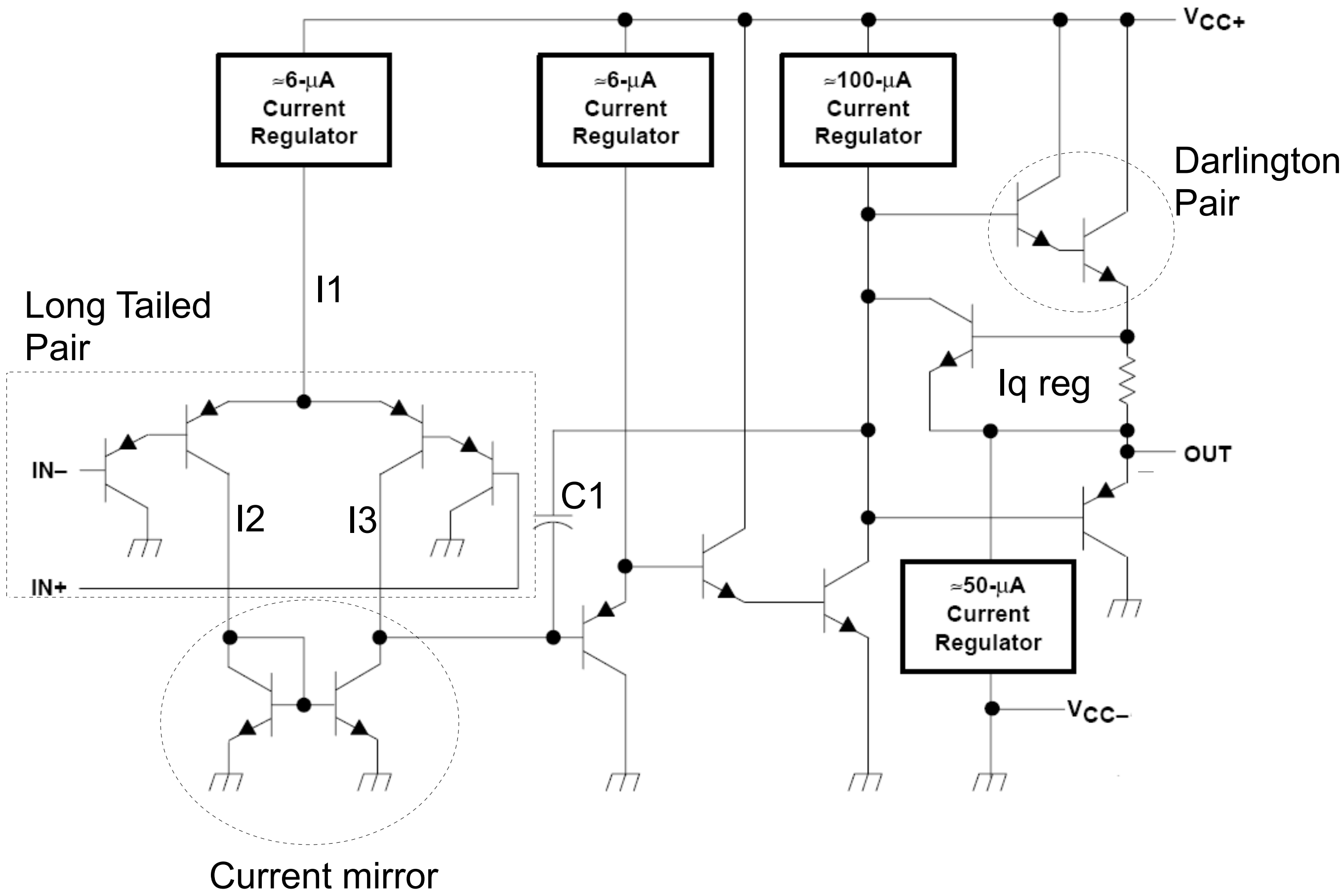
Source: Author's own diagram, (2013).

What is the input impedance of the non-inverting amplifier circuit?
What is the input impedance of the inverting amplifier circuit?

Typical Opamp Internal Circuit

It is important to have a basic understanding of what goes on inside an opamp. This will help one understand the differences between different types of opamps, which will later be discussed in detail, and it will also give one an understanding of the limitations of opamps in general.

Here is the simplified internal diagram for an LM358 opamp: (The annotations have been added).



Source: *Schematic of LM358 Dual Operational Amplifier* 2002, Texas Instruments, LM158, LM158A, LM258, LM258A, LM358, LM358A, LM2904, LM2904Q Data Sheet, p.3, (amended by author).

The input stage consists of a long tailed pair, a current mirror and a current source. Remember, the current source (in its simplest implementation) is just a transistor-zener circuit.

Exercise: Draw a circuit diagram for the -6uA current source shown in the input stage. Why is the current shown as negative?

The current source forces the current I1 to be 6uA. The current mirror circuit "mirrors" current I2 with I3, so these two currents are the same; ie, 3uA. It achieves this by varying the voltages across its transistors to force these two currents to equalise. If I2 > I3 then the base currents on the transistors will increase, causing I3 to increase, which then makes I2 = I3.
If the inverting input is taken to a lower voltage (closer Vcc-) then the current I2 will try to increase. The current mirror will attempt to compensate by increasing I3. It does this by pulling its output closer to Vcc-, and vice-verse. The inverse therefore also applies. Thus, the output of the input stage is proportional to the difference between the inverting and non-inverting inputs. This feeds to the next stage of the amplifier.

The next stage of the amplifier is essentially an emitter follower circuit. Remember that a current source may be replaced by its Thevenin equivalent, which is an infinite voltage source with an infinite resistance in series. In practice, the Thevenin equivalent of a current source is a finite voltage source (Vcc) with a high value resistor in series. Remember that this is a PNP transistor. If one is ever confused by the PNP device, first sketch an emitter follower using an NPN device, then swap to PNP and invert all the voltages. The emitter follower provides current gain but no voltage gain.

The third stage is another emitter follower, using an NPN transistor.

The fourth stage is a common emitter amplifier. This stage provides a very high voltage gain. The transistor is biased to operate in its linear region. The voltage on the collector will be an inverted amplified replica of the voltage on the base. The 100uA current source acts like a very high value collector resistor giving very high voltage gain.

The fifth and final stage (the "output stage") of the amplifier is a push-pull stage. It provides current gain but no voltage gain. The NPN Darlington pair sinks current into the load connected to the output. The PNP output transistor allows the load to source current to the negative rail. When the output voltage is changing the current flow will change from one output transistor to the other, depending on the slope of the waveform and its voltage level. In order to make this transition smooth it is common practice to always have a bit of current flowing in both devices. This inactive current is controlled by the circuit labelled Iq reg. The current consumed while the circuit is inactive is known as "Quiescent Current".

Normally opamps are used in feedback configurations. Any closed loop system must be carefully designed in order to ensure stability. The design of stable closed loop systems has a different course dedicated to it. In order to improve the stability of the amplifier, C1 is added. This feedback capacitor rolls the gain of the amplifier down at higher frequencies. Because of the inversion of one of the stages in the amplifier the capacitor is a negative feedback element. The capacitor is a high-pass element, resulting in more negative feedback at higher frequencies, thus lower gain, thus better stability. This stability comes at a price: the bandwidth (useful frequency range) of the amplifier is reduced. Opamps which have this capacitor are referred to as "compensated" opamps.

Exercise: Show that the inverting and non-inverting inputs are really the right way around.

Opamp parameters and departure from ideal

As one might expect, opamps are not quite perfect. While they approximate the ideal opamp model quite closely in some respects, they are lacking in other respects. To a great extent the large range of different opamps on the market is because of the compromises that need to be made when designing an opamp. Different opamp manufacturers specify some of the performance parameters slightly differently from others, so sometimes one will encounter different terminology to that used here, although the concepts remain the same.

The following items are typically specified in an opamp datasheet. The values shown in brackets are for the LM358 opamp.

- Input Offset Voltage:

An opamp in a negative feedback circuit will ideally drive its inputs to equality. In reality, the input transistors will be slightly mismatched. This means that the inputs are never driven quite to equality. Input offset voltage specifies the maximum voltage difference between the inputs when the opamp is in negative feedback mode. This parameter is specified at DC. (9mV)

- Input Bias Current:

The input transistors need to have current fed into their bases in order for collector current to flow. This current is the input bias current. Because the current into the two inputs might not be exactly the same this is specified as the average of the two inputs. This is a departure from the ideal case, which is zero bias current. (500nA)

- Input Offset Current:

This is the maximum difference in bias current between the inputs. (150nA)

- Common Mode Input Range:

This is the allowable range of voltages on the input pins of the opamp. Note that exceeding this range can cause some strange effects. Early opamps would sometimes self-destruct at the slightest violation of this specification. Because of the PNP input stage of the LM358, its inputs can go to low voltages, but as the input voltage approaches Vcc the input transistors go out of conduction (remember they need 0.6V across each base-emitter junction) and so the input range is (Vcc-2V) maximum and -Vcc minimum.

- Common Mode Rejection Ratio:

Ideally, the opamp should only amplify the voltage difference between its inputs. In reality, the actual voltage level of the inputs does also make a difference. A model which takes this into account is:

$V_{out}=A_v(V_+-V_-)+0.5A_{cm}(V_++V_-)$

CMRR is measured as the ratio of differential mode amplification (Av) to common mode amplification (Acm). Because it is a ratio it is often specified in Decibels. The effect of CMRR is to impose an offset on the output. This offset changes as the DC level of the inputs changes, while the difference voltage between them remains the same. (80dB, which is a ratio of 10000 times)

- Large Signal Differential Voltage Amplification:

This is also called open loop voltage gain. The ideal opamp has infinite voltage gain. Practically, this cannot be made. This specification gives the voltage gain at DC. (100V/mV = 100 000 = 100dB).

[For further details on open loop gain and how it relates to closed lop gain see Appendix 1].

- High Level Output Voltage:

As you can see from the above circuit diagram the output stage is made up out of transistors. The transistors that pull the output towards Vcc are in a Darlington configuration, and the base of that pair is driven from a current source. A Darlington pair used as it is here will always have a minimum of 0.6V across it and the base of the pair can never be driven to Vcc because the current source always needs some voltage across it. (Vcc-1.5V)

- Low Level Output Voltage:

This is the minimum output voltage that can be output when the bottom output transistor is turned fully on. (20mV)

- Output Current:

How much current the opamp can source or sink from its output. (20mA)

- Supply Current:

How much current the opamp needs to run. This excludes current flowing in the output pin. This parameter is important in low power battery operated devices. (2mA)

- Slew Rate:

How fast the output of the opamp can change (how fast it can slew). The output of a real opamp has parasitic series impedance and stray capacitance to ground as well. These limit the maximum rate of change that the output can achieve. Low slew rates limit performance in two ways. Firstly, they cause square wave outputs to become triangular (poor rise and fall times) and secondly, they limit the amplitude of output signals at higher frequencies. For sinusoidal outputs the slope is steepest at the zero crossing (remember calculus?) and the slew rate must be sufficient to accommodate this or distortion will result. (0.3V/us)

Exercise: Calculate the slew rate required to accurately produce a 10Vpk-pk sinusoidal waveform with a frequency of 1kHz

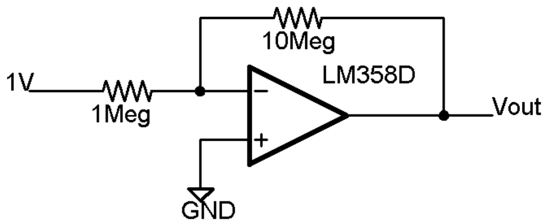
- Gain Bandwidth Product:

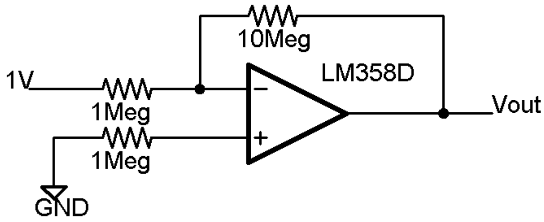
The common abbreviation is GBWP. The product of open loop gain and operating frequency is roughly constant for conventional opamps. As the operating frequency goes up the open loop gain goes down proportionately. In most linear applications we operate opamps in negative feedback mode, which sacrifices gain for precision. In order for our assumptions about input equality to be approximately correct we need substantially more open loop gain than closed loop gain (a rule of thumb is 10 times, but this depends on the final gain accuracy required). As an example, if the GBWP is 1MHz and we want to make an amplifier for the audio band extending to 20kHz, we then only have an open loop gain of 50. This only allows us to design a closed loop gain of 5 per amplifier stage. (1MHz)
[For further details on gain bandwidth product and where the 10x rule of thumb comes from see Appendix 1 and Appendix 2.]

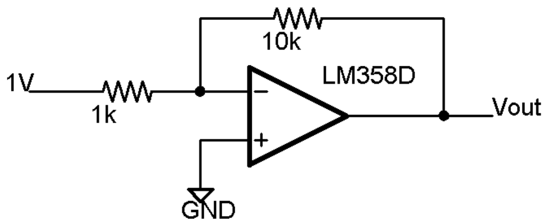
- Equivalent Input Noise Voltage:

This specifies the noise level generated by the opamp. We will discuss this in detail at a later stage. (40nV/sqrt(Hz))

Exercises:

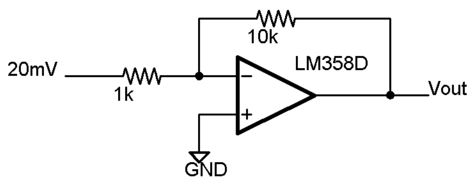




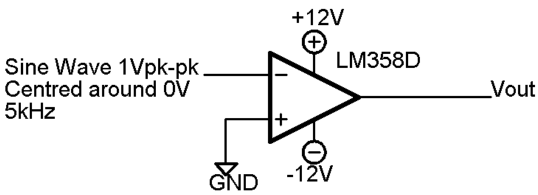


For each of these circuits calculate the output voltage. First assume that the input bias currents are zero. Then do a worst-case calculation assuming that the currents are 500nA, flowing into the inputs.
Source: Author's own diagram, (2013).

Calculate the output for this circuit. First assume that the input offset is zero. Then do a worst case calculation and apply the datasheet limit of 9mV. Assume that the inverting input is 9mV higher than the non-inverting input.



Draw the waveform produced by this circuit. First assume a perfect opamp, then assume a slew rate of 0.3V/us



Source: Author's own diagrams, (2013).

<u>Types of Opamps, single supply, rail to rail, decompensated etc.</u>

In many applications there are a few specifications which are vitally important and others which are less so. The following are a few examples.

1. A battery powered device, running off two AA batteries, totaling 3V:
   In this case the power supply specifications are important. Firstly, the opamp must be able to run off 3V (many opamps need a higher supply voltage). Secondly, the power consumption of the opamp should be low (the LM358 will give about 1000 hours of operation, or a bit

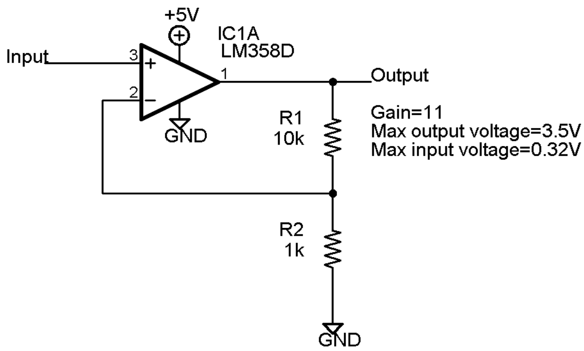
over 1 month), which is not acceptable in many applications. Thirdly, the output swing will only be 1.5V, which would lose 1 bit of resolution if it was fed into a microcontroller ADC with a 3V reference. Other specifications might, however, not be critical.
Exercise: Explain why a 1.5V signal would result in 7 bits of resolution if fed into a 3V, 8 bit ADC

2. An amplifier used to condition the output of a Wheatstone bridge application, possibly a bridge consisting of strain guages:
   In this case the differential signal will typically be millivolts, but this will ride on a large DC offset. This calls for good CMRR, as the ability to reject the DC offset is important. Furthermore the input offset voltage is important. If the input differential voltage is only several millivolts then the input offset voltage must be appreciably lower than this.
3. A microphone amplifier for use with a high quality moving coil microphone:
   Here we need a high gain, typically 1000 at frequencies of up to 15kHz. This calls for a fairly high GBWP. In order to get a good clean signal at the output we will also need low equivalent input noise voltage.

In order to accommodate these requirements a wide range of optimised opamps is available. Some common tweaks are the following

- *Single Supply/Dual Supply:*

The LM358 is called a single supply opamp. Nevertheless these devices can run off a single supply or a split rail. The difference is that single supply opamps have inputs which operate all the way to the most negative supply rail of the opamp. Consider this circuit:



Source: Author's own diagram, (2013).

This is a very standard non-inverting amplifier. The input range of the amplifier is from 0V to 0.32V in order to give the maximum output swing that the amplifier is capable of. Clearly the performance of the opamp's inputs near ground is important. If you compare the common mode input range of the LM358 to that of the LM741 you will see that the single rail device will work, while the dual rail 741 will not.

Exercise:
Look online for the datasheet for the 741 from National Semiconductor. Notice that they have a different way of specifying the Common Mode Input range. What range of inputs could it accept if it was run from a single rail 12V supply? Look at the internal circuit in the datasheet and explain the reason for this.

- *Rail to Rail:*

Sometimes it is really useful to have an opamp with inputs and outputs which swing all the way to both supply rails. As an example, suppose that you have a circuit running off a 3V battery. The circuit has an opamp feeding into the ADC of a microcontroller like the GT16A. The input range of the GT16A is 0-3V when running off a 3V power supply. If you use an LM358 the output range will only be 0-1.5V. In this case you have three options. The first option is to power the microcontroller off a lower supply voltage. In the case of the GT16 the lowest supply voltage is 1.8V. This does not quite solve the problem. In addition, the clock speed will need to be reduced. The second option is to power the opamp off a higher voltage. This will require a voltage converter to convert the available 3V to at least 4.5V. This adds cost and complexity, and increases power consumption. Option three is to use a rail to rail opamp instead. Then the output swing is wide enough to take full advantage of the ADC's input range. Low cost rail to rail opamps such as the MCP6002 are now freely available. They have MOSFET output stages instead of bipolar transistor outputs.

- *JFET input:*

JFET inputs offer very much lower input bias currents than bipolar inputs. An inexpensive JFET input opamp, the LF351, offers a worst-case input bias current of 200pA. Some transducers only output tiny currents, and so low bias currents are a considerable advantage. A common application of this would be in photodiode amplifiers, where the input current is often less than a microamp. The error caused by the bias currents on a standard opamp would be significant.

- *Chopper Stabilised:*

Chopper stabilisation is a modulation technique that is applied to some opamps for high precision applications. These opamps have very small input offsets and also sometimes reduced noise figures. The MAX420 has in input offset that is guaranteed to be less than 5uV, about $1/1000^{th}$ of the LM358's. Because of the improvements in conventional precision opamps these devices seem to have lost popularity in more recent designs.

- *Low Noise:*

In the example given before in these notes:
A microphone amplifier for use with a high quality moving coil microphone needs a high gain, typically 1000 at frequencies of up to 15kHz.
If we use the LM358 the equivalent input noise wold be:

$$V_n = 40\text{nV}/\sqrt{Hz} = 40\text{nV} * \sqrt{15000} = 40\text{nV} * 122 = 4.9\text{uV}$$

The output noise would be:
1000*4.9uV = 4.9mV.
This noise would sound like a hiss on the loudspeakers of the system.
In contrast, the NE5534 opamp has an equivalent input noise voltage of 3.5nV/sqrt(Hz). This would result in an output noise of 429uV, which is more than 10 times better.
It is important to note here that we assume that the bandwidth of the amplifier is only as wide as we actually need for our application. If the bandwidth was wider we would also get more noise. For this reason it is often important to restrict the bandwidth of amplifiers by adding capacitors across the feedback resistors.

- *High Gain Bandwidth and Decompensated Opamps:*

The low GBWP of the LM358 limits the applications to which it can be applied. It cannot be used at all for signals above about 100kHz, and even there it cannot really provide any gain. The LF351 offers a GBWP of 4MHz at a similar price, but this too is quite limited. The LMH6609 has a GBWP

of 900MHz.
As discussed above, the compensation capacitor inside the opamp reduces its bandwidth while improving stability. If the value of that capacitor is reduced then the bandwidth will go up, for no extra cost. The only drawback is that the designer must ensure that the opamp is stable in its feedback loop. Opamps which have this reduced capacitor are called “decompensated” opamps. They are generally not “unity gain stable”, which means that they will oscillate if put into low gain amplifier circuits. As an example, the LM6364 is only stable down to a closed loop gain of 5.

Generally these opamps will also feature very high slew rates, as required for large amplitude high frequency signals.

- *Precision:*

As implied by their name, precision opamps have a focus on implementing their governing equations as precisely as possible over a limited range of conditions. Generally, precision opamps will have very low input offset voltages and temperature drift. They will generally have relatively low bias currents, although not as good as JFET. The LT1097 has input offset voltages lower than 50uV and this will drift by less than 1uV per degree of temperature change. The bias currents are a few hundred picoamps, and the drift on this is less than 4pA per degree. They are usually quite expensive and often fall short in areas such as GBWP, so only use them where needed.

- *Zero Drift:*

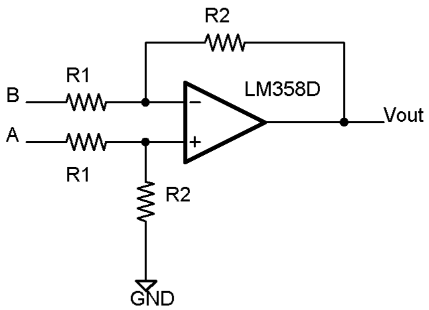
When designing precision circuits it is often possible to calibrate out the effects of input bias current, input offset and gain-related errors, sometimes by using trimmer potentiometers, digital potentiometers or software algorithms. This technique does not usually compensate for long term drift (change over time) in these parameters so the circuit may be highly accurate while in assembly but become inaccurate over time, or with temperature change. Zero Drift amplifiers consist of several (usually two) precision opamps combined with switches and sampling capacitors in a configuration which cancels long term change in amplifier parameters by sampling them and subtracting their effects from the output of the amplifier. As an example, the AD8628 from Analog Devices boasts an offset voltage of 10uV under worst-case conditions and drift by less than 0.02uV per degree of temperature change.

Differential Amplifiers and Instrumentation Amplifiers

The equation describing the following circuit is:

$$Vout=(V_a-V_b)*(\frac{R_2}{R_1})$$

Exercise: Derive this.

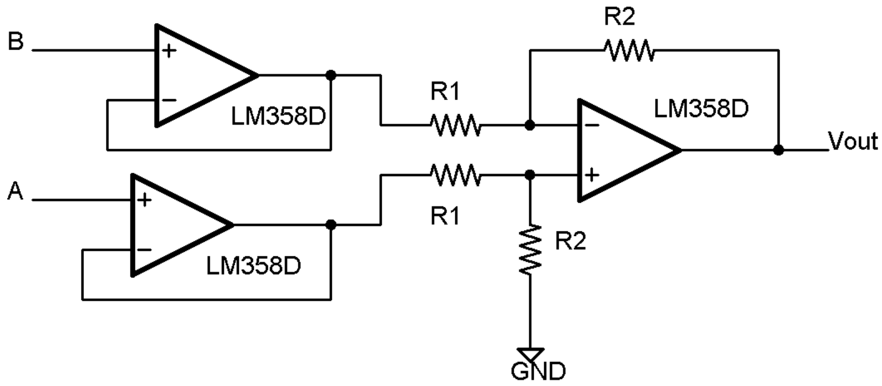


Source: Author's own diagram, (2013).

As you can see this amplifier subtracts one voltage from another and multiplies that by a gain. This is the simplest form of differential amplifier.

The impedances seen by the circuit feeding this amplifier are going to be different. Input A sees an impedance of R1+R2 while the impedance seen by input B will vary. This is problematic in some applications. Specifically, the amplifier's uneven loading will impose differential mode signals on the high impedance outputs and the common mode noise rejection of this amplifier is sub-optimal.

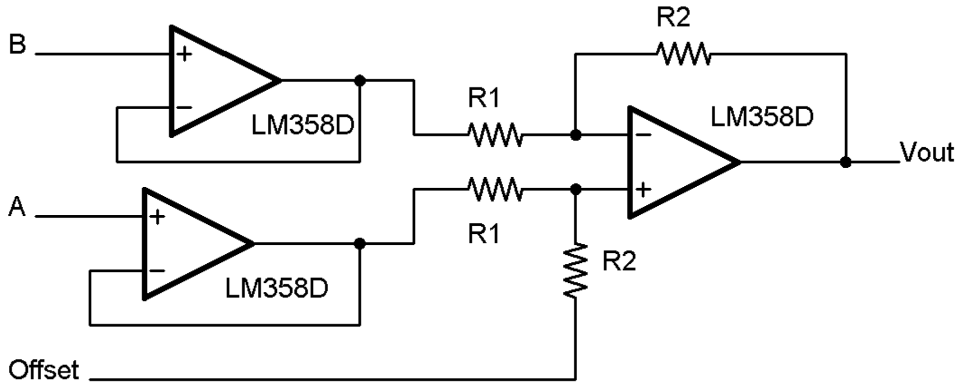
This problem can be fixed reasonably well by buffering the inputs with unity gain followers:



Source: Author's own diagram, (2013).

Now A and B are both high, and equal, impedances. Notice the use of unity gain followers in the input stage, rather than amplifiers with gain. If gain is used then gain mismatches between the two amplifiers would lead to poor common mode performance.

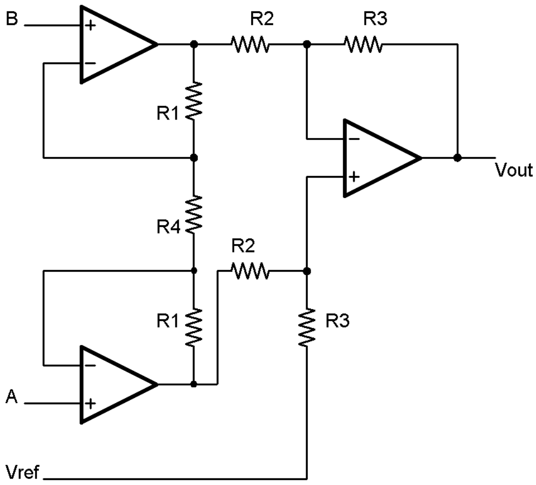
An offset feature can be added as follows:



$$Vout=(V_a-V_b)*(\frac{R_2}{R_1})+V_{offset}$$

Source: Author's own diagram, (2013).

In order to change the gain of this circuit two resistors need to be changed. Furthermore, they need to be changed to the same value. This is a nuisance if, for example, the gain needs to be trimmed to a precise value using a potentiometer. For that reason (and another), the following circuit is often preferred:



$$Vout=(V_a-V_b)*(1+(\frac{2R_1}{R_4}))*(\frac{R_3}{R_2})+V_{ref}$$

Source: Author's own diagram, (2013).

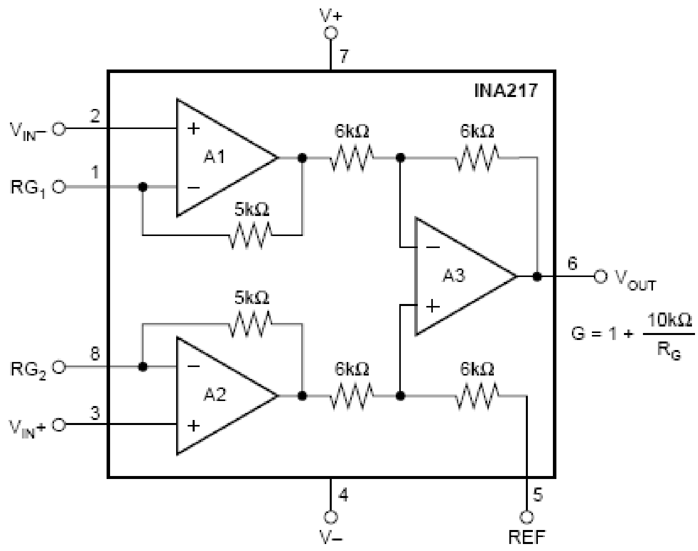
And now the gain can be set by simply changing the value of R4.

There is another, more subtle, advantage to this circuit. The input stage of the simpler buffered differential amplifier has a common mode gain of 1 and a differential mode gain of 1. Essentially, it does nothing to subtract one signal from the other, it just buffers them. This means that the differential amplifier stage is left to do all of the subtraction. This means that the resistor matching for high CMRR is critical. The slightly more complex *input stage* has a gain given by:

$$G_{sig}=1+\frac{2R_1}{R_4}$$
$$G_{common\,mode}=1$$

The input stage also rejects some of the common mode signal, thus reducing the need for high CMRR in the output stage. Setting R4 such that the amplifier has substantial gain will improve the overall CMRR of the amplifier.

Because precision differential amplification is such a useful function precision differential amplifiers are sold as integrated circuits and in that format they are called "Instrumentation Amplifiers".



Source: *INA217 Instrumentation Amplifier* 2002, Burr Brown Products (Texas Instruments Incorporated), Low-Noise, Low-Distortion Instrumentation Amplifier p.1.

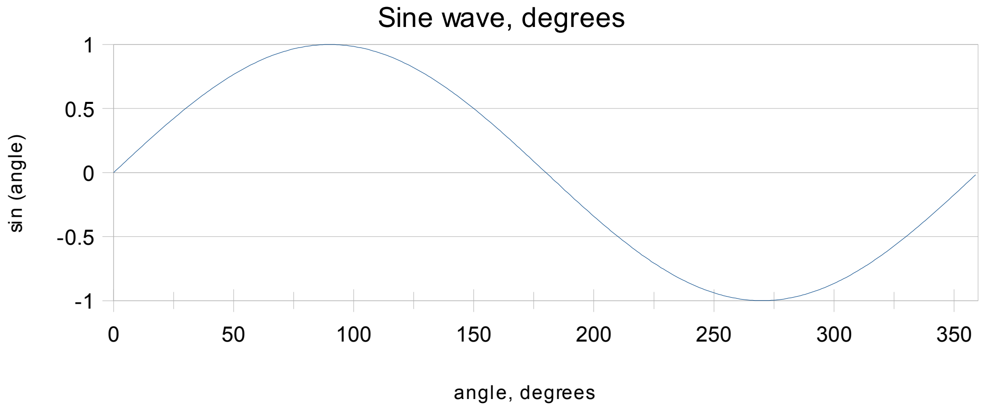
One will see that this is exactly the diagram from above, except that R4 has been brought out to external pins. When using the instrumentation amplifier R4 is added on externally (as a component on the circuit board) in order to allow the circuit designer to set the gain as needed.
Good quality instrumentation amplifiers achieve much better matching of their resistors than is feasible using discrete components in a production environment. Some instrumentation amplifiers have laser trimmed internal resistors to ensure extremely close matching. They tend to be expensive, but are much better for precision applications.

Power Amplifiers

If one looks at the datasheets for most opamps one can see that they can typically deliver about 20mA maximum output current and have a maximum supply rail of about 30V. This is fairly limiting for driving actuators, such as motors, loudspeakers etc. For higher output power we need to look at power amplification.
There are high powered opamps available, such as the mighty OPA541 from Texas Instruments, but they are very expensive and offer a limited number of options.
The vast audio amplifier market has produced numerous audio power amplifier microchips. These are often good for audio applications, but seldom good for other applications.
There is a growing market for miniature high efficiency audio amplifiers for use in cellular telephones, music players and other portable applications. (The principles of these amplifiers will be discussed in a later chapter).
In this subsection we will focus purely on linear power amplifiers built from discrete components.
Amplifiers are divided into classes. The two most common classes are called Class A and Class B, with a Class AB sharing characteristics of both. Class D amplifiers are becoming more popular, but they will be mentioned in a later section. The distinguishing feature of Class A versus Class B amplifiers is the conduction angle of the output stage transistors.

