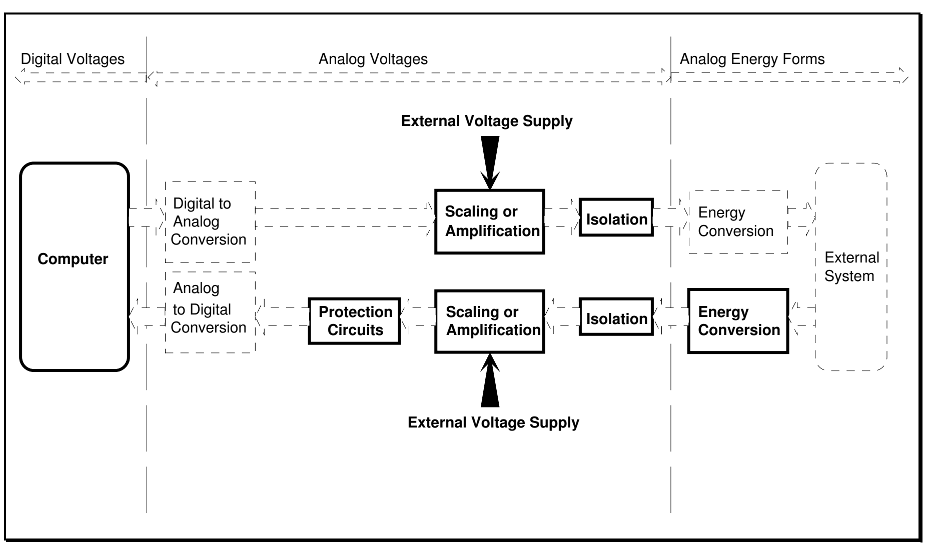
Chapter 3

Fundamental Electrical and Electronic Devices and Circuits

A Summary...

An overview of the electrical and electronic devices that are the basis of modern analog and digital circuits. Basic analog devices including diodes, Bipolar Junction Transistors (BJTs) and Field Effect Transistors (FETs). Diode based circuits including regulators and rectifiers. Simple analog transistor amplifier circuits. Operational amplifier circuits. Transistors for digital logic. Interfacing circuits to one another - input/output characteristics. Thyristors and thyristor based circuits.



3.1 Introduction to Electronic Devices

Most people can readily relate to the concept of resistance in electric circuits and to the concept of energy storage and release through inductance and capacitance. Our understanding is greatly enhanced by the fact that relatively straightforward and systematic techniques can be used to model circuits with these elements. In the so-called "time-domain" we can use the simple interrelationships between voltage and current in these devices to analyse these passive circuits:

$$
\begin{aligned}
&v = iR \quad \text{(Resistor Relationship - Ohm's Law)}\\
&i = C\frac{dv}{dt} \quad \text{(Capacitor Relationship)}\\
&v = L\frac{di}{dt} \quad \text{(Inductor Relationship)}
\end{aligned}
\qquad ...(1)
$$

We couple these relationships with the use of Kirchoff's voltage and current laws in order to analyse circuits. As circuits become more complex, we introduce the traditional mathematical approaches to the solution of differential equations in order to determine the transient and steady-state conditions of circuits. The *LaPlace Transform* technique and the *Phasor-Method* technique are, respectively, the two most common (and interrelated) methods used to solve for transient and steady-state circuit conditions.

When we introduce common electronic devices, such as diodes, transistors, etc. into our circuits, analysis becomes far more complicated - particularly when circuits are used for analog applications. Not only do we have to contend with all of the above analysis techniques, but we additionally need to consider dependent voltage and current sources that add substantially to analysis problems. Moreover, analysis of circuits with devices such as diodes and thyristors requires the use of intuition in order to simplify circuits that are otherwise unwieldy. It is therefore much more difficult to develop systematic techniques for analysing and understanding the operation of such circuits.

It is not only the analysis of analog electronic circuits that causes problems. Implementation introduces a whole range of complex problems with which we need to contend. It is often said that the implementation and testing of analog electronic circuits is composed of 10% design and 90% trouble-shooting. This rule-of-thumb arises because of the parasitic characteristics of each of the electronic devices that we will be examining in this chapter.

Most of the devices which we shall look at in this chapter are fabricated on semiconductor materials (Group IV in the Periodic Table) that have been doped with Group III and Group V impurities. The interaction of doped regions within the semiconductor gives rise to the valuable properties of each particular device and also leads to other parasitic or non-ideal behaviour patterns.

As a result of these parasitic (non-ideal) characteristics, it is often difficult to justify the cost of engineers designing and debugging analog electronic circuits from first-principles. In addition, the staggering growth in digital computing since the 1960s has led to a need for circuits that can co-exist in heterogeneous analog/digital circuits. For these reasons, a number of interesting trends have arisen:

(i) An enormous range of commonly used electronic circuits are normally available in a modular form in single-chip Integrated Circuit (IC) packages

(ii) IC packages are normally designed in family groups so that a range of different devices can be put together in "building-block" fashion to create new systems

(iii) Analog devices are often made compatible with digital circuits in order to facilitate bridge building between computers and continuous external signals.

In terms of power electronics (ie: the conversion of low energy electronic signals to high-voltage and/or high-current outputs) it is also necessary to note specific trends that have arisen in electronic devices. Firstly, there is the ability of small, single-chip semiconductor devices to absorb, supply and switch high currents and voltages. Secondly, there has been a trend away from the traditional analog approach to circuit design. As we shall see later in this text, devices such as transistors can form far more energy-efficient amplification circuits when they are used as digital switches in Pulse Width Modulation (PWM) based circuits, rather when they are used as analog (linear) amplification devices.

There are many issues that need to be examined in detail before one can carry out any electronic circuit design that will have industrial relevance in terms of reliability and accuracy. The objective of this chapter is not to make you an expert in electronic circuit design, but to assist you in understanding the basic phenomena involved, so that you can make intelligent decisions in the analysis and application of the semiconductor modules and electronic interfacing devices required in modern systems design.

3.2 Diodes, Regulators and Rectifiers

3.2.1 Fundamentals and Semiconductor Architecture

Diodes are the most basic of electronic devices and are an important part of any electronic circuit design because they (and their controlled derivatives such as thyristors) are used for:

- Providing uni-directional current paths through a circuit
- Regulating and limiting voltages
- Power supplies
- Converting a.c. signals into d.c. (rectification)
- Converting d.c. signals into a.c. (inversion).

Modern diodes are formed through the p-n junction which can be created by doping intrinsic (pure) silicon with Group V elements (giving n-type semiconductor) and group III elements (giving p-type semiconductor). This is shown in Figure 3.1

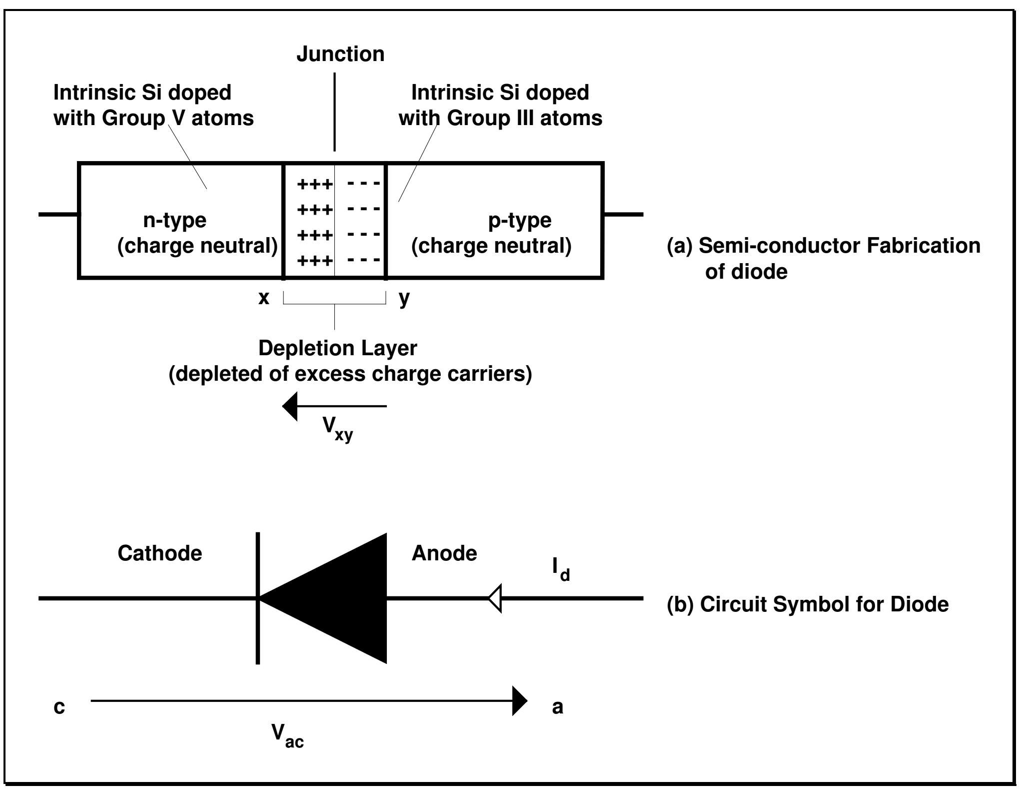


Figure 3.1 - The Semiconductor Diode

Intrinsic (pure) Silicon is charge neutral (equal number of protons and electrons) and is a poor conductor because it exists in a "covalent lattice" form, with all its valence shells complete. However introducing charge neutral Group V elements also introduces additional electrons free for conduction (n-type semiconductor). Introducing charge neutral Group III elements reduces the total number of electrons and thereby introduces "holes" for conduction (p-type semiconductor).

When we have p-type semiconductor butted against n-type semiconductor, we have a region of instability. At the junction, the excess electrons in the n-type semiconductor (majority n-type carriers) recombine with the excess holes in the p-type semiconductor (majority p-type carriers). The junction region is therefore normally deplete of excess carriers and is referred to as the "depletion region". The n side of the junction is deplete of electrons and therefore has a net positive charge and because the p side of the junction is deplete of holes, it has a net negative charge. In other words there is a barrier potential formed across the junction (referred to as V_{xy} in Figure 3.1). In a silicon based diode, the junction potential V_{xy} is in the order of 0.7 volts. In a germanium based diode, the junction potential is in the order of 0.3 volts.

If we apply a positive external voltage supply (V_{ac}) to the diode (as shown in Figure 3.1), then the potential of the anode is higher than that of the cathode. Majority carriers in the p-type material (holes) are repelled from the positive side of the supply towards the junction, thereby replenishing the depletion region on the p-side. Similarly majority carriers in the n-type material are repelled from the negative side of the supply towards the junction, thereby replenishing that region. Provided that the applied voltage (V_{ac}) is greater than the opposing barrier potential (V_{xy}), the diode can freely conduct current.

If we apply a negative external supply (V_{ac}) to the diode, such that the cathode potential is higher than the anode potential, then majority carriers are attracted away from the junction, thereby increasing the depletion layer width and thus prohibiting the flow of current. The diode acts as an open circuit. In practice a small leakage current still flows due to the presence of minority carrier holes and electrons in the vicinity of the depletion region. However, if we make the negative supply extremely large, then the potential across the junction is sufficient to force carriers across the depletion region. The structure of the junction effectively breaks down because new electron-hole pairs are created and conduction can once again occur. This is referred to as "avalanche breakdown". Provided that power dissipation in the diode is limited, the avalanche breakdown is not destructive.

A first-order approximation of diode behaviour is to say that the diode is a perfect conductor (short-circuit) whenever it is forward biased (ie: V_{ac} positive) and a perfect insulator (open-circuit) whenever it is reverse biased (ie: V_{ac} negative).

A second-order approximation of the behaviour of the diode structure (of Figure 3.1) is given in Figure 3.2. It shows a device which is an ideal conductor (short-circuit) when forward biased, provided that the opposing barrier potential has been exceeded. It also shows a device which behaves as an ideal insulator (open-circuit) when reverse biased - provided that the reverse breakdown voltage is not exceeded. After the reverse breakdown voltage is exceeded, the diode becomes an ideal conductor in the reverse direction.

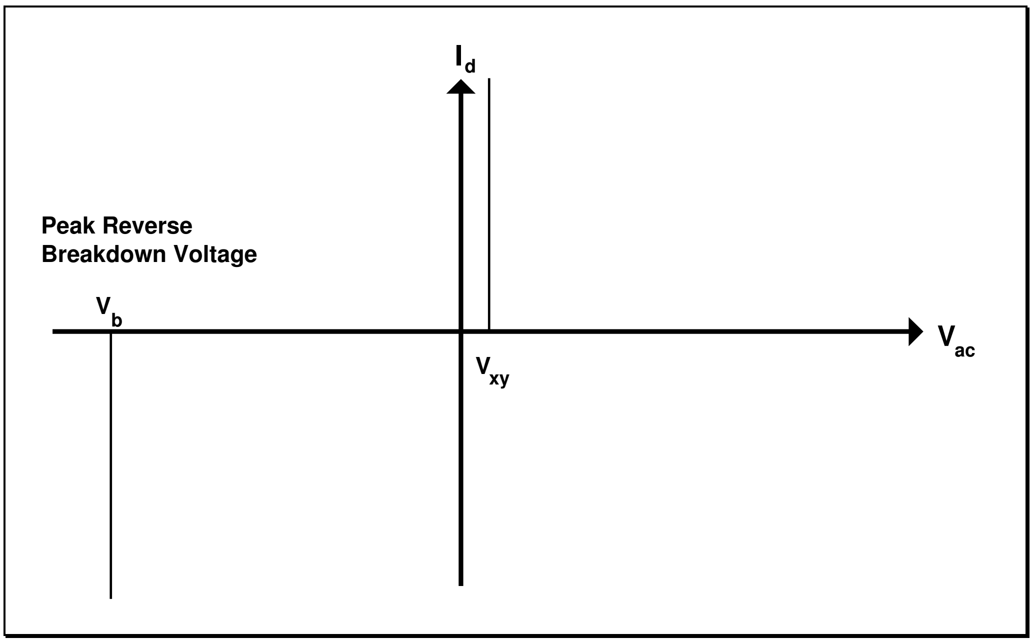


Figure 3.2 - Second-order Approximation of Diode Behaviour

A third-order approximation of diode behaviour takes into account the reverse leakage current of the diode and the resistance of the bulk semiconductor material (which can never be an ideal conductor or short-circuit). The resistance of the semiconductor material contributes to what is termed the "bulk resistance" of the diode. The third-order characteristic for a Silicon diode is shown in Figure 3.3, using different scales for the forward and reverse bias regions.