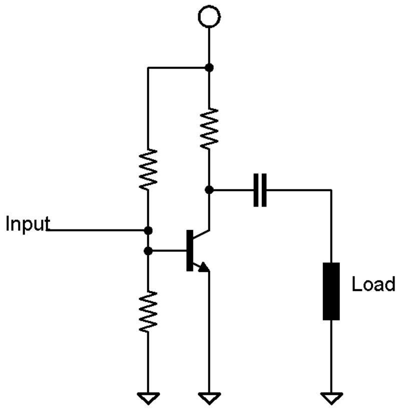
Class A Amplifiers

Class A amplifiers have a 360 degree conduction angle. This means that the power device conducts some current all of the time. The concept of conduction angle is shown here:



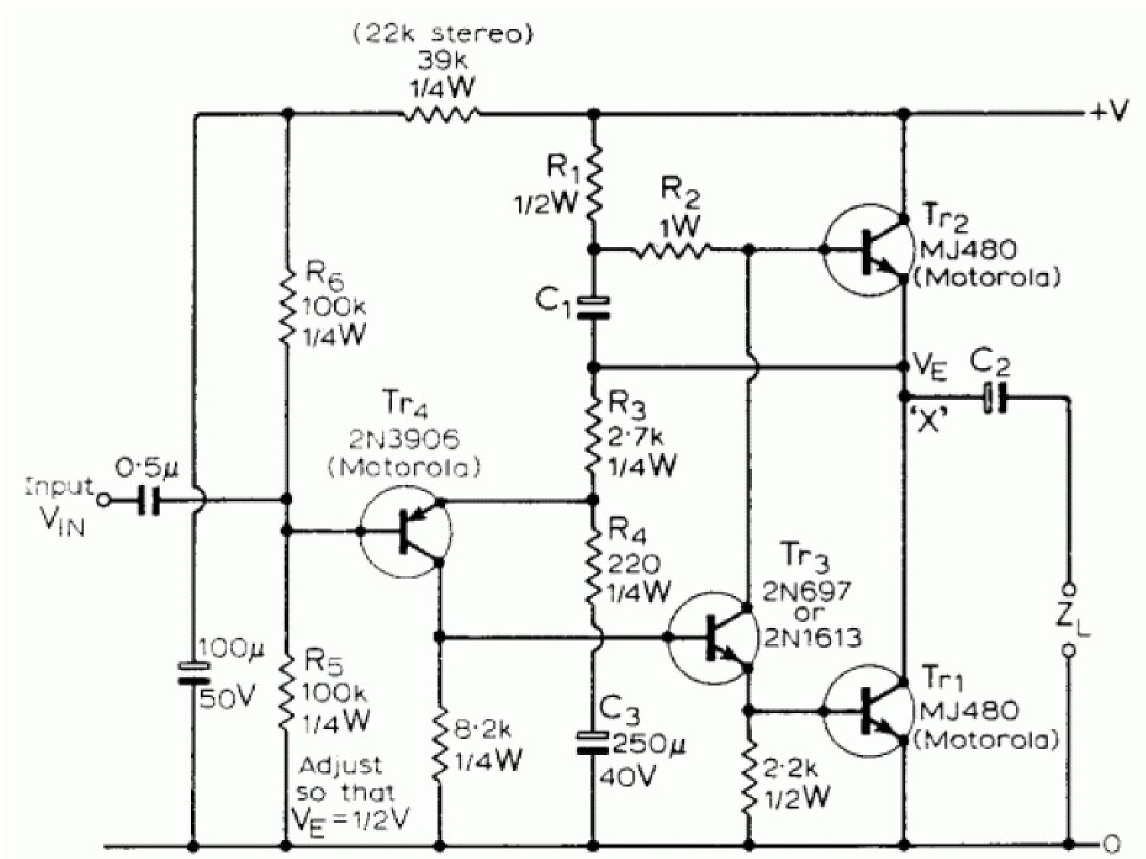
Source: Author's own diagram, (2013).

If the power device(s) conducts all of the time we say that it has a 360 degree conduction angle. A simple Class A amplifier circuit is shown here:



Source: Author's own diagram, (2013).

The base of the transistor is biased so that under all input conditions the transistor allows some current to flow. Essentially, this amplifier works by "throwing away" the power that is not going to the load. This makes it very inefficient. Class A amplifiers enjoyed a brief period of popularity among the ranks of audiophiles in the early days of transistor amplifiers, but the practical problems of making usefully powerful audio amplifiers make them generally unpopular. Class A amplifiers are the most linear amplifiers.
Suppose that a simple Class A amplifier is designed to deliver 20Watts into a 4 ohm load. This would be typical of a modest high fidelity audio amplifier. A 20W sine wave going into a 4 ohm load has an RMS voltage of 9V. This implies a peak-to-peak voltage of 25V. Thus, the supply rail of the amplifier must be 25VDC or preferably a bit more. The quiescent current in the amplifier must be greater than the peak current in the load. For a 20W sinusoid into a 4 ohm load this RMS current is 2.23A, translating into a peak current of 6.32A. The quiescent power dissipation of the amplifier is therefore 6.32A*25V=158W. Even when there is no signal coming out of the amplifier it would dissipate this. This massive inefficiency is usually unacceptable.
The circuit diagram for the power stage of a classic Class A audio amplifier is shown here:

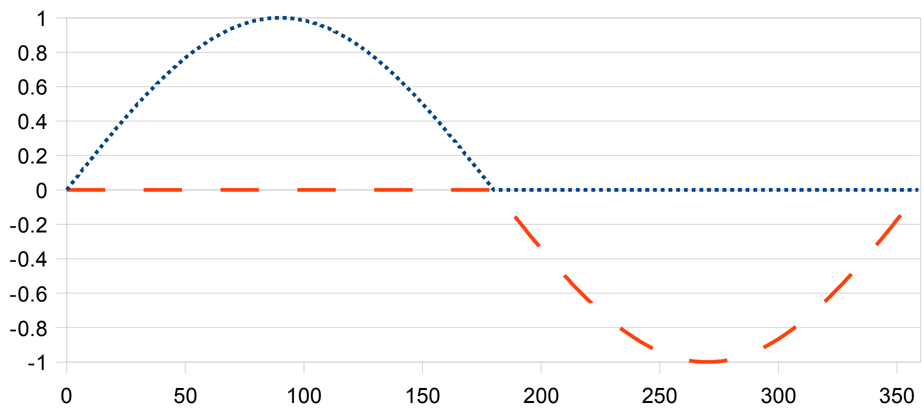


Source: *Class A Audio Amplifier* 1969, "Wireless World", John Linsley Hood.

Despite the fact that there are two transistors in the output stage they are both always in conduction, rendering this a Class A amplifier. TR2 is an active version of the collector resistor shown in the simpler diagram above.

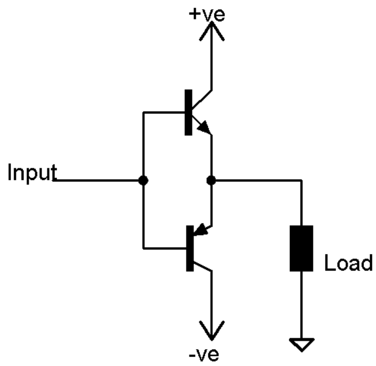
Class B Amplifiers

In a Class B amplifier there are two output devices and they each have a conduction angle of 180 degrees. This is illustrated in the waveform below. The fine dotted line illustrates the current flow through one of the output devices, while the dashed line shows the current flow through the other device.



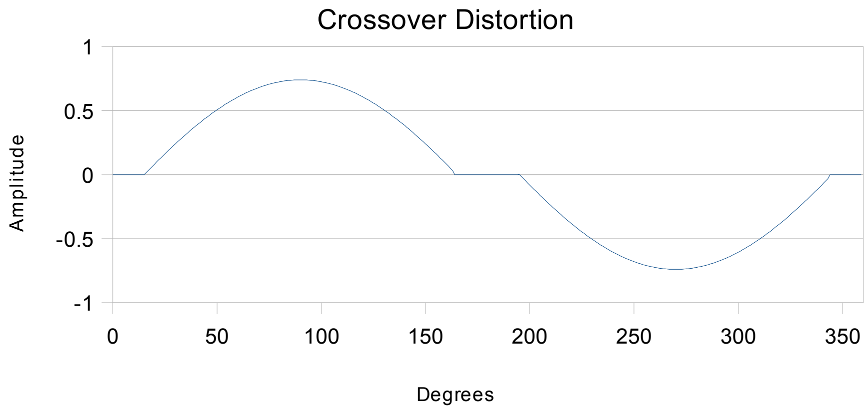
Source: Author's own diagram, (2013).

To understand the "push-pull" Class B amplifier, think of it as being two emitter followers connected together, biased so that only one of them is operational at any time.
The simplest Class B amplifier schematic is shown here:



Source: Author's own diagram, (2013).

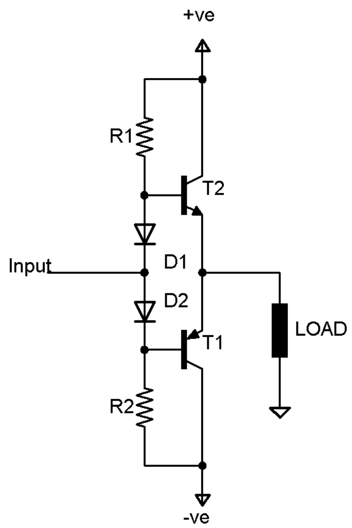
When the input is above ground potential, the top transistor will then conduct. When the transistor is below ground potential, the top transistor will then turn off and the bottom transistor will conduct. In reality, the transistors will only start to conduct when the input is about 0.6V above ground and the bottom transistor will only start to conduct when the input is more negative then 0.6V below ground. This means that the conduction angle of the transistors is actually a bit less than 180 degrees. If a sinusoid is fed into the input the following waveform will result:



Source: Author's own diagram, (2013).

This "notching" around the zero crossing is called "crossover distortion". It is extremely unpleasant to listen to in audio amplifiers and could result in jerky movement in actuator systems.

In order to more closely approximate ideal Class B operation we can bias the transistors so that they are just about to conduct when the input is at ground potential. This can be done by biasing the transistors like this:

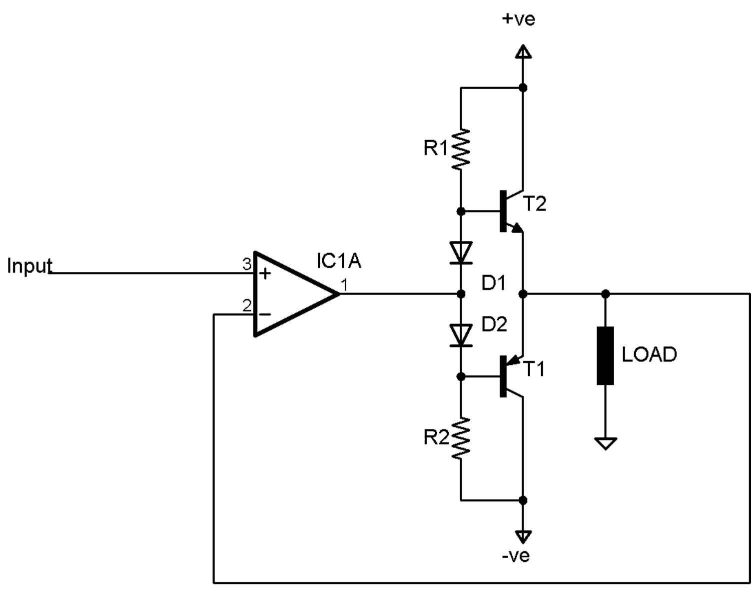


Source: Author's own diagram, (2013).

In theory, the diodes' voltage drops are the same as the turn on voltages of the transistors. Then, when the input is at ground, the base of T2 is at 0.6V and the base of T1 is at -0.6V. Theoretically, any deviation of the input will result in one of the transistors suddenly going into conduction.

In practice, this does not work out as neatly as one would like. The turn on voltages of the diodes and transistors are not perfectly matched. One of two things will happen: either a slight quiescent current will flow if the transistors conduct for slightly more than 180 degrees, or the transistors will still be off and will need a small input voltage before they conduct, resulting in residual crossover distortion.

The amplifier can be improved by putting the power stage in a feedback loop. One possible arrangement is shown here:

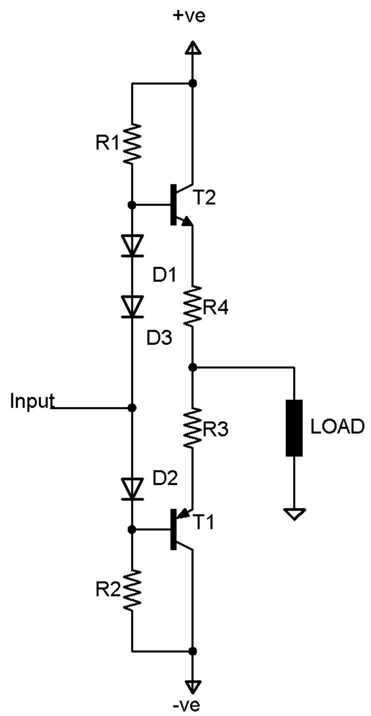


Source: Author's own diagram, (2013).

In this circuit the power stage has been included inside the feedback network of the opamp. The opamp compares the input to the output of the power stage and forces them to be equal. In theory, this would perfectly correct and crossover distortion (even if D1 and D2 were not present) because when the power stage is in the "crossover zone" the output of the opamp should immediately swing to correct the situation.
In reality, however, the opamp's slew rate is limited, resulting in crossover distortion (albeit reduced) even in this configuration.

Class AB Amplifiers

As their name implies these amplifiers have characteristics in common with both Class A amplifiers and Class B. The power devices conduct for slightly more than 180 degrees, and much less than 360 degrees. This means that the power devices allow a small (relative to the maximum) current through them even when the input signal is removed. This overlap in conduction angles eliminates crossover distortion.
A simple Class AB amplifier is shown here:

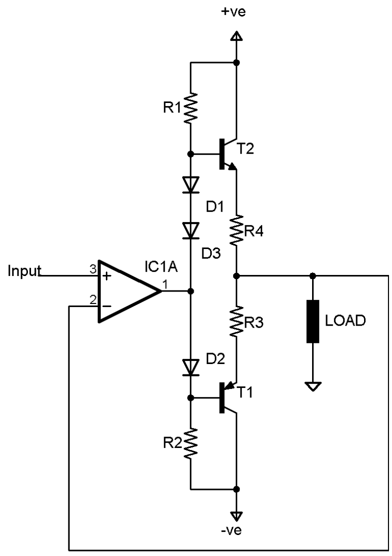


Source: Author's own diagram, (2013).

The three diodes establish a 1.8V drop across the bases of the transistors. Only 1.2V is required to turn the transistors on, so current flows through them. This current is limited by the emitter resistors, R3 and R4. The standing current (quiescent current) is approximately: Iq=0.6V/(R3+R4).
The forward voltage drops of silicon junctions are heavily dependent on their temperature. For this reason, the quiescent current will change slightly as the power devices warm up. Some circuits feature more sophisticated circuits in place of the third diode. These circuits often use resistor/transistor/zener combinations and often have a potentiometer so that the quiescent current can be adjusted.

Exercise: Draw a version of the Class AB amplifier with Darlington transistors.

The load impedance is effectively in series with either R3 or R4. The amplifier's output impedance is therefore not as low as it could be. Also, the amplifier is not perfectly linear because the simple emitter follower model is not perfectly accurate. Both of these issues can be improved by putting the power stage in a feedback loop, as before:



Source: Author's own diagram, (2013).

Suppose that we want to make a Class AB amplifier which can deliver 20W into a 4 ohm load. As in our analysis of the Class A amplifier, we need a total supply voltage of 25V. In this case, for convenience we will use a dual 13V supply.
The amplifier is (roughly) symmetrical. To a first approximation (because the quiescent current is small compared to the peak current) we can analyse the positive-going half of the amplifier in the same way as the negative half.
Each transistor has half the supply voltage across it. Thus, each transistor has a maximum of 13V across it.
The amplifier has no energy storage elements, so the RMS current supplied to the load must be the same as the RMS current supplied by the supply. The RMS current in the load is given by $P=I^2R$. This gives 2.23A. The power supplied to the amplifier by the power supply is then 13*2.23=29W. Of this 29W, 20W goes to the load. The remaining 9W is shared among the two transistors, so each transistor dissipates 4.5W as heat.
Unlike the Class A amplifier, the Class B or AB will dissipate less power as the output signal is reduced.

Maximum Power and Power Dissipation
There are several factors limiting the power that a power amplifier can deliver to the load.

Supply voltage:
Suppose that the amplifier is running off a dual 12V supply. Because the output stage of the Class AB amplifier that we have discussed is an emitter follower, the maximum output swing is approximately 11V above and below ground. If the amplifier is outputting a sinusoid then it follows that the maximum amplitude of that sinusoid is 11V. The RMS voltage of that waveform is 11/sqrt(2), which is 7.8V.
The power delivered to the load is V^2/Z_{load}. In many amplifiers the supply rail is the limiting factor for the output power.
In the case of a 4 ohm loudspeaker this amplifier could deliver a maximum of 15W RMS.

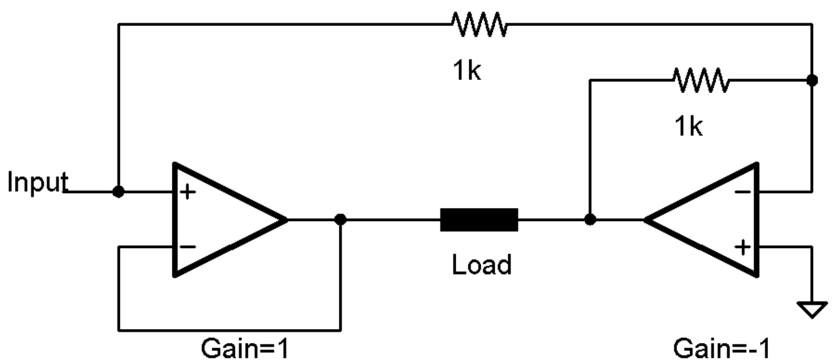
Maximum current rating of power devices:
Power transistors have a maximum collector current rating. This cannot be exceeded or they will be destroyed. Care must be taken; the specifications for power transistors often give a maximum average (RMS) current, as well as a maximum peak current. The lower of these ratings will determine the maximum power.
The calculation of maximum power is then based on the equation $P=I^2R$.

Maximum power dissipation of power devices:
The amount of power that the power transistors (or collector resistor, in the case of the simple Class A amplifier) can dissipate is limited by the temperature rise of those devices. Power transistors in amplifiers are often mounted on heatsinks in order to reduce the temperature rise per watt of dissipated power. (This topic will be dealt with later).

Power supply capabilities:
Naturally, the amplifier depends on its power supply to deliver sufficient supply to it in order to drive the load. This is a limiting factor in many practical applications.

Bridge Amplifiers

Consider the scheme shown here:



Source: Author's own diagram, (2013).

When the input is positive the output of the amplifier on the left also becomes positive. The output of the amplifier on the right becomes negative. This means that the load will see twice the voltage swing that it normally would.

What is the gain of this circuit?
If the circuit was running off a dual 15V rail and the input is a sinusoid, what is the maximum RMS voltage that can be put across the load? How does this compare to the previous amplifiers that we have seen?

Bridge amplifiers are popular for applications where the supply voltage constrains the maximum output power of the amplifier. The load is not ground referenced, which may be a problem in some cases, (think of a stereo headphone socket- why would this be problematic?) but often the advantages outweigh the problems. Class B and AB power amplifiers can also be used in this configuration.

Chapter 7: Noise and Frequency Response

Types of Noise

Electrical noise can come from a variety of sources and exhibit a variety of different characteristics. Understanding these is useful in reducing their effect.

- External Noise

  This is noise that is radiated into our system by other electronic/electrical systems. The most common example of this is 50Hz noise which is radiated by power conductors.
  There are two ways in which noise radiates in to a system:

    Capacitive coupling of electric fields will occur when an external source radiates an electric field.

    Inductive coupling of magnetic fields will occur when an external source radiates a magnetic field.

  In general, external noise can be minimised by shielding your system, including cables, connectors and circuitry. One can buy small metal boxes that solder down onto circuit boards to shield sensitive circuitry from external noise sources.
  The layout of the system (including circuit boards) can also help minimise noise ingress.
  Also, one's system must not radiate excessive noise or it will interfere with other systems. The mutual good behaviour of electronic systems is known as Electromagnetic Compatibility (EMC) and is regulated by law in some countries and in some product sectors in other countries.

  Another good example of external noise is noise generated by nearby switching elements, such as power electronics and relays. This noise tends to be very impulsive ('spiky') and is thus rich in high frequency components that propagate easily through systems.

- Internal Noise

  This is noise generated by the components from which our circuit is built. Some commonly found types include:

    Component Noise

    Thermal Noise (Johnson Noise)

      At any temperature above absolute zero thermal agitation will cause the electrons in a material to move. This results in thermal noise. This type of noise can be reduced by cooling the system down, and for this reason extremely low noise systems are sometimes cooled. An example of this is the 'ultra low noise, high gain' amplifier used on the antenna of a radio telescope.

Thermal noise voltage across a resistor is given by the equation:

$$V_{(noise[rms])}=\sqrt{(4\text{kTRB})}$$

where:

k is Boltzmann's constant ($1.38*10^{-23}$)

T is the temperature in Kelvin

R is the resistance of the resistor

B is the bandwidth (in Hz) that your system can pass.

Example: Suppose that one has a noise free amplifier with a gain of 1000 and a bandwidth of 15kHz. Suppose that the transducer feeding into this amplifier has a source impedance of 10 kilo ohm. What will the output noise voltage be if the circuit is run at 300 kelvin?

$$V_{noise\,rms}=1000*\sqrt{4\text{k}*300*10000*15000}=1.6\text{mV}$$

This is another reason for avoiding unnecessarily high value resistors in circuitry. What was the other reason?

- Shot and Partition Noise

These types of noise are caused by random fluctuations of electron movement in semiconductor material. When we consider a current flow we abstract it to a number which represents the average movement of electrons. At a 'smaller' level electrons behave in a much more chaotic way. This leads to random current fluctuations, visible to us as noise. The noise has a flat power spectrum.

- 1/f Noise

This is also called “flicker noise” and “excess noise” and “current noise”. This is caused by fluctuations in the conductivity of materials. Examples are to be found in resistors (where different constructions exhibit different levels of noise) and in semiconductor materials. The spectrum of this noise is inversely proportional to frequency, hence the 1/f name.

- System Noise

This is noise introduced by the system into its own signals.

Intermodulation Distortion

If the sum of multiple sine waves is fed into a perfectly linear system the output of that system will only contain spectral components at the input frequencies. If that same input is fed into a non-linear system then the output will contain spectral components at the input frequencies as well as “beat” frequencies, i.e. sum and difference components of the input frequencies. This is called intermodulation distortion as it is the result of input frequency components modulating each other. While not strictly noise, this effect often has much the same effect on a system.

Noise Colour

Often the spectral characteristics of noise are described by referring to them as a colour. The phrase "white noise" has entered popular language to describe all sorts of disturbance, however, it has a strict scientific definition.

> White Noise:
> This is noise having a flat spectrum. The noise power is constant over frequency.
> Pink Noise:
> This is noise having a 1/f spectrum, which means that as you examine higher frequency bands of the same width you get less noise power.

Because of the different spectral characteristics of different noise sources the dominant noise type will change as the operating frequency of one's system increases. This can have design implications for parameters such as resistor values, quiescent current choices etc.

Signal to Noise Ratio

In many applications noise is only a problem if it is large compared to the signal of interest. As a demonstration of this imagine taking a small transistor radio and tuning it to a "blank" frequency; one would hear a hiss. Turn on some soft music in the same room. One would notice that the hiss is an irritation. Now turn up the music's volume. One will notice that the hiss vanishes into the background and is no longer a problem. For this reason we often discuss the Signal to Noise Ratio (SNR) of a signal rather than the amplitude or power of a noise source. Because this is a ratio it will often be specified in decibels.

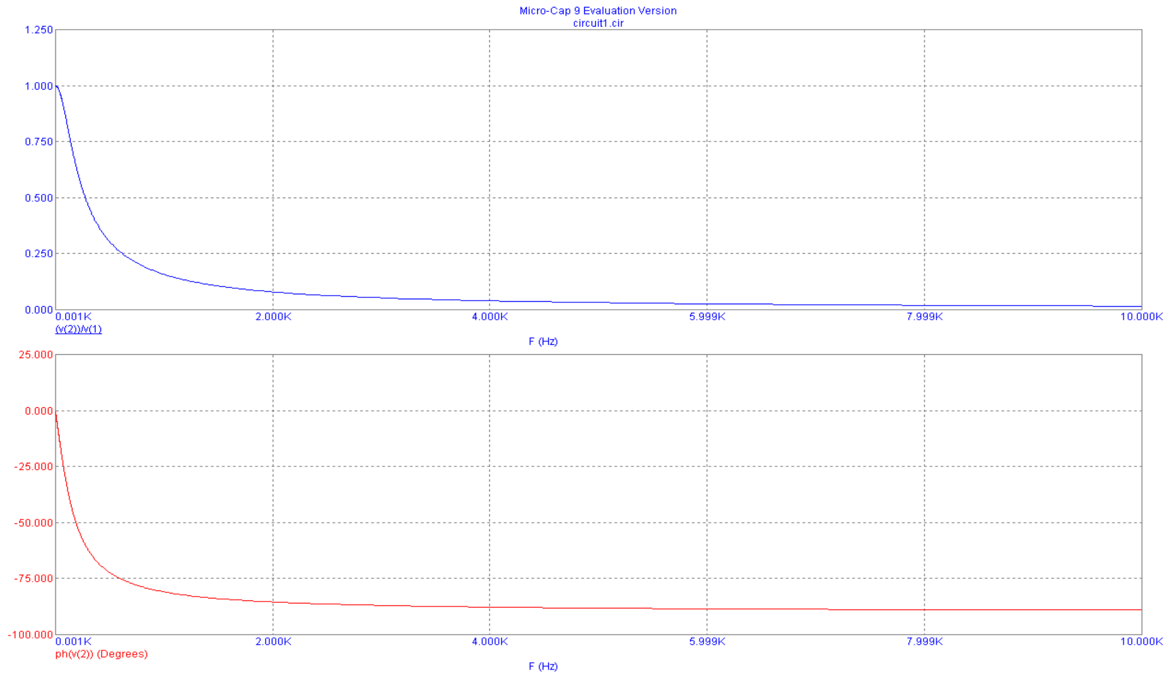
The noise power present in a signal will depend on the bandwidth under consideration. The wider the bandwidth the more noise will be present. For this reason it is sometimes useful to refer to noise-voltage density, V_n .
Then $RMS\ noise\ voltage = V_n \sqrt{Bandwidth}$
Notice that the RMS noise voltage increases with the square root of the bandwidth because of the uncorrelated nature of noise.
Noise is also specified using units of temperature. This is sometimes used to compare the noise performance of systems. A noisy system can be replaced (conceptually) by a noiseless system fed from a real resistor at some temperature. The noise power produced by that resistor depends on its temperature. The temperature to which that resistor must be raised to give the correct overall system noise power is the "noise temperature".

Exercise: A signal of 1V RMS has 10mV RMS of noise present on it. Express the SNR of the signal in dB.

Frequency Response

All circuits change their behaviour as their operating frequency changes. Some circuits do this deliberately, such as filters, and some circuits do it as an unavoidable consequence of their construction, such as opamps. In order to characterise this change we are often interested in the frequency versus various other parameters. Most commonly, we are often interested in frequency versus gain (or attenuation) and frequency versus phase angle, with some fixed reference, often the input signal.

A simple first order low pass filter has the following frequency response curve. The top plot is the output amplitude as a fraction of the input amplitude. The bottom plot is the output phase angle with reference to the input waveform. Both are plotted over frequency.

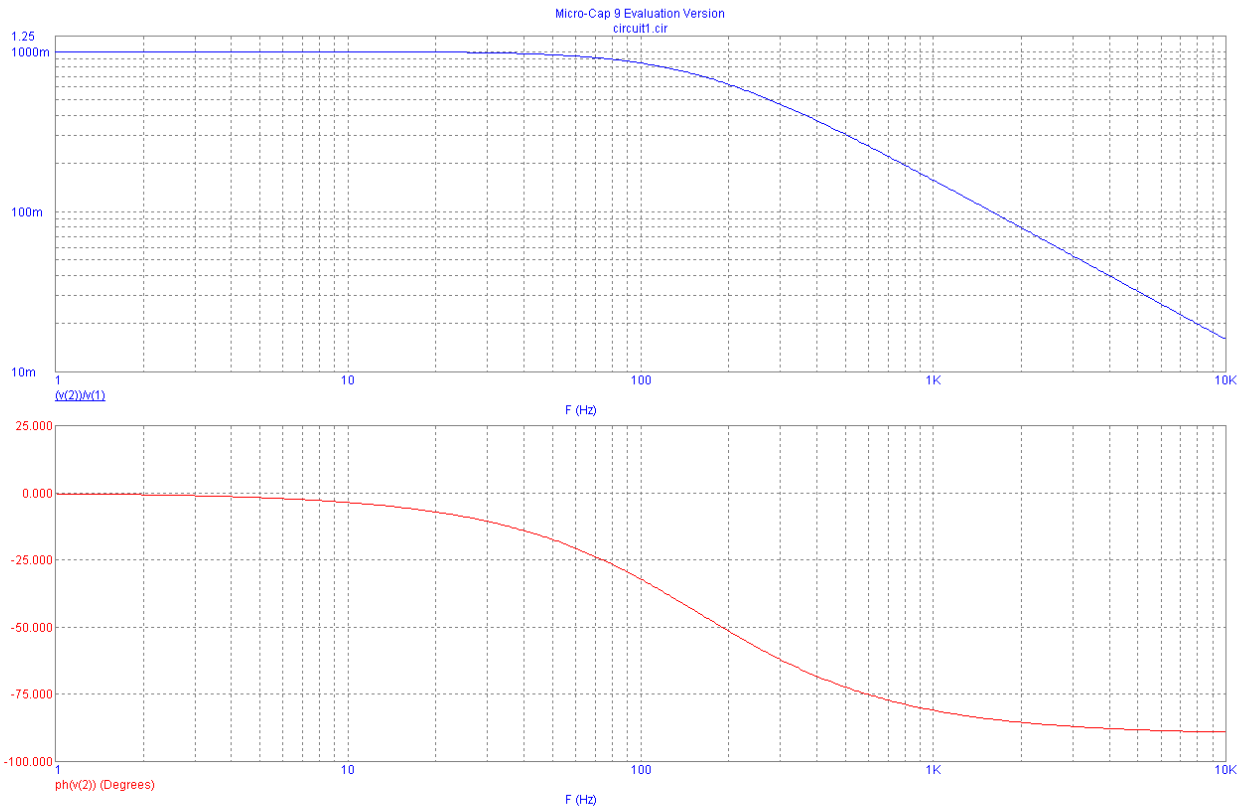


Source: Author's own diagram, (2013).

Linear scale plots are not good for representing very wide dynamic ranges. In the above graphs most of the space was taken up by a long “tail”.

Bode Plots

If we plot frequency response over a logarithmic scale often we get a very convenient result. Here are the same graphs as above, but on a log frequency scale:



Source: Author's own diagram, (2013).

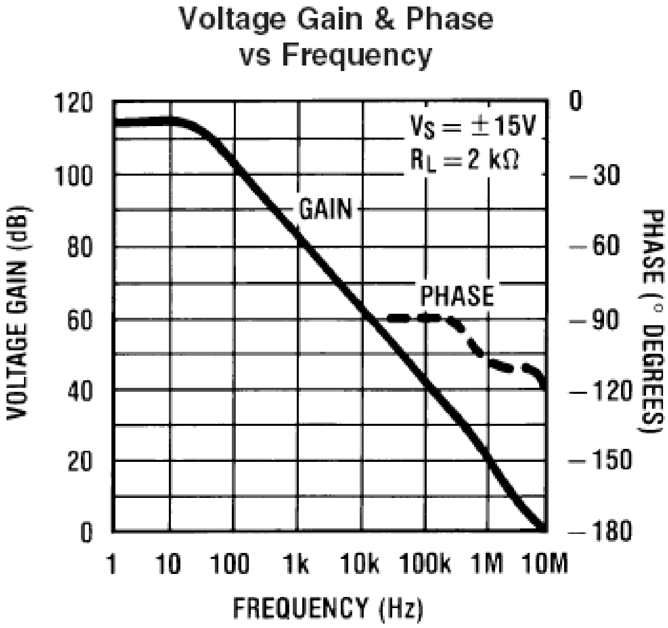
The magnitude's dependent axis is logarithmic, and the phase response is measured on a linear axis. A Bode plot is simply a frequency response plot with logarithmic axes.
The magnitude plot reveals that the filter has a fairly flat frequency response up to about 100Hz. If the vertical axis is calibrated in decibels one would see that the -3dB point is at 159Hz. One would also see that after that frequency the filter's attenuation increases by 20dB/decade of frequency, which corresponds to a straight line on the Bode plot. 20dB is a ratio of 10:1 in voltage.

If one knows the cut-off frequency of a filter, the type of filter (high pass, low pass, etc.) and the slope of its “roll-off” then one can easily sketch bode plots by hand. The roll-off will generally be 20dB/decade per cascaded first order filter stage, assuming that the stages are prevented from interacting.

Exercise:

> *Sketch the Bode plot for a low pass filter consisting of two cascaded stages, one with a 10k resistor and 100nF capacitor and the second stage with a 1Megaohm resistor and a 1 nanofarad capacitor.*

Here is an example of an instance where it would be preferable for the gain of the circuit to remain constant over frequency:

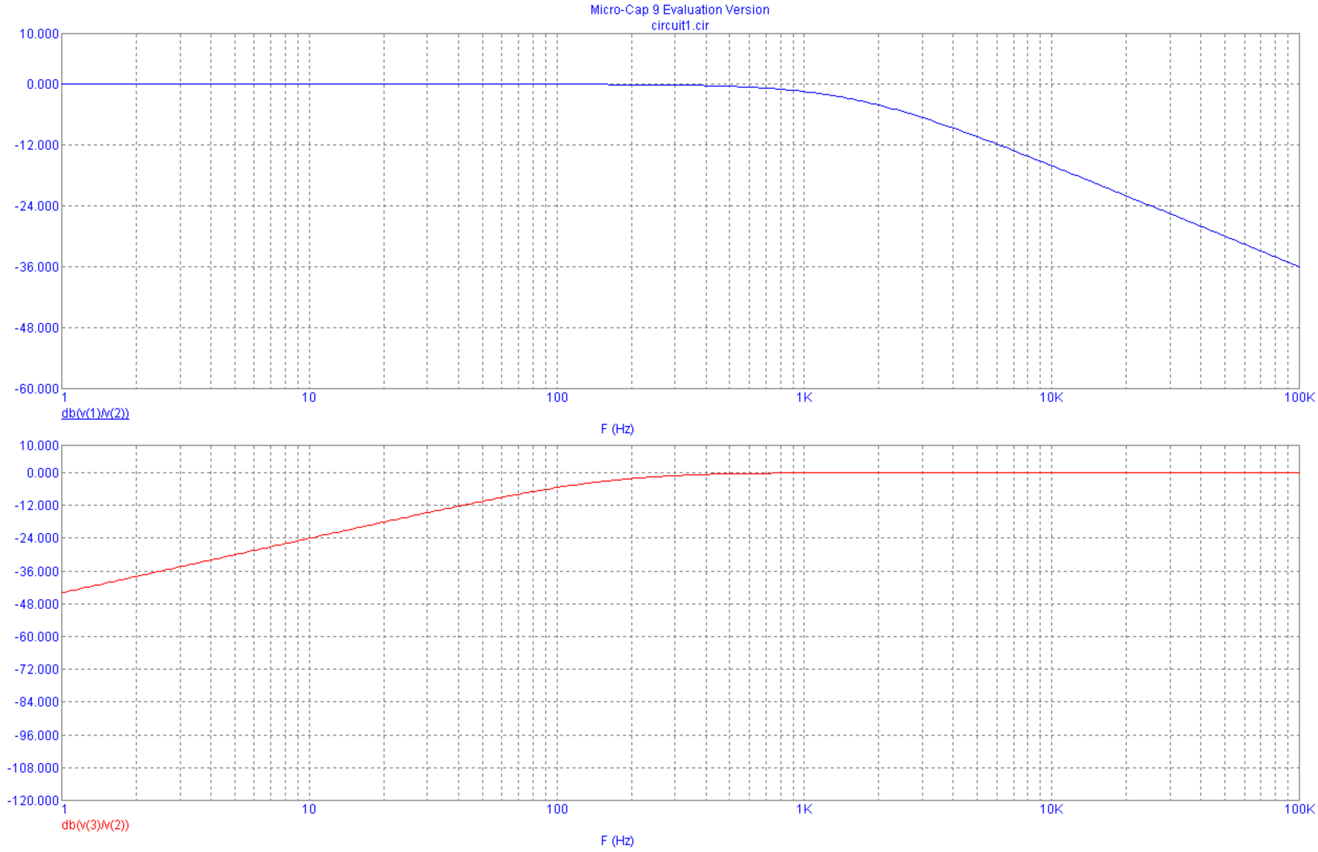


Source*: LM833 Audio Operational Amplifier* 2003, National Semiconductor, "Dual Audio Operational Amplifier" Datasheet, p.5.

This is the bode plot of the open loop gain of the LM833 "Audio Operational Amplifier" made by National Semiconductor. Note the 20dB/decade roll-off, consistent with a single pole caused by the compensation capacitor inside the opamp.

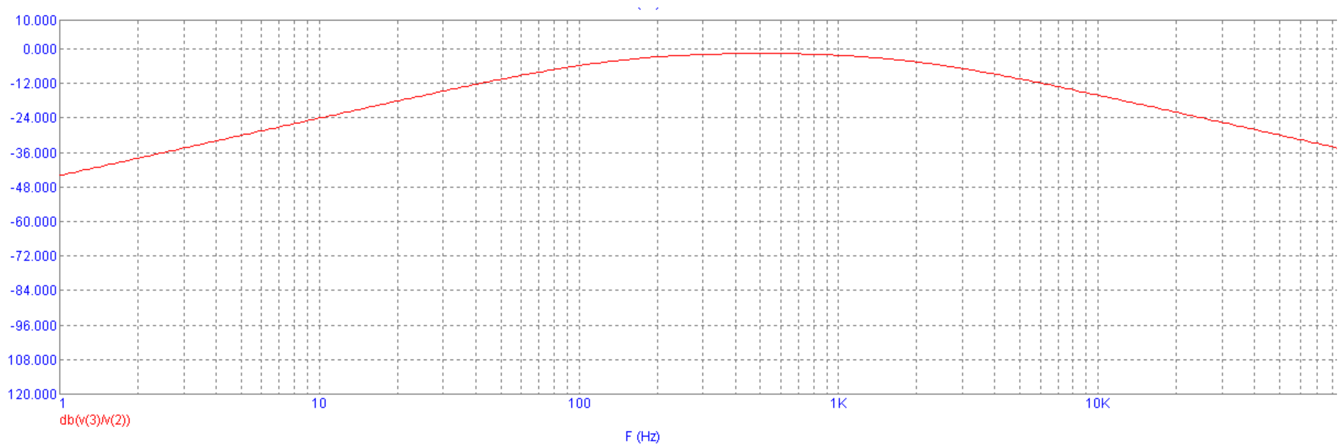
Phase is plotted against a linear scale on the right vertical axis. Notice that the phase shift at unity gain is -120 degrees. The datasheet specifies a "phase margin" of 60 degrees, which is an indication of the maximum phase shift that the feedback circuit can have while maintaining stability. (More detail on this will be given later).

The transfer functions of cascaded circuit stages multiply together. This is quite easy to imagine intuitively. We know that log (AB)=log (A)+log(B). This property leads to another useful feature of Bode plots; the ability to get a system's transfer function easily by adding the plots of the subsystems. As an example of this, suppose that we have a filter consisting of a low pass stage followed by a high pass stage. The Bode plots for the two stages are shown here:



Source: Author's own diagram, (2013).

The cascaded filter's Bode plot looks like this:



Source: Author's own diagram, (2013).

(We will use Bode plots extensively in further sections).

Chapter 8: Filters

The purpose of a filter

A filter's function is to process signals based on their frequency. Generally filters modify signals by attenuating them based on frequency or by rotating their phases based on frequency, and most filters perform a combination of these two.

Implementation Techniques

It is possible to build equivalent filters in a variety of ways. Four of these methods are described here. For each of these methods there are strengths and weaknesses, often relating to operating frequency, size and cost of implementation. When designing a filter it is important to consider which filter properties are important and make the design decision based on the feasible methods.

Passive

Passive filters consist of passive components, which are resistors, capacitors and inductors. Generally speaking, the higher the operating frequency the lower the inductances and capacitances in the filter will be. Conversely, the lower the frequency the higher these values will be. As an example consider the LC resonant circuit, with resonant frequency given by:

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

For a filter with a band-pass at 100MHz (used to tune an FM radio for example), typical values of C and L are 10pF and 250nH. These are both small and cheap components. On the other hand, if the resonant frequency must be 50Hz the C and L are 10uF (which is the biggest non-polarised capacitor that is easily available) and 1H, which is both physically big and expensive and difficult to obtain.

Another problem is that passive filters have no gain. This leads to problems in designing complex filters because of loading effects when stages are cascaded. A low pass filter made of a 100nF capacitor and a 10k resistor has a 20dB/decade roll-off and a -3dB frequency of 159Hz. Cascading two of these will not give a 40dB/decade roll-off from 159Hz because of the loading on the first stage caused by the second stage. Generally, an amplifier or buffer will be needed between stages, although careful component choice can help a bit.

In general, passive filters are a good choice at higher frequencies and sometimes at lower frequencies for lower order filters.

Active

Active filters are built up out of passive components and active components; typically resistors, capacitors and amplifiers. Generally they lack inductors, which is a big advantage at low frequencies. Opamps do not often offer good performance at high frequencies, with the result that it is difficult to make active filters with high operating frequencies. This limitation, combined with the decreasing size of inductors with increasing frequency, means that the operating frequency will often determine whether an active or passive filter will be most suited to an application.

Switched Capacitor

It is possible to create filter circuits using capacitors, analog switches and amplifiers. Essentially the switched capacitors emulate resistors. These filters are available as microchips with complete high order filters. They have strong advantages in terms of tunability, stability and ease of implementation. They have significant disadvantages in that they have high noise floors and switching noise invariably breaks through to the output signal, causing contamination. Their operating frequencies are fairly limited.

Digital

Another way of filtering a signal is to sample it using an analog to digital converter. A software filtering algorithm is applied to the data captured and the filtered signal is then sent out via a digital-to-analog converter. The frequency response is limited by the speed of data conversion and the processing power available to perform the calculations. Extremely high order filters with moderate frequency responses can be built. (This method is in the realm of digital signal processing (DSP) and will not be dealt with further here).

Filter Types

A filter is classified by its frequency response. There are a few important parts to the response of a filter:

- Passband. In this frequency range the attenuation of the filter is small; the filter passes the signal through. A certain maximum attenuation is defined (depending on the application) and the filter will have less attenuation than this. For many simple filters this level is set at 3dB, which means that the filter's output magnitude is greater than $1/\sqrt{2}$ of the input magnitude. This translates to a half power level.
- Stopband. In this frequency range the attenuation of the filter is large; the filter stops the signal from passing. A certain minimum attenuation is defined and the filter will have more attenuation than this.
- Transition Band. This is the band of frequencies between the passband and the stopband.

Highpass

This filter allows high frequencies through and removes low frequencies.

Exercise: Draw the bode plot for a high pass filter showing the passband, stopband and transition band.

Lowpass

This filter allows low frequencies through and removes high frequencies.

Exercise: Draw the bode plot for a low pass filter showing the passband, stopband and transition band.

Bandpass

This filter allows a band of frequencies through and stops all other frequencies. As an example suppose that one wants to "pick out" a signal of approximately 40kHz (useful for ultrasonics). One would pass the signal through a bandpass filter with a passband from 38kHz to 42kHz to remove extraneous noise from the signal. In the ideal case one would want a very narrow passband, but in reality the passband frequency will depend on the tolerances of the filter's components and using too narrow a passband might well result in the desired signal falling outside of the real-world filter's response. It should be noted that different filter types are differently sensitive to component tolerances.

Exercise: Draw the bode plot for a band pass filter showing the passband, stopbands and transition bands.

Notch

The notch (also called bandstop) filter removes a range of frequencies from a signal. As an example, it is often desirable to remove 50Hz interference from signals; a notch filter can be used to do this.

Exercise: Draw the bode plot for a notch filter showing the passbands, stopband and transition bands.

Allpass

The allpass filter modifies the phase of a signal depending on frequency. The attenuation is constant over the whole frequency range. This can be used to correct phase shifts in signals which have undergone phase distortion. One interesting circuit that is closely related to the all-pass filter is the phase sequence network, shown on page 295 of "The Art of Electronics". This circuit can produce quadrature outputs from a pair of inverted sine waves and it has a flat magnitude response over frequency.

Transfer Functions and Complex Impedance

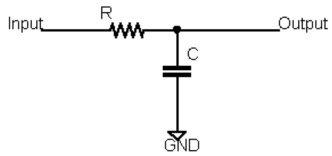
All filters can be characterised by their complex transfer function. The transfer function of a filter is useful in that it allows us to calculate and express the entire magnitude and phase response. It also allows us to determine at a glance if the filter is high pass or low pass and what the order of the filter is.

The concept of complex impedance should be familiar from work that was covered previously. It should be known that:

$X_c = \frac{(1)}{(sC)}$ for a capacitor

$X_l = sL$ for an inductor

Using these complex impedances we can apply normal network theorems to get filter responses.
Consider this circuit:



Using Ohm's law and complex impedance we know that the current flowing from the input to ground (gnd) is

$$i = \frac{V_i}{R + X_c}$$

$$i = \frac{V_i}{(R + (\frac{1}{(sC)}))}$$

$$i = \frac{sCV_i}{sRC + 1}$$

The output voltage (with respect to gnd) is the product of current in the capacitor and its complex impedance:

$$V_o = iX_c$$

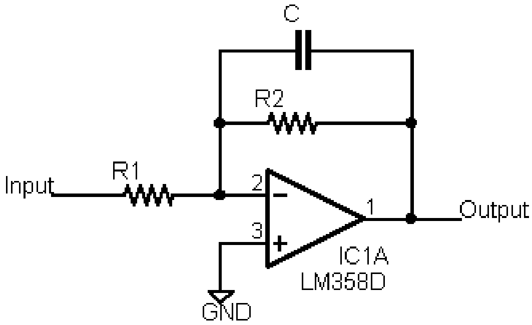
$$V_o = \frac{sCV_i}{sRC + 1} \frac{1}{(sC)}$$

$$\frac{V_o}{V_i} = \frac{1}{sRC + 1}$$

Imagine that *s=jw* (for a limited set of cases). One will see from the above expression that as *s* gets larger the gain of the filter drops. We expect this behaviour from a low pass filter. For very low frequencies sRC<<1 and the filter response is approximately 1. At higher frequencies sRC is

significant and gain drops at a rate of 1/f which is equivalent to the 20dB/decade roll-off. The output is never zero, even at infinite frequency.
One can derive the magnitude response from this filter by calculating the modulus of the complex response at any frequency. Multiplying the numerator and denominator of the response by the denominator's complex conjugate is a useful technique.

Here is your first active filter circuit:



Source: Author's own diagram, (2013).

Two parallel impedances result in:

$$X=\frac{1}{\frac{1}{X_1}+\frac{1}{X_2}}$$

Thus, the parallel combination of C and R2 is:

$$X=\frac{1}{\frac{1}{R_2}+\frac{1}{1/sC}}=\frac{R_2}{1+s\,R_2\,C}$$

and the gain of the circuit is:

$$\frac{V_o}{V_i}=\frac{-X}{R_1}=\frac{-R_2}{R_1+s\,R_1\,R_2\,C}$$

One can see from the transfer function for this filter that this is also a low pass filter. The DC gain is set by the ratio of R1 and R2, as would be expected. The output impedance of this filter is low, because of the opamp's low output impedance. This is a substantial advantage over the passive version of the filter, which has a higher output impedance. Cascading active filters is often more practical than cascading their passive equivalents.

We can cascade first order sections to create a wide variety of filters. Cascading active sections such as this can create filters with extremely steep roll-offs. However, they are not particularly good for creating sharp “knees” at the cut-off frequency. We will shortly see a technique for dealing with this.

Here is a table of transfer functions for some common filters. We assume a knee point of 1 rad/sec.

Filter Type	First Order Transfer Function	Second Order Transfer Function
Lowpass	$\frac{1}{as+c}$	$\frac{1}{as^2+bs+c}$
Highpass	$\frac{s}{as+c}$	$\frac{s^2}{as^2+bs+c}$
Bandpass	*Exercise: Why is this block blank?*	$\frac{s}{as^2+bs+c}$

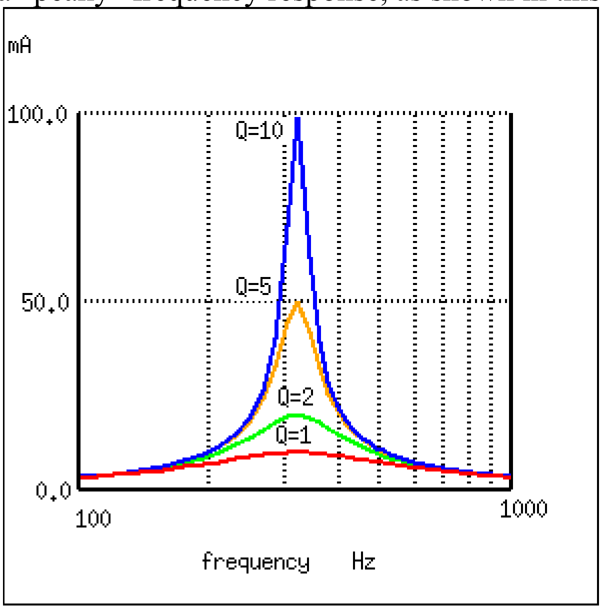
Quality Factor

The Quality factor, or "Q" of a circuit, is a measure of energy loss in a circuit per cycle. In terms of energy: $Q=2\pi\frac{Maximum\ energy\ stored}{energy\ lost\ per\ period}$

In terms of transfer functions the coefficient 'b' in the above table is the reciprocal of Q. b=1/Q.

The -3dB bandwidth of a resonant filter is $f_{resonant}/Q$

Circuits with high Q have a "peaky" frequency response, as shown in this set of curves,

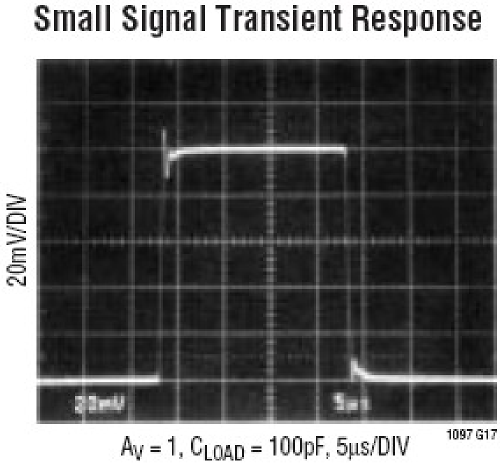


Source: "Q and bandwidth of a resonant circuit", *http://www.allaboutcircuits.com/vol_2/chpt_6/6.html*

In the past we have only seen high Q factors in RLC circuits, but we wish to eliminate inductors as far as possible. Our active filter circuits can produce high Q factors using only resistors, capacitors and amplifiers.
In the time domain we say that circuits with a high Q are "underdamped". These circuits will exhibit "ringing" if they are excited with a step pulse (such as a digital signal). As a very rough (but

useful) estimate the number of ringing cycles seen on the oscilloscope screen after a sharp edge is roughly equal to Q.

This oscilloscope trace is taken from the datasheet for a Linear Technology LT1097 precision opamp datasheet:



Source: *Typical Performance Characteristics of the Precision Opamp* 1989, Linear Technology, LT1097 Low Cost, Low Power Precision Opamp Datasheet, p. 7.

Note the overshoot and ringing at the edges.

Circuits that have too much energy loss per cycle are overdamped. They have the effect of slowing edges down in the time domain.
For preservation of pulse shapes in the time domain we usually want a critically damped circuit.

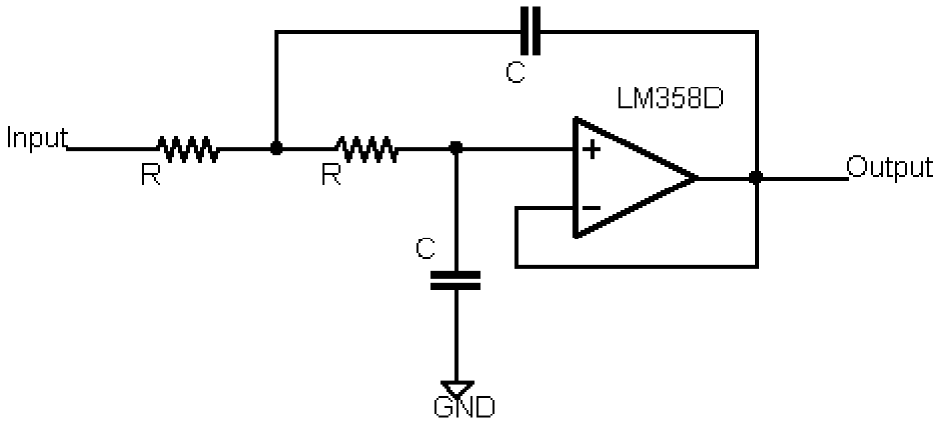
Common Active Filter Configurations

Sallen and Key

The Sallen and Key filter is an active second-order filter which has a sharper transition from passband into transition band (also called the filter's "knee") than a simple cascaded filter. It can be made high pass or low pass or band pass by exchanging resistors and capacitors.
The -3dB knee frequency is given by:

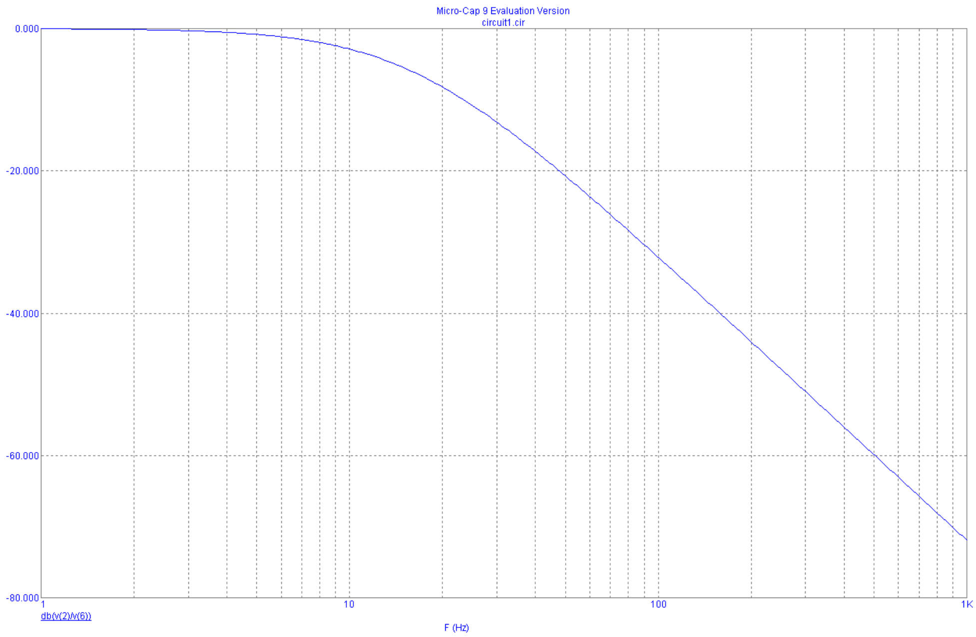
$$f = \frac{1}{2\pi RC}$$

Here is the circuit for a low pass version:



Source: Author's own diagram, (2013).

Here is a Bode plot of its response:

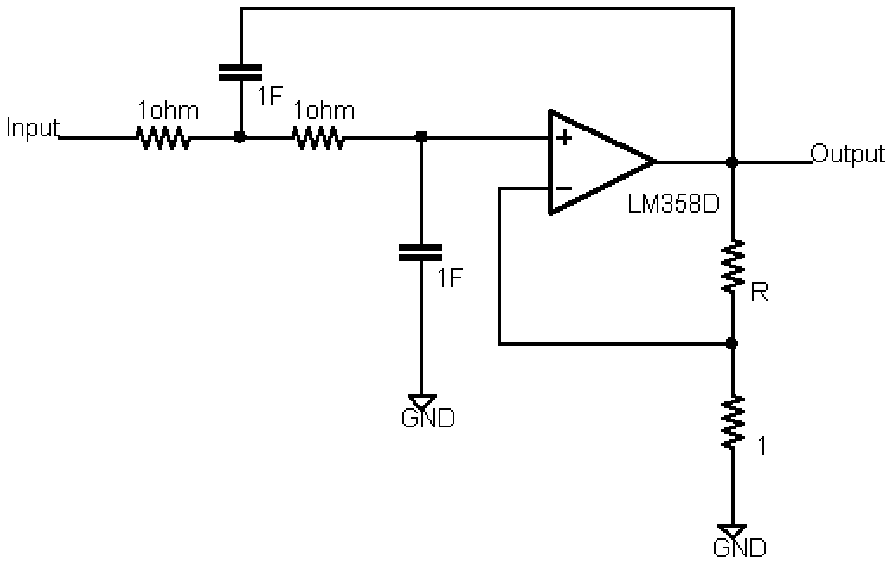


Source: Author's own diagram, (2013).

Note the -40dB/decade roll-off. The knee is sharper than two cascaded filters.

Exercise: Derive the transfer function for this filter. You may wish to start your analysis by working from the junction of the two resistors.

Voltage Controlled Voltage Source



Source: Author's own diagram, (2013).

The Sallen and Key filter can be made somewhat more flexible by giving the amplifier a bit of gain. This results in the VCVS filter.

Here the amplifier has a gain of K=R+1.
The filter has a cut-off frequency of 1 radian/sec. Any other frequency can be achieved by scaling the 1 Farad capacitors and 1 Ohm resistors. More practical components can be selected this way.

The transfer function for this filter is:

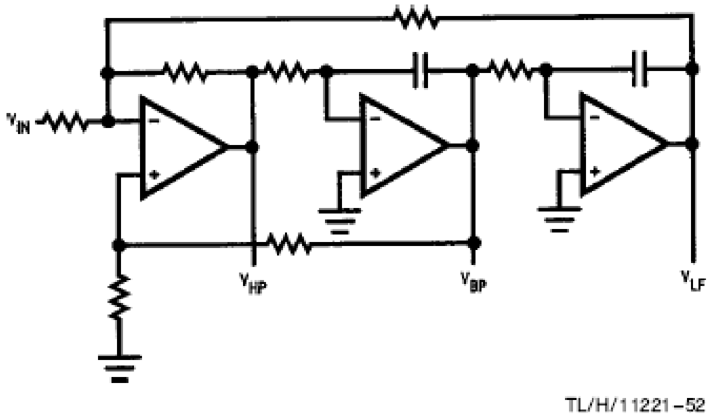
$$\frac{V_o}{V_i}=\frac{K}{s^2+(3-K)s+1}$$

The Q of this circuit is 1/(3-K). If K>3 then the circuit will break into oscillation. If K is close to 3 then the Q of the circuit will be very sensitive to component tolerances, which limits the usefulness of the filter. In many cases we only need moderate Q's from our filters, so this is a good solution.

State Variable Filters

The state variable filter is a complex filter circuit that overcomes some of the component sensitivity of the VCVS filter. It typically uses either three or four opamps per stage, where each stage is second-order. They can achieve high Q, unlike the VCVS filter. A typical circuit can be found in "The Art of Electronics", on page 277.
The four opamp versions can provide highpass, lowpass and bandpass outputs from the same filter, so it may be useful in channeliser applications.

Here is an example of a state variable filter. This version produces high pass, low pass and band pass outputs. Its flexibility can be improved by adding another amplifier to combine various of the outputs.



(d) Universal State-Variable 2nd-Order Active Filter

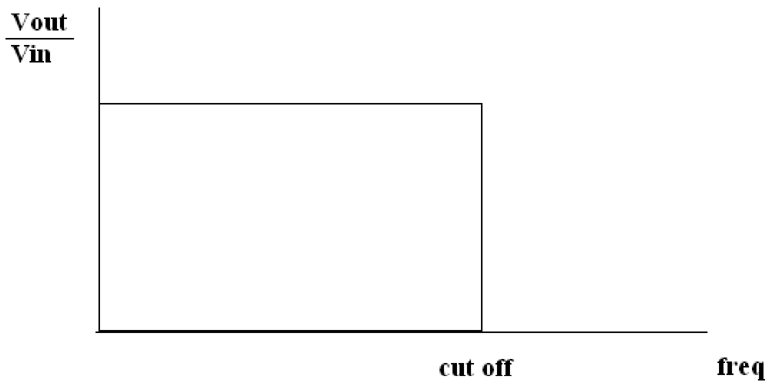
Source: *Universal State-Variable 2nd-Order Active Filter 2010*, National Semiconductor Application Note 779 A Basic Introduction to Filters - Active, Passive,and Switched Capacitor P. 18.

Normalisation

In previous examples we sometimes used resistors of 1 ohm and capacitors of 1F in order to simplify our calculations. For simplicity we often design filters with cut-off frequencies of 1 radian/sec. When we want to convert the filters to have more useful cut-off frequencies we can simply scale the RC product to the correct value.
Bear in mind that there is much less flexibility when selecting capacitors than resistors. Capacitors are available in a much smaller set of "standard values" than resistors are. In addition, large value capacitors are physically bulky, often polarised and relatively inaccurate.

Approximations to Brick Walls

In many cases, the ideal filter is one with an infinitely steep transition band. Such a filter would allow desired signals through without any distortion and would totally stop undesired frequencies from passing. The frequency response of such an ideal low pass filter is shown here:



Source: Author's own diagram, (2013).

This ideal response is called a "brick wall" filter.
It is not possible to create a brick wall filter. This is not because of component limitations, rather it is because it is fundamentally impossible. Brick wall filters violate the principle of causality.

From Fourier analysis (see Introduction to Fourier Analysis by UCT's Dr N Morrison, p. 262) we see that the inverse Fourier transform of a brick-wall function in the frequency domain is a scaled (sin (x))/x function in the time domain. The (sin (x))/x function stretches to infinity in both directions on the time axis. Thus, making a brick wall filter would entail making a circuit which started responding infinitely long *before* the input arrived at the filter's input. This violates our concept of cause and effect so we say that brick wall filters are “non causal”.

Although we can never make an ideal filter we can approximate the ideal as closely as necessary. There are a number of different approximations that optimise different aspects of filter performance. A few of these approximations are discussed here.

<u>Butterworth</u>

This filter type is also called the “maximally flat approximation”. Butterworth filters optimise the flatness of the passband. This is important if a flat frequency response is required from the system, such as might be the case in a spectrum analyser. The price that is paid for the flatness of the passband is that the knee is not as sharp as some other filters. While the ultimate roll-off rate is still 20dB/decade/pole the initial roll-off is not as good as some other filter types.
The Butterworth filter consists of cascaded filter stages which are often made up of VCVS filters. Each of these stages has the same cut-off frequency but the Q of each stage varies so that peaks in the frequency response of some stages cancel the dips in the frequency response of other stages.
The phase response of Butterworth filters is not optimised, so the time-domain performance of Butterworth filters is not very good.
The more stages in the filter the flatter the passband becomes.

<u>Bessel</u>

The Bessel filter optimises time-domain performance. It does this by ensuring that each cascaded filter stage has a time delay that changes linearly with frequency. The trade-off is that the transition band response of the Bessel filter is very poor.

<u>Chebychev</u>

The Chebychev filter is a very pragmatic trade-off between the flatness of the passband, time domain distortion and steepness of transition band. The idea behind the Chebychev filter is that often some non-uniformity of gain in the passband is acceptable and this can be traded off for improved transition band steepness. Bear in mind that component tolerances are likely to cause a usually-practical Butterworth filter to be non-ideal in practice. The Chebychev filter allows the designer to specify how much “passband ripple” is acceptable and then to design a filter which does not exceed that figure and gives a steeper transition band. High ripple Chebychev filters can actually give *more* than 20dB/decade/pole in the transition band, although ultimately they are limited to this roll-off in the stopband.

<u>Elliptic</u>

The Elliptic filter gives even sharper transition band drop-off than the Chebychev filter. It allows ripples in the passband and the stopband. All of the other filter types mentioned give monotonic roll-off in the stopband. Once again the roll-off can be chosen by selecting the allowable ripple in pass and stop bands.

Table Based Filter Design

The filter types mentioned above can all be built up out of cascaded filter stages. The design question that must be solved is how to choose the parameters of those cascaded stages.

Each of the filter types described above is specified using a polynomial in 's'. There are tables of constants for these polynomials in most books on filters. The process of synthesizing a circuit from the polynomials must then be undertaken to produce a filter.

“The Art of Electronics” makes this process a lot simpler:

TABLE 5.2. VCVS LOW-PASS FILTERS

Poles	Butter-worth	Bessel		Chebyshev (0.5dB)		Chebyshev (2.0dB)	
	K	f_n	K	f_n	K	f_n	K
2	1.586	1.272	1.268	1.231	1.842	0.907	2.114
4	1.152	1.432	1.084	0.597	1.582	0.471	1.924
	2.235	1.606	1.759	1.031	2.660	0.964	2.782
6	1.068	1.607	1.040	0.396	1.537	0.316	1.891
	1.586	1.692	1.364	0.768	2.448	0.730	2.648
	2.483	1.908	2.023	1.011	2.846	0.983	2.904
8	1.038	1.781	1.024	0.297	1.522	0.238	1.879
	1.337	1.835	1.213	0.599	2.379	0.572	2.605
	1.889	1.956	1.593	0.861	2.711	0.842	2.821
	2.610	2.192	2.184	1.006	2.913	0.990	2.946

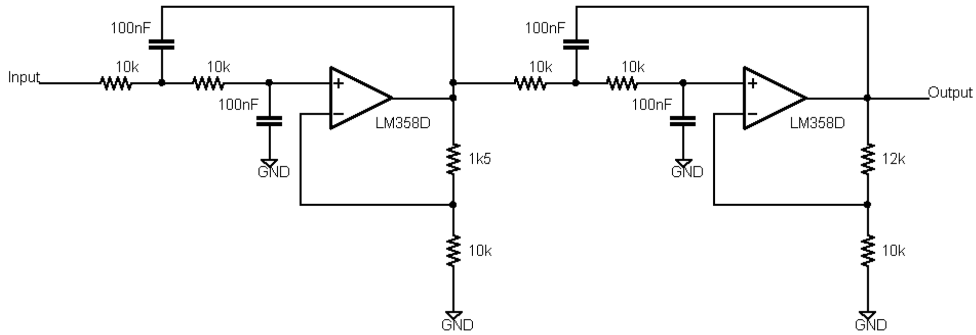
Source: *The Art of Electronics* (Second ed.), Paul Horowitz and Winfield Hill(1989), Cambridge University Press, ISBN 978-0-521-37095-0 p. 274

This table is very useful for filter design because it simplifies the design process. It assumes that one's filter will be made up out of cascaded VCVS stages. Because each VCVS stage is a second-order filter this table lists filters of 2, 4, 6 and 8 poles. These correspond to 1,2,3 and 4 opamp circuits.