When analysing circuits containing diodes, we intelligently select the diode approximation model which is best suited to our level of analysis. The circuit model for each order of approximation is shown in Figure 3.4.

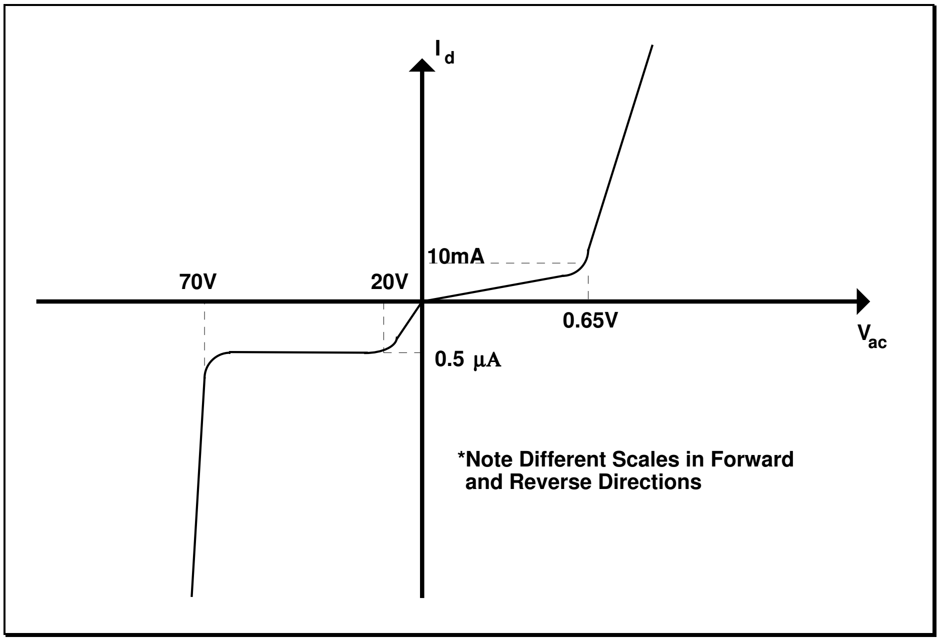


Figure 3.3 - Third-order Approximation of Silicon Diode Behaviour

None of the approximate models fully describe the characteristics of the diode, particularly in the so-called "knee" regions where the diode starts to conduct. An accurate model is not generally necessary and makes any practical analysis of circuits extremely difficult. In realistic situations, where we wish to accurately determine currents in a circuit, we generally measure and plot a voltage-current characteristic and then use the graphical "load-line" technique to determine exact operating points.

The technique for analysing circuits with diodes is relatively straightforward. One generally starts with the first order approximation of the diode and redraws the circuit diagram at least twice - once for the condition where the diode is forward biased and once for the condition where the diode is reverse biased. A third diagram is required if the diode is likely to go into reverse breakdown. The operation of the circuit can then be traced through via normal network analysis principles. Once the general operation of the circuit is understood, a more accurate picture can be obtained by substituting second and third-order models.

The power dissipation in a diode is simply calculated by multiplying the operating voltage and current together. In situations where voltages and currents are time varying, the power consumption is also time-variant and therefore the average power needs to be derived through integration. The average power determines the heating in the diode and therefore its susceptibility to damage. Note that the r.m.s. (root mean square) value of a power waveform has no physical significance whatsoever in engineering terms and should never be used for calculations.

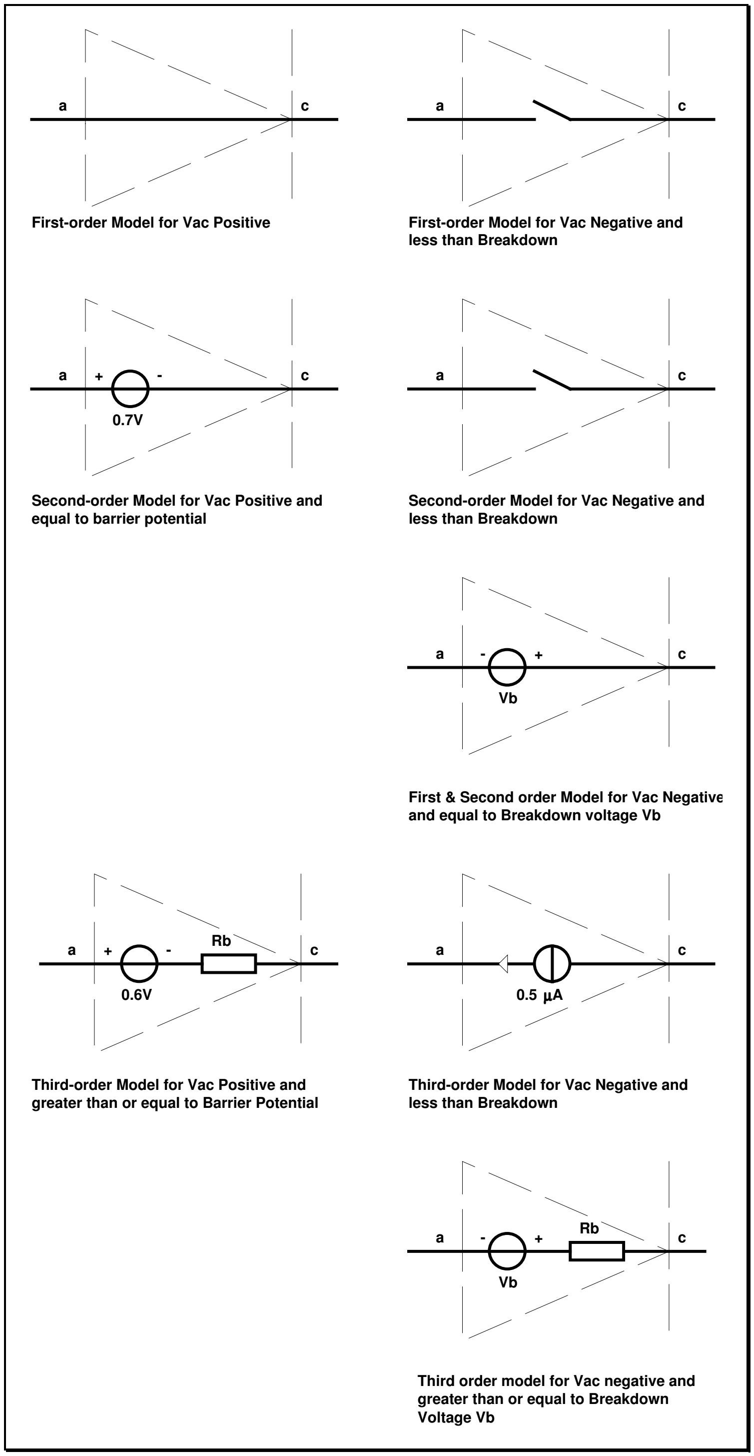


Figure 3.4 - Circuit Approximations for Silicon Diodes

3.2.2 Zener Diodes for Voltage Regulation

Zener diodes are a special type of p-n junction diode which are specifically designed for operation in the reverse breakdown region. These diodes do not suffer from avalanche breakdown, but rather from a phenomenon known as "Zener" breakdown. The doping in the p and n regions of a Zener diode is much higher than in a normal diode. This creates a far smaller depletion region at the junction and subsequently, a lower reverse-bias voltage will cause the junction to break down. Zener breakdown is not destructive in diodes, provided that the power dissipation within the device is kept within defined limits. The forward characteristic of the Zener diode is similar to the traditional diode. However Zener, diodes are seldom operated in the forward active region because it is their reverse characteristic that is of value.

The reverse breakdown voltage on a Zener diode can be well defined by the semiconductor manufacturer and varies little with current. End-users can purchase Zener diodes with reverse breakdown regions ranging from a couple of volts, through to hundreds of volts. These features make the Zener diode ideal for voltage regulation, since the voltage drop across the diode can be selected and varies little with the current flowing through it. The reverse breakdown characteristic of the Zener diode is therefore of prime importance and the forward active region is seldom discussed at length. The second and third order approximate circuit models for Zener diode are shown in Figure 3.5. Note that the circuit symbol for a Zener diode is slightly different to that of the traditional diode (small wings are drawn on the cathode side). A typical characteristic for a Zener diode is shown in Figure 3.6.

The voltage cited as the reverse breakdown voltage of the diode is quoted at a particular test current. Looking at the characteristic, it is clear that if we wish to achieve the rated breakdown voltage, then we need to ensure that we operate the diode at the appropriate current rating. Connecting a Zener diode across a component (ie: in parallel) not only helps protect the component from voltage spikes, but additionally regulates the voltage across that component and maintains it at approximately the reverse breakdown level of the diode.

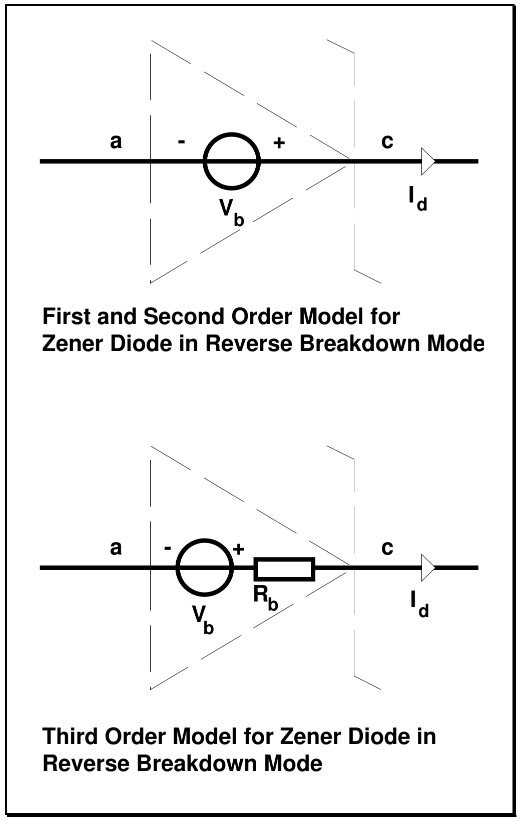


Figure 3.5 - Approximate Models for Zener Diode in Reverse Breakdown

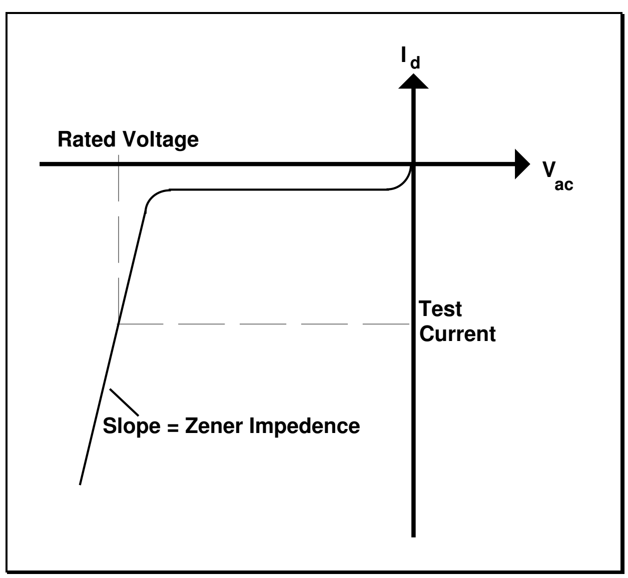


Figure 3.6 - Typical Zener Diode Characteristic

3.2.3 Diodes For Rectification and Power Supplies

Those who are not familiar with modern semiconductor technology could be forgiven for believing that semiconductors can only be used for the fabrication of low power diodes, transistors and digital circuits. In fact a good proportion of modern semiconductor applications are in high power circuits and there are a wide range of devices that can handle both high currents and voltages.

Diodes in particular are used in high powered circuits for conversion from a.c. to d.c. (rectification) and complete converters are also commonly available as a single-module solid-state device. Rectification is one of the most important functions that diodes are used to perform because there is an enormous demand for d.c. power supplies in engineering design - particularly with the overwhelming emphasis on digital circuit technology and computing.

When we talk of designing a power supply, there are essentially three basic blocks that we need to look at:

(a) The transformer
(b) The rectification
(c) The regulation.

(a) The Transformer

The transformer is used to convert an incoming a.c. waveform (normally from a general purpose power outlet) to a suitable level for rectification. Since power supplies can be either single-phase or three-phase, we need to have transformers for both applications. However, for most analyses, the three-phase transformers are essentially treated as three, single-phase transformers. The basic construction of a single-phase transformer is shown schematically in Figure 3.7 (a). This is composed of a laminated ferromagnetic core (that provides a low reluctance magnetic flux path) and a primary and secondary winding. The core is laminated to reduce power losses due to the circulation of unwanted "eddy currents".

When we model transformers, we begin with a concept known as the "ideal transformer", which takes no account of losses in real systems, and then add the parasitic "loss" elements. The ideal transformer is shown in the shaded region of Figure 3.7 (b) and its characteristics are as follows:

- The ratio of the secondary voltage to the primary voltage is the turns ratio (voltage transformation):

$$v_2 = \frac{N_2}{N_1} \cdot v_1$$

- The ratio of secondary current to primary current is the inverse of the turns ratio (current transformation):

$$i_2 = \frac{N_1}{N_2} \cdot i_1$$

- An impedance placed on the secondary side of a transformer (Z_2) has a value that can be measured on the primary side as Z_1. Z_1 is said to be the value of secondary impedance "referred" to the primary side and is equal to the secondary impedance multiplied by the square of the turns ratio (impedance transformation):

$$Z_1 = \left(\frac{N_1}{N_2} \right)^2 \cdot Z_2$$

The losses in a transformer include:

- Resistances of the primary and secondary windings
- Flux leaking from the core so that the same magnetic flux does not couple both windings
- Power losses due to "eddy currents" circulating in the core
- Power losses arising from the magnetisation and de-magnetisation of the core through the application of a.c. voltages - that is, "hysteresis" losses.

A complete transformer model is difficult to work with in an analytical sense, and so a number of minor approximations are made to create a working model. The approximate "working" model for a single-phase transformer is shown in Figure 3.7 (b). This circuit lumps together primary and secondary resistances and leakage reactances into single elements, "referred" to the primary side. The approximate model also includes a shunt resistance to represent hysteresis and eddy current losses and a shunt inductance to represent the magnetising current required for the transformer to operate even when no load current is flowing. This model is adequate in practical terms and makes analysis considerably easier than the complete model.