The filter types that allow one to design are the Butterworth, Bessel, and two types of Chebychev filter. The difference between the 0.5dB Chebychev filter and the 2dB Chebychev is the size of the allowable passband ripple. (One can see the book for graphs showing the compromise between passband ripple and transition band steepness).

The table assumes a cut-off frequency of 1 rad/sec. Any other frequency can be obtained by denormalisation.
For each VCVS stage the two selectable parameters are frequency and K.

When designing a Butterworth filter each VCVS stage has the same cut-off frequency. The K of each stage varies however. As an example, a 4^{th} order Butterworth filter consists of a stage with K=1.152 followed by a stage with K=2.235.
Here is the circuit diagram of a 4^{th} order Butterworth low pass filter with a cutoff of 159Hz:



Source: Author's own diagram, (2013).

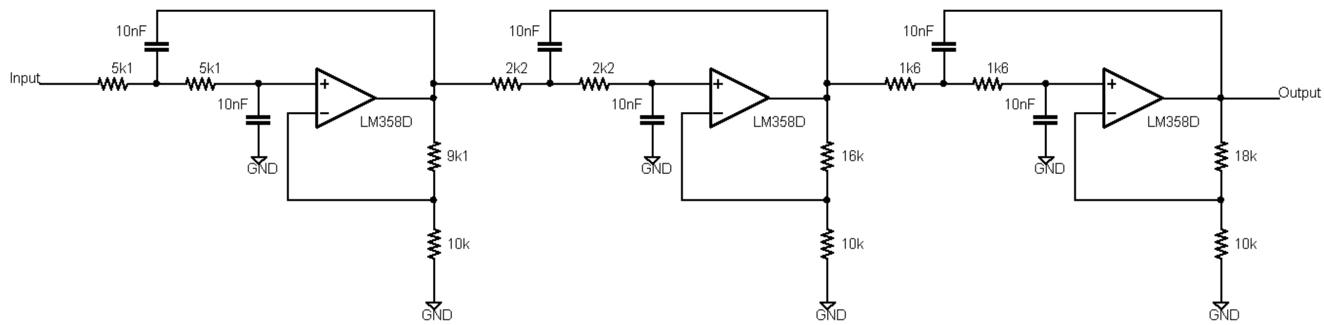
Notice that we have not achieved the precise values of K mentioned above. This is because the diagram shown here has been drawn using E12 standard value resistors. In reality, this is one of the major constraints regarding how close to ideal our real filters can come. This Butterworth filter will have ripples in the passband because of this non-ideal reality.

When designing a 2dB Chebychev filter the filter stages have slightly different cut-off frequencies as well as different K values.

A 6^{th} order 2.0dB Chebychev low pass filter with a cut-off of 10kHz will have the following parameters:

Stage 1: freq=3.16kHz K=1.891
Stage 2: freq=7.3kHz K=2.648
Stage 3: freq=9.83kHz K=2.904

The circuit diagram for this filter is:



Source: Author's own diagram, (2013).

The higher order filters shown in this table call for high values of K. This is problematic from a stability point of view. If needed, these filters could be implemented using VCVS stages for the low-K stages and state-variable stages for the high-K stages.
High pass filters can also be designed using this table. Simply swap the positions of the resistors and capacitors and use cut-off frequencies given by $1/f_n$, rather than f_n as used in low pass filters.

Exercise: Design an 8 pole Bessel high pass filter with a knee point at 2kHz.

Chapter 9: Oscillators

The purpose of an Oscillator

An oscillator is a circuit which produces a voltage which changes periodically over time. It does this by using a combination of positive and negative feedback.

Types of Oscillator

For the purposes of this discussion we will categorise oscillators into two categories; those which produce sine waves and "all the rest".

Sine Wave Oscillators

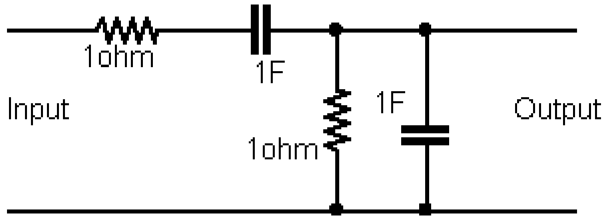
Imagine the following: an amplifier has been built. Imagine feeding in a 1kHz sine wave. Imagine that the amplifier had a gain of precisely 1 and a phase shift of zero degrees (or some integer multiple of 360 degrees). The output would be a precise replica of the input.
Now imagine throwing a switch which caused the output of this amplifier to be fed back into the input. What would happen?
The circuit would continually output an identical sine wave! This is *not* perpetual motion; for that to happen the power supply would have to supply energy to the system, this is in fact self-sustaining oscillation.

From this we can establish a few criteria for oscillation. These are called the Barkhausen Criterion.
1. The gain of the circuit must be at unity at the desired frequency of oscillation.
2. The phase shift around the loop must be an integer multiple of 360 degrees.

Note: Some references ("Opamps for everyone" page 15-2 is an example) use a phase shift of 180 degrees, but they establish this using an inverting amplifier, so the overall result is the same.

Wein Bridge

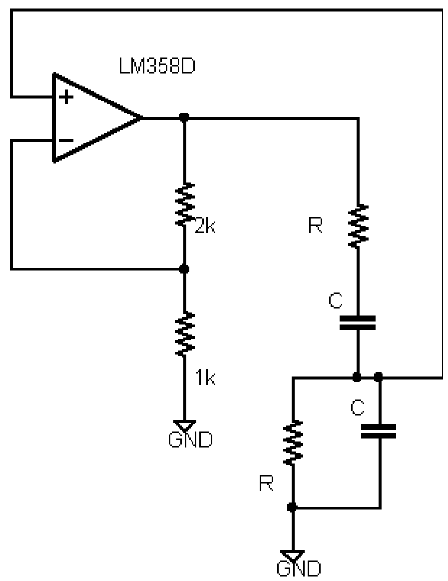
The Wein Bridge circuit is shown here:



Source: Author's own diagram, (2013).

This filter has a gain of 1/3 at f=1/2piRC and at that frequency the output is in phase with the input.

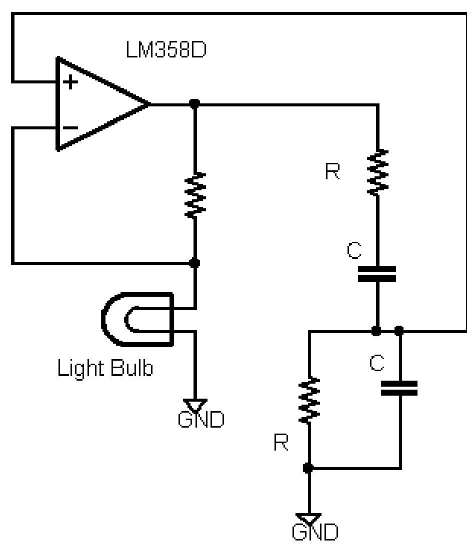
If we put this filter into a circuit along with an amplifier with a voltage gain of 3 then we get an oscillator:



Source: Author's own diagram, (2013).

How does this oscillator start up? Where does the initial sine wave come from? How can we be sure that the loop gain of the oscillator will be precisely 1?

If the amplifier gain is more than 1 then the oscillations will grow until they saturate the output of the amplifier and square wave output results. If the gain is less than 1 then the oscillations will never start. We design the amplifier to have a gain of more than 1 and implement a feedback system to bring the gain down if the output gets too large. Noise in the circuit is enough to start the oscillations because noise will contain the final frequency, and that frequency will get amplified every time it "circulates" through the amplifier.
Here is a common implementation:



Source: Author's own diagram, (2013).

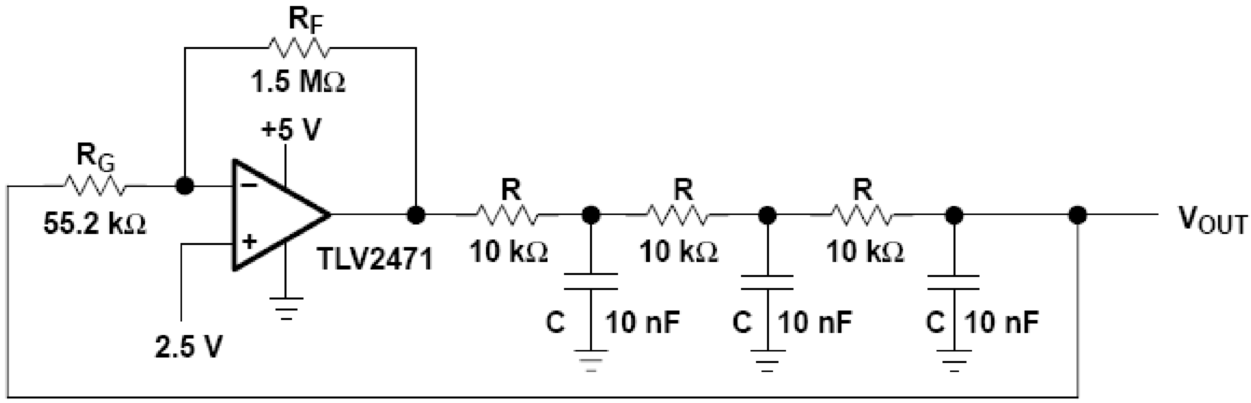
The above light bulb might seem strange, but it is correct!
A light bulb heats up substantially when operating (unlike a resistor). As the bulb heats up its resistance rises. As the resistance of the bulb rises, the gain of the amplifier drops. Thus, as the output of the oscillator grows the bulb gets hotter. The result is that the output amplitude stabilizes, resulting in a very pure sine wave output from this circuit.

In order to avoid using light bulbs many implementations of this circuit use JFET's to stabilise the gain of the amplifier.

Exercise: Revise the operation of the JFET.

Phase shift Oscillators

Study the following oscillator circuit:



Source: *Figure 15-14 Phase Shift Oscillator (Single Op Amp)* 2002, Ron Macini (ed.), Texas Instruments “Opamps for Everyone” Design Regerence, p. 15-15.

Each of the low pass filters in this circuit produces a phase shift. The phase shifts a total 180 degrees, with the inverting amplifier also contributing 180 degrees. The amplifier's gain is designed to compensate for the voltage lost as the sine wave passes through the filter.

Some implementations of this amplifier use a nonlinear element, for example two back-to-back zener diodes, in the opamp's feedback loop. Simpler implementations (as shown above) rely on the opamp becoming nonlinear when heavily driven to stabilise the loop gain to prevent distortion caused by clipping.

Many implementations of the phase shift oscillator use opamp buffers between each filter stage in order to prevent the filters from loading each other. If the filter is broken down into four stages then quadrature outputs will be available. This is useful for applications such as some modulation schemes.

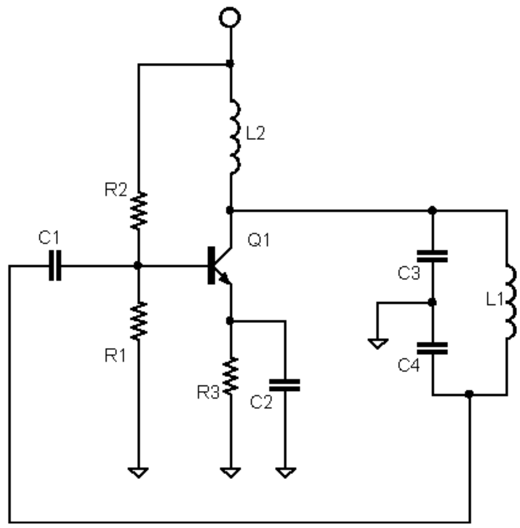
Exercise: Look back to the phase and gain Bode plots for a first order low pass filter.
- From the phase plot read off the frequency at which the phase shift is 45 degrees.
- Now take that frequency and from the gain plot work out the attenuation in each section.
- Now work out what amplifier gain would be needed to satisfy the Barkhausen criteria.

Colpitts, Hartley

The Hartley and Colpitts oscillators are LC oscillators. Because of the size/ cost/ quality constraints on inductors these amplifiers are most commonly seen in high frequency applications.

As with the other sine wave oscillators these are amplifiers with a positive feedback network. Because of the high operating frequency most implementations of these amplifiers use transistor amplifiers rather than opamps. There are a wide variety of circuits, but only two will be shown here.

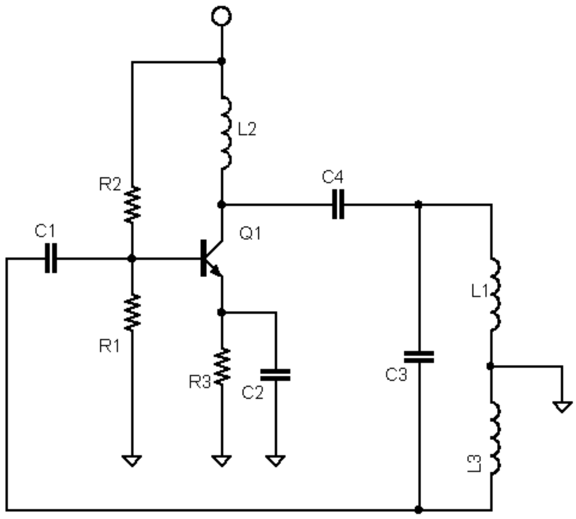
Here is the Colpitts Oscillator:



Source: Author's own diagram, (2013).

Q1, R1, R2, R3, C2 and L2 form a normal common emitter amplifier. The only slightly unusual feature is that the collector has L2, which is designed to have high impedance at the operating frequency and low impedance at DC. One may recall from previous courses that this amplifier's gain is determined by the ratio of collector impedance to emitter impedance, so this is a high gain amplifier at the operating frequency.
C1 is designed to have a low impedance at the operating frequency. It is only used to decouple the base bias voltage from the feedback circuit.
L1, C3 and C4 make the feedback network.

Here is the Hartley Oscillator:



Source: Author's own diagram, (2013).

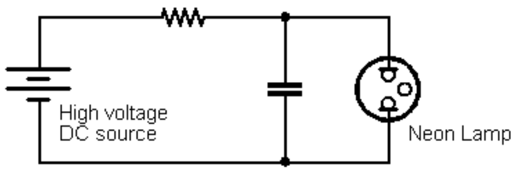
The amplifier is as before. C4 blocks the DC level on the collector from being short circuited to ground through L1. The feedback network consists of C3, L1 and L3.

Non- Sine Wave Oscillators

Often we need a circuit which generates a square wave. These oscillators are discussed here.

Relaxation Oscillators

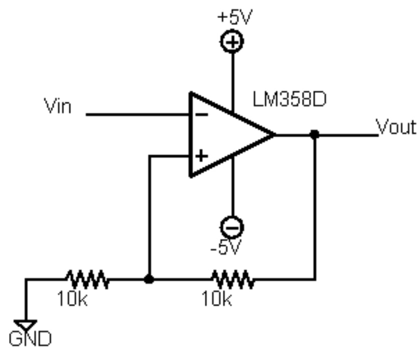
The classic relaxation oscillator circuit is shown here:



Source: Author's own diagram, (2013).

Internally the lamp is a set of electrodes in a gas. As the voltage across it is raised (capacitor charging), no current flows through it. When the voltage gets to some threshold suddenly the lamp goes into conduction. This causes the capacitor to discharge. The lamp stays in conduction until the voltage across it drops to *below* the turn-on threshold. This property is called hysteresis.

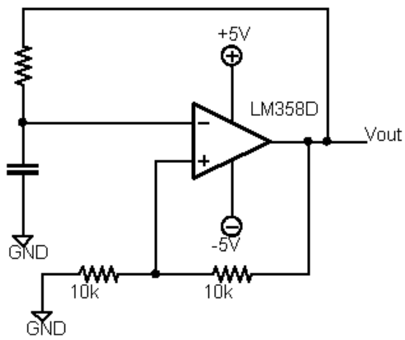
A more modern circuit which exhibits hysteresis is the comparator with positive feedback:



Source: Author's own diagram, (2013).

Assume that when power is turned on to the circuit the input is at 0V, the output is at +5V, and that the opamp output swings to the rails.
As the input is increased above +2.5V the output swings to -5V. This causes the threshold voltage to jump to -2.5V. If the input is decreased then it will have to go below -2.5V to change the output. Note that the threshold changed from +2.5V to -2.5V.

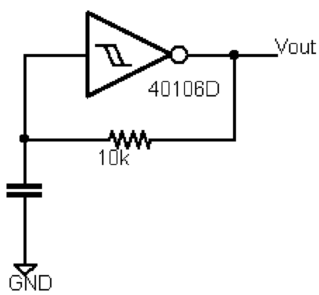
We can use this circuit to make an oscillator:



Source: Author's own diagram, (2013).

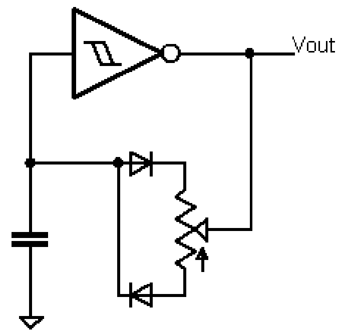
To understand this circuit imagine that the capacitor is discharged on turn-on. Assume that the output is 5V and that the opamp output swings to the rails.
Because the output is high the capacitor charges through the series resistor. As it charges past 2.5V the opamp output swings to -5V. This causes the capacitor to discharge until it gets to -2.5V. When the capacitor gets to -2.5V the output swings to +5V and the cycle starts again.

The hysteretical comparator is also called a Schmitt trigger. Logic gates with Schmitt trigger inputs are available and can be used for simple square wave oscillators:



Source: Author's own diagram, (2013).

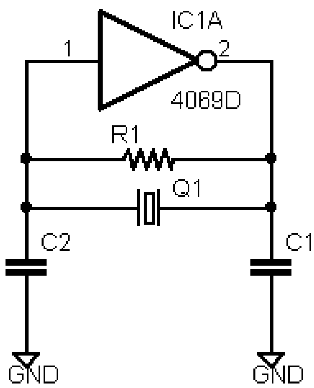
Here is a circuit which gives a constant frequency with a variable duty cycle:



Source: Author's own diagram, (2013).
Exercise: Figure out how this circuit works.

<u>The Pierce Oscillator</u>

This type of oscillator is actually a sine wave oscillator which is most commonly operated in such a way as to saturate the output. This is the most common form of quartz-crystal based oscillator. This is the circuit used in the vast majority of micro-controller circuits.



Source: Author's own diagram, (2013).

Some implementations use a resistor in series with the crystal in order to give an additional phase shift. This is mostly needed at low frequencies, such as the very popular 32.768kHz.

R1 forces the inverter into linear operation. For CMOS gates (CD4000 or 74HCxx) 10 MegaOhms is usually used. The crystal resonates at a particular frequency. One can select the frequency when purchasing the crystal.

<u>Waveform Generator IC's</u>

Waveform generation is a widely used circuit function. Because of the complexities of setting up an oscillator based around individual amplifiers and discrete components, sometimes it pays to use a dedicated waveform generator microchip.

- RC based

The best known RC oscillator is the 555 timer. This device has been dealt with in previous courses

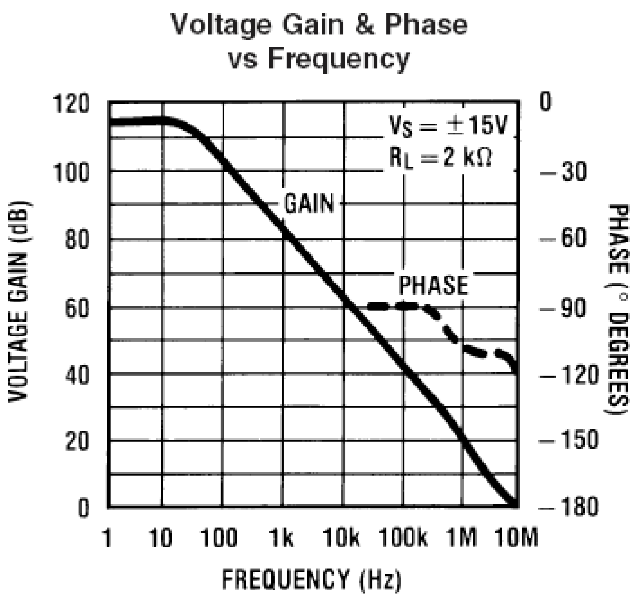
and will not be mentioned further.
There are logic family oscillators, such as the CD4047, which can be used to implement square wave oscillators with logic level outputs.
There are voltage controlled oscillators (VCO's) where the frequency of oscillation is controlled by an externally applied voltage. Popular VCO's include the LM566 and LM331.
The XR2206 is a waveform generator which produces sine, square and triangle waves. The square wave is integrated to produce the triangle output and the sine wave is formed by shaping the triangle wave with a non-linear diode-based circuit. The sine wave is not as pure as the waveform produced by a good sine wave oscillator.

- *Direct Digital Synthesis (DDS)*

It is possible to produce high quality waveforms by storing samples in a memory device, retrieving them when needed and using digital to analogue conversion and filtering to reproduce the waveform. This technique is widely used today. There are DDS microchips, such as the AD9834 from Analogue Devices which work on this principle. They are crystal based, offering very high stability (much better than RC circuits) and they often have digitally programmable output frequencies.

Opamp Stability – Amplifiers Sometimes Oscillate.

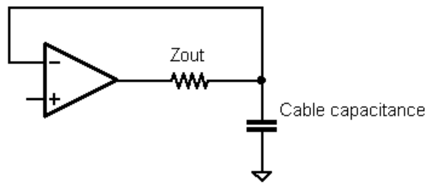
We assumed in our previous discussion that the opamps that we were using introduced no phase shift into their circuits. Actually, this is not true. As discussed before, an opamp has an internal capacitor which gives it a low pass characteristic. In addition, there are stray capacitances and resistances which add other breakpoints.
We have seen the phase vs frequency graph of the LM833 opamp before, but we need to examine it in more detail:



Source: *Voltage Gain & Phase vs Frequency* 2003, National Semiconductor, LM833 Opamp, p.5.

This graph shows that at 10 MHz the gain of the amplifier drops to 0dB unity, and the phase shift is 120 degrees. If this opamp is put into an inverting amplifier and the feedback loop introduces phase shift then there is a danger that the amplifier will oscillate. The feedback loop can introduce up to 60 degrees of phase shift before oscillation. Thus, we say that this amplifier has a 60 degree phase margin.

One very common source of feedback network phase shift is the combination of the opamp's output impedance and load capacitance. Consider an opamp driving a long cable:



Source: Author's own diagram, (2013).

Here the Z_{out} element represents the opamp's output impedance. The capacitor represents the capacitance of a long cable. These two elements combine to form a filter with phase shift. If this phase shift reaches 60 degrees then oscillation will occur.
In extreme cases parasitic capacitance on the circuit board can cause oscillations.

If a long piece of cable is to be driven, there are a few common solutions. These include putting a series resistor between the output of the amplifier and the cable (assuming that a high frequency signal will not be sent down the cable), using an opamp with lower output impedance (avoid rail-to-rail output types), or using an open loop buffer chip to drive the cable.

Chapter 10: Linear Power Supplies

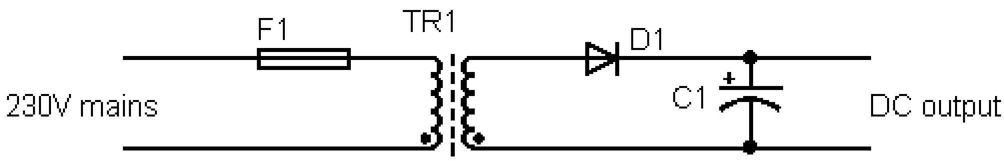
This section will discuss the circuits needed to supply power to electronic apparatuses.
We will discuss the use of high voltage AC mains in this section.

[MAINS CAN BE LETHAL! TAKE APPROPRIATE PRECAUTIONS WHEN DEALING WITH LIVE HIGH VOLTAGE SUPPLIES]

Bear in mind that touching mains is dangerous, and also remember that incorrect circuits can explode, leading to injury from from flying fragments.

Raw DC Supplies

The simplest form of mains-based power supply looks like this:

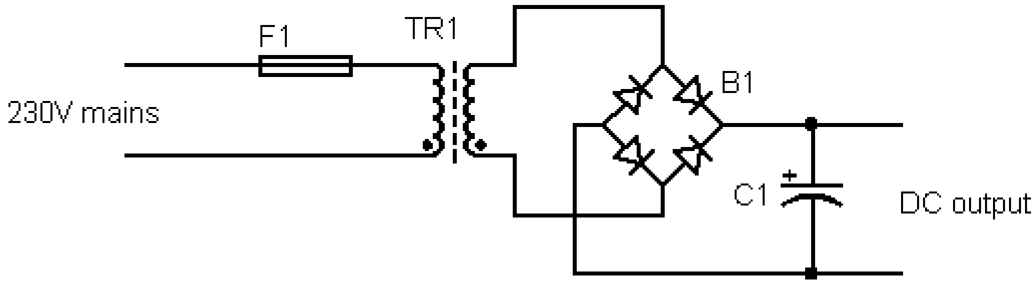


Source: Author's own diagram, (2013).

TR1 is a step-down transformer.
F1 is a fuse.

The mains passes through the fuse into the transformer, where it is galvanically isolated and stepped down to a suitable voltage for our circuitry. The diode allows the positive half of the waveform through and we call this half-wave rectification. The capacitor charges up during the diode's conduction period and discharges into the load during the diode's blocking period. As the capacitor charges and discharges the output voltage changes, resulting in ripple.
The secondary winding of the transformer only carries current during half the waveform. This means that in order to supply a specified load the transformer must supply twice as much current for half as much time. This is inefficient because the resistive losses in the transformer increase with the square of the current.
In addition, the capacitor needs to supply the load with current for most of the cycle. This calls for a large amount of capacitance, which is large and expensive.

The situation can be improved with a full wave rectifier:



Source: Author's own diagram, (2013).

The rectifier works by routing the AC through two different paths, which path the current takes is determined by its polarity.

Exercise: Show the paths of secondary current.

The transformer now carries current for up to 360 degrees.
The size of the capacitor determines the amount of ripple present on the output. The bigger the capacitor, the less ripple will be present.

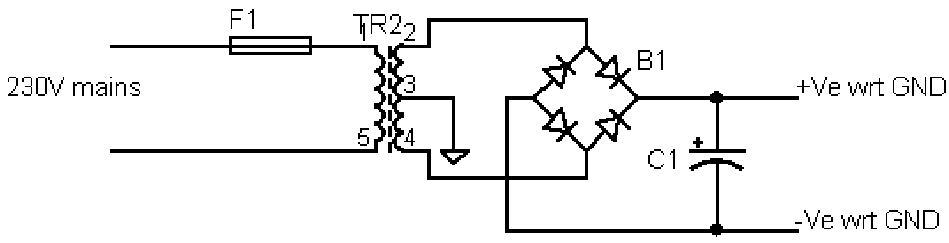
The ripple voltage is approximately:

$$V_{ripple} < \frac{IT}{C}$$

where I is the load current,
T is the time between capacitor charges, for 50Hz this is 10ms,
C is the value of the smoothing capacitor.
This formula is approximate but conservative.

It is also important not to overdo the smoothing capacitor. As the capacitor gets bigger the conduction angle of the rectifier gets smaller and the peak current gets bigger. This causes increased heating of the transformer and rectifier.

If a split supply (dual rail) is needed then this is a good option:



Source: Author's own diagram, (2013).

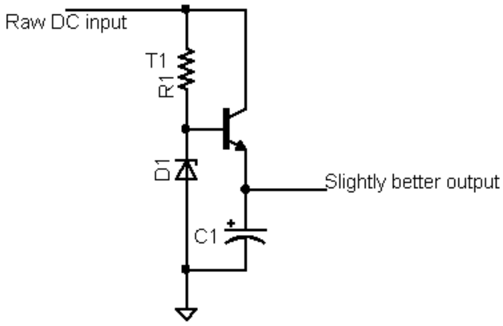
When choosing the transformer remember that there is a fair bit of variation in the mains supply voltage. This will directly affect the secondary voltage of the transformer.

It is also important to remember to take the rectifier diodes' drops into account when calculating output voltages.

Voltage Regulator Circuits

The raw DC supplies shown above are useful for getting approximately the correct DC supply, but they do not produce a clean enough DC supply for many circuits. The supply voltage will have ripple and it will vary with changes in the supply voltage. Voltage regulators can be used to produce a precise and stable DC output.

A crude scheme like this is already an improvement:



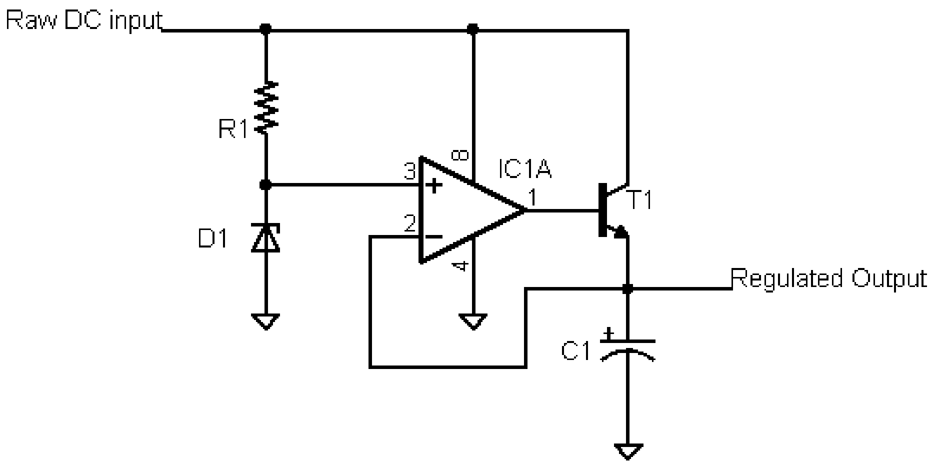
Source: Author's own diagram, (2013).

The zener diode produces a roughly constant voltage (the exact voltage depends slightly on zener current), and the emitter follower will provide current gain. The big weakness of this circuit is that the Vbe of the transistor will drop as the transistor warms up, so the output is not very well regulated.

The key to making very good voltage regulators is to replace the zener with a better voltage reference and to compare the output voltage to that reference and then correct any error.

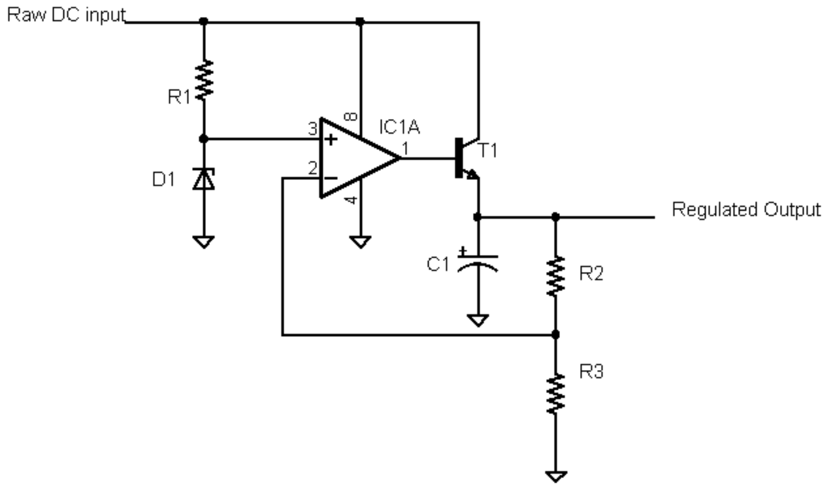
The zener will give more consistent performance if it is driven with a constant current. Precision voltage references are also available (the LM385 is one of our standard types).

Here is a simple circuit which corrects for changes in the emitter follower's Vbe:



Source: Author's own diagram, (2013).

We can adjust the output voltage by feeding a proportion of the output back:



Source: Author's own diagram, (2013).

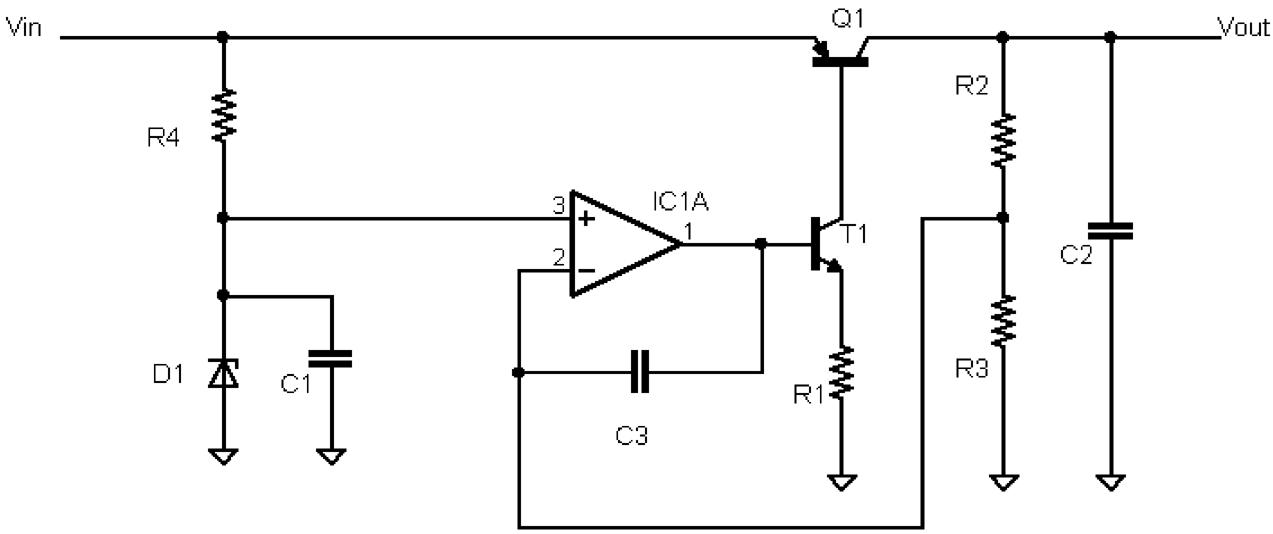
The output voltage will be determined by the ratio of R2 and R3 as well as the zener (or reference) voltage.

The DC input will have to be bigger than the desired output voltage by just over 2 Volts. This is because the opamp's output can swing to about 1.5V below its supply and the emitter follower's output can swing to about 0.6V below this. The term used to describe this voltage gap is “headroom”.

Many systems can tolerate large amounts of headroom, but some systems benefit from reduced headroom. As an example of a system which works better with a smaller headroom, consider a circuit built out of 5V logic circuitry which must run off a battery. If the regulator has a headroom of 2V then the supply voltage must be at least 7V, calling for five 1.5V batteries. If the headroom could be reduced to, for example, 0.1V then we could run the system off four batteries. Alternatively, we could run the system off the same number of batteries and get more life out of them because the system will run off more heavily depleted batteries (flatter batteries put out less voltage).

A regulator with small headroom is called a *“Low Dropout Regulator” (LDO).*

The secret to making an LDO is in the pass element. Here we use a PNP transistor connected so that it can saturate. When a transistor is saturated, the voltage across it will often be less than 0.1V.



Source: Author's own diagram, (2013).

In this circuit the zener, D1, provides the voltage reference. If the output voltage, as seen through the R2/R3 divider is too small then the opamp output will increase. This will increase the base drive to T1, which will increase the current flowing in its collector. The increased T1 collector current will come from the base of the pass element, Q1. Increasing the base drive of Q1 will result in the pass element going further into conduction, thus increasing the output voltage.

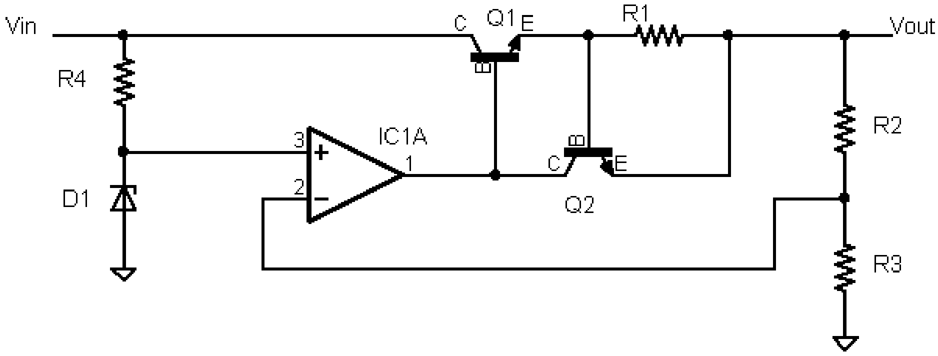
The opamp output does not have to approach the supply voltage to cause Q1 to saturate; we therefore avoid large headroom.

Because of the extremely high voltage gain of the saturated-switch pass element, low dropout regulators need to be designed carefully in order to achieve stability. C3 and C2 are essential to stabilise the control loop. C3 will typically be about 1nF while C2 will often be large and will often have a ceramic bypass capacitor to ensure low impedance at high frequencies.

Power Supply Protection

Overcurrent Protection is an important feature of many power supplies. If the output of the regulator is overloaded or short-circuited it is preferable that the power supply survives the experience. In some power supplies the output impedance of the supply is high enough that no damage results. Some other power supplies have fuses or resettable fuses to protect them.

A more sophisticated approach is to sense the output current and to reduce the output voltage if the current is excessive. The easy way to sense current is to pass it through a resistor and compare the voltage across the resistor to a threshold to determine if the current is excessive.
Here is a regulator circuit with this built in:



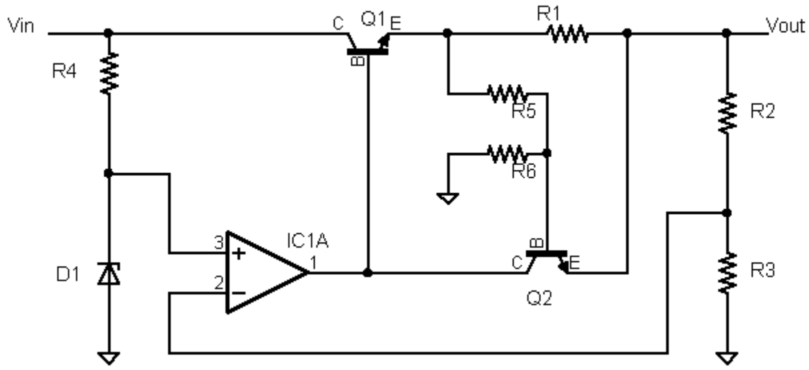
Source: Author's own diagram, (2013).

This is the same circuit as above, but with two extra components, R1 and Q2, added. R1 is chosen so that at the maximum allowable current the drop across R1 is approximately 0.7V. (For example, if Iout must be limited to 1A then R1 would be 0.68 ohms). If the current is below the maximum then Q2 has less than 0.7V across its base-emitter junction and it is turned off, so it has no effect. If the current gets to the limit then Q2 goes into conduction and allows current to flow from the collector to the emitter. This collector current is "stolen" from the base of Q1. The opamp can only supply a limited amount of current (remember the opamp's internal current limiting) so if Q2 steals enough current then Q1 will be deprived of base current. The output current of the regulator is controlled by Q1. If Q1 is deprived of base current then it will limit the current that it passes through its collector, thus limiting the current from the regulator.

An important subtlety here is that the feedback network, R2 and R3, is *after* the current limiting circuit. This is important because it means that the output voltage will not drop as the voltage across R1 increases.

Simple current limiting is very useful for some regulators and is commonly used for laboratory power supplies, such as those in our labs. It does have a weakness however. Suppose that the output is overloaded so that the output voltage is forced to drop to half of its nominal value in order to enforce the maximum load current. The current through the pass transistor will then be at its design maximum. The problem is that the voltage across the pass transistor is greater than its design maximum. This means that the regulator is dissipating more power than it was designed to handle; this may cause the regulator to fail.

The solution is to make the current limit value related to the actual output voltage. If the output voltage drops (because of overload) then the current limit reduces the amount of output current that it allows. This is called foldback current limiting.

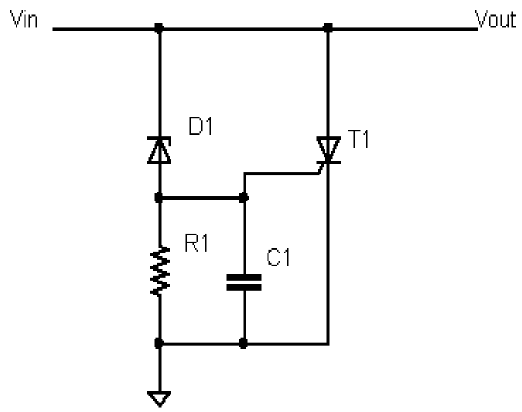


Source: Author's own diagram, (2013).

The difference between this circuit and the simple current limiter is in the voltage divider made of R5 and R6, and the turn on voltage for Q2 is constant. The voltage lost across R5 varies with Vout (as it is a fraction of Vout). This means that at high Vout R5 'loses' a lot of voltage (to give a high current limit) and when Vout is closer to GND, R5 only has a small voltage across it, making the current limiter more sensitive.

When designing a current limiting circuit remember that the output load might have large amounts of capacitance across the rails. If a normal current limiter is used then this will cause the output voltage to rise slowly on start-up. This is often not a problem, although it may cause some circuits to start up incorrectly. Foldback protection is more difficult because it may never start correctly into a capacitive load.

Crowbar protection is a form of overvoltage protection. Imagine that one is building an expensive and delicate electronic system which needs to be powered from an external power supply. As a prime example, think of a cellphone or laptop computer. There is a good chance that someone may plug in a power supply of the wrong voltage. If the input voltage is too low then there is usually no damage. But, if the input voltage is too high then it is likely that there will be expensive damage. A crowbar circuit is useful for preventing this. Here is a simple crowbar circuit:



Source: Author's own diagram, (2013).

Under normal conditions zener D1 is in reverse bias but does not have enough voltage across it to cause breakdown. There is no current through R1 and so the gate of the thyristor, T1, is at ground. Because the gate and the cathode of T1 are at the same potential there is no trigger current and T1 remains switched off. If the input voltage rises too high then the zener goes into conduction. This causes the gate of T1 to be pulled up, supplying a trigger current. The trigger current flows into the gate of T1. T1 goes into conduction and short-circuits the input voltage to ground. It acts like a crowbar being dropped across the supply rail. Hopefully the power supply feeding this circuit has current limiting, or else it might suffer damage, but the expensive circuitry inside our electronic device will not suffer damage. C1 is a small capacitor which gives a very short time delay to prevent the thyristor from triggering when the correct power supply is plugged in. If this is omitted then the stray capacitance in the zener might occasionally cause false triggers. If the crowbar is triggered it may be reset by unplugging the offending input voltage.

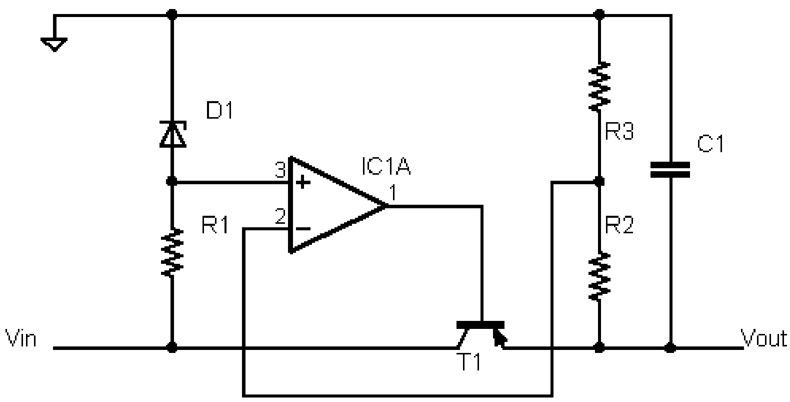
Power Dissipation

The pass element in a linear regulator will dissipate power and that dissipated power will be the product of current through the regulator and voltage across it.
If the regulator is fed from a rectified AC waveform then the amount of ripple is an important factor. The voltage fed into the regulator at the "bottom" of the ripple must still be large enough for the regulator to work correctly. If the ripple is large the pass transistor will need to dissipate large amounts of power because of the high peak voltage across it at the peak of the input voltage.
Variable output voltage regulators also pose some challenges. The input voltage must be larger than the maximum output voltage. When the regulator's output voltage is adjusted to be small the excess voltage is across the pass transistor. In order to keep dissipation at reasonable levels, many laboratory power supplies use a switched transformer secondary. The "click" that one hears as one adjusts the voltage on many lab-bench power supplies is caused by a relay switching in higher or lower voltage taps from the mains transformer into the rectifier.

Exercise: Draw a circuit which switches a different transformer secondary tap in to the circuit when the output voltage exceeds 12V.

Negative regulators

Circuits which run off dual rails call for regulators which can regulate a voltage which is negative with respect to 0V. The circuits for these are generally identical to the positive regulators, with one exception; the polarity of the pass elements needs to be reversed. In the case of the simple regulator here is the circuit:



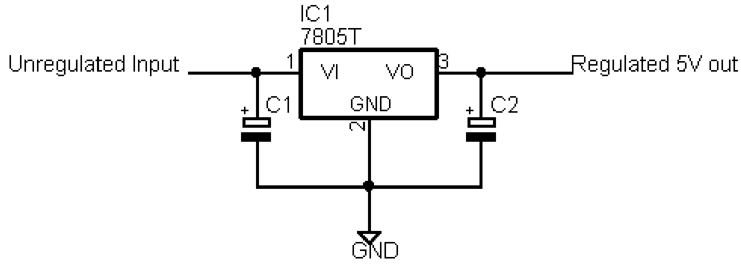
Source: Author's own diagram, (2013).

Voltage Regulator Chips

Because there is such a huge need for voltage regulators there is a large selection of voltage regulator chips. A few of these are discussed here.

Fixed Voltage 3 terminal regulators
These are used to produce common system voltages, such as 5V, -5V, +12V, -12V. The positive regulators are the 78xx series, where xx is the output voltage. The most common are the 7805 and 7812. They are also known as LM340, depending on manufacturer and date of manufacture. They

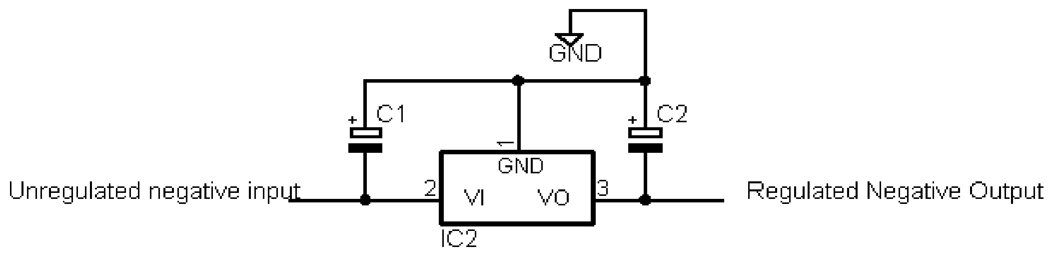
are used as follows:



Source: Author's own diagram, (2013).

The capacitors are used to remove high frequency noise from the inputs and outputs.
These regulators come in two different case styles, TO92 for small output currents and low dissipation applications, and T0220 for large output currents (up to 1.5A) and higher dissipation. The maximum input voltage for these regulators is 30V.

The negative regulators are the 79xx series. They are used as follows:

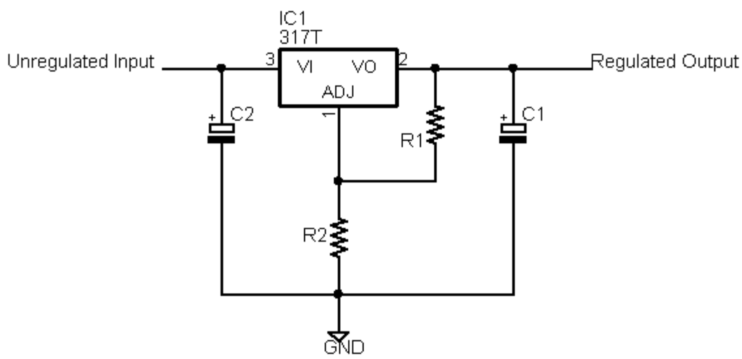


Source: Author's own diagram, (2013).

Adjustable Voltage 3 Terminal Regulators

These regulators allow you to select the output voltage by means of external resistors. The most common positive output regulator is the LM317 and it has a negative counterpart, the LM337.

The LM317 is used as follows:



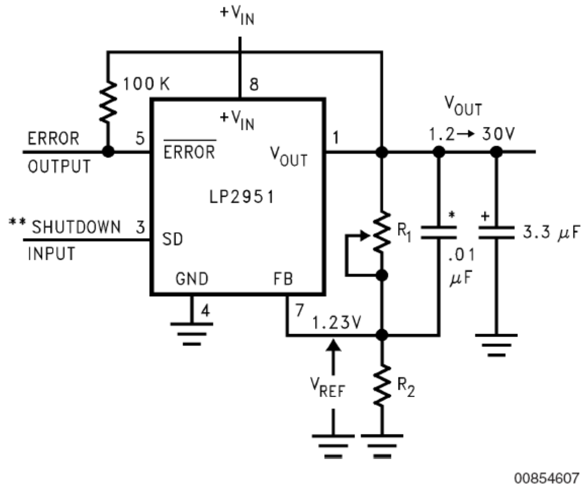
Source: Author's own diagram, (2013).

The LM317 establishes a 1.25V drop across R1. It will push as much current through R1 as it needs in order to force the drop across this resistor to 1.25V. The ADJ pin has minimal current flowing into it (typically 50uA). Thus, the current in R1 is forced to flow in R2. This causes a voltage drop across R2. The output voltage is the sum of the drop across R1 and R2. R1 is typically 200-330 ohms.

When designing low-power circuits bear in mind that the regulator also draws current. The LM317 typically draws around 5mA and low quiescent current regulators are available.
All of the three terminal regulators mentioned here have current and temperature limiting, which makes them fairly robust. The thermal protection is not perfect and operating these regulators at excessive dissipation for extended periods will still result in failure.

Low Dropout Regulator Microchips

There is a large range of integrated low dropout regulators available. The LP2951 is one example. The typical application diagram is this:



e Application Hints

$$V_{out} = V_{Ref}\left(1 + \frac{R_1}{R_2}\right)$$

Source: *LP2950/LP2951 Series of Adjustable Micropower Voltage Regulators* 2004, National Semiconductor Datasheet, LP251, p.16.

This LDO is optimised for battery powered equipment. It has low quiescent current and a worst-case headroom of 600mV (although this is a somwehat pessimistic estimation). Because it is optimised for small battery powered equipment it can only put out 100mA.
The ERROR output is used to inform external circuitry that the battery is flat. It is active when the input voltage is too low for the regulator to put out the correct output voltage.
The Shutdown input is used to turn the regulator on and off from a logic signal. When the regulator is off it wastes only a few microamps.

Heatsink Design

When any object dissipates power its temperature will rise. Electronic devices have a specified "thermal resistance", measured in K/W. A device with a thermal resistance of 1K/W will experience a temperature rise of 1 degree Celsius for every Watt of power that it dissipates. In order to decrease the temperature rise a piece of metal is often bolted onto power-dissipating devices. These "heatsinks" are made of aluminium or copper because they are good conductors of heat.

Here is a picture of a typical heatsink arrangement for a TO220 device.

Source: “Index of /wp-content/gallery/electronics-packages” http://cladlab.com/wp-content/gallery/electronics-packages/component-package-to-220ab-mounted-on-heatsink.jpg accessed 10 December 2013

The transistor is mounted on the heatsink with an M3 machine screw. Between the tab of the transistor and the heatsink is an insulator. This is often necessary because multiple devices are frequently mounted on the same heatsink and the tabs connect to the middle lead of the transistors. Between each layer a silicone-based “heatsink paste” is used to improve the thermal conductivity of the joint.
The specifications for power devices give a maximum “junction temperature”. This is the maximum temperature that the silicon can withstand. This is usually around 150 degrees Celcius. When calculating the maximum temperature rise that a device can withstand the maximum ambient temperature must be subtracted. Remember that many electronic devices get hot inside, either from their own dissipation or from environmental conditions such as sunlight.

Example:
A transistor has a maximum junction temperature of 150 degrees Celsius. The maximum ambient temperature inside the product's casing is 70 degrees Celsius. The device is dissipating 10 Watts.
Calculate the maximum allowable thermal resistance:
Maximum rise = 150-70 = 80 degrees.
Maximum thermal resistance = 80/10 = 8 K/W.

The thermal resistance of a heatsink assembly is the sum of its thermal resistances. Typically:

$$R_{th\,total} = R_{th\,device} + R_{th\,insulator} + R_{th\,heatsink}$$

Example:
A TO220 package typically has about 2 K/W junction to case. (This varies from device to device).
The Rth_insulator is typically about 0.5 K/W

To achieve a total thermal resistance below 8K/W the heatsink needs to have a thermal resistance of less than 5.5K/W. The lower the thermal resistance of the heatsink the bigger it gets.

Because the dissipation versus temperature difference is actually non-linear, the above calculations are approximations. Most heatsink datasheets give graphs of dissipation versus temperature rise. Complex circuit boards with multiple heat sources can be modeled with special software packages.

If the heatsink size gets too large to be feasible or economic it is possible to reduce the ambient temperature by using cooling fans. Extreme heat dissipation problems sometimes call for liquid cooling or Peltier effect thermoelectric coolers. If refrigeration is used attention must be paid to the possibility of condensation forming in the product.

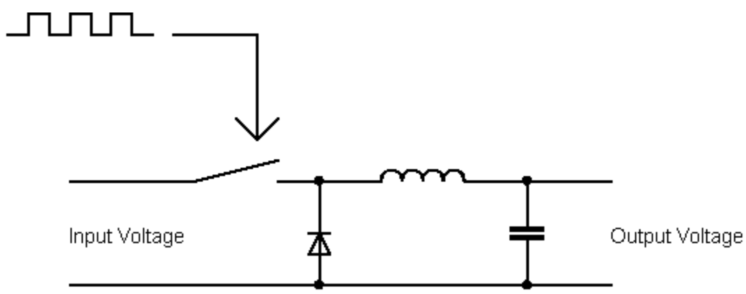
Exercise: A voltage regulator has an average input of 12V and an output of 5V. The maximum current output of the regulator is 1A. For a transistor with a maximum junction temperature of 120 degrees Celsius, and a maximum ambient temperature of 50 degrees Celsius, what is the maximum thermal resistance of the combined heatsink assembly?

Chapter 11: Switched Mode Circuits

Introduction

The voltage regulator and amplifier circuits that we have seen all operate by throwing away a portion of their input (supply) voltage to produce a reduced output voltage which is regulated or controlled in some way.
There are a few consequences of this.
- The output voltage is always lower than the input voltage.
- The output voltage is always of the same polarity as the supply voltage.
- The power devices always dissipate power, even in the theoretically ideal case.
- The efficiency is limited.

Instead of throwing away power by dissipating it as heat one could also lower the output power of a regulator by switching the output on and off. As an example, think of a 50% duty cycle square wave with an amplitude of 5V. The average voltage is 2.5V.
Exercise: prove this, using integration.
If this square wave was passed through an averaging circuit then the output of that circuit would be 2.5V. If we wanted a higher output voltage then we could increase the duty cycle and if we wanted a lower voltage we could decrease the duty cycle.
We want a (theoretically) lossless averaging circuit, so a normal RC low pass filter is not suitable, but an LC low pass filter is perfect. Here is the scheme:



Source: Author's own diagram, (2013).

The switch is turned on and off by the square wave. When the switch is on, current flows through the inductor into the capacitor and the load. When the switch is off, the capacitor supplies current to the load. The diode is needed to prevent inductive kick. The current through an inductor cannot change infinitely fast, so when the switch is turned off the inductor still needs a current path; the diode provides this. If the switch is switched fast enough then the voltage fluctuations across the capacitor will be small. The LC circuit is operated very far from its resonant frequency. The output voltage can be varied by changing the switch's duty cycle.

Theoretically, the power dissipated by the switch is zero. When the switch is open there is no current through it, hence no power dissipation. When the switch is closed there is (ideally) no voltage across it, so no power is dissipated.
The current drawn from the input only flows while the switch is closed. Ideally (since the circuit is loss-less), the power drawn from the supply is the same as the power supplied to the load.

Step Down Regulators

The circuit above is called a step down regulator because it produces an output voltage lower than the input voltage. There are a few practical details.
The switching frequency has an effect on the size of the inductor and capacitor. The higher the frequency the smaller these become. Because high value inductors and capacitors are bulky and expensive these converters have to be operated at high frequencies to make them practical. Switching frequencies are seldom below 10kHz, and some regulators operate in the MHz region. Interestingly, low output current step down regulators with on-chip inductors have started to appear in the market.

Because there are fast-changing currents in this circuit all switch mode regulators emit some radio frequency interference (RFI). The choice of inductor value and construction will affect the amount of RFI emitted by the circuit. Slightly larger values of the inductor will lead to lower current fluctuations, hence less RFI. Toroidal inductors will emit less RFI than inductors wound on bobbins.

The diode must operate at the switching frequency. The peak diode current is equal to the maximum inductor current. The average diode current is given by:

$$I_d = I_{out} * (1 - Duty\ Cycle)$$

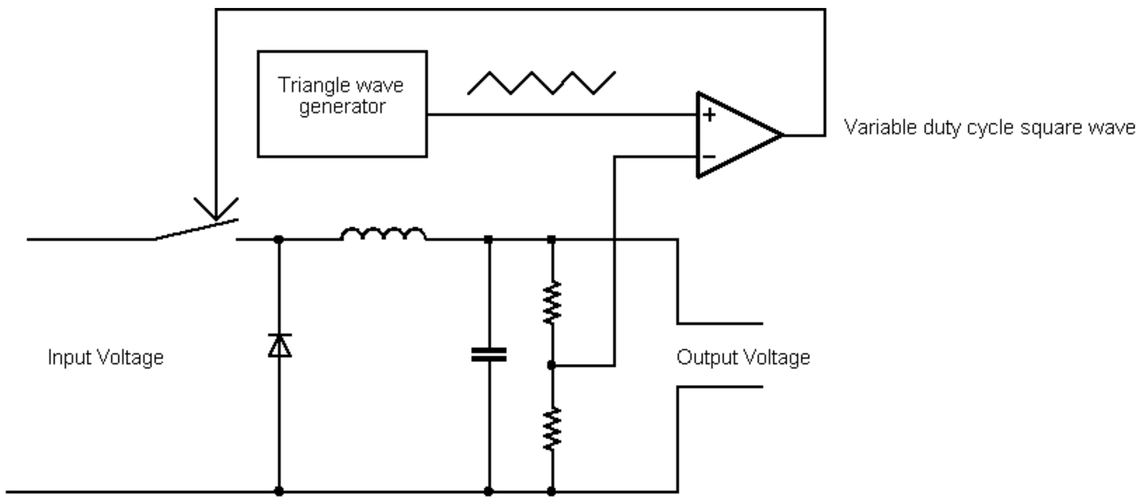
This means that we require a high-speed high-current diode. Normal rectifier diodes are not suitable because their reverse recovery time is too long. Schottky diodes are therefore often used because they are generally fast. Schottky diodes generally have low peak inverse voltage, so they are only useful for low input voltage regulators. The PIV of the diode must be greater than the input supply voltage. Fast Recovery Diodes are often used here. They perform the same function as a normal diode but are optimised for fast changes from conduction to blocking. The diode represents a departure from the ideal in that it conducts current during part of the switch's “off” time. When the diode is conducting it dissipates power because it has voltage across it.

The switching device can be a bipolar transistor, MOSFET or IGBT. If a BJT is used it should be used as a saturated switch, so a PNP device would be the natural choice. The voltage across the transistor when it is on represents a departure from the ideal and, as a result, all real switching devices do dissipate some power. In addition, there is finite time during which the switch is changing from open to closed. During that time the voltage across the switch drops from (Vin-Vout) to Vsaturated. This leads to additional losses during the transition period; these additional losses are called “switching losses”.
The power device must be able to block the full supply voltage and it must be able to carry the load current. The current carried by the switch will have peaks, which must not exceed the switch's peak current handling capability. The larger the inductor the flatter these peaks will be.

A MOSFET is most commonly used for the switch. Modern MOSFETS offer very low resistance and consequently give high efficiencies when used as the switching element.

Some step-down regulators use a second switch in place of the diode. This improves the efficiency because the switch will dissipate less power in conduction than the diode. The two switches are carefully synchronized so that they are never on together. This version is sometimes called a “synchronous switched mode step down regulator”.
As the load varies the voltage drop across the switch and across the resistance of the inductor will vary. Generally, a regulation scheme is needed to give a stable output voltage. A very simple

scheme is shown here:



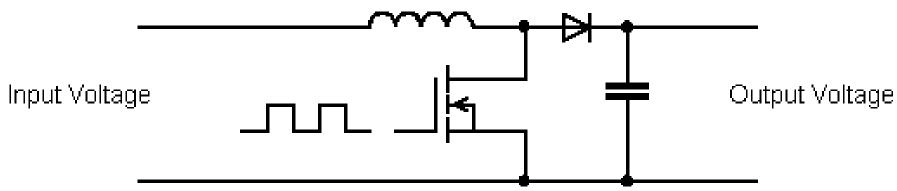
Source: Author's own diagram, (2013).

The triangle wave generator produces a triangle wave, often by integrating a square wave. This is compared to a threshold to produce a square wave. If the triangle wave is above the threshold at any instant then the output of the comparator is high. The threshold is derived from the output of the regulator. If the output voltage goes up then the output of the comparator is high for a shorter amount of time (lower duty cycle), which causes the output voltage to drop. This is a negative feedback loop. In practical circuits more complex arrangements are often used.

Step down regulators are often called Buck Regulators.

Step Up Regulators

Consider this circuit:



Source: Author's own diagram, (2013).

When the MOSFET is turned on, a current will build up in the inductor which is supplied by the input supply. The “on” voltage across the MOSFET is smaller than the forward voltage of the diode, so no current flows through the diode.
When the MOSFET is turned off, the inductor attempts to keep the current through it flowing. This results in a high voltage appearing across the inductor. The MOSFET end of the inductor shoots up to a high voltage, which causes the diode to start conducting. This causes the capacitor to charge from the inductor current. This, therefore, causes the capacitor voltage to be *greater* than the supply voltage. This is the first circuit that we have seen which allows us to boost a voltage higher than the supply. This circuit is called a Boost Regulator.

For boost regulators the output voltage is theoretically given by:

$$V_{out} = \frac{V_{input}}{1-D}$$

where D is the duty cycle of the MOSFET.

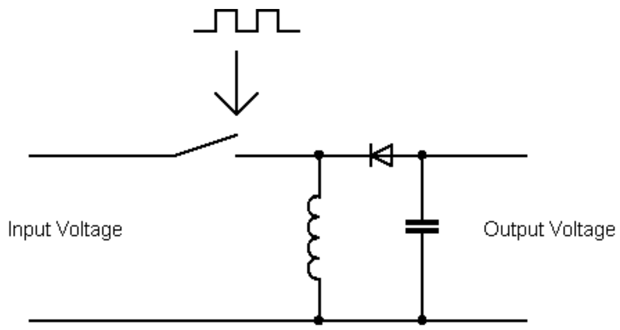
In practice, there is a limit to the output voltage that can be obtained. This is often limited by the quality of the inductor. Inductors have parasitic capacitance between windings which limits the voltage that the inductor can produce.
The MOSFET must carry a maximum average drain current of Iout/(1-D) and it must be capable of withstanding a maximum Vds of Vout. Also, a NPN transistor can also be used in place of the MOSFET.

The diode must be fast and must carry the full output current, so either a Schottky or fast recovery diode is preferred. When the MOSFET is on, the diode will experience Vout across it in reverse; it must therefore be capable of withstanding that inverse voltage.

Usually, the output voltage will be regulated. This is accomplished by increasing the duty cycle if the output voltage falls too low.
The boost circuit allows a wide variety of circuitry to be operated off low voltage battery supplies.

Inverting Regulator

Here is a circuit that can convert a positive voltage into a negative one:



Source: Author's own diagram, (2013).

When the switch is closed, current will build up in the inductor. When the switch is opened the inductor will attempt to keep current flowing. Note carefully the direction of current flow that existed during the "on" time. This must be maintained, and the only way that this can happen is if the non-ground end of the inductor goes to a negative voltage, as predicted by Lenz's law The diode goes into conduction and charges the capacitor. The voltage across the capacitor will be of the opposite polarity to the input voltage.

$$V_{out} = \frac{-V_{input} * D}{1-D}$$

The switch and diode must withstand a voltage Vds = Vin-Vout,
The average switch current = Iout / (1-D),
The average diode current Id=Iout.

This circuit is also called the Buck Boost regulator.

Isolated Regulators

All of the converters that we have mentioned up until now are non-isolated. This means that they have a direct connection (at least part of the time) between the input the output. In some applications this poses a safety hazard. Any device that is powered from the mains needs to have galvanic isolation between the mains and the user.
Simple transformer based power supplies are inherently isolated. The transformer consists of two coils of wire, which can be insulated from each other quite simply. The weight and price of 50Hz transformers often makes them impractical. As an example, a typical desktop computer has a 300VA power supply. A 300VA transformer weighs just over 3kg. One would then have to add a rectifier, smoothing capacitors, regulator, and heatsink to produce a power supply. By contrast, an isolated switched mode power supply typically weighs under 1.5kg.