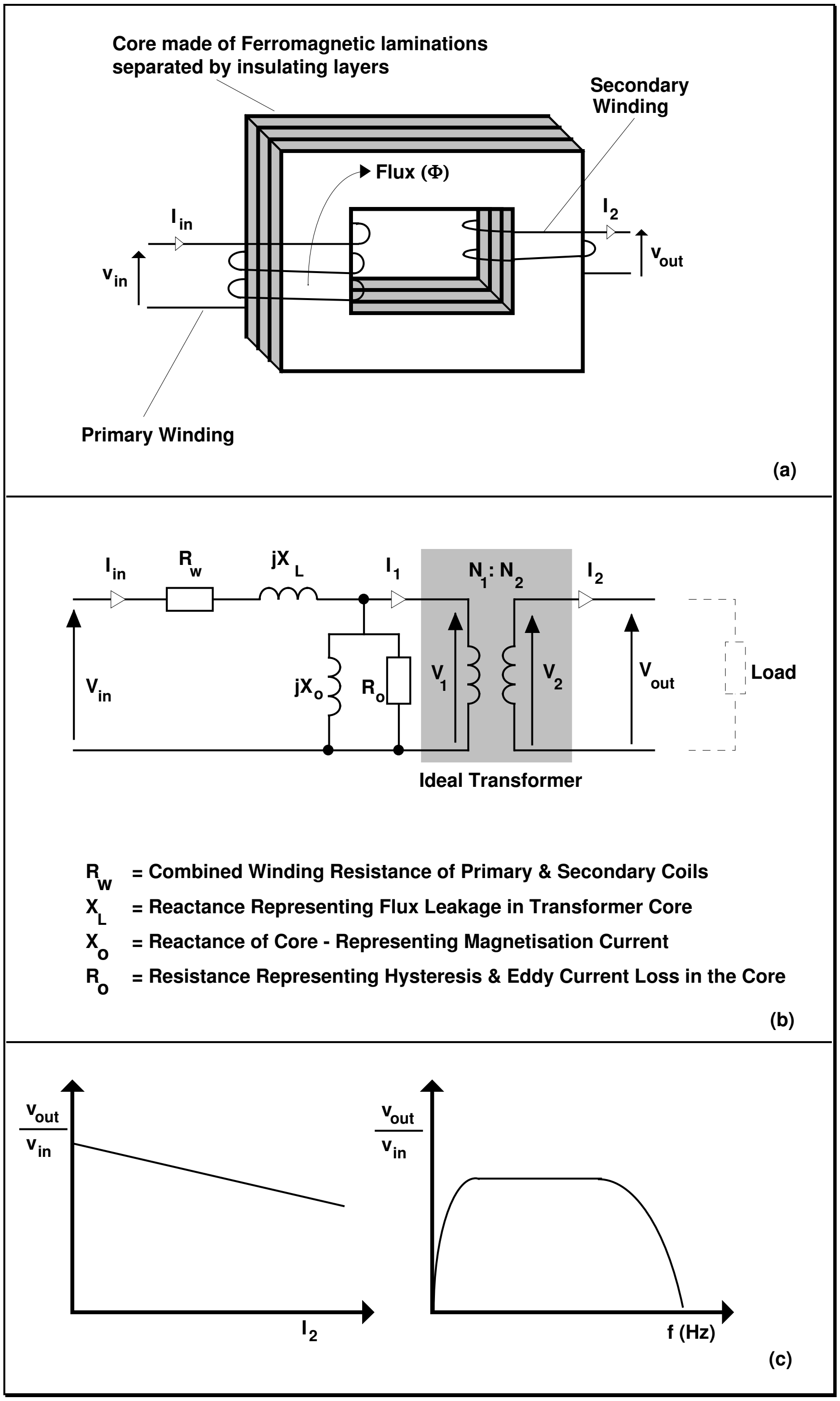


Figure 3.7 - (a) Schematic of Transformer Construction
(b) A Manageable Circuit Model for a Transformer
(c) Transformer Characteristics for Varying Load Current and Operating Frequency

The approximate circuit model of the transformer reveals the effects of the basic loss elements in the following way:

- As the secondary (load) current, I_2, increases, the voltage drop across the winding resistance and flux leakage reactance increases, resulting in a decreasing secondary voltage (for a constant primary voltage) - this is referred to as regulation and is responsible for the characteristic shown in Figure 3.7 (c)

- At zero frequency, the transformer is essentially short-circuited except for the winding resistance. Transformers therefore are unable to transfer d.c. voltages from the primary side to the secondary side

- As frequency increases from zero, the impedance due to the leakage reactance increases ($j\omega L$) and the output is attenuated, ultimately to zero as frequency tends towards infinity.

(b) Rectification

Nearly all modern large-scale electricity generation is a.c. in nature. This is primarily because the generation and transformation of a.c. voltages was originally far more practical than the d.c. alternative. In some instances the end-use of the electricity generation process is also very efficient in the a.c. form - particularly in the case of a.c. machines (motors). In other situations, both a.c. and d.c. are equally suitable - for example in resistive heating or incandescent lighting. However, there are a great number of applications for which a.c. is not well suited. These are primarily in the field of small-scale electronics (both digital and analog), metal smelting (furnaces, etc.) and historically in motor speed-control (servo applications). In recent years, however, in the case of motor speed-control, a.c. technology has also reached comparable levels to d.c.

In the case of electricity transmission, it has been suggested in recent years that losses can be minimised by transmission of d.c. rather than a.c. and hence that there needs to be conversion from a.c. to d.c and back to a.c. again.

Regardless of the relative merits of either system, the result of the differing end uses for electricity is that we need to be able to convert from a.c. to d.c. and from d.c. to a.c. Diodes are the primary mechanism for conversion from a.c. to d.c. and the associated process is referred to as "rectification". The reverse process (d.c. to a.c.) involves the use of triggered diodes (called thyristors) and is referred to as "inversion". We will briefly look at the process of inversion later in this chapter.

Rectification is based upon the use of transformers to provide a suitable level of source voltage which can then be converted to the required d.c. level. The rectified d.c. output voltage normally contains some ripple which is eliminated via two techniques:

- For low load currents, a capacitor is connected across the load to reduce output ripple

- For high load currents, an inductor (choke) is placed in series with the load to reduce output voltage ripple.

The simplest circuit is the single-phase, half-wave rectifier shown in Figure 3.8, with both low and high-current filtering for minimisation of ripple. The circuit is analysed empirically in several stages:

(i) The capacitance or inductance filtering is ignored and the turns ratio of the transformer is ignored

(ii) The output voltage waveform is determined for the situation where the transformer secondary voltage has a positive polarity (in Figure 3.8, this means the diode is approximately a short-circuit)

(iii) The output voltage waveform is determined for the situation where the transformer secondary voltage has a negative polarity (in Figure 3.8, this means that the diode is approximately open-circuit)

(iv) The effects of filtering components are then included. The net effect of the inductance or capacitance is to store and release energy in such a way as to minimise the change in current or voltage, respectively

(v) The load current is determined by dividing the load voltage by the load resistance - the waveform shapes are identical for resistive loads.

The results of the multi-stage analysis of the rectifier circuit are shown in Figure 3.9. Note the smoothing effects of the capacitance or inductance, which introduce an exponential decay into the output waveform (with a time-constant of $R_L \cdot C$ or L/R_L) during periods where the output would otherwise have been zero. The final stage of analysis is of course to determine the load current, which is simply obtained by dividing the load voltage by the load resistance. The output voltage waveform is uni-polar but is still time-variant. We therefore quantify such values by referring to their average or rms, rather than peak level.

We can also approximately consider the effects of diode voltage drop, by subtracting a value of say, 0.7 volts from the load voltage waveform while the diode is turned on.

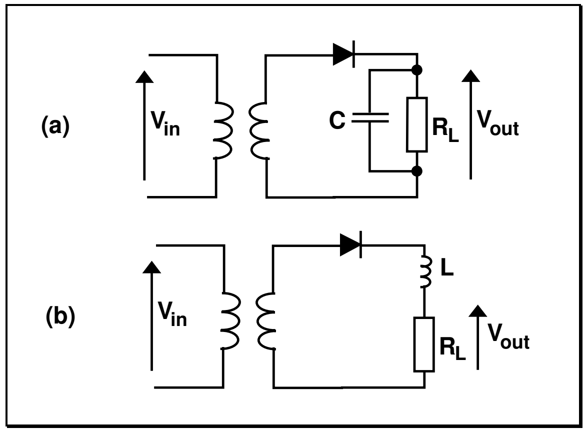


Figure 3.8 - Simple Half-Wave Rectifier Power Supply with
(a) Capacitance Smoothing for Low Load Currents
(b) Inductance (Choke) Smoothing for High Load Currents

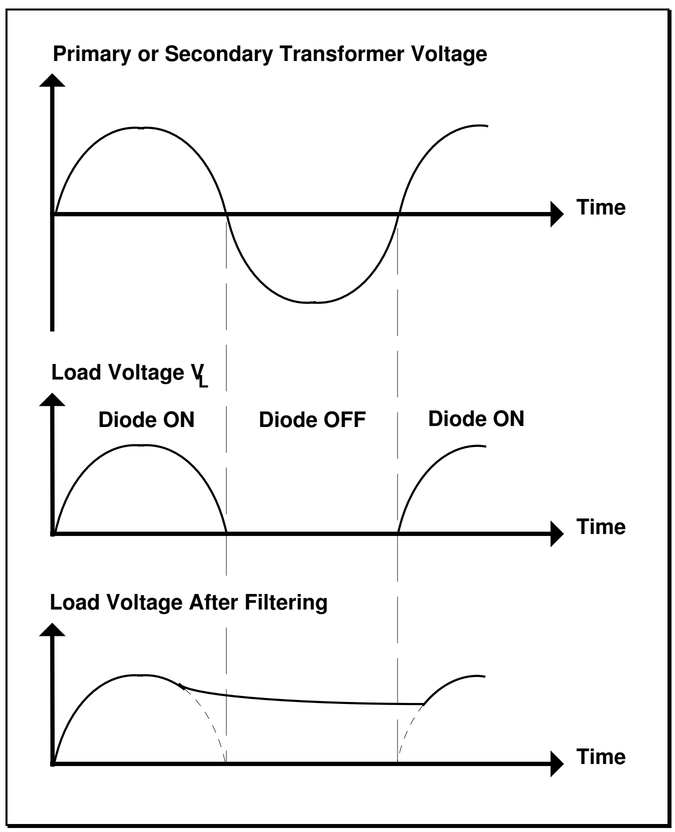


Figure 3.9 - Analysis of Half-Wave Rectifier With Capacitance or Inductance Filtering

One of the most common single-phase rectifiers is the bridge-rectifier, which is shown schematically in Figure 3.10. The diagram is shown in three parts. The first shows the total circuit and the other two diagrams consider the conditions where the output from the transformer (v_{ab}) has a positive and a negative polarity respectively. The net effect of the bridge is to provide a uni-polar output voltage and current independent of the input voltage polarity. We can again make the analysis more accurate by accounting for diode voltage drop and we can also analyse the effects of filtering (via capacitors, as shown in Figure 3.10, or via an inductive choke).

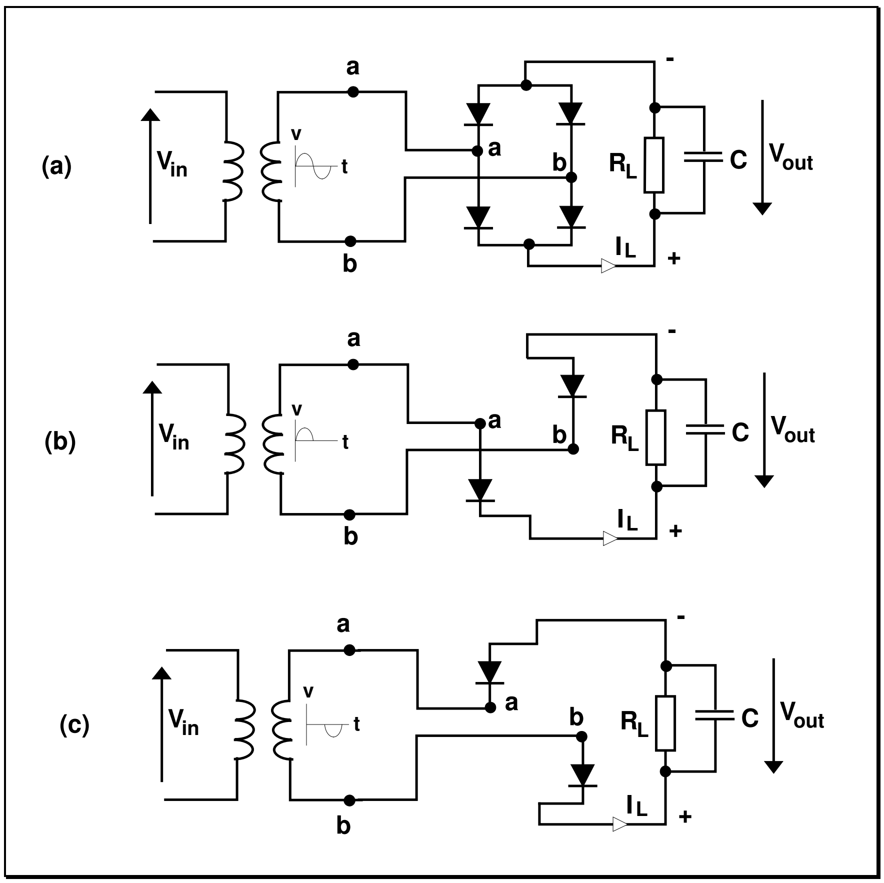


Figure 3.10 - Single-Phase Bridge Rectifier with Capacitance Filtering
(a) Circuit Diagram
(b) Effective Circuit Diagram for Voltage $v_{ab} > 0$
(c) Effective Circuit Diagram for Voltage $v_{ab} < 0$

The bridge rectifier is most commonly represented in circuits as a diamond of four diodes. In this text however, it is displayed as in Figure 3.10 in a rectangular shape. The reason for this is so that its functional similarity to the three-phase bridge rectifier is more apparent. The three-phase bridge rectifier is shown in Figure 3.11, connected to the secondary (star) windings of a three-phase star-star transformer. Note that while four diodes are required in order to fully rectify a single-phase sinusoidal voltage, only six diodes are required to rectify a three-phase voltage waveform. Note also that in Figure 3.11, an inductive smoothing choke is shown rather than the parallel capacitance alternative. The reason for this is because three-phase rectifiers are predominantly used with high load currents and hence inductive smoothing is the only practical option for such systems.

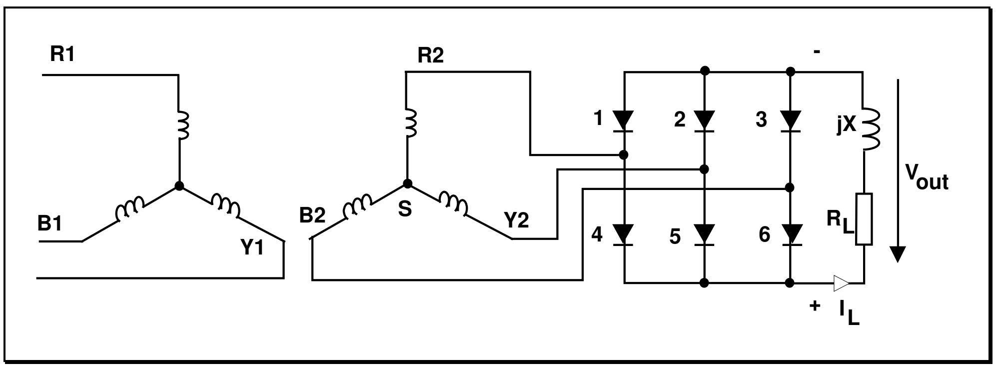


Figure 3.11 - Three-Phase Bridge Rectifier

The three-phase rectifier of Figure 3.11 is a little more complex to understand than its single-phase equivalent, but is functionally similar. Its operation is best understood by looking at the three-phase waveforms generated on the secondary side of the transformer. These are shown in Figure 3.12. Each of the six diodes can only conduct (turn on) whenever the magnitude of its corresponding phase voltage is greater than that of the other two phase voltages. In Figure 3.12, conduction regions are shown on the voltage waveforms with heavy lines. The output voltage at any time is the difference between the voltage on the top half of the bridge and the bottom half of the bridge and this is shown in the second part of the diagram in Figure 3.12, highlighting the six-phase ripple that is produced.

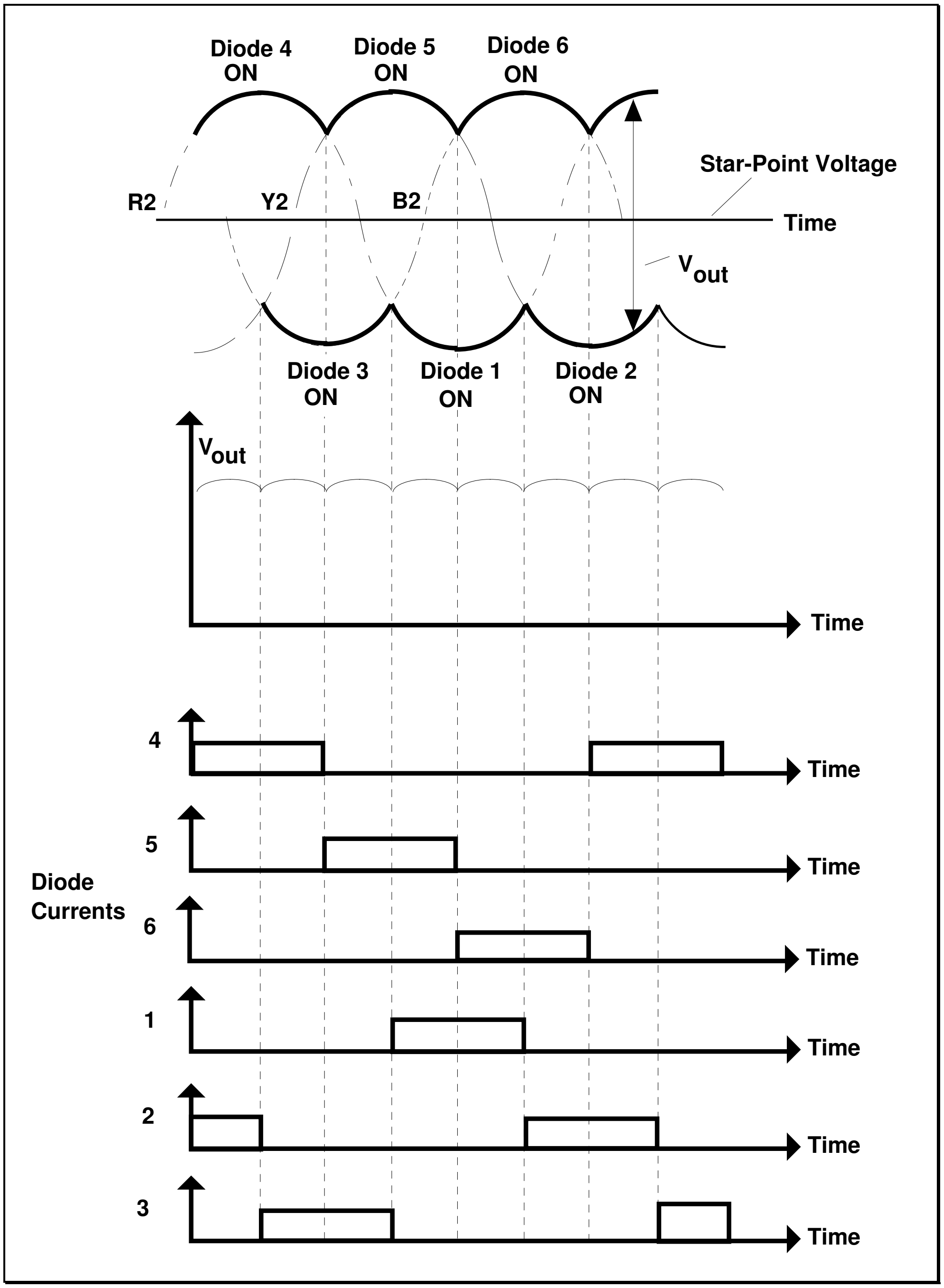


Figure 3.12 - Operation of Three-Phase Bridge Rectifier

Although the output (V_{out}) contains a six-phase ripple, we assume that in an ideal bridge rectifier, the output inductance is infinite and hence the d.c. output current is time invariant (ie: constant). This constant waveform is the sum of the currents in diodes 1 to 6. The current in each of the diodes is therefore a rectangular pulse, whose duration is one third of the phase-voltage frequency. These diode currents are shown in sequence in Figure 3.12. The phase currents flowing from the secondary side of the transformer can be determined by graphically summing the appropriate diode currents (eg: $I_{R1} = I_4 - I_1$). If you carry out this computation you will note that at every instant in time, these phase currents sum to zero. This means that the three phase bridge rectifier can be used with both Delta-Star and Star-Star transformer configurations.

We can again improve our analysis by accounting for the traditional diode voltage drop on each of the six diodes.

There are many other configurations of single and three-phase rectifier circuits, generally less common than the bridge circuits discussed thus far. However, their analysis is carried out in an analogous manner using a combination of graphical and analytical techniques as shown here.

(c) Regulation

Thus far we have seen little that could lead us to designing a variable d.c. power supply. There are in fact a number of different techniques that can be used. The most obvious method is to use a variable resistance across the d.c. output and to tap from the wiper arm the output voltage. The problem with this technique is of course that it wastes a lot of energy and subsequently generates a lot of heat, neither of which are desirable traits in power supplies.

Another technique is to use a variable transformer (traditionally known by one of its early trademark names "Variac") to vary the a.c. voltage supplied to the rectification stage of the system. A Variac simply has a mechanical knob that is used to position a wiper, that effectively varies the number of turns across which the secondary voltage is extracted. This is then fed into a rectifier and ultimately to a load.

A third technique is to use a closed-loop amplifier to vary the a.c. input or d.c. output voltage of a power supply. We shall look at these devices later in this chapter. Finally, the digital technique is to "chop" (switch on and off) the output d.c. voltage so that its average value increases or decreases according to the duty cycle (on:off ratio). If the output voltage waveform is filtered by an inductive choke then the net result is that a variable d.c. output has been obtained.

There are however many instances where one does not need a variable power supply and the objective is to design a circuit which simply provides a very stable nominally-defined output voltage. As we have seen in (b) above, the rectifier circuit (in its own right) does not provide a pure, time-invariant output waveform and hence capacitive filtering or inductive choking are sometimes used. These techniques still leave some ripple in the output waveform. The solution to this problem is to use Zener diodes to regulate the output waveform and to minimise the ripple to a known range.

Figure 3.13 (a) shows the output of a single-phase bridge rectifier being fed into a load via a limiting resistor (R_L) and regulated by a Zener diode of a known reverse breakdown voltage. The load on the circuit could be a simple resistance or the front end of some other, more complicated circuit. We know that as long as the output voltage from the rectifier is lower than the reverse breakdown voltage of the Zener diode then the diode is effectively an open circuit and plays no part in the circuit. However, when the rectifier output voltage is greater than the reverse diode breakdown voltage then the diode begins to conduct and draws current away from the load. The voltage is then effectively tied to the reverse breakdown voltage of the Zener diode.

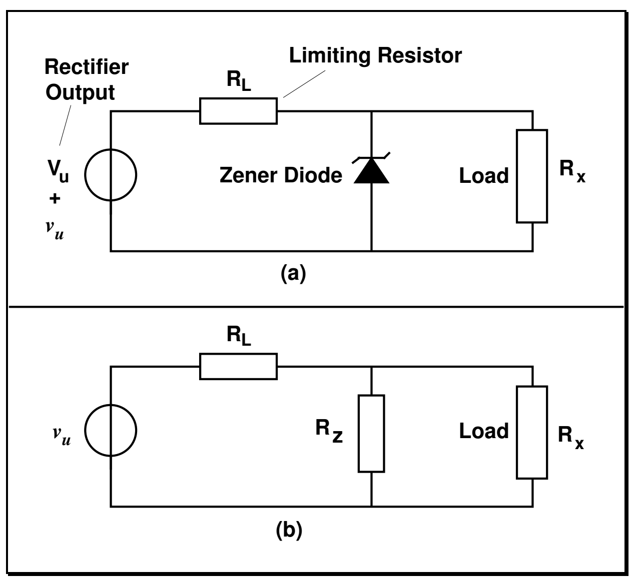


Figure 3.13 - (a) Zener Diode Regulating Output Voltage V_u from a Rectifier (b) Equivalent Circuit Replacing Diode With its a.c. Resistance R_Z

The unregulated output voltage from the rectifier circuit is the sum of two components - a d.c. offset and an a.c. ripple (in this case $V_u + v_u$). The principle of superposition tells us that we can always analyse such circuits by examining the effect of each voltage acting in isolation and then adding the results for the total solution. Figure 3.13 (a) is adequate for analysing the d.c. effects of the problem, but for the effect of the a.c. ripple, we need to look at the a.c. resistance of the Zener diode (R_z). The a.c. resistance of a Zener diode (normally the value quoted by manufacturers) is the slope of the reverse V-I characteristic of the diode obtained for a constant junction temperature (this is normally much lower than the d.c. value measured by users as in Figure 3.6). The equivalent circuit for a.c. operation is shown in Figure 3.13 (b). From this circuit we can determine that the output voltage fluctuation v_l caused by the a.c. component (ripple) from the rectifier is given by:

$$v_l = \frac{R_z}{R_z + R_L} \cdot v_u \qquad \text{...(2)}$$

Equation (2) is derived by simple voltage division, making the assumption that R_Z is much lower than R_L which is normally the case in practical circuits. The output ripple can therefore be minimised by making the limiting resistance much larger than the a.c. resistance of the Zener diode. These sorts of circuits only work with low load currents and nominally constant load voltages. In other situations, transistorised voltage regulators need to be designed for greater stability.

3.3 Basic Transistor Theory and Models

3.3.1 Introduction

In terms of their wide-ranging operational characteristics, transistors are by far the most complex of all the discrete semiconductor-based electronic components. They are also the basis for an enormous range of analog and digital integrated circuits. There are essentially two major groups of devices to be examined herein - that is, the Bipolar Junction Transistors (BJTs) and the Field Effect Transistors (FETs). The FETs also have a sub-group known as the Metal Oxide Semiconductor Field Effect Transistors or MOSFETs, which have similar characteristics to the FETs, but some operational advantages under certain conditions. As with p-n junction diodes, transistors can be analysed at a number of different levels. A complete discussion and analysis of the behavioural characteristics of transistors is beyond the scope of an entire book, much less a chapter section such as this. The purpose of this section is therefore to provide a brief overview of the basic transistor technologies, their functionality and typical circuits in which they are used.

Firstly we will examine the use of transistors in digital circuits where they perform simple switching functions. Subsequently we will examine the use of transistors in analog circuits. In the discussions on transistor devices we again need to refer to the principle of superposition. We treat all d.c. voltages and a.c. (signals) separately in what are referred to as the "large signal" and "small signal models", respectively. In transistor theory, we are generally not greatly interested in the total solution (d.c. offset + a.c. signal) but rather the individual components. In particular, the d.c. components are of primary interest in digital circuits and the small signal (a.c.) components are of primary interest in analog circuits. Moreover, we need to understand that in analog circuits the d.c. voltages are only used to place transistors into a linear (analog) mode so that the a.c. signals can be amplified.

There are many differing opinions on how transistor theory should be introduced. Some authors prefer to avoid the complexities of the so-called "small-signal" model, whilst others get so carried away with the semiconductor physics that the readers tend to lose sight of any practical applications of the transistors themselves. In this book, we will endeavour to tread the middle ground in order to provide as much insight into the design and analysis of analog circuits as is practical in this short treatise.

3.3.2 Bipolar Junction Transistors (BJTs)

The Bipolar Junction Transistor (BJT) is one of the earliest forms of transistor. It looks deceptively simple and yet its true functionality is quite complex. It can be formed by successively doping intrinsic (pure) Silicon with "n-type" (Group V) impurities then a higher concentration of "p-type" (Group III) impurities, then again with a still higher concentration of "n-type" impurities. This creates a device with three regions and two p-n junctions in close proximity to one another.

Simplistically, the transistor can be considered as two p-n junction diodes back to back. Figure 3.14 schematically shows the so-called "grown" construction of the "npn" transistor, together with its symbolic circuit representation. The three "doped" regions within the transistor are connected to the outside world via conductors and the three resulting terminals of the device are referred to as the "collector", "base" and "emitter. In this chapter we shall concentrate on the npn transistor although the "pnp" transistor is also readily available as a complementary device. The only major difference between the two types is that the d.c. voltages (biases) applied to pnp devices need to be of opposite polarity to those applied to the npn devices described herein. In both transistor types, we normally refer to the base (B) as the input side and the collector (C) and emitter (E) as the output side.