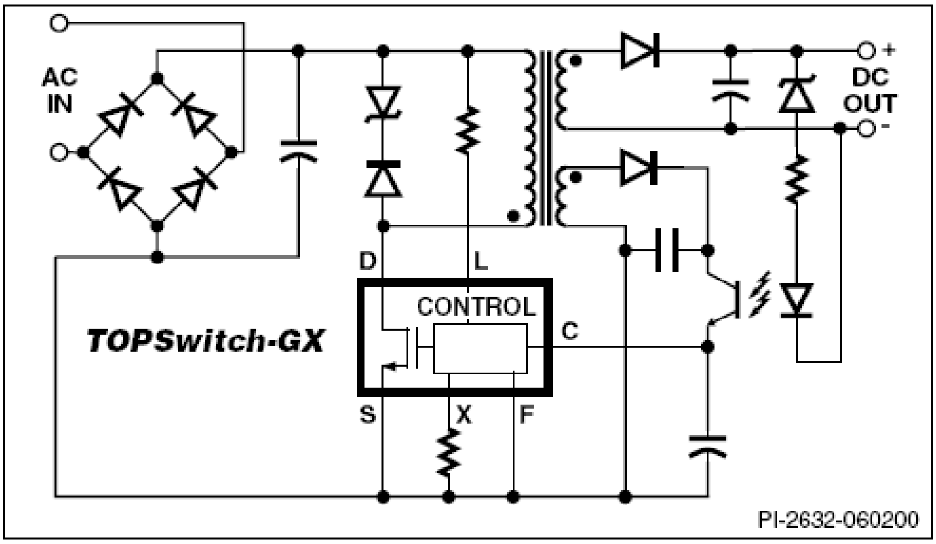
The basic idea behind isolated switched-mode power supplies is simple. The mains is fed directly into a rectifier and smoothing capacitor. The capacitor needs to be rated for high voltage, but only small capacitances are needed because of the (relatively) low currents required. For example, 300W at 230V is only 1.3A, but at 12V it is 25A.
This produces a high voltage DC supply, often called a "DC Bus". The DC bus is then "chopped" at high frequency by a high voltage MOSFET or IGBT. The losses in the power device are generally low because of the low currents, which translate to small on-state losses.
The chopped output is fed into a transformer. Because the transformer operates at high frequency, it will be significantly smaller, lighter and cheaper than a 50Hz transformer. The high frequency transformer will typically be wound on a ferrite core to reduce core losses, which is important because laminated steel cores are very lossy at high frequencies.
The output of the transformer is rectified and smoothed. The smoothing capacitor does not need to be large because of the high frequency of operation.
The smoothed output is compared to a reference voltage. If the output is too high then the output of the comparator is fed back via an opto-isolator to the high voltage input stage.
Many power supplies have multiple output voltages. This is easily accommodated by having multiple windings on the high frequency transformer.

When working on high voltage switched mode power supplies please remember that the input capacitor is charged to typically 320VDC during operation. It is very important to remember that *this voltage is often present after the power supply is switched off!*

Because this sort of power supply is very common there are a wide variety of microchips made for the application. The most common device in larger power supplies, such as computer power supplies, is the TL494. This integrates many of the circuit blocks needed to implement a power supply.

"The Art of Electronics" shows a complex example, which braver readers may examine.

Here is a circuit diagram of a simple mains powered isolated SMPS. This diagram is taken from the datasheet for "Power Integrations" TOPSwitch family of switch mode control microchips.



Source: *Simple mains powered isolated SMPS* 2005, Power Integrations, Top242-250 TopSwitch-GX Family extended power, design flexible, EcoSmart, Integrated off-line switcher, p.1.

Note the key components in this system are as described. There is a rectifier and capacitor to produce a DC bus. The MOSFET is integrated into the control chip. The diodes across the primary of the transformer limit the inductive kickback when the MOSFET turns off. The transformer has two secondary windings, and one of those windings is rectified and supplies DC to the load. When the load voltage exceeds the breakdown voltage of the zener diode plus the forward voltage on the opto-isolator LED, the opto-isolator's transistor will go into conduction. The second secondary winding provides a DC supply to the feedback loop.

Where does the control chip get its power from? On initial startup the “Drain” pin is pulled to a high voltage through the transformer's primary winding. This provides a startup supply. Once the supply is running the chip is powered via the “C” pin from the second secondary winding.

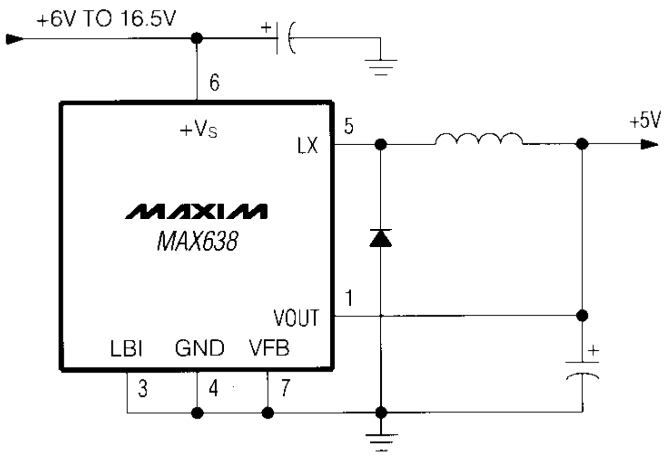
The “L” pin is used to sense the mains voltage. This is done so that undervoltage and overvoltage conditions can be detected. The control chip also features feed-forward control.

The “X” pin is used to set a threshold for current limiting. Current is limited by comparing the MOSFET's Vds in the 'on' state to a threshold. Note that the current limiting is applied on the high voltage side, which is more meaningful in terms of preventing damage to the MOSFET and more practical in view of the fact that the control chip 'floats' with respect to the low voltage output.

All three windings must be isolated from each other; between the output winding and the other two windings a high degree of safety is required.

Integrated Switched Mode Regulator Devices

We have just seen an example of an integrated circuit for controlling isolated switch mode supplies. There is a vast range of control devices for controlling non-isolated supplies as well. Some of these devices make designing buck, boost, buck-boost and other topologies easy.
As an example, here is a vastly simple 5V output buck regulator using Maxim's MAX638, which is stocked in the White Lab store:



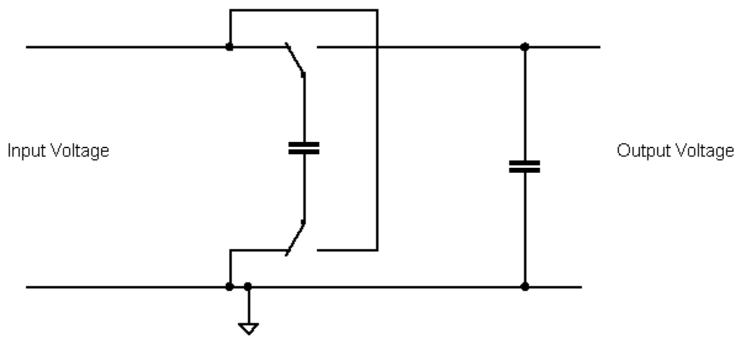
+5V OUTPUT STEP-DOWN REGULATOR

Source: *Typical operating circuit of +5V Output Step-Down Regulator* 1996, Maxim Integrated Products +5/Adjustable CMOS Step-Down Switching Regulator, MAX638, p.1.

The LM3578, also stocked, is much more flexible. It can implement all three of the topologies mentioned above. The datasheet gives circuit diagrams as well as tables of component values to make design very simple.

Charge Pump Voltage Multipliers

Charge pump voltage multipliers are useful for producing an integer multiple of the input voltage if only a small current is needed. They are also useful for inverting the polarity of a supply.
Conceptually, they work by charging up capacitors from the power supply and then switching the capacitors around. Consider this circuit diagram:



Source: Author's own diagram, (2013).

When the switches are in the position shown in this diagram the capacitor charges up from the incoming supply. When the switches are both changed over to the other position the charged capacitor is connected in series with the incoming power supply. This double voltage is used to charge up the output capacitor which maintains the load voltage during the charging phase.

One of the classic charge pump chips is the ICL7662. This chip can be used to double the input voltage (as shown above) or to invert the power supply voltage. Here is a voltage inverter scheme:

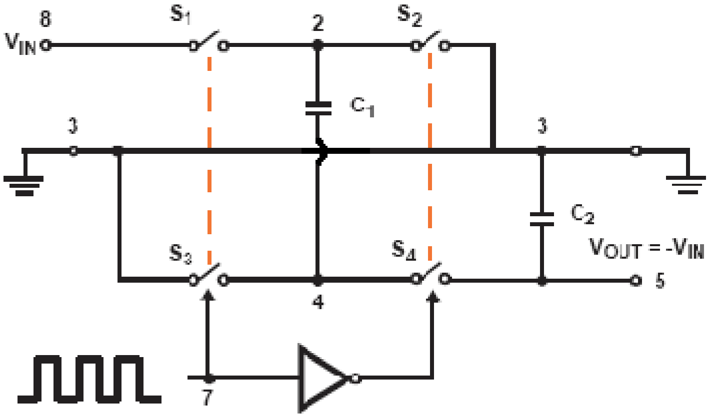
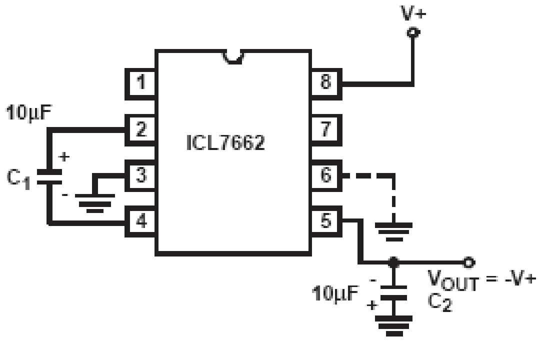


FIGURE 15. IDEALIZED NEGATIVE CONVERTER

Source: Figure 15: *Idealized Negative Converter* 1999, Intersil Datasheet, ICL7662, p. 7.

and below is the circuit diagram:



Source: Figure 18A: *Simple Negative Converter and its Output Equivalent* 1999, Intersil Datasheet, ICL7662, p. 8.

Class D Amplifiers

Audio amplifiers are traditionally linear, most commonly Class AB. As electronic systems are pressured into becoming smaller and more efficient the limitations of a linear amplifier become more constrictive. The best example of this is in cellular telephones. Size and weight are of key importance, but so are features like high-volume polyphonic ring tones and speakerphone functions.

One will recall that the average value of a pulse-width-modulated signal is proportional to the duty cycle. The averaging circuit is simply a low-pass filter with a time constant much longer than the period of the PWM. Note that everything is relative. If the PWM has a period of only one microsecond then a filter time constant of 100 microseconds is considered long. Medium quality audio signals have maximum frequencies of about 10kHz, with a period of 100 microseconds.
Class D amplifiers take the audio signal and use it to modulate a high frequency PWM waveform. Because PWM only involves digital switching it is very efficient. The mechanical inertia of the loudspeaker's cone acts as a mechanical low pass filter, thus reproducing the original audio signal.

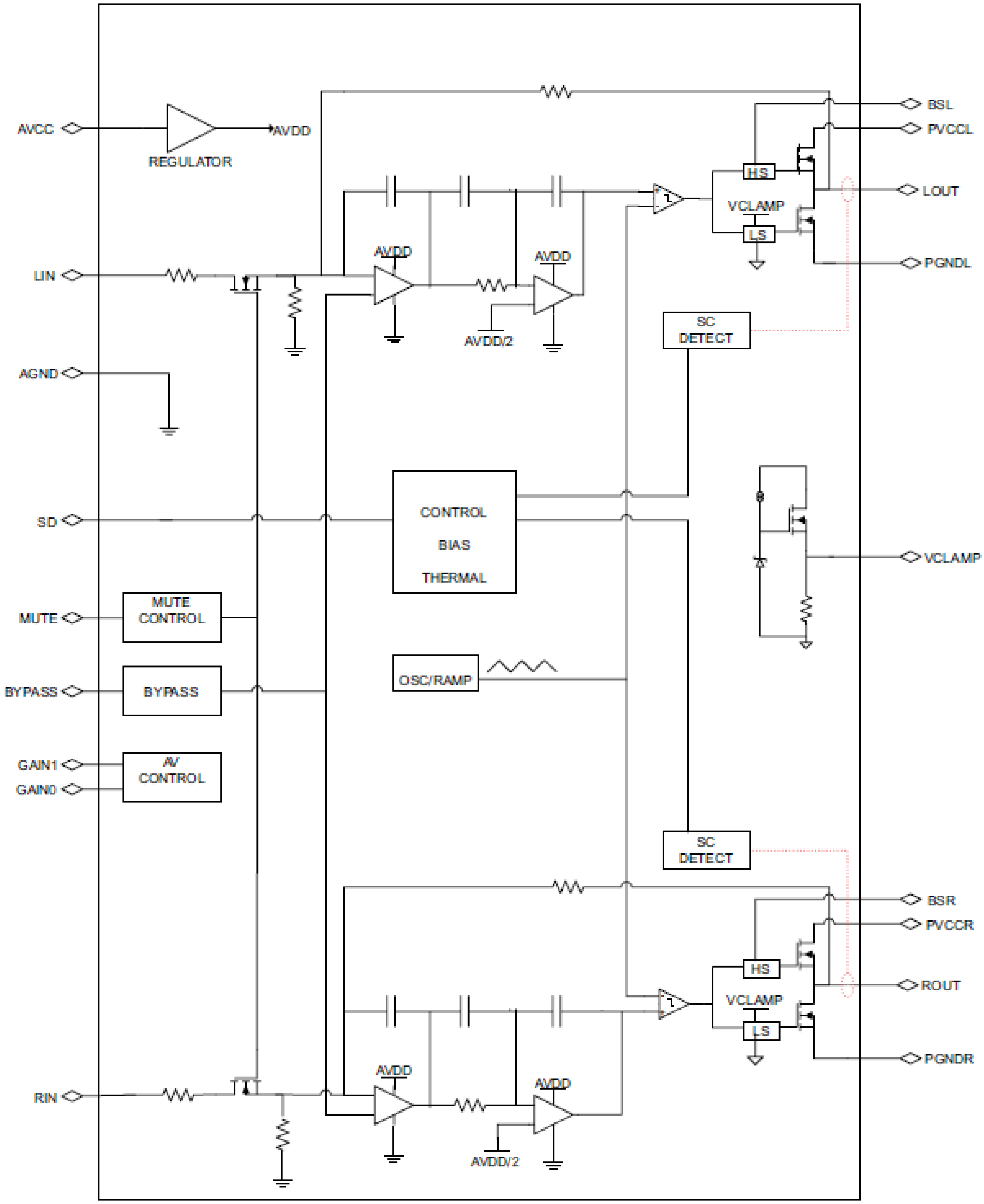
The audio may be presented to the amplifiers input in two different ways. Many audio sources are stored in digital format. It is easy to (digitally) produce a PWM waveform modulated by digital samples. This can be done by most microcontrollers. The signal can thus travel through the system entirely in digital format, with the conversion to analog happening at the loudspeaker.

If the audio signal originates from an analog signal then it can be converted to a PWM signal by comparing it to a triangle wave.

This process is illustrated in the following diagram. This is a stereo device, so it has two identical amplifier stages in it. In addition, it has a mute input, which mutes the audio signal and a two bit digital gain input to select the amplifier gain.

Class D amplifiers can attain efficiencies of 90% or better. This means that they generate very little heat and need relatively small heatsinks. In addition, they will consume less battery power in portable applications. Because of the growing popularity of this class the price of the devices is dropping and they are now entering the main stream audio market.

FUNCTIONAL BLOCK DIAGRAM



Source: *Functional Block Diagram* 2007, Texas Instruments Incorporated, TPA3122D2 Class D Audio Amplifier, p.5.

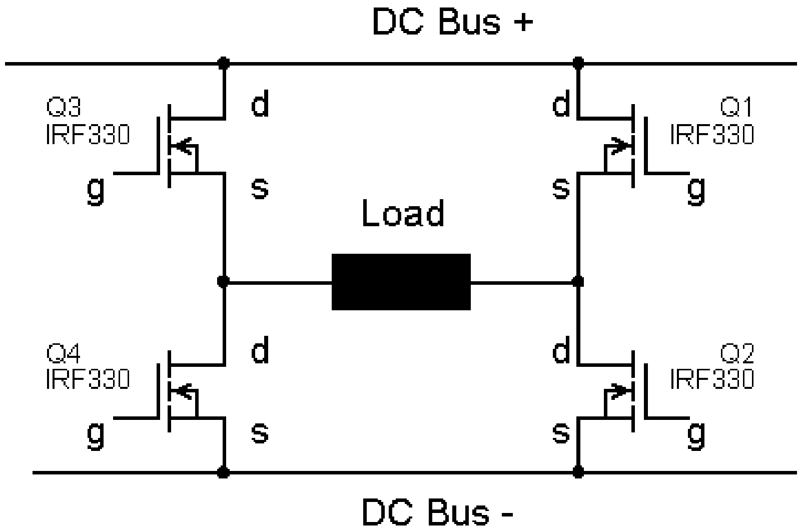
MOSFET Drivers

Most MOSFETs require that their gates are pulled high (typically around 12V) with respect to their source terminal. The gate of a MOSFET has capacitance.
In order to minimise switching losses it is important to switch the MOSFET on and off as fast as possible. This requires fast charging and discharging of the gate capacitance, which calls for a high drive current to the gate. Once the MOSFET has switched on no appreciable current is required to hold the device in its state.

In full bridge applications there are two MOSFETs per "leg" of the bridge. One of these MOSFETs is connected to the low voltage rail. This MOSFET is called the "low side MOSFET" and the other MOSFET is connected to the positive rail. This MOSFET is called the "high side MOSFET".

Because of the superior performance of high N channel MOSFETs, most applications use N channel MOSFETs for both the high side and low side switches.

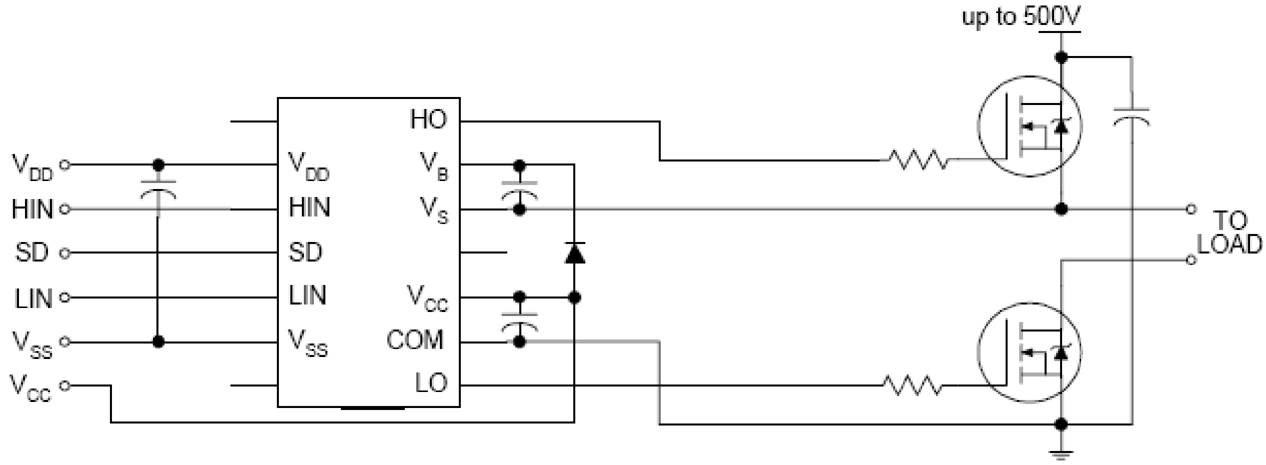
The circuit diagram for a MOSFET full bridge with N channel FETs is as follows:



Source: Author's own diagram, (2013).

The negative DC bus rail is normally grounded. The gate drive for the low side devices is a simple high-current output. The high side poses two problems. Firstly, the gate must be pulled high with respect to the source terminal. The source terminal's voltage changes every time the devices are switched, which makes this requirement tricky. The second problem is that when the high side devices are on, the gate terminal needs to be at a higher potential than the DC bus. This is often unavailable.

Very often, these problems are solved by means of a charge pump circuit. Here is the application diagram for the very popular IR2113 half bridge driver chip, taken from the datasheet:



Source: *Typical Connection* 2003, IR International Rectifier Datasheet, IR2113 High and Low side Driver, p. 1.

The Vcc pin is also connected to a 12VDC supply. Usually the two "to load" terminals are connected together. When the bottom FET is on (while the top one is off), then the capacitor between Vb and Vs charges through the diode. To turn the top FET on, the capacitor is connected by the driver across the top FET's gate and source. The diode is in reverse bias so the capacitor has no discharge path.

Safe Operating Area of Power Devices

All electronic components have specified limits of operation. Power devices normally specify their maximum ratings by means of graphs. The area inside the graph is called the "Safe Operation Area" (SOA). Exceeding the SOA will result in the destruction of the device.

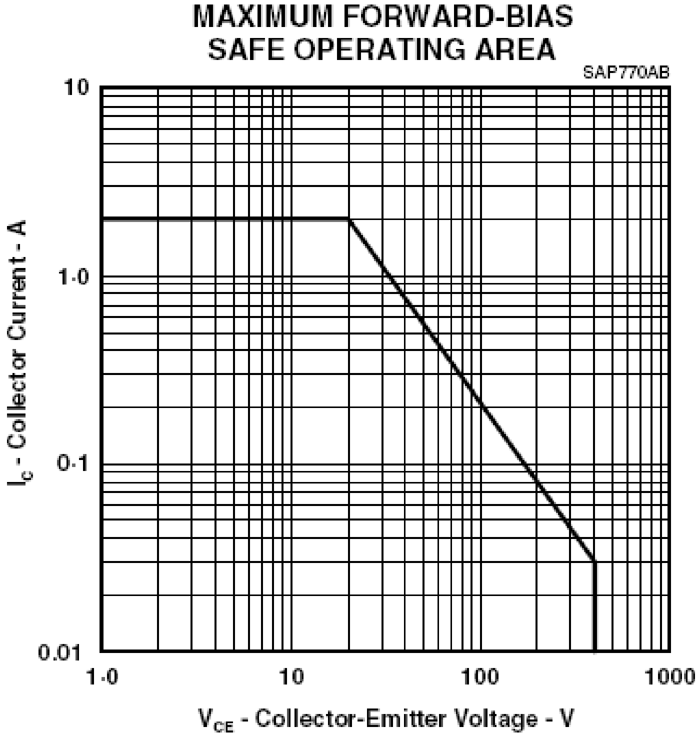
For any power device there are three fundamental limitations. These are:

- Blocking voltage: The maximum voltage that the device can have across it when it is turned off.
- Forward Current: The maximum current that the device can allow through before melting its internal circuitry.
- Power Dissipation: The amount of power that the device can dissipate before it is destroyed by its own heat.

These three parameters lead to three lines that any SOA graph will have. One vertical, one horizontal and one constant-power line. If the voltage and current axes are logarithmic then the constant power line is straight.

Many power devices have further operating restrictions, which reduce the area enclosed by these lines.

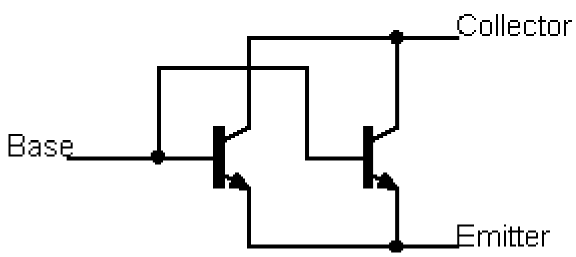
The Safe Operating Area (SOA) for a high voltage power transistor, the BUX84 is shown here:



Source: Figure 5 *Maximum Safe Operating Regions* 2002, Bourns, BUX84 NPN Silicon Power Transistor, p.4.

The horizontal line shows that this device can handle a maximum of 2A flowing in the collector. The vertical line shows that this device can block up to 400V across its collector/emitter. This is unusually high for a BJT.

The diagonal line is not a constant-power line. The dissipation of the transistor at 2A, with a V_{ce} of 20V is 40W. The dissipation of the transistor at 400V is and Ic of 0.03A is only 12W. Thus, the sloped line is at a steeper gradient than you would expect and it encloses less “safe area”.
This sloped line represents the “secondary breakdown” phenomenon. This phenomenon is caused by the formation of localised hot spots inside the transistor. This is caused by the negative temperature coefficient of the PN junctions. A similar problem called “thermal runaway” occurs on a larger scale.
Imagine trying to build a circuit which needs more collector current than one transistor can handle. If one decides to use parallel BJTs, the obvious, but flawed, circuit would look like this:



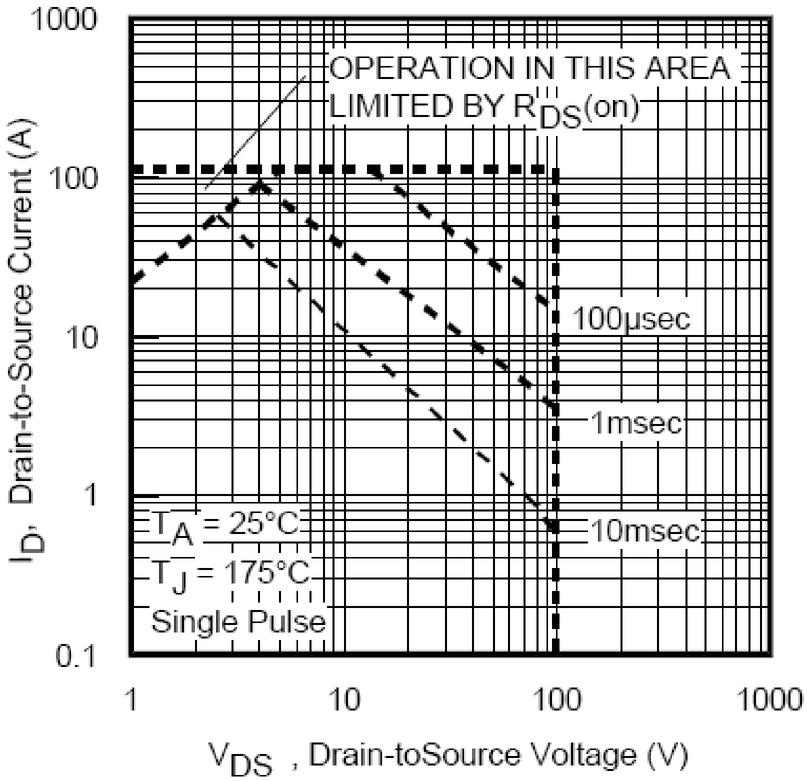
Source: Author's own diagram, (2013).

Suppose that the left transistor has a slightly lower Vbe than the right transistor. This will result in a higher Ib in that transistor. The higher Ib will result in higher Ic through the left transistor. Because the left transistor has higher Ic it will heat up more than the right transistor. A PN junction's voltage

drop *decreases* as it gets hotter. This causes the Vbe of the left transistor, causing even more unbalanced current distribution, leading to more heating, leading to lower V_{be}... until destruction of the transistor occurs.
If BJTs need to be put in parallel they need series resistors in their base and collector leads.

MOSFETs have a positive temperature coefficient and thus are immune to secondary breakdown. As a MOSFET heats up the voltage across it will increase (opposite to a PN junction) which provides self-limiting, preventing runaway conditions.

The SOA for an IRF540 MOSFET is shown here:



Source: Figure 8 *Maximum Safe Operating Area* 2001, IR International Rectifier, IRF540N HEXFET Power MOSFET, p.4.

This device is rated to carry a maximum of 110A for a short period. It can block 100V across it and has a maximum resistance of 0.044 ohms when turned on. The horizontal and vertical lines show these limits. The graph shows a family of dissipation curves. The reason for having four curves is that the MOSFET takes time to heat up. The shorter the pulse that it conducts the less the temperature will rise; hence the device being able to dissipate more power for short pulses.
The area labeled "Limited by Rds(on)" is the area which is limited by the 'on' resistance of the MOSFET. If the MOSFET's resistance is 0.044 ohms and 2V is put across it a maximum of 45A can flow (by Ohm's law), as indicated on the graph.

MOSFETs are very commonly used in parallel. Because of their positive temperature coefficient they need no series resistance, the parallel devices will share the load fairly evenly. As more MOSFETs are paralleled the gate capacitance will increase, increasing the importance of a good gate driver system.

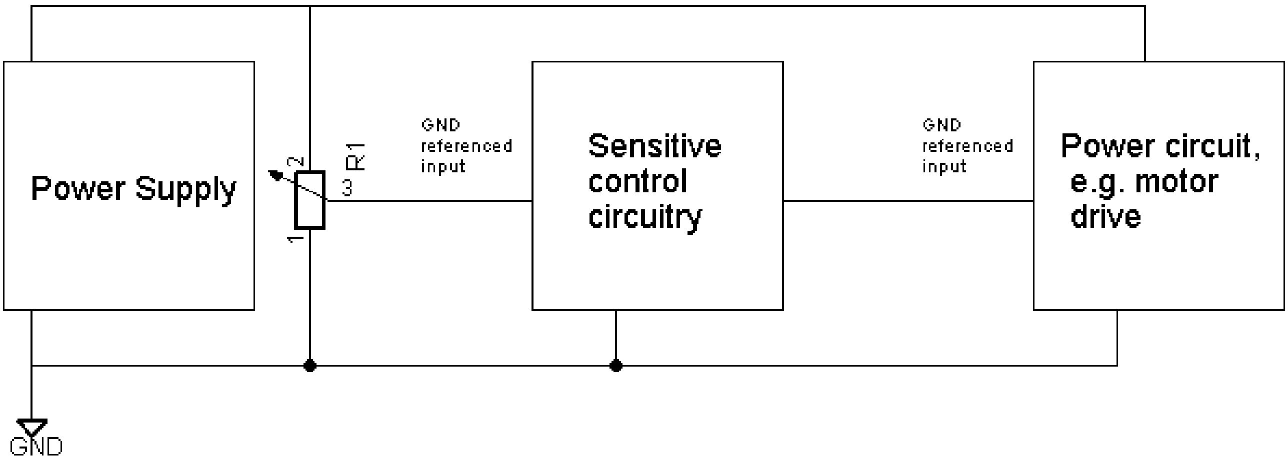
Chapter 12: Unintentional Circuits

This section is called "unintentional circuits" because it describes exactly that. Sooner or later one will move from simple single function lab circuits into building full systems. This will often introduce additional complications which need to be dealt with. Common sources of complications include:

- The mixture of digital circuitry and analogue circuitry. Digital systems often have high speed impulsive waveforms which do not mix well with the low level high impedance signals common in analogue systems,
- The mixture of high power circuitry and low signal levels,
- The physical size of the system increases.

Wiring

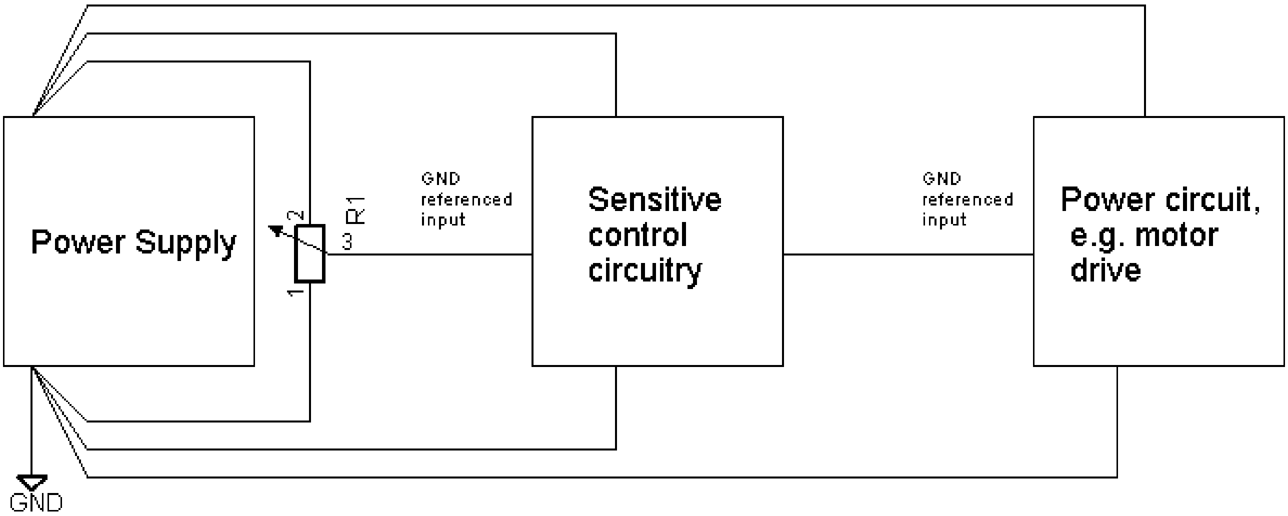
Consider the following system:



Source: Author's own diagram, (2013).

In reality, the interconnecting wires all have resistance and inductance. Both of these cause a voltage drop along the interconnecting wires when the power circuit switches the load on. This has two effects on the control circuitry. Firstly, the overall power supply drops. The control circuitry will experience dips and peaks in its supply rail. To some extent this could be dealt with by adding additional supply regulation within the control circuitry block, however, this is fairly wasteful in terms of circuitry used. The second issue is that the ground level, as seen by the control circuit, will vary. Many circuits use their incoming ground point as a reference voltage. Varying ground levels are often a big problem because they will often behave like signals which have been added to the input. (Although this also depends on the physical location of connectors and similar factors). If this was a closed loop control system there is a fair possibility that this would cause oscillation.