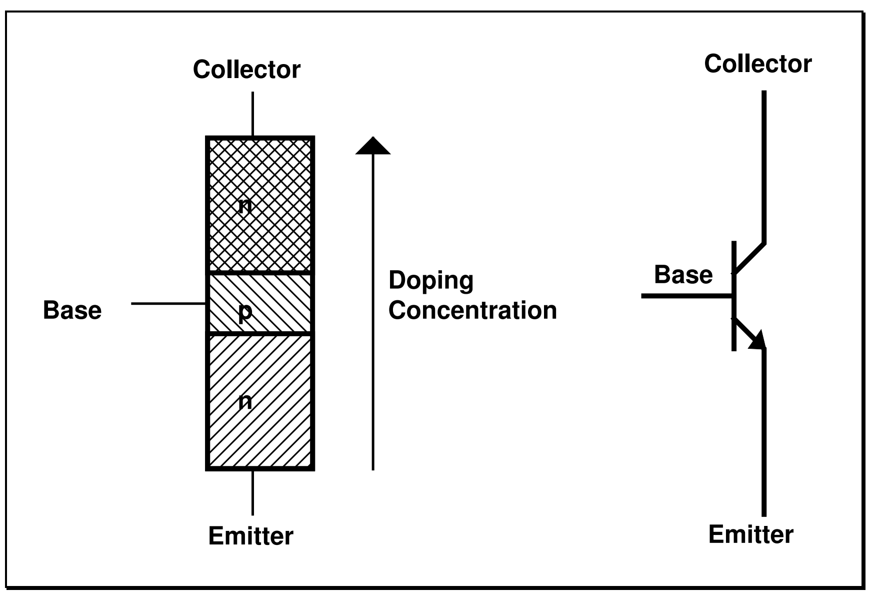


Figure 3.14 - Grown (npn) Transistor Schematic and Circuit Symbol Representation

There are a number of points to note about the BJT. Firstly it is not a symmetrical device. Like the diode, the transistor is fabricated on a piece of intrinsic (pure) semiconductor, successively doped with increasing concentrations of impurities. The impurity concentration of "n-type" dopant in the collector is much higher than that in the emitter. Secondly, the width of the base region of the transistor is of critical importance. In Section 3.2, we briefly examined the p-n junction. In the BJT there are two p-n junctions and the width of the base region is such that these two junctions can interact with one another thereby producing a variety of different effects. Moreover, it needs to be noted that the width of the depletion region at each of the p-n junctions in the transistor is dependent on a number of factors, particularly the voltage applied across the junction itself (the "biasing" of the transistor). As the width of the depletion regions varies, so too does the effective base width of the transistor and hence a number of different effects can be achieved.

The simplistic, two-diode approximation of the transistor is shown in Figure 3.15.

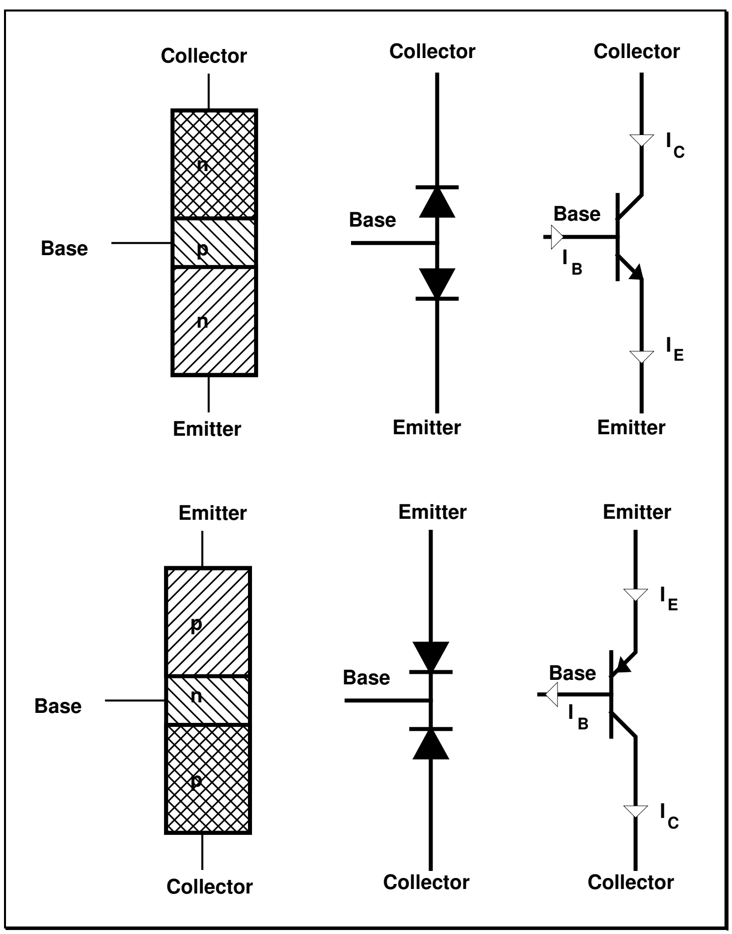


Figure 3.15 - Transistor Diode Analogy for "npn" and "pnp" Devices

We can begin to examine the operation of the BJT by establishing the basic relationships between the currents flowing within the device (using the conventions shown in Figure 3.15). From Kirchoff's current law, we know that:

$$I_C + I_B = I_E \quad ...(3)$$

The transistor is designed such that there is a large "forward current gain" or Beta (β) and hence the collector current:

$$I_C = \beta \cdot I_B \quad ...(4)$$

Since typical values of β can range from 100 to 1000, the base current is normally negligible compared to the collector current and hence I_C is approximately equal to I_E. However, the base current effectively flows into the emitter and hence the emitter current is actually slightly higher than the collector current.

Transistors can be placed into a variety of different modes depending upon the voltages applied to their terminals - that is, the d.c. "biasing". These modes are known as:

(i) Cut-off
(ii) Forward Active
(iii) Saturation
(iv) Reverse (Inverse) Active

We will not enter into a discussion of the physical phenomena that occur within the semiconductor as the d.c. voltages (biases) applied to the outside terminals are altered. Our objective herein is to summarise the electrical characteristics of the transistor for a range of different biasing conditions described in (i) - (iv). However, in order to understand the concept of forward and reverse biasing of junctions, it is necessary to refer back to our earlier discussions on diodes (section 3.2.1). A p-n junction is said to be forward biased when the potential of the "p" side is greater than the potential of the "n" side. The simple way to remember this is to think of the junction as being forward biased when "p" is connected to positive and "n" is connected to negative. The junction is reverse biased when this polarity is reversed. With these basic concepts in mind, the four operational modes of the BJT are summarised as follows:

(i) Cut-off Mode

A BJT device is cut-off when the emitter-base junction and the collector-base junction are both reverse biased. Looking at the diode representation of Figure 3.15, we can see that this will be achieved when V_{BE} has a value less than the barrier potential voltage required to cause forward conduction across the emitter-base junction (typically 0.7v). In this condition, no base current flows and hence no collector current flows. If we view the transistor as a switch that facilitates current flow from collector to emitter, then we can say that a cut-off transistor is effectively open circuit between collector and emitter. Cut-off mode is used in digital circuits.

(ii) Forward Active Mode

A BJT device is placed into forward active mode when the emitter-base junction is forward biased (and greater than the junction potential) and the collector-base junction is reverse biased. When d.c. voltages are applied to the terminals of a transistor to establish this forward active mode, then a.c. signals can be superimposed onto the base to be amplified at the collector. When there are no a.c. signals input into the transistor, then the circuit is said to be in the "quiescent" state. However, when small signals are applied to the base, then the transistor can be used to amplify them in a linear fashion and hence forward active is the operational mode of BJTs in analog circuits. This mode is also referred to as the linear region for the device.

(iii) Saturation Mode

A BJT device is placed into saturation mode when the emitter-base junction is forward biased and the collector-base junction is forward biased. In an "npn" transistor, this normally occurs when a relatively large d.c. voltage is applied to the base, thereby forward-biasing both transistor junctions. The net effect of this can be understood by examining Figure 3.15, where it can be seen that, between the collector and emitter terminals, the transistor becomes almost short-circuited because the two junctions behave like two forward biased diodes. In practice, there is a small voltage drop across both diodes - approximately 0.2v (much less than for individual p-n junction diodes). Saturation mode can be compared to the short-circuit mode of a switch and hence forms the complementary digital circuit function to cut-off mode.

(iv) Reverse (Inverse) Active Mode

One may intuitively feel that a transistor is a device which will operate equally well in both directions, since its semiconductor structure is essentially an "npn" or "pnp" sandwich. However, the difference in doping levels between the collector and emitter means that the BJT is not symmetrical. Endeavouring to operate a transistor with the collector-base junction forward biased and the emitter-base junction reverse biased will place the transistor into inverse active mode. While the transistor will still function, the forward gain will be greatly reduced and the transistor will be relatively inefficient. This mode is normally achieved by accident, when the emitter and collector terminals are inadvertently mistaken by a developer.

The circuits that are used in order to create the different transistor modes are called biasing circuits. There are many different types of biasing circuits and their names are often somewhat confusing to novices in the field. Biasing circuits actually have two major functions:

- To provide quiescent voltages that will place the transistor into an appropriate mode of operation

- To provide feedback that will stabilise the gain of the transistor.

We have already covered the need for the first function in our discussion of operating modes. The second function is of particular importance in analog circuits. When we wish to amplify a signal, we generally need to be sure that the gain will be well defined. However, the forward gain of a BJT device (the Beta) is an ill-defined value and is subject to variation due to the manufacturing process. The variation in the β value between a number of transistors (of the same type) can be as high as three to one. For this reason it is rather futile designing circuits whose amplification is dependent upon this parameter. The solution to the problem is extracted from classical control theory, where we feed a proportion of the output signal back into the input circuit for closed-loop control. This is shown schematically in Figure 3.16.

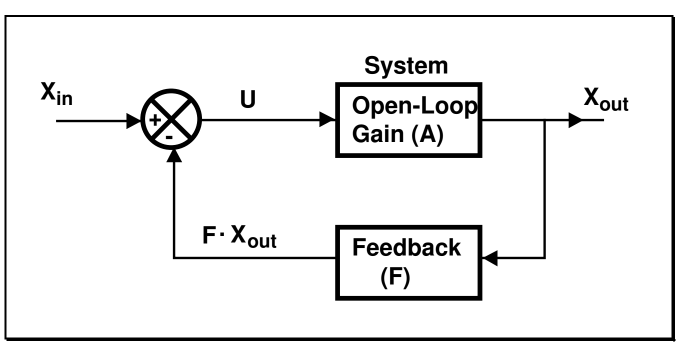


Figure 3.16 - Classical Closed-Loop Control System

In Figure 3.16, the open-loop gain is defined as the amplification of a device acting in isolation (that is, with no feedback). The circle with the cross and polarity signs is the common symbol for a summing junction - in the case of Figure 3.16, the summing junction contains a "+" and a "-" and hence the output of the junction (U) is the difference between the two inputs. The whole system is referred to as a negative feedback arrangement and causes the device with an open-loop gain of "A" to amplify the difference between the input and the output. Analytically this system is described as follows:

$$
\begin{aligned}
&X_{out} = A \cdot U \\
&U = X_{in} - F \cdot X_{out} \\
&X_{out} = A \cdot \left(X_{in} - F \cdot X_{out} \right)
\end{aligned}
$$

$$\Rightarrow \frac{X_{out}}{X_{in}} = \frac{A}{1 + A \cdot F} = \frac{1}{\frac{1}{A} + F} \qquad \text{...(5)}$$

Equation (5) simply tells us that if the open-loop gain of the system is large enough (that is, "A" tends to infinity), then the ratio of output to input in under closed-loop control will be inversely proportional to the feedback. In an analog circuit, the device with an open-loop gain of "A" could simply be a transistor with a forward gain of β. The feedback arrangement is normally achieved through a simple network of resistors. The net result is that we can design bias circuits that fulfil both the mode selection and feedback roles and thereby provide us with the basis for stable amplifiers.

It is important to keep the classical control system model in mind when examining transistor circuits. Not only does a basic understanding of this model assist in analysing amplifier circuits, but it also assists in understanding the nomenclature used in regard to transistor circuits. For example, some circuits are referred to as "Common Emitter (CE)" or "Common Base (CB)". This terminology refers to the fact that the emitter or base (respectively) are common to both the input and output circuits. A Common Emitter circuit is shown in Figure 3.17. This particular circuit can have a number of different roles. Firstly, it can be used to measure the output characteristics of a transistor, which show the dependence of I_C upon V_{CE} for a range of different base currents. A typical Common-Emitter output characteristic is shown in Figure 3.18 for a range of different base currents.

There are however, several points to note about the circuit shown in Figure 3.17. Firstly, we see the base and the emitter circuits grounded at the same point (common). Secondly, the resistance connected to the collector is referred to as the load. Since we know that the transistor can amplify the base current by a factor of β, we know that there is potential to use the transistor to drive a higher current load than the voltage source V_{in} may be able to provide. Thirdly, we note the presence of the d.c. supply rail V_{cc}.

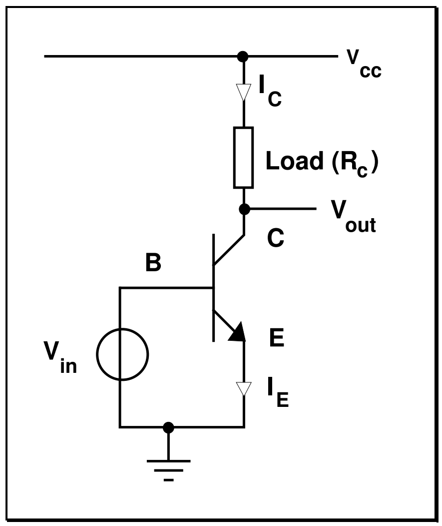


Figure 3.17 - Common Emitter Circuit

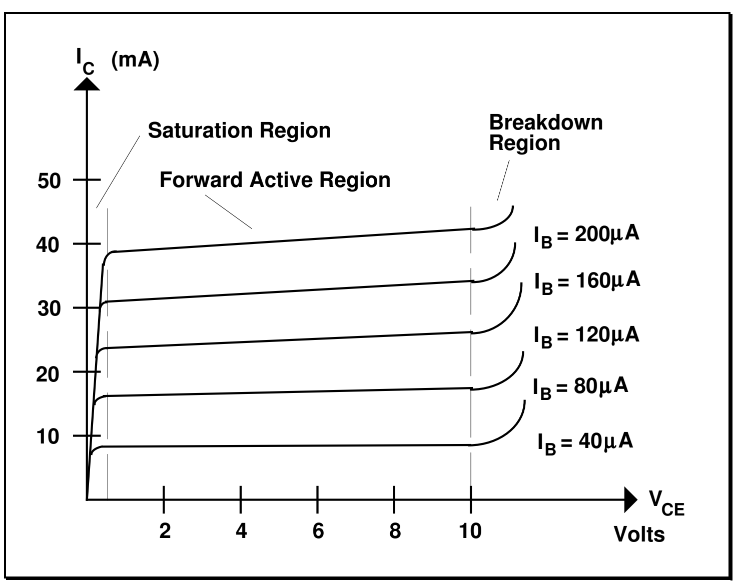


Figure 3.18 - Typical Common-Emitter Output Characteristic of an npn Transistor

Most text books tend to understate the importance of the supply rail V_{cc}. It is referred to as a supply rail for two reasons. Firstly, because it is normally connected directly to the d.c. power supply for the system and secondly, because it supplies all the energy required for amplification of signals. In other words, the transistor is really a mechanism for controlling the energy obtained from the power supply rails. In digital circuits the transistor acts as a switch whose open/closed status is dependent upon the emitter-base voltage. In analog circuits the transistor acts as an infinitely variable valve (or resistance), controlled by the emitter-base voltage, that regulates current flow from the collector through to the emitter.

Kirchoff's current law tells us that because some base current is flowing into the transistor the collector and emitter currents cannot be identical. There is an additional parameter, which like β is also referred to as a current gain parameter. This other parameter "α" is defined as follows:

$$\alpha = \frac{I_C}{I_E} \qquad \text{...(6)}$$

It is less useful than β because the base current in a transistor is normally much smaller than the collector current and hence the collector current and emitter current are almost identical for practical measurements.

Thus far, we have only examined the d.c. characteristics of the transistor circuit shown in Figure 3.17 and the output characteristic of Figure 3.18 which shows some of the typical operating regions of the transistor. The d.c. voltages applied to the transistor terminals as a result of biasing circuits such as that in Figure 3.17 are referred to as quiescent operating conditions and it is conventional to represent all quiescent (d.c.) voltages and currents with upper-case letters. We shall later see that we can superimpose a.c. signals (represented by lower-case letters) onto the base terminals of the transistor so that it can be used as an analog amplifier. However, we first need to look at the role of the transistor as a digital switch, where we are only interested in using the device to switch on and off quiescent voltages and currents.

Referring again to Figure 3.17, which illustrates the common-emitter transistor configuration, we can see that if we set the input voltage V_{in} to zero, then the emitter-base junction is less than the barrier potential and hence the junction is reverse biased - as is of course the collector-base junction. As a result, the transistor is said to be "cut-off" (open-circuit) and hence no current flows from the collector to the emitter. As a result, the quiescent output voltage V_{out} is equal to V_{cc}:

$$V_{out} = V_{cc} - I_c \cdot R_c = V_{cc}$$

In other words, if we input a low voltage, then we obtain a high voltage as an output from such a circuit.

On the other hand, if we input a high quiescent voltage into the base of the transistor (the same size as V_{cc}, for example), then we force the transistor into saturation mode, where the collector and emitter are effectively short-circuited together (in fact a 0.2 volt drop across V_{CE} is typical). As a result, the output voltage is close to zero. In other words, a high input voltage results in a low output voltage. This circuit referred to as a Boolean inverter because the output has the inverse status of the input.

For practical reasons, an actual inverter is fabricated onto a single piece of semiconductor and is typically made up of a number of transistors. A realistic circuit diagram for an inverter fabricated from BJTs is shown in Figure 3.19 and typically, a number of inverters would be integrated onto the same piece of semiconductor for economic reasons. Boolean logic circuits fabricated from BJTs are the oldest digital circuits and are referred to as "Transistor to Transistor Logic" or "TTL", typically operating with a value of V_{cc} equal to around 5 volts.

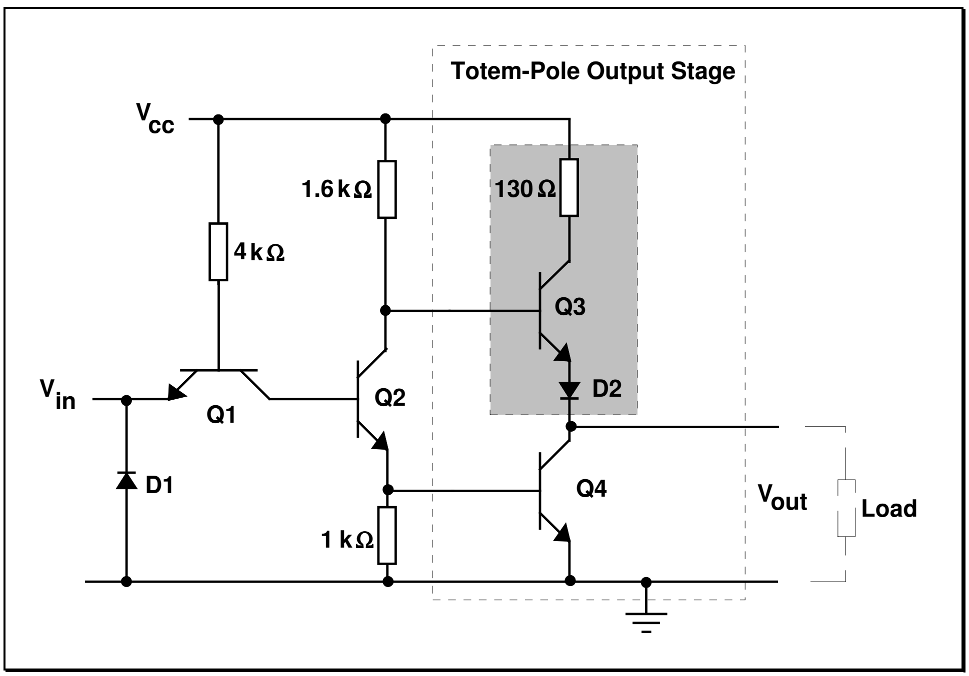


Figure 3.19 - Circuit Diagram for a realistic TTL Inverter Circuit

The circuit of Figure 3.19 can readily be analysed if one always assumes that the transistors therein can only be in either cut-off mode or saturation mode. For example, if V_{in} is low, then Q1 saturates (because both the collector and emitter junctions are forward biased), thereby short-circuiting the base of Q2 to low. Q2 is therefore cut-off and hence the base of Q3 is high. Q3 saturates and the output diode is short-circuit, so therefore the output voltage is high. The reverse analysis can be applied for a high input.

Another feature of the TTL circuit is the output stage, composed of two transistors (Q3 and Q4) connected together in what is referred to as a "totem-pole". This feature is common to a number of TTL circuits and is designed such that only one transistor or the other conducts at any one time.

The circuit of Figure 3.19 also illustrates another important point about digital circuits integrated on a single piece of semiconductor material. Transistors switching from cut-off to saturation (and vice-versa) only consume a small amount of energy. However, it is evident that there are also resistors fabricated into the digital circuit. These linear elements dissipate energy and generate heat. As long as the current flowing through these devices is small, then there is no problem - however, if the current flowing through the resistors is large, then they become a potential source of device failure.

Let us assume that the output of the circuit in Figure 3.19 is used to drive a load (eg: a resistor or light-emitting-diode between the output terminal and ground). When the transistor Q3 is saturated, then the current flows from the supply rail, through the 130Ω resistor, through Q3 and through the external resistive load. If the load is large (ie: low resistance) then the current flowing through Q3 and the 130Ω resistor is also large. It is the 130Ω resistor that can cause problems, since the current flowing through it generates heat in the small semiconductor. For this reason, such a circuit can only provide (source) a very limited amount of output current.

The level of current available for a load (eg: light-emitting diode, relay, etc.) can be increased by using a different form of TTL chip that eliminates the "heat-generating" totem-pole output resistor. This is known as "Open-Collector TTL", because the resistor, npn transistor and diode in the normal totem-pole output (shown in the shaded region of the inverter of Figure 3.19) are omitted. Instead, the load is connected between the supply rail and the collector of the pnp output transistor (Q4 in Figure 3.19), via an external, current-limiting resistor. The external resistor can of course be rated at a higher power level than the semiconductor one that is normally fabricated in the totem- pole. The other advantage of the open-collector gates is that the outputs from a number of gates can be "tied" together in order to generate a Boolean "AND" function. This is referred to as "wired" logic and can not be achieved with standard TTL.

There are a number of different digital circuit devices that can be fabricated into Small Scale Integrated (SSI) circuits in order to implement common Boolean functions. The generic name for such devices is gates, and we will examine these and their logic further in Chapter 4. At this stage however, it is necessary to examine the operation of the BJT as an amplifier. This often causes some confusion and so we will take a step by step approach to the analysis. The key point to remember is that amplifier circuits are analysed by using the principle of superposition. Firstly we analyse the quiescent circuit, assuming that all signal voltages are zero and then we analyse the small-signal circuit, assuming that all the d.c. voltages are zero. The actual voltage at any point in the circuit is the sum of the two components. However, we normally only concern ourselves with either one or the other.

Figure 3.20 shows the common-emitter circuit of Figure 3.17 in the way it would normally be modified for amplification purposes. The d.c. voltage source at the base of the transistor is actually replaced with a resistive ladder arrangement and the values of the resistors are chosen to create a situation where the emitter-base junction is forward biased and the collector-base junction is reverse biased, thus setting the transistor into its normal mode of operation. A resistor, R_e, has been placed between the emitter terminal and the ground to provide another negative feedback element that helps stabilise the circuit. If the collector current starts to increase in this circuit, the voltage drop across R_e increases and hence the emitter voltage increases, hence the base-emitter voltage decreases, hence the base current decreases and hence the collector current decreases. The overall circuit is called an "emitter-feedback" circuit.

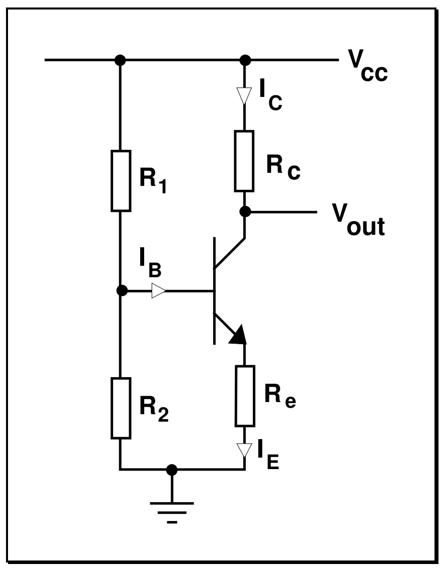


Figure 3.20 - Emitter-Feedback Circuit Commonly Used for Amplification

Figure 3.21 shows the same circuit, with the resistive ladder replaced by its "Thévenin Equivalent" circuit which is composed of a voltage source in series with a resistance R_b (refer to section 3.4). The effective values of the voltage source and resistance are simply calculated from the original circuit of Figure 3.20 as follows:

$$V_B = \frac{R_2}{R_1 + R_2} \cdot V_{cc}$$

$$R_b = \frac{R_1 \cdot R_2}{R_1 + R_2} \qquad ...(7)$$

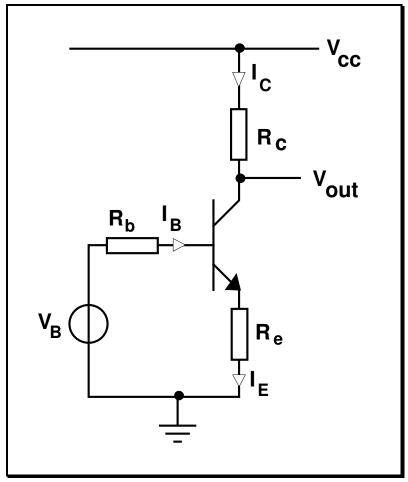


Figure 3.21 - Emitter-Feedback Circuit with Thévenin Equivalent of Base Biasing

Given the circuit of Figure 3.21, the output characteristic of the transistor (as shown in Figure 3.18) and the fact that the base-emitter voltage must be greater than approximately 0.7 volts in order for the transistor to be placed into active mode, we can calculate all the quiescent voltages and currents in the circuit. We can then look at applying small a.c. signals to the base of the transistor and determining the response of the circuit at the collector terminal. In order to do this we need to place one capacitor between the a.c. source and the existing base circuit and one capacitor between the collector terminal and the a.c. output. The capacitors "de-couple" (separate) the a.c. and d.c. components and are normally selected so that they have no impedance at normal operating frequencies. The combined bias and signal circuit is shown in Figure 3.22.

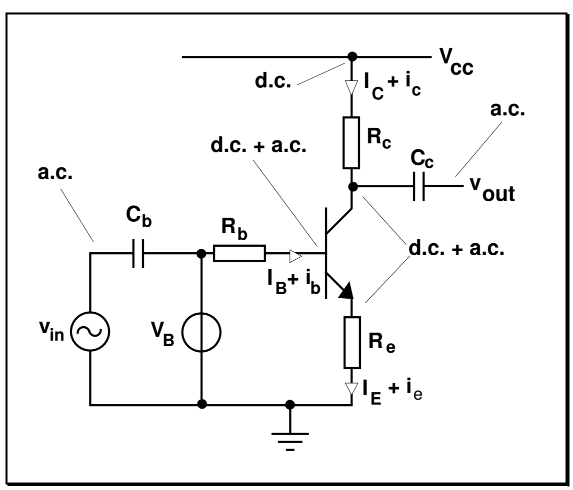


Figure 3.22 - Emitter-Feedback Circuit with a.c. Input Signal Applied

Figure 3.22 shows the complete circuit and the distribution of a.c. signals and d.c. voltages at different points in the system. The circuit is now treated in two separate components and since we have already looked at the d.c. components, we now need to examine the a.c. influences in the system. We do this, following the principle of superposition, by setting all the d.c. voltages in the system to zero.

When we began to analyse the BJT, we were only concerned with its characteristics under large-signal (quiescent / d.c.) conditions and we did our analysis using a very simplistic model, which we did not derive ourselves, but rather assumed to be correct from our limited overview of the semiconductor structure. It is interesting to note however, that when we wish to apply a.c. signals to the transistor, then we need a more sophisticated model. In particular, we are concerned with what happens when we make small a.c. perturbations about the quiescent operating points established by bias circuits, such as the emitter-feedback arrangement. A thorough derivation of the a.c. (small-signal) model is outside the scope of this book, and so it will need to be taken as read, that the model we have presented is correct. The small-signal or so-called "hybrid-π" model for the BJT (operating in forward-active mode) is shown in Figure 3.23.