In this configuration the problem has been removed with a simple change in wiring:



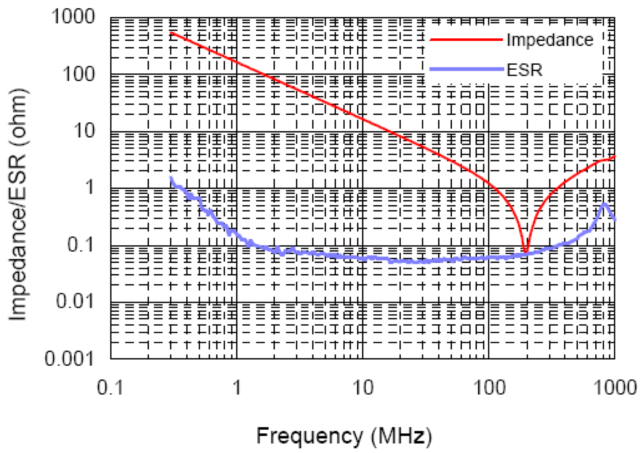
Source: Author's own diagram, (2013).

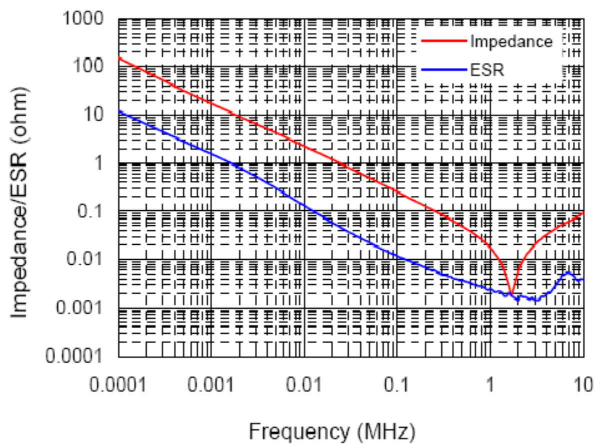
The current that is drawn by the large load has a minimal effect on the supply, as seen by the control circuitry. If the power supply is regulated the regulator will compensate (within limits) for drops in supply voltage caused by changes in the load.

Power Supply Decoupling

Each of the load modules in the diagram above should also have capacitance across its supply lines. This capacitance supplies short-term load peaks. The inductance of the supply lines opposes changes in current flow, thus limiting the ability of the power supply to supply transient current peaks. By locating a capacitor close to the load the inductance between the capacitor and the load is minimised.

Very often, circuits will have two decoupling capacitors in parallel. Typically, there will be a 10uF in parallel with a 100nF capacitor. This is seemingly pointless but examination of these graphs below will reveal that all capacitors have strengths and weaknesses:





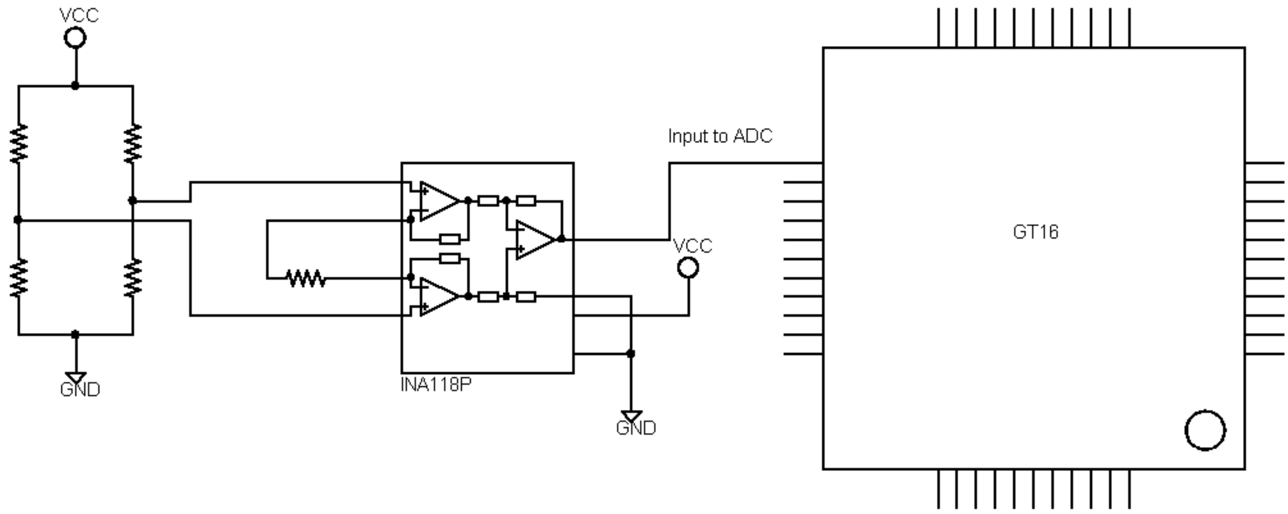
Source (left image): *Impedance/ESR - Frequency* (2006), Murata, Chip Monolithic Ceramic Capacitor Electrical Characteristics Data 0402 COG 1000pF 50V Murata Global Part No. GRM1555C1H102J, p.1

Source (right image): *Impedance/ESR - Frequency* (2007), Murata, Chip Monolithic Ceramic Capacitor Electrical Characteristics Data 0805 X5R 10uF 16V Murata Global Part No. GRM21BR61C106K, p.1

These graphs plot the impedance of two different types of ceramic capacitor versus frequency. Ideally the impedance of a capacitor should only depend on the value of the caopacitor and its operating frequency and thus the impedances of the capacitors should be a straight lines on the bode plots. However, because of the construction of the capacitors and the properties of the dielectric materials used, all capacitors have a maximum operating frequency. Thus, the 1nF COG ceramic capacitor behaves correctly up to about 100MHz, whereas the 10uF X5R ceramic capacitor only behaves correctly up to about 1MHz. Electrolytic capacitors are seldom useful above 100kHz.

Physical Layout

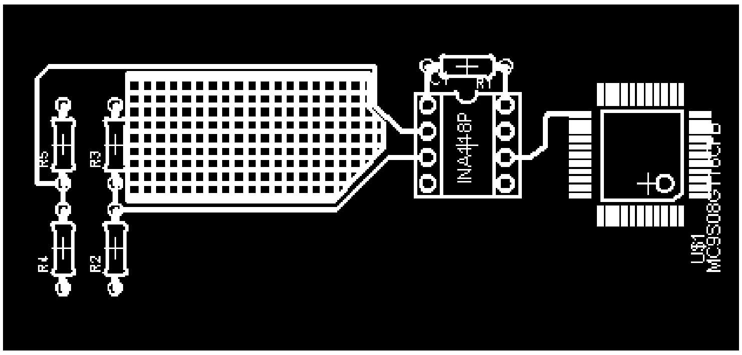
Imagine that you had the following circuit:



Source: Author's own diagram, (2013).

The INA118 is an instrumentation amplifier. This circuit would be typical of strain gauge applications. Often, the physical layout of the circuit board would be constrained by the product into which it must fit. As an example, it is usually convenient for the connectors to be along the edge of the circuit board.

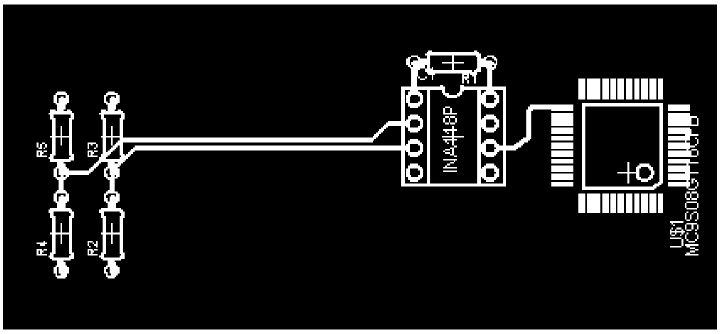
One might well decide to lay the circuit out on a circuit board like this:



Source: Author's own diagram, (2013).

The area shown with the white cross-hatching poses a problem. As one should be aware, the area enclosed by a loop will have a (linear) effect of the EMF induced in that loop by a changing magnetic field. This layout is poor because it has a large area for EMF to be induced in. The induced EMF will be amplified along with the signal and it will cause interference on the output.

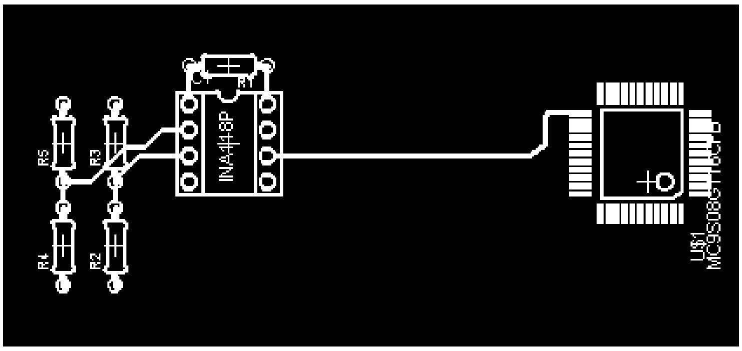
One simple improvement that can be made is to re-route the tracks as follows:



Source: Author's own diagram, (2013).

This circuit will perform much better in the presence of interference because it has a much smaller enclosed area cutting any interfering magnetic field.

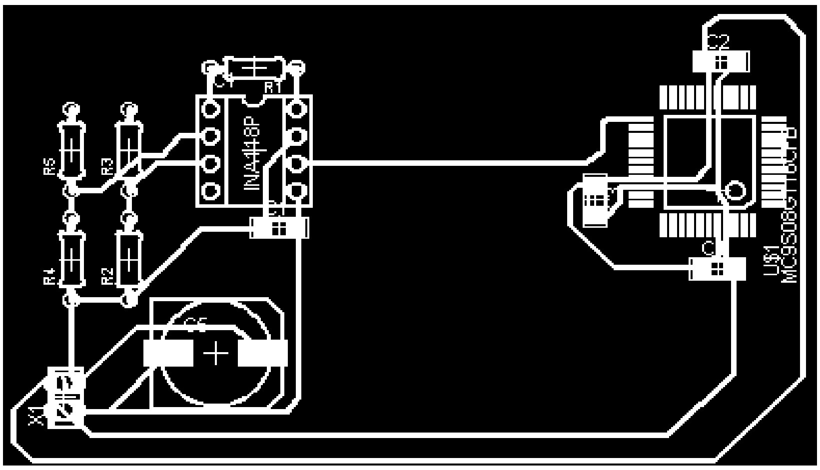
Another improvement would be as follows:



Source: Author's own diagram, (2013).

This is an improvement because it has an even smaller sensitive loop area. In addition, we have optimised the layout by running the less sensitive post-amplifier signal over the long distance and kept the more sensitive signal as short as possible.

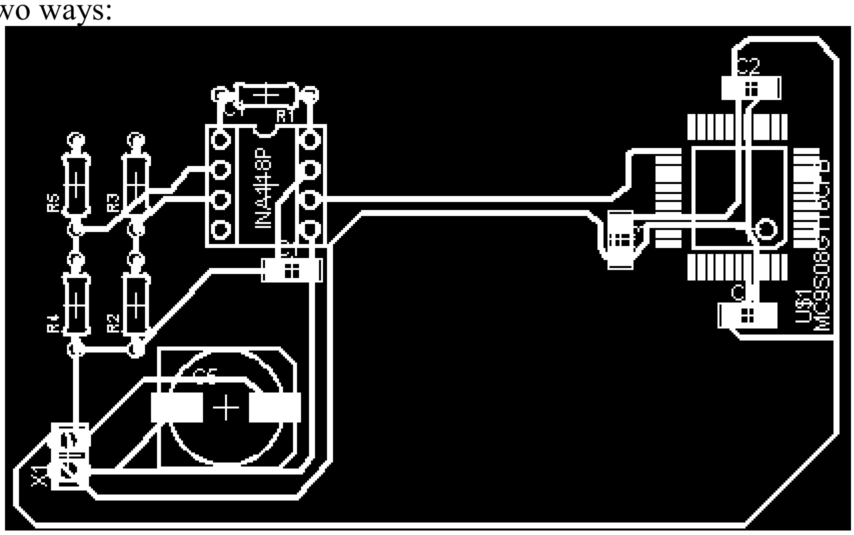
If the earlier ideas about power supply connection and decoupling are included, this is the result:



Source: Author's own diagram, (2013).

Notice that we have added a small decoupling capacitor close to each set of power supply pins (there are three sets on the GT16) and one large decoupling capacitor to ensure that the overall supply to the circuit board is stable even if the wires carrying power to the board are long.
This example also shows how a single point earthing on a circuit board can be used. The digital circuitry (GT16) and the analogue circuitry (instrumentation amplifier) each have their own supply lines running to the incoming power supply connector (X1) at the edge of the circuit board. This is not ideal, as the board is typically supplied through long wires, however, it is the best available power connection point on the board.

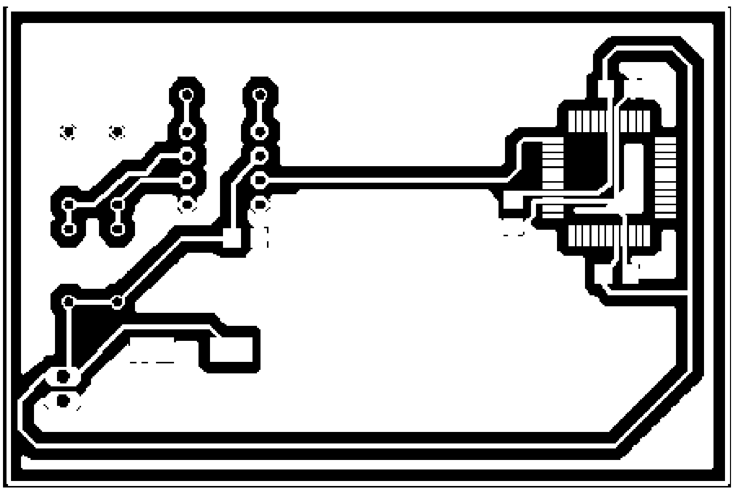
While the layout is the above figure is much better than before, it is still not optimal. The loop area between the ADC input and the ground signal that it is referenced to is larger than it should be. This can be fixed in two ways:



Source: Author's own diagram, (2013).

In this example the ground track to the microcontroller has been re-routed. In order to do this the positive track also had to move slightly.

Another approach is to use a "ground plane". This involves filling unused areas of the board with copper. When done properly this results in a lower impedance ground connection to the components and greatly reduced loop areas. Here is this board with a ground plane added:


Source: Author's own diagram, (2013).

Note that adding a ground plane does not always guarantee good results. It must be done intelligently. If the tracks are badly routed so that the ground plane is forced into a narrow strip then it will not help. Sometimes ground planes are "split" in order to separate the grounds of digital and analogue circuitry.

When the circuit board is complex it is often not possible to route all of the tracks on one side of the circuit board. In that case double-sided or multi-layer circuit boards are used. Connections between layers are made by metal plating through the holes in the board. Double sided and multilayer circuit boards open up a range of opportunities for good as well as bad design.

Using a ground plane on one side of a double sided board can improve two very important parameters.
Firstly, it can be used to reduce the amount of noise picked up by the circuit. The ground plane can be used to shield sensitive signals from interference.
Secondly, placing a ground plane under a track will reduce the inductance of the track. This can be a useful feature when fast rise and fall times are needed for the signal traveling down the track.

Routing noisy tracks under sensitive circuitry is a common way in which double sided boards are abused. This will cause noise to propagate from the noisy signal into the sensitive signal. If a noisy track and a sensitive track need to pass over each other, keep them perpendicular and, failing that, keep the length of the pass-over as short as possible.

Multi layer boards can offer very good performance because they allow a lot of control over the routing strategy used. An example of this would be a board which has digital control lines controlling analog circuitry. The problem is that the digital control lines are likely to interfere with the analog signals. It is not possible to simply route the digital lines away because they need to go to the same chips.

If one can use a four layer board it would be possible to do this:

Top layer/ Layer 1:	Analogue
Layer 2:	Analogue/ power supply
Layer 3:	Ground plane
Bottom/ Layer 4:	Digital lines

The ground plane prevents contamination of the analogue signals by the digital signals. In addition, because the entire layer is devoted to the ground plane it is easy to achieve a very low impedance ground connection. The order of layers is known in industry as the "stack up".

Exercise: How would you stack layers on a four layer board if optimal shielding from external interference was needed and no power plane was required?

Transmission Lines

Many digital circuits operate in the hundreds of MHz to GHz band. The trace on the circuit board is no longer a simple piece of wire. The lumped circuit element model no longer applies to components either.

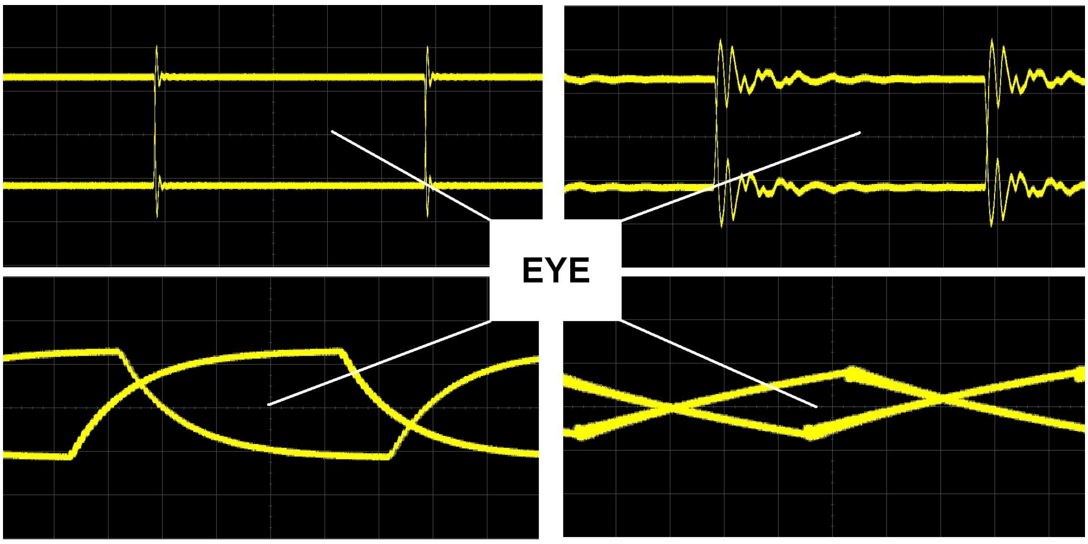
As a rule of thumb, when working at high frequencies the components should be as physically small as possible. Connecting leads onto the components should be as short as possible. This is one factor driving the use of surface mount components.

Very high speed circuit board tracks are designed using microstrip techniques. A signal traveling down a mismatched line (where the characteristic impedance of the line differs from the driving and receiving circuitry) will suffer distortion. Microstrip techniques are used to achieve impedance matching of circuit board traces in order to attain distortion-free high speed digital signals. There are a range of online “microstrip calculators” which allow you to enter the board parameters (such as the dielectric coefficient of the PCB material, thickness of the board etc.) and calculate the width of the track for a given impedance.

Differential signaling techniques are also used in the digital domain. Low Voltage Differential Signaling (LVDS) is one popular method. In this method, the transmitting side forces a current through a pair of conductors. The receiver has a resistor for the current to flow through and it examines the voltage across the resistor. The current is forced in one direction for a logic high and the other direction for a logic low. The receiver is sensitive to the polarity of the voltage across the resistor. Because only small voltage swings need to be present across the resistor, the slew rate of the system is a much smaller problem than with single ended systems. The current-carrying conductors are always run very close to each other. Because the currents in the conductor have a vector sum of zero the system radiates little noise and because the loop area is small the system is fairly noise immune.
LVDS systems often operate at gigabits per second.

As one may see from the above discussion, when a digital signal propagates through a channel the effects that it suffers are analogue in nature. Eye diagrams are an important tool in analysing and specifying the amount and nature of distortion that a signal may suffer. An eye diagram is generated by allowing the digital waveform to trigger an oscilloscope on either edge. The oscilloscope is put into “infinite persistance” mode and the waveforms accumulate on the screen. Here are some examples of eye diagrams:

The top left diagram shows the eye diagram for a good quality square wave. The “eye” is the central part. The bigger the area of the eye, the easier it is to distinguish the logic levels. The top right illustration shows the eye diagram for a digital signal passing through a slightly inductive channel. The bottom diagrams show a digital signal passing through capacitive channels.



Source: Author's own diagram, (2013).

USB3.0 signals pass through long cables. Here is an eye diagram showing distortion before and after the addition of anti-static protection devices on the signal lines.

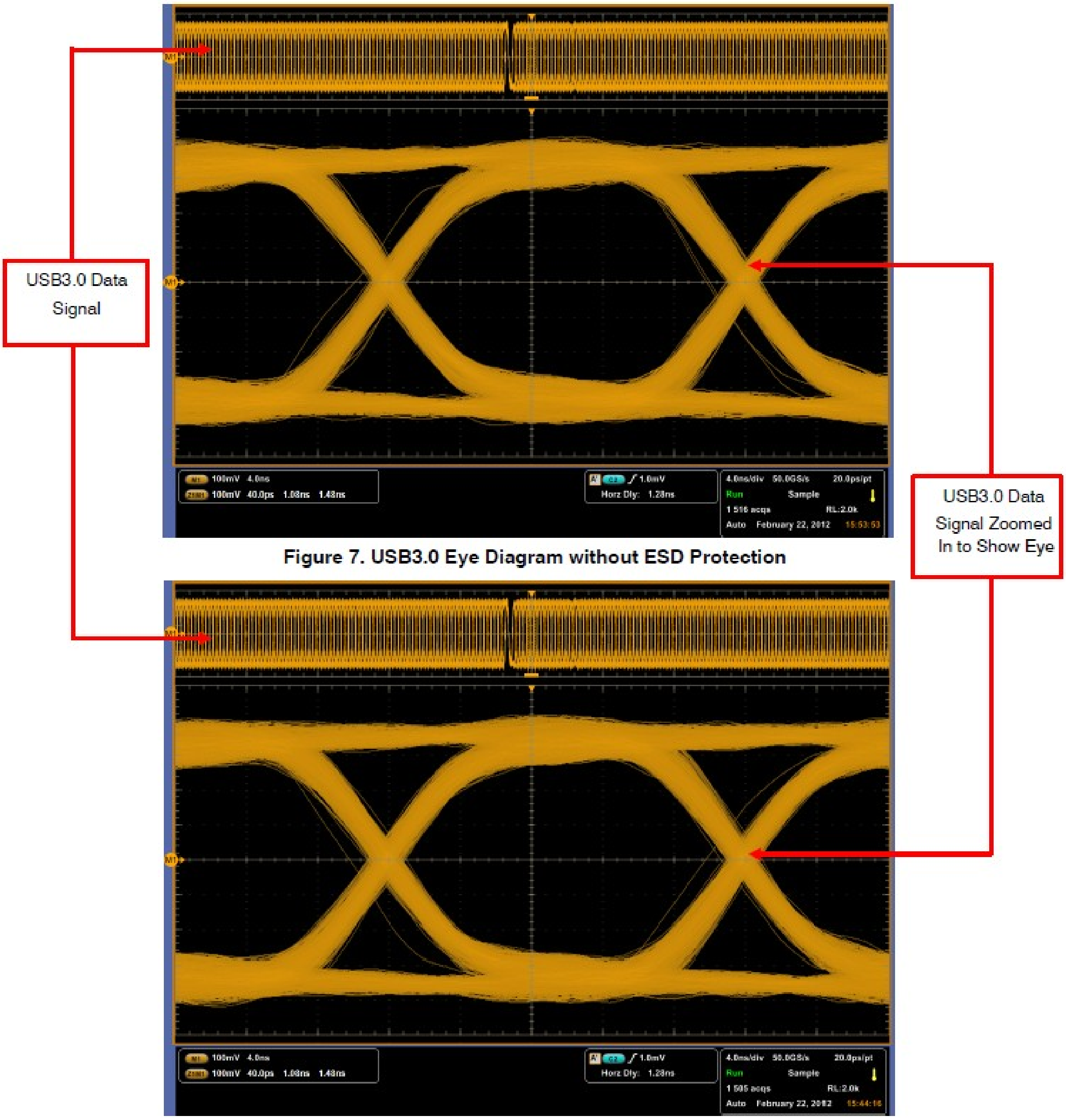
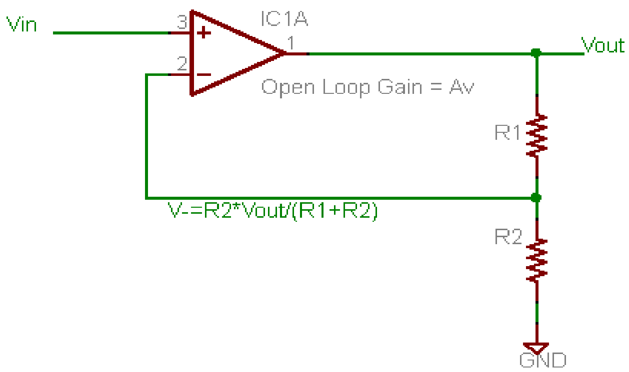


Figure 7. USB3.0 Eye Diagram without ESD Protection

Source: Figure 8. USB3.0 Eye Diagram with ONsemi ESD7016 2012, ON Semiconductor, “AND9075/D Understanding Data Eye Diagram Methodology for Analyzing High Speed Digital Signals” p. 6

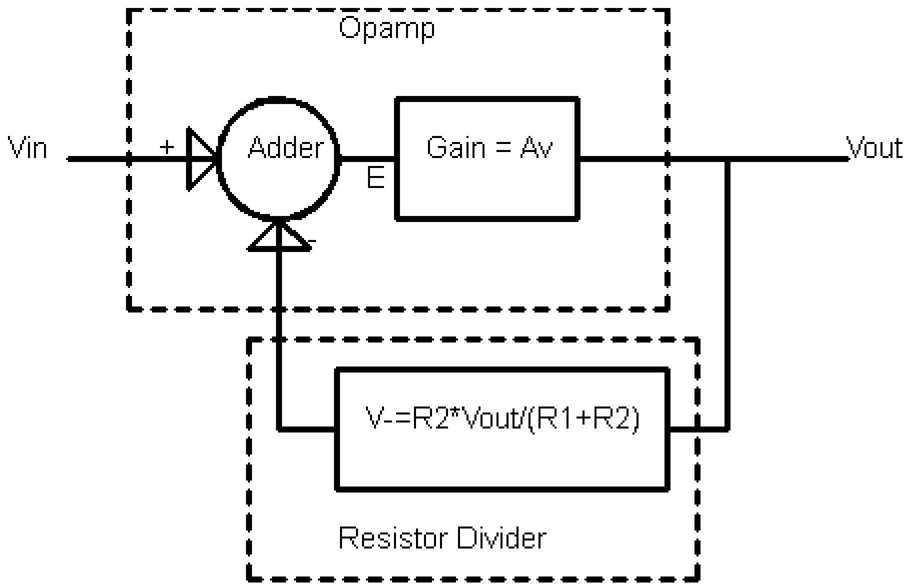
Appendix 1 – Open Loop versus Closed Loop Gain

Consider this circuit:



Source: Author's own diagram, (2013).

In block diagram form, this circuit can be modelled as follows:



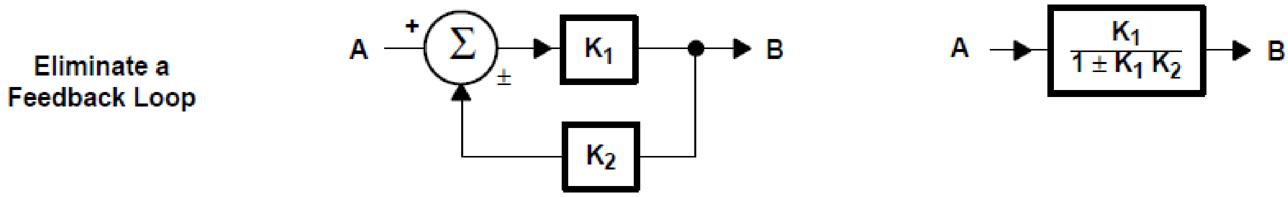
Source: Author's own diagram, (2013).

In the diagram above, the "E" is the error term. $E=\frac{V_{out}}{A_v}$

Clearly, in the ideal case E would be zero, but practical limitations of the opamp (limited gain) prevent this from happening.

Furthermore, from block diagram algebra (from "Opamps for Everyone", p.5-5) we see:

(Note that there is an error in the book with regard to the sign in the denominator).



Source: *Block Diagram Math and Manipulations* 2002, Ron Mancini "Opamps for Everyone: Design Reference", Block Diagram Transforms, 5-5, p.79.

From this one can see that the gain is not simply $G=\frac{R1+R2}{R2}$, but is actually:

$$G=\frac{K_1}{1+K_1K_2}=\frac{A_v}{1+A_v\frac{R_2}{R_1+R_2}}$$

If $A_v\frac{R_2}{R_1+R_2}\gg 1$ then indeed the closed loop gain is $G\approx R_1+\frac{R_2}{R_2}$

What happens when this is not true?

For the condition $A_v\frac{R_2}{R_1+R_2}\gg 1$ to be true, $A_v\gg\frac{R_1+R_2}{R_2}$. Effectively, if we want an accurate transfer function then Av must be much bigger than the desired closed loop gain.

Here is a table showing the effect of this. For now, one assume that all of the calculations are at DC, that the resistors are perfectly accurate, and that the opamp has a gain of 100dB or 100 000.

Gain, as set by R1 and R2	Gain, as achieved when Av=100 000	Error
1	0.99999	0.001%
10	9.999	0.01%
100	99.9	0.10%
1000	990	1%
10000	9090	9%

The next idea might be to simply compensate for the slightly lower gain by increasing the ratio of the resistors. This, however, fails for one simple reason: the specified gain of the opamp is a *minimum,* so the manufacturer may give as much Av as they want without informing people.

If one needs more gain than can be provided within the required error margin there are a few possibilities:

One could buy a better opamp. Many precision opamps provide higher Av specifically to enhance their DC gain accuracy. An example would be the LTC1150 which guarantees Av>135dB.

One could also cascade amplifier stages.

As an example, if one were using an LTC1150 and desired a gain of accuracy of 1% one can achieve a maximum DC closed loop gain of around 57000.

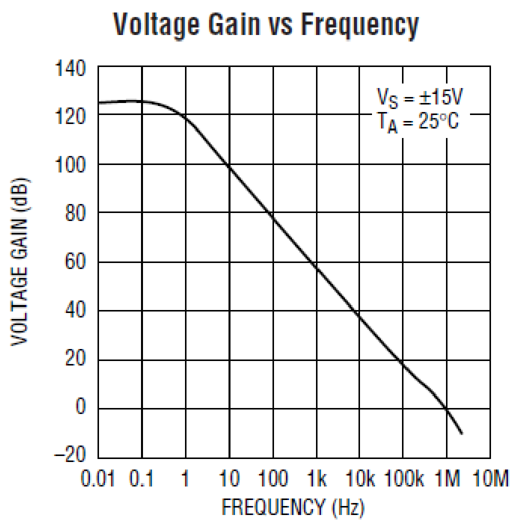
It should also be noted that it is somewhat pointless to specify 0.1% tolerance resistors as one's circuit is not capable of achieving this because of Av limitations.

Appendix 2 – Gain Bandwidth Product

This appendix requires a knowledge of filters, as given in the chapter on that subject in this work.

A real opamp has an internal capacitor as shown earlier in the notes to improve stability when the amplifier is placed in a negative feedback loop. The concepts of stability will be explored in a later appendix.

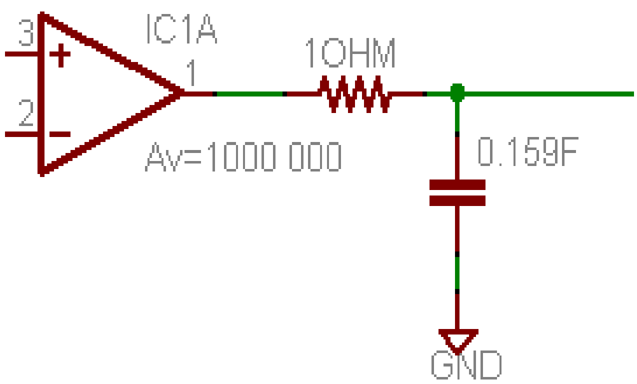
These "compensated" opamps do not have constant Av over their operating frequency range. A better model would be an ideal amplifier followed by a low pass filter. At low frequencies we can consider this filter to be first order, as shown in this graph:



Source: *Typical Performance Characteristics* 1989, Linear Technology Corporation Datasheet, LT1097 Low Cost, Low Power Precision Opamp, p.6.

When examining the graph one will see that the -3dB point is at approximately 1Hz and that the slope of the graph is -20dB/decade so this amplifier behaves like a block with a gain of about 1000 000 followed by a first order 1Hz low pass filter.

In circuit form we get this (assuming that the opamp in the following circuit has a perfect frequency response):



Source: Author's own diagram, (2013).