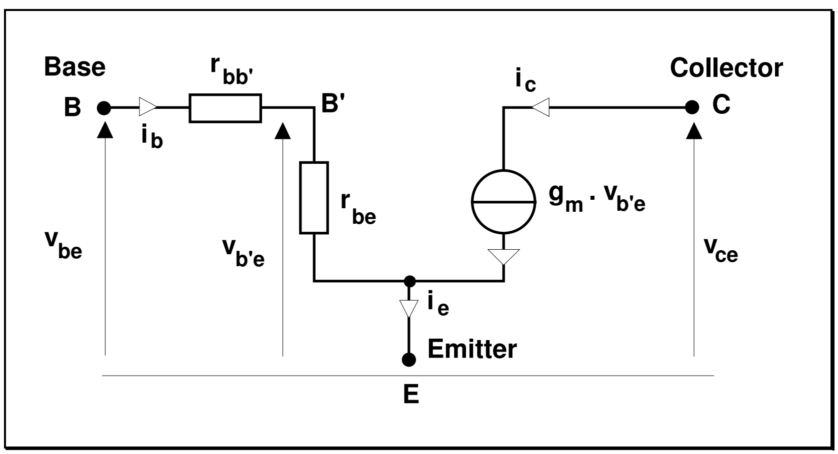


Figure 3.23 - Small-Signal (Hybrid-π) Model for BJT

The small-signal circuit model is composed of a number of different elements which are explained as follows:

- The element $r_{bb'}$ represents the resistance of the contact between the actual base terminal of the transistor (on the semiconductor) and the outside world. Typically $r_{bb'}$ is small (50 - 200 Ω) and is sometimes neglected in analysis

- The quantity g_m is referred to as the mutual conductance or transconductance of the transistor and is obtained by taking the partial derivative of collector current with respect to base-emitter voltage, for a constant collector-emitter voltage. The value of g_m depends upon the quiescent (d.c.) current on the transistor (the operating point) and the temperature at which the transistor is operated. The following expressions are used to derive the transconductance of a transistor:

$$g_m \approx \frac{|I_C|}{V_T}$$

$$V_T = \frac{\bar{k} \cdot T}{q}$$

$\bar{k}$ is Boltzmann's Constant in Joules/ Degree Kelvin

q is the electron charge

T is the temperature in Kelvin

From the above expressions, we can determine that the transconductance of a transistor at room temperature (25C) is given by:

$$g_m = \frac{|I_C|}{25.8} \quad (\text{mA} / \text{V}) \qquad \text{...(8)}$$

- The resistance between the semiconductor base and emitter terminals, r_{be} is dependent upon both the β and the transconductance of the transistor and is defined as follows:

$$r_{be} = \frac{\beta}{g_m} \qquad \text{...(9)}$$

- The current source in the collector circuit is a dependent source whose value is determined by the transconductance and the base-emitter voltage:

$$i_c = g_m \cdot v_{b'e}$$

$$v_{b'e} = i_b \cdot r_{be} = i_b \cdot \frac{\beta}{g_m}$$

$$\Rightarrow i_c = \beta \cdot i_b \qquad \text{...(10)}$$

Equation (10) illustrates that the small signal and large signal current gains from base to collector are identical and either form of the relationships described in this equation can be used for analysis purposes.

In order to analyse the emitter-feedback circuit of Figure 3.22 in terms of its performance as a small signal amplifier, we set the d.c. supply voltages to zero. In this case, the only voltage we need to concern ourselves with is V_{cc} (since V_b is not actually a voltage source, but rather a Thévenin equivalent element). We also replace the circuit symbol for the BJT in Figure 3.22 with the small signal model of Figure 3.23. Finally, we can work on the assumption that C_b and C_c are large capacitances and hence their impedance is assumed to be zero. The end result of these three actions is the circuit shown in Figure 3.24.

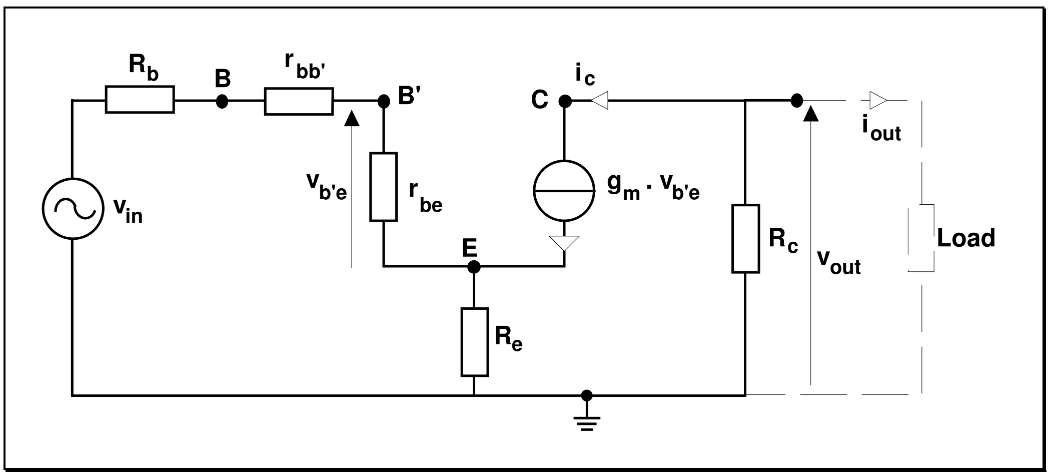


Figure 3.24 - Complete Circuit Model for Emitter-Feedback Amplifier Incorporating Small-Signal Model for BJT

The circuit shown in Figure 3.24 will only provide us with the small (a.c.) signals existing in the amplifier. The original diagram in Figure 3.22 shows that at some points in the system both a.c. and d.c. voltages will exist and hence the total voltage is calculated by adding the small signal voltage to the quiescent d.c. values obtained from the large-signal model in Figure 3.22. The results are shown in Figure 3.25.

Note well how the V_{cc} voltage is fixed at some d.c. level because it represents a d.c. power supply and hence no signal can exist at that point in the circuit. An important question that arises therefore is what happens when the signal at the collector and its quiescent voltage add up to a value higher than the V_{cc} supply rail. The simple answer is that they can't. If, for example, the signal waveform at the collector is sinusoidal, as shown in Figure 3.25, and its level increases as a result of increasing input, then it will ultimately become distorted (flattened off to the V_{cc} level). This is referred to as clipping. Similarly, if the waveform at the collector is so large that its bottom end attempts to go below the emitter voltage, then it too will distort into the shape of the waveform already existent at the emitter. In fact, the emitter and collector voltages can never be separated by less than approximately 0.2 volts (the voltage at saturation).

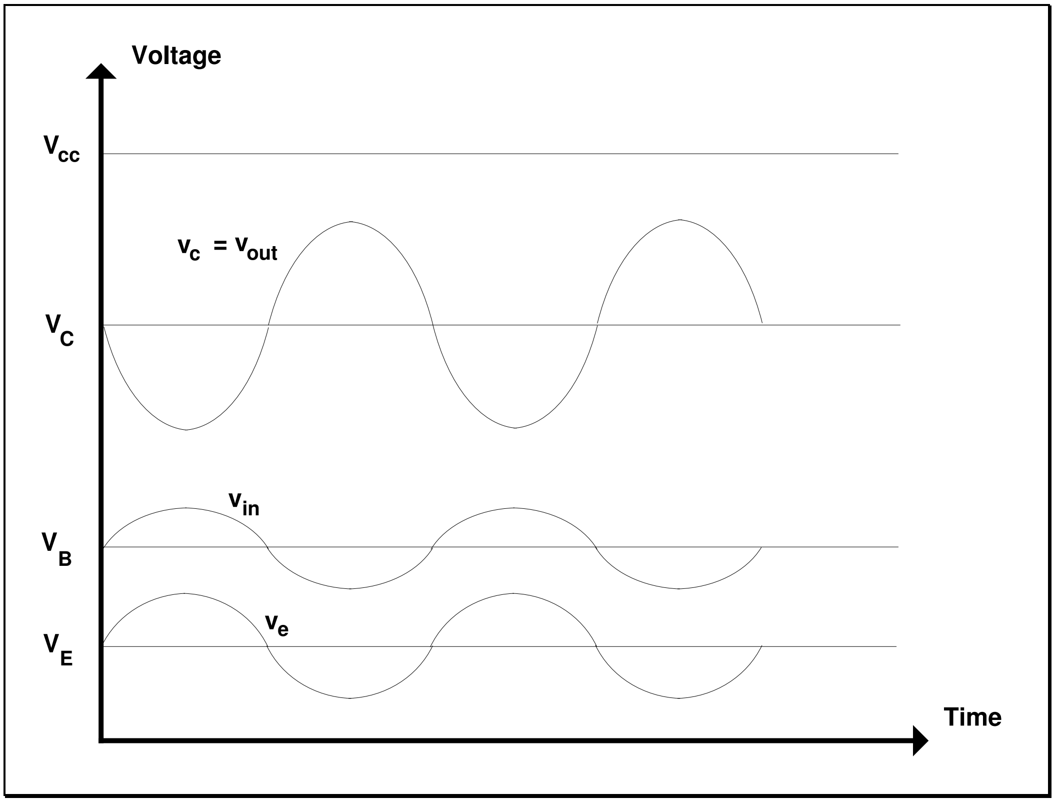


Figure 3.25 - Total Voltage Waveforms (Quiescent + Signals) in an Emitter Feedback Amplifier Circuit

Figure 3.24 shows that the emitter-feedback amplifier circuit can be analysed just like any other simple network, using Kirchoff's voltage and current laws and Figure 3.25 shows how the principle of superposition enables us to get a total solution for all node voltages. The only point to remember however, is that in general, the results will tend to somewhat complicated in an algebraic sense, since the circuit contains a current source which is dependent upon the true base to emitter voltage.

As an amplifier, our main concern is to ensure that the signal input at the base of the system is amplified at the collector and so generally we wish to develop an expression for the voltage gain of the system (v_{out}/v_{in}). We also need to ensure that the voltage gain is independent of ill-defined values such as β. However, we are additionally interested in ensuring that the input impedance of the system (v_{in}/i_b) is high and that the output impedance (v_{out}/i_{out}) is low. All these parameters can be adjusted by varying the size of resistive components in the emitter-feedback circuit.

In a realistic amplifier, the values of input and output capacitance are finite and hence these need to be included in the model of Figure 3.24. The fact that these have imaginary impedance values ($1/j\omega C$) means that the signal voltage gain will be frequency dependent and the output voltages will (in general) be phase-shifted from the inputs. We therefore also need to concern ourselves with determining and adjusting the frequency bandwidth in which the amplifier will give us a well-defined signal gain.

The emitter-feedback amplifier is one of the most fundamental amplifier circuits. However, there are many other BJT circuits that can be used for amplification and many of these rely upon multiple transistors. These circuits are typically designed so that the stability of gain is not compromised by minor variations in component tolerances, transistor β, etc. They are also specially designed to provide higher levels of input impedance or lower levels of output impedance, etc. The analysis of all these amplifier circuits is carried out on the same basis described for the emitter-feedback circuit, and with experience, it is possible to simplify this analysis by making a range of different assumptions. Our next step however, is to move on to a slightly different transistor technology and to evaluate its effectiveness in digital and amplifier circuits.

3.3.3 Field Effect Transistors (FETs)

There are two generic types of Field Effect Transistor (FET). These are the Junction Field Effect Transistor (JFET) and the Metal Oxide Semiconductor Field Effect Transistor (MOSFET). FET devices are quite different in semiconductor structure to BJTs and yet, in terms of their electrical models, they can ultimately be used in similar applications to the BJTs.

JFETs and MOSFETs offer both advantages and disadvantages over the BJT architecture, and it is in digital circuits and switching where the FET devices offer their greatest potential. In particular, FETs are much more compact than BJTs and have a higher input impedance, thereby making them suitable as loads on circuits with a high "fan-out". FETs can also be configured to act as resistive loads and hence integrated circuits can be designed using only FETs, without a need for other devices. The BJT, on the other hand, has the capacity to provide a higher gain over a much larger frequency range (bandwidth) than a FET device and hence it is more useful as an analog amplifier. The BJT can also switch at higher speeds than a FET in digital circuit applications.

The semiconductor structure of the so-called "n-channel" JFET is shown schematically in Figure 3.26. Like the BJT, the JFET is composed of three, doped semiconductor regions. However, in the FET, two of these regions have the same doping and are connected together to form a terminal known as the gate. The third region is called a channel and one terminal is connected to each end. One end of the channel is referred to as the source and the other end of the channel is known as the drain..

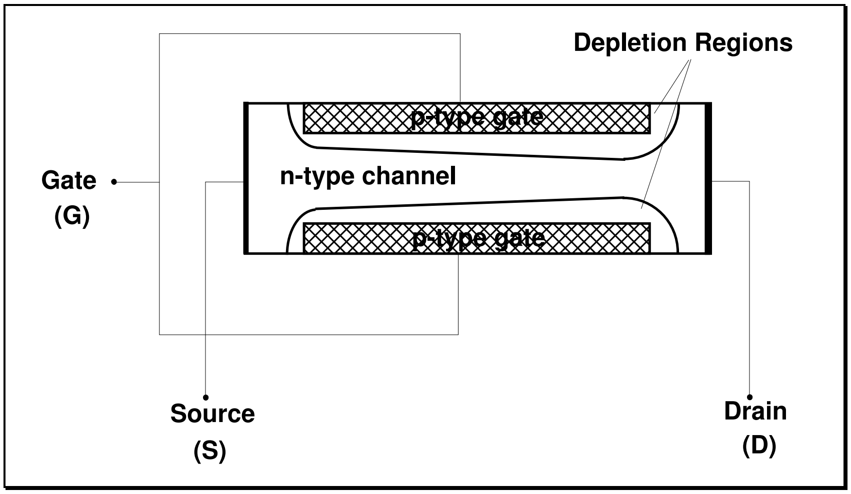


Figure 3.26 - Schematic of "n-channel" JFET Structure

An examination of Figure 3.26, reveals that the JFET is a three-terminal device which relies upon the flow of majority carriers from the source to the drain. JFETs can be designed with either "n" or "p" channels and hence rely upon either electron or hole conduction (respectively) in the channel. The flow of majority carriers from source to drain is facilitated by providing an electric field, which can be generated by applying a voltage between the source and the drain.

In section 3.1.1, we looked at the p-n junction diode and we examined the phenomena that occurred at the junction when "p" and "n" type materials were butted together. In particular, we noted the depletion region that this generates for some distance on either side of the physical junction. The same also applies in the BJT and FET devices. Moreover, in all these devices, the external potential applied to the junction can be used to vary the width of the depletion region. In the JFET, if we vary the width of the two depletion regions, we can effectively vary the width of the channel through which majority carrier conduction can occur and hence control the flow of current from source to drain. We again note that the depletion region is so named because it is an area deplete of majority carriers for conduction purposes. The width of the junction is controlled by varying the gate potential with respect to the source. If the potential is sufficiently large, then the junction will effectively widen to the extent where conduction in the channel is restricted to a value which is independent of the drain to source potential. This is referred to as the "pinch-off" condition.

In addition to the JFET devices there are Metal Oxide Semiconductor (MOS) devices, of which three different types exist. These are the:

- Enhancement MOSFET device
- Depletion MOSFET device
- Complementary MOSFET (CMOS) device.

In fact, the MOS devices are of greater significance to us than the JFET device since they are far more prolific in both digital electronic circuits and power electronics circuits. The structure of the Enhancement MOSFET and the Depletion MOSFET are both shown schematically in Figure 3.27 (a) and 3.27 (b) respectively.

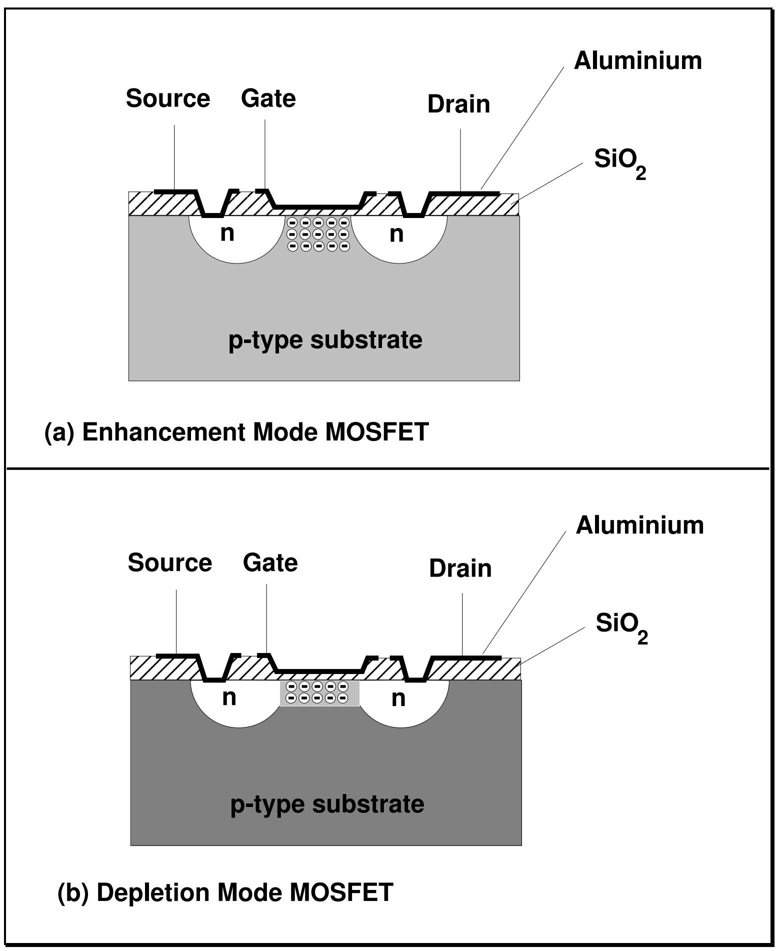


Figure 3.27 - Schematic of MOSFET Construction
(a) Enhancement Mode n-Channel MOSFET
(b) Depletion Mode n-Channel MOSFET

The Enhancement MOSFET is unlike the JFET in the sense that no physical channel exists from source to drain until the transistor is biased. If a positive voltage is applied to the gate of the n-channel device shown in Figure 3.27 (a), while the source, drain and p-type substrate are all connected to ground (say), then the majority carriers in the substrate (positively charged holes) move away from the surface and are replaced by carriers (electrons) from the two neighbouring regions in the source and drain, thus creating an inversion layer, just below the Silicon Dioxide coating. A conducting channel is therefore formed in the substrate, joining the source and drain. The channel current is enhanced by the positive gate voltage, and hence the name of the device. The current flowing through the channel varies with the potential difference between the drain and the source.

If the gate voltage is kept constant as the drain to source voltage increases, then the voltage of the gate, with respect to the drain decreases, thus reducing the size of the inversion channel, until a point is reached where the channel is effectively "pinched" off. Despite the physical difference between the enhancement MOSFET and the JFET, this characteristic is similar.

Figure 3.27 (b) shows the Depletion MOSFET (also n-channel), which already has a channel diffused between the source and the drain. Current can be made to flow through the channel simply by applying a voltage between the drain and the source and maintaining the gate to source voltage at zero. If the gate voltage on the Depletion MOSFET is made more negative, then positive charges will be induced into the channel (from the p-type substrate), thereby recombining with the holes and hence, diminishing the number of free majority carriers available for conduction in the channel. Eventually the channel is pinched off. This type of MOSFET clearly derives its name from the fact that the channel conductivity is depleted as a negative gate voltage is applied. It is interesting to note, however, that the Depletion MOSFET can also be used in enhancement mode, provided that the gate to source voltage is kept positive.

Despite the apparently more complex semiconductor structure of the MOSFET over the BJT, it is interesting to note that even in the mid 1970s (when the devices were introduced) it was possible to build MOSFET devices in a semiconductor area less than one twentieth of that required by the BJT. As a result of the smaller size of devices such as the enhancement MOSFET, it was decided at an early stage to combine two complementary devices (one n-channel and one p-channel) onto the same chip. These devices are called Complementary MOSFETs or CMOS and are commonly used in digital circuits. The semiconductor structure of a CMOS pair is shown in Figure 3.28.

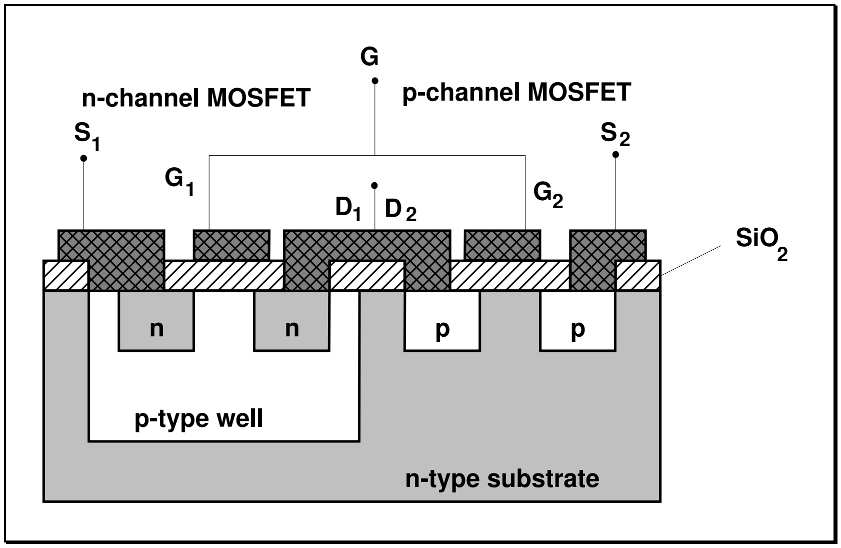


Figure 3.28 - Complementary n and p channel MOSFETs (CMOS)

The circuit symbols for the n and p channel versions of the JFET, the Enhancement and Depletion mode MOSFET devices are shown in Figure 3.29. Note that because the Depletion MOSFET can be used in Enhancement Mode, there are two possible circuit representations for the Enhancement MOSFET (as shown in Figure 3.29 (b) and (c)). Note also that the symbols for the CMOS devices is the same as for normal MOSFETs since a CMOS pair is simply two transistors fabricated within the same piece of semiconductor material.

A typical output characteristic for a FET device is shown in Figure 3.30. Note that the normal mode of operation for a FET is in the so-called saturation (or pinch-off) region, where drain current is approximately constant with changing drain-source voltages. This differs somewhat from the BJT characteristic of Figure 3.18 where the device is not "saturated" in its normal mode of operation.

Despite the physical differences between all the FET devices and the BJT it is interesting to note that they all share a similar output characteristic - that is, a linear region which knees into an approximately constant voltage-current region that knees into a breakdown region when high output voltages are applied to the end terminals of the devices. Regardless of their similarities, it turns out that all the devices have strengths in different applications. BJTs, for example, are used where high digital switching speeds are required or in wide-bandwidth, general-purpose linear amplifiers. JFETs, on the other hand, are typically used in special-purpose low-noise amplifiers and in circuits requiring a high input-impedance. JFETs can also be used as voltage controlled resistors. MOSFETs are predominantly used in power-switching circuits and in digital circuits where low power consumption is required and where high input impedances are required.

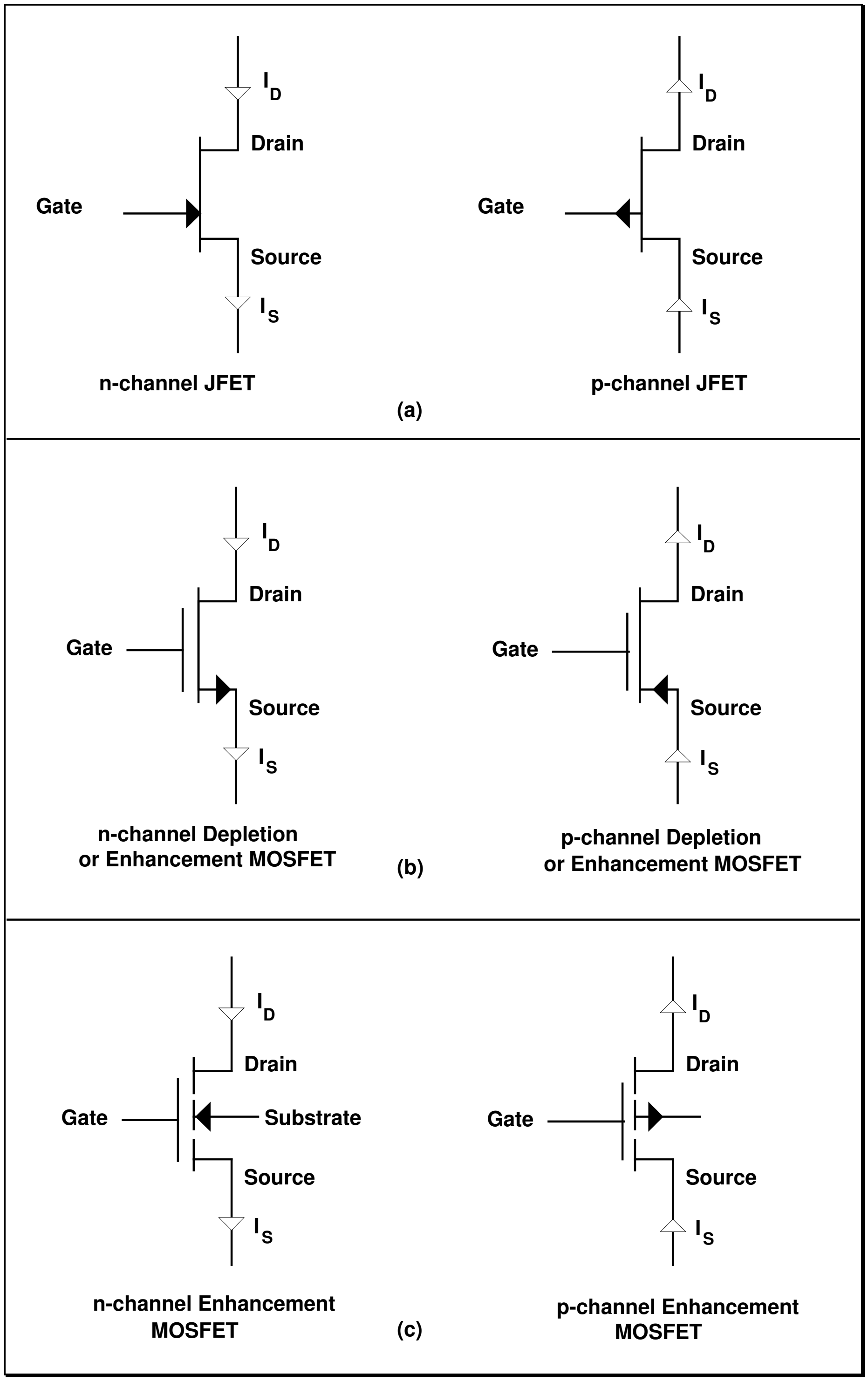


Figure 3.29 - Circuit Symbols for Different FET Devices

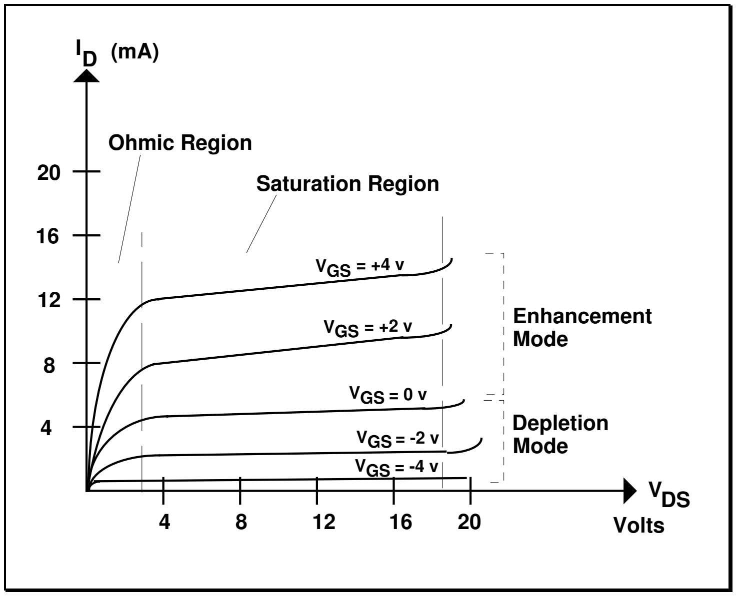


Figure 3.30 - Typical Output Characteristic for an n-channel MOSFET Operating in Both Enhancement (V_{GS} Forward Biased) and Depletion (V_{GS} Reverse Biased) Mode

Unlike BJTs, JFETs are bi-directional devices because either end of the channel can be used as the source or the drain and the direction of current flow is only determined by the channel type and the polarity of the voltage applied across the channel. For an n-channel JFET, however, a negative gate-source voltage needs to be applied in order to cause conduction (compared with the npn transistor, where a positive base-emitter voltage needs to be applied to activate the device). The source of the transistor is always connected in keeping with the channel type - for example, the source on an n-channel FET is connected to the negative side of the circuit, while the source on a p-channel FET is connected to the positive side of the circuit.

All FET devices can readily be used as variable resistive loads, by virtue of the fact that below the pinch-off region the channel current is dependent upon the drain to source voltage. This is an important feature of FETs because a normal resistor, diffused into semiconductor material (such as the 130Ω