As a transfer function for the overall equivalent circuit we can derive that

$$H(s)=\frac{A_v}{1+RCs}$$

For derivation consider that this is a first order low pass filter with a 1Hz -3dB point. This, in more familiar frequency terms, along with R=1ohm and C=0.159F, gives us:

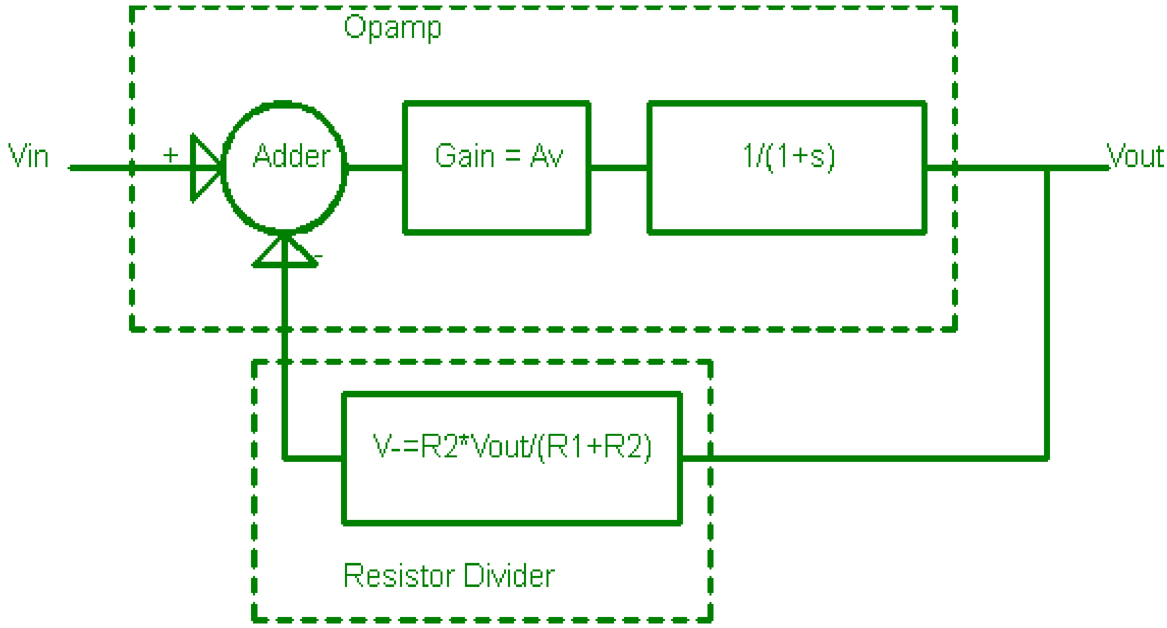
$$H(f)=\frac{A_v}{1+jf} \textit{ where } j=\sqrt{-1}$$

One can observe that effective open loop gain drops approximately linearly with increasing f, resulting in a constant product of gain and operating frequency, hence the approximately constant “gain bandwidth product”.

One will also observe that, because of the 'j' term, that opamps do also create a phase shift which increases with frequency. At very low frequencies the output moves in phase with the non-inverting input and 180 degrees out of phase with the inverting input. At 1Hz the phase shift of the opamp is 45 degrees, so the output waveform will be at 45 degrees to the non-inverting input and 135 degrees to the inverting input.

What is the effect of Gain Bandwidth Product on an amplifier's ability to handle high gains at high frequencies? We have shown, in Appendix 1, that there is a relationship between open loop gain and achievable closed loop gain.

For the circuit in Appendix 1 the block diagram for the loop is as follows, assuming the use of LT1097:



Source: Author's own diagram, (2013).

The closed loop transfer function is now:

$$G=\frac{A_v}{1+s+A_v\frac{R_2}{R_1+R_2}}$$

For accurate operation we want the term (1+s) in the denominator to be insignificantly small. This requires that both of the following conditions are met:

- AvR2/(R1+R2) >>1   so we are constrained at DC, as explained in appendix 1
- |AvR2/(R1+R2)| >>|s| which means that there is a frequency dependence as well, since |s| increases with increasing frequency.

To show the effects of this on accuracy, below is a table. The rows show different possible gains, whilst the columns show different possible frequencies. We have assumed the use of the LT1097 with its 120dB DC open loop gain and 1Hz -3dB point.

	0Hz	1Hz	10Hz	100Hz	1kHz
Gain=1	G=0.9999999 Φ=0deg	G=0.999999 Φ=0.0004deg	G=0.999999 Φ=0.004deg	G=0.999999 Φ=0.036deg	G=0.999979 Φ=0.36deg
Gain=10	G=9.9999 Φ=0deg	G=9.9999 Φ=0.004deg	G=9.9999 Φ=0.036deg	G=9.9997 Φ=0.36deg	G=9.98 Φ=3.59deg
Gain=100	G=99.99 Φ=0deg	G=99.98 Φ=0.036deg	G=99.98 Φ=0.36deg	G=99.79 Φ=3.59deg	G=84.67 Φ=32.1deg
Gain=1000	G=999 Φ=0deg	G=998.98 Φ=0.36deg	G=997 Φ=3.59deg	G=846 Φ=32.1deg	G=157 Φ=80.9deg
Gain=10000	G=9900 Φ=0deg	G=9881.9 Φ=3.56deg	G=8407 Φ=31.9deg	G=1571 Φ=80.9deg	G=159 Φ=89.1deg

As can be seen from this table, this precision opamp can give extremely accurate results at moderate gains and low frequencies. At high frequencies and gains the accuracy, both in gain and in phase, can rapidly become appallingly bad. All opamps display this tendency although similar devices, called current feedback amplifiers (CFAs), are a good substitute when gain bandwidth product is a problem.

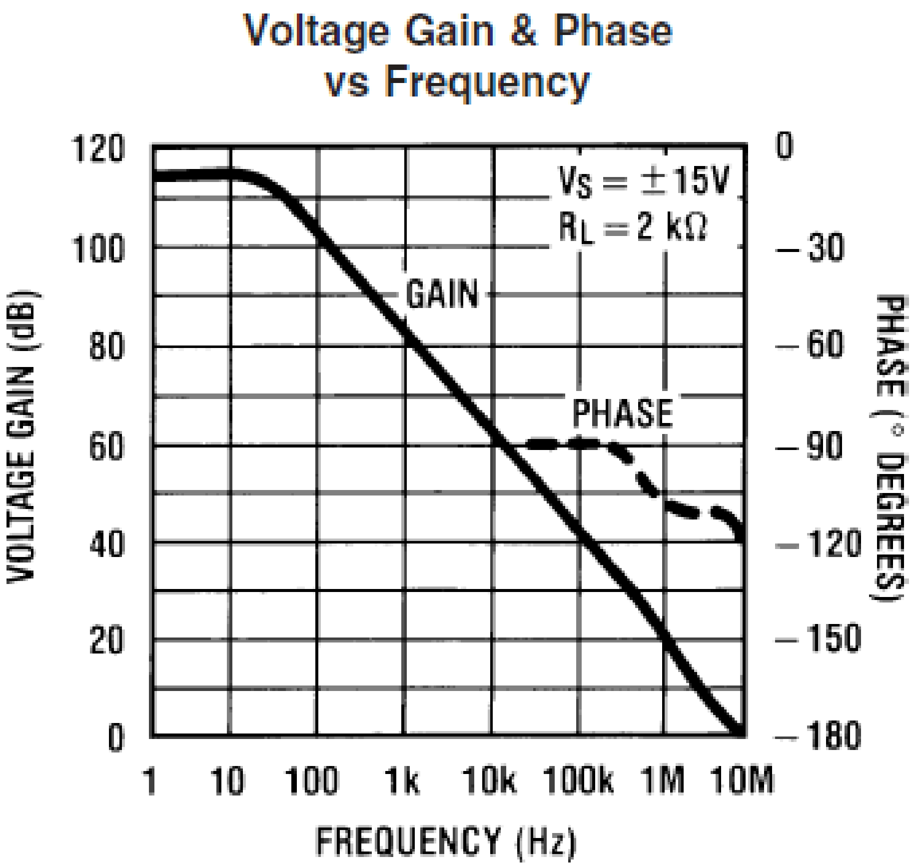
The gain bandwidth product of the LT1097 is specified as 700kHz. As in the table above, our rule of thumb of “GBWP>10 x Closed Loop Gain x Highest Frequency” results in a gain accuracy better than 10% and a phase accuracy better than 30 degrees.

Appendix 3 – Stability of Opamp Circuits

Not all feedback loops are stable. A badly designed feedback loop will oscillate which, in the case of amplifiers, is an undesired effect. The Barkhausen criteria determine whether any feedback loop will oscillate or amplify and unintentional satisfaction of these criteria will cause unwanted oscillation.

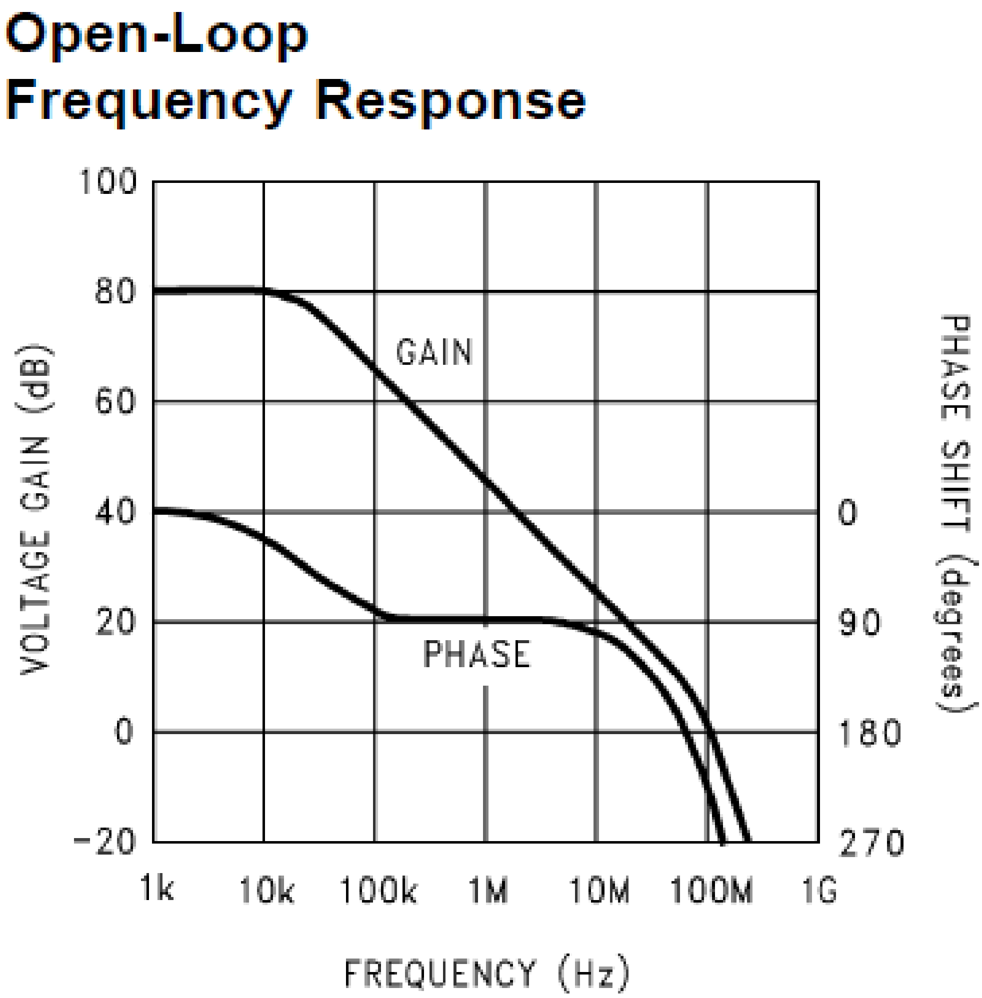
In the case of an amplifier feedback loop, the negative feedback to the inverting input provides 180 degrees of phase shift. If the accumulated phase shift in the amplifier and feedback loop amounts to a further 180 degrees while the loop gain is greater than unity the Barkhausen criteria will be satisfied.

Compensated opamps have an internal capacitor which causes their gain to roll off so that they reach unity gain while their phase shift is still comfortably below 180 degrees. As an example, below is the gain/phase plot of the LM833. One will see that at unity open loop gain (0dB) it has only 120 degrees of phase shift, so it is stable if placed in a feedback loop:



Source: *Typical Performance Characteristics* 2003, National Semiconductor, LM833 Dual Audio Operational Amplifier, p.5.

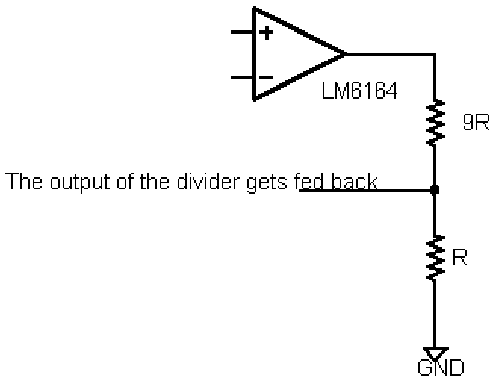
High speed opamps, however, do not have this convenient property; they trade off stability for high frequency performance. An example is the LM6164, which has a gain/phase plot as shown here:



Source: *Typical Performance Characteristics* 1999, National Semiconductor, LM6164/LM6264/LM6364 High Speed Operational Amplifier, p.7.

At 100MHz the gain reaches unity, but at that frequency the phase shift caused by the opamp is about 225 degrees. Thereofore, this opamp, if put in a unity gain voltage follower circuit, will oscillate.

Under what conditions will the LM6164 be stable in a feedback loop? Imagine that it had *less* gain, and suppose we reduced the opamp's gain by a factor of 10, which is equivalent to subtracting 20dB from the graph above. The phase shift at unity gain would then be about 110 degrees and the feedback loop would be stable. We could do this with the following circuit:



Source: Author's own diagram, (2013).

The resistive divider effectively reduces the open loop gain and makes the loop stable. The bigger the voltage division is, the more phase margin the loop will have.

One should notice that putting a voltage divider into the feedback loop also *increases* the closed loop gain. This seems counter-intuitive; as one *increases* the closed loop gain one *improves* the stability of this feedback loop. As strange as it may seem, it is correct.

> “The LM6164 family of high-speed amplifiers exhibits an excellent speed-power product in delivering 300V per µs and 175 MHz GBW (stable down to gains as low as +5) with only 5 mA of supply current. Further power savings and application convenience are possible by taking advantage of the wide dynamic range in operating supply voltage which extends all the way down to +5V” (National Semiconductor Corporation 1999, “LM6164/LM6264/LM6364 High Speed Operational Amplifier” Datasheet, p.1).

This device is guaranteed to be stable in feedback as long as the closed loop gain is *greater* than 5.

As noted in the chapter on oscillators, it is also possible for output loads to cause the opamp circuit to oscillate. Consider the following extract:

FEATURES

Single-supply operation
- **Output swings rail-to-rail**
- **Input voltage range extends below ground**
- **Single-supply capability from 3 V to 36 V**

High load drive
- **Capacitive load drive of 500 pF, G = +1**
- **Output current of 15 mA, 0.5 V from supplies**

Excellent ac performance on 2.6 mA/amplifier
- **−3 dB bandwidth of 16 MHz, G = +1**
- **350 ns settling time to 0.01% (2 V step)**
- **Slew rate of 22 V/µs**

Good dc performance
- **800 µV maximum input offset voltage**
- **2 µV/°C offset voltage drift**
- **25 pA maximum input bias current**

Low distortion: −108 dBc worst harmonic @ 20 kHz

Low noise: 16 nV/√Hz @ 10 kHz

No phase inversion with inputs to the supply rails

Source: *Features information* 2007, Analog Devices, AD 823: Dual, 16MHz, Rail-to-Rail FET Input Amplifier Datasheet, p.1.

(Note specifically the capacitive drive capability).

As well as this graph showing open loop gain versus the load placed on the output of the opamp

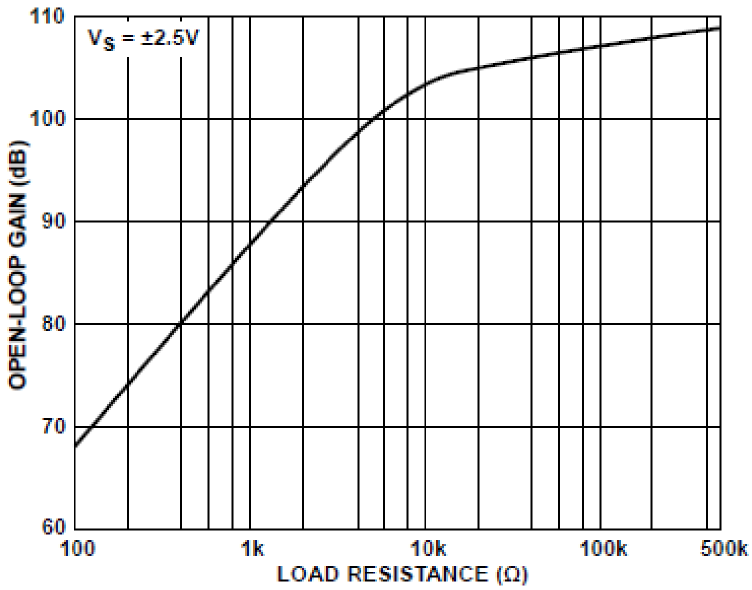


Figure 11. Open-Loop Gain vs. Load Resistance

Source: Figure 11 *Typical Performance Characteristics* 2007, Analog Devices, AD 823: Dual, 16MHz, Rail-to-Rail FET Input Amplifier Datasheet , p.8.

This gives us a clue about the internal impedance of the opamp's output. It is often necessary to rely on graphs such as these because output impedance is seldom specified explicitly. From the slope of the graph above it is easy to calculate that the internal impedance of this opamp is about 7 kilo-ohms.

This graph shows the opamp's phase margin (here the phase margin is shown rather than phase itself) and gain versus frequency:

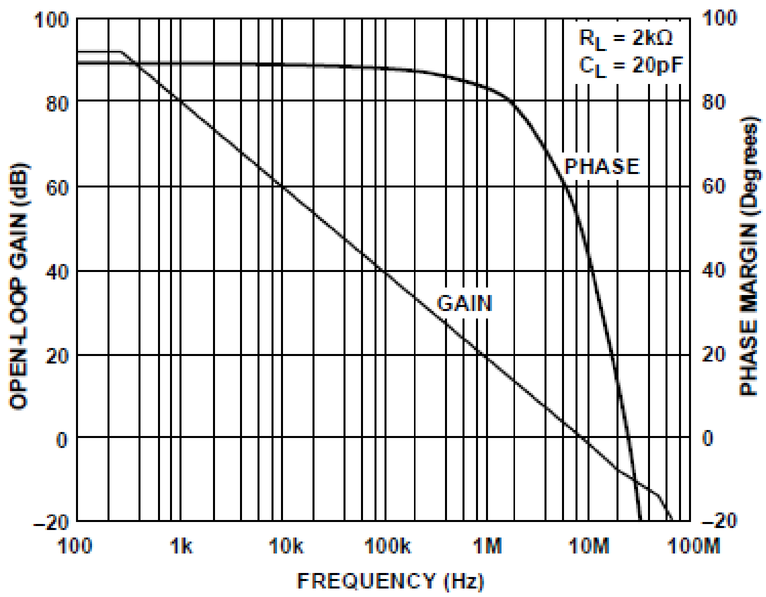
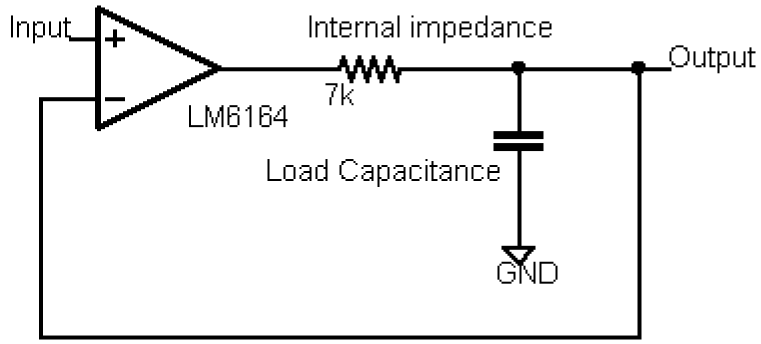


Figure 15. Open-Loop Gain and Phase Margin vs. Frequency

Source: Figure 15 *Typical Performance Characteristics* 2007, Analog Devices, AD 823: Dual, 16MHz, Rail-to-Rail FET Input Amplifier Datasheet , p.8.

One should note that this opamp has a phase margin of around 50 degrees at unity gain, so it is unity gain stable.

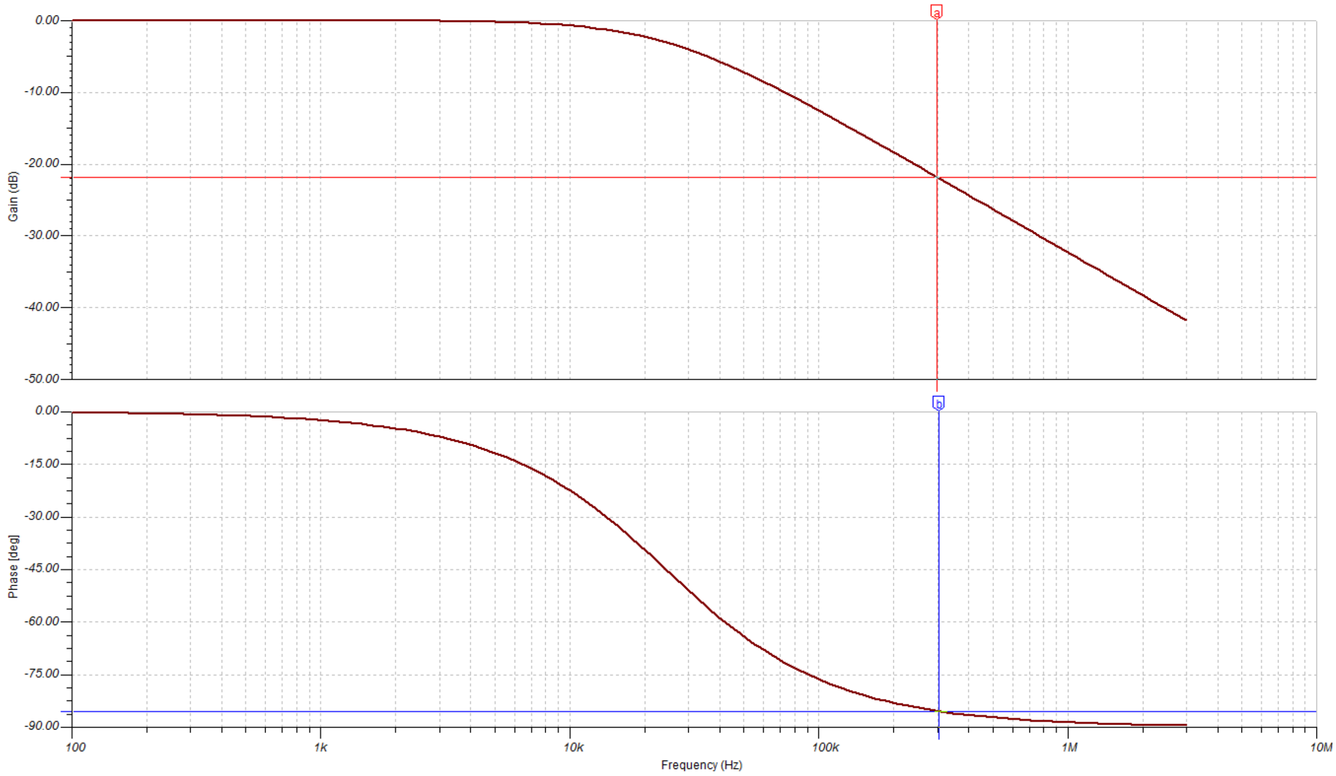
If we introduce a capacitive load on the output pin of the opamp then the combination of the opamop's internal impedance and the load capacitance adds additional phase shift (and high frequency rolloff) to the feedback loop. To give a concrete example consider the problem of using an AD823 to drive 10m of RG-58/U coaxial cable (You'll have handled this in the lab, it's commonly used for signal generator output leads) which typically has a capacitance of 93.5pF/m.



"Unity Gain" follower with capacitive load and internal resistance shown

Source: Author's own diagram, (2013).

From this we can calculate that the load capacitance is 0.935nF, and it is simple to calculate (or simulate) the gain and phase response of the filter formed by the internal impedance and the load capacitance.



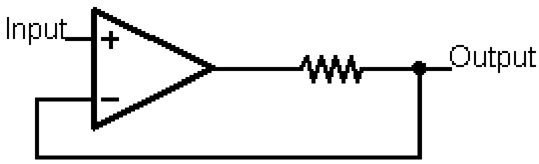
Source: Author's own diagram, (2013).

If we combine this filter response with the opamp's bode plot from the datasheet we find that the phase margin drops to zero when the available gain is about 8dB at around 300kHz. This system will oscillate when we connect the output cable.

Appendix 4 – Output Impedance of Opamp Circuits

In Appendix 3 the graph of open loop gain versus output impedance was shown for the AD823 and it was stated that the output impedance of this device is around 7 kilo-ohms. From a similar graph shown in the LM6164 datasheet it can be calculated that the output impedance of that device is around 3500 ohms. This would seem worryingly high because it implies that a small load current would lead to a large resistive drop inside the opamp; and hence poor accuracy of the overall circuit. Fortunately, the graphs shown are for the device in open loop configuration and placing the opamp in a feedback loop improves the situation dramatically.

First, one needs to understand why the feedback loop improves the output impedance. Here is a model of an opamp in a feedback circuit, with its parasitic internal impedance shown explicitly:



Source: Author's own diagram, (2013).

Notice that the feedback comes from *after* the internal resistance. This means that any voltage drop across the internal resistance is compensated for by the feedback mechanism. This implies, in an ideal situation, that as the load causes voltage drop across the internal impedance it is compensated for by the opamp raising its output voltage by exactly the same amount. If this worked perfectly then the output impedance would be zero, since Zout = ΔVout/ΔIout and ΔVout would then be zero.

In reality, the feedback loop is limited in two ways: firstly it has limited loop gain and secondly it has limited bandwidth. Both of these have an effect.

Feedback theory gives the following equation for closed-loop output impedance:

Zout = Rinternal/(1+AvB) where Rinternal is the internal resistance of the opamp, Av is the open loop gain of the opamp, and B is the gain of the feedback network.

Thus, for an AD823 in a closed-loop configuration with a gain of 10 the output impedance is about 2.2 ohms. This is rather more usable. If the AD823 is placed in a unity gain configuration it will give an effective output impedance of 0.22 ohms.

Because Av drops with increasing frequency the effective output impedance of an opamp circuit does not stay constant over frequency. Generally, it is to be expected that the output impedance of an opamp circuit will increase (worsen) as the frequency of operation increases.

References

Chapter 3

- *Shielded SMT Power Inductors LPS3008 Series* 2009, CoilCraft Document 438-2, p.2.

Chapter 4

- *High-Speed Diodes (1N4148; 1N4448)* 2004, Koninklijke Philips Electronics, Philips Semiconductors Product Specification Data Sheet, p.4.
- *Zener Voltage Regulator* 2009, BZX84B4V7LT1, BZX84C2V4LT1 Series: Zener Voltage Regulators, On Semiconductor Data Sheet for BZX84 Diodes, p. 2

Chapter 5

- *DC Current Gain* 2005, 2N3055(NPN), MJ2955(PNP): Complementary Silicon Power Transistors, On Semiconductor Data Sheet, p. 3
- *NPN General Purpose Amplifier* 2010, *PN2222A/MMBT222A/ PZT2222A NPN General Purpose Amplifier*, Fairchild Semiconductor Data Sheet for PN222A/MMBT2222A/PZT2222A p. 1
- *N-channel 525 V, 1 Ω*, 5 A, D²PAK, DPAK, TO-220FP, TO-220 SuperMESH3™ Power MOSFET 2011, STB6N52K3,STD6N52K3 STF6N52K3, STP6N52K3: Supermesh Power MOSFET, ST Microelectronics Data Sheet P. 1
- *NPN Epitaxial Darlington Transistor Equivalent Circuit* 2001, Fairchild Semiconductor Data Sheet, TIP 120/121/122: Medium Power Linear Switching Applications, p.1.

Chapter 6

- *Schematic of LM358 Dual Operational Amplifier* 2002, Texas Instruments, LM158, LM158A, LM258, LM258A, LM358, LM358A, LM2904, LM2904Q Data Sheet, p.3, (amended by author).
- *INA217 Instrumentation Amplifier* 2002, Burr Brown Products (Texas Instruments Incorporated), Low-Noise, Low-Distortion Instrumentation Amplifier p.1.
- *Class A Audio Amplifier* 1969, "Wireless World", John Linsley Hood.

Chapter 7

- *LM833 Audio Operational Amplifier* 2003, National Semiconductor, "Dual Audio Operational Amplifier" Datasheet, p.5.

Chapter 8

- "Q and bandwidth of a resonant circuit", *http://www.allaboutcircuits.com/vol_2/chpt_6/6.html*
- *Typical Performance Characteristics of the Precision Opamp* 1989, Linear Technology, LT1097 Low Cost, Low Power Precision Opamp Datasheet, p. 7.
- *Universal State-Variable 2nd-Order Active Filter 2010*, National Semiconductor Application Note 779 A Basic Introduction to Filters - Active, Passive,and Switched Capacitor P. 18.
- *The Art of Electronics* (Second ed.), Paul Horowitz and Winfield Hill(1989), Cambridge University Press, ISBN 978-0-521-37095-0 p. 274

Chapter 9

- *Figure 15-14 Phase Shift Oscillator (Single Op Amp)* 2002, Ron Macini (ed.), Texas Instruments "Opamps for Everyone" Design Regerence, p. 15-15.
- *Voltage Gain & Phase vs Frequency* 2003, National Semiconductor, LM833 Opamp, p.5.

Chapter 10

- *LP2950/LP2951 Series of Adjustable Micropower Voltage Regulators* 2004, National Semiconductor Datasheet, LP251, p.16.
- "Index of /wp-content/gallery/electronics-packages" http://cladlab.com/wp-content/gallery/electronics-packages/component-package-to-220ab-mounted-on-heatsink.jpg accessed 10 December 2013

Chapter 11

- *Simple mains powered isolated SMPS* 2005, Power Integrations, Top242-250 TopSwitch-GX Family extended power, design flexible, EcoSmart, Integrated off-line switcher, p.1.
- *Typical operating circuit of +5V Output Step-Down Regulator* 1996, Maxim Integrated Products +5/Adjustable CMOS Step-Down Switching Regulator, MAX638, p.1.
- Figure 15: *Idealized Negative Converter* 1999, Intersil Datasheet, ICL7662, p. 7.
- Figure 18A: *Simple Negative Converter and its Output Equivalent* 1999, Intersil Datasheet, ICL7662, p. 8.
- *Functional Block Diagram* 2007, Texas Instruments Incorporated, TPA3122D2 Class D Audio Amplifier, p.5.
- *Typical Connection* 2003, IR International Rectifier Datasheet, IR2113 High and Low side Driver, p. 1.
- Figure 5 *Maximum Safe Operating Regions* 2002, Bourns, BUX84 NPN Silicon Power Transistor, p.4.
- Figure 8 *Maximum Safe Operating Area* 2001, IR International Rectifier, IRF540N HEXFET Power MOSFET, p.4.

Chapter 12

- *Impedance/ESR - Frequency* (2006), Murata, Chip Monolithic Ceramic Capacitor Electrical Characteristics Data 0402 COG 1000pF 50V Murata Global Part No. GRM1555C1H102J, p.1
- *Impedance/ESR - Frequency* (2007), Murata, Chip Monolithic Ceramic Capacitor Electrical Characteristics Data 0805 X5R 10uF 16V Murata Global Part No. GRM21BR61C106K, p.1
- Figure 8. USB3.0 Eye Diagram with ONsemi ESD7016 2012, ON Semiconductor, "AND9075/D Understanding Data Eye Diagram Methodology for Analyzing High Speed Digital Signals" p. 6

Appendix 1

- *Block Diagram Math and Manipulations* 2002, Ron Mancini "Opamps for Everyone: Design Reference", Block Diagram Transforms, 5-5, p.79.

Appendix 2

- Typical Performance Characteristics 1989, Linear Technology Corporation Datasheet, LT1097 Low Cost, Low Power Precision Opamp, p.6.

Appendix 3

- *Typical Performance Characteristics* 2003, National Semiconductor, LM833 Dual Audio Operational Amplifier, p.5.
- *Typical Performance Characteristics* 1999, National Semiconductor, LM6164/LM6264/LM6364 High Speed Operational Amplifier, p.7.
- *Features information* 2007, Analog Devices, AD 823: Dual, 16MHz, Rail-to-Rail FET Input Amplifier Datasheet, p.1.
- Figure 11 *Typical Performance Characteristics* 2007, Analog Devices, AD 823: Dual, 16MHz, Rail-to-Rail FET Input Amplifier Datasheet , p.8.
- Figure 15 *Typical Performance Characteristics* 2007, Analog Devices, AD 823: Dual, 16MHz, Rail-to-Rail FET Input Amplifier Datasheet , p.8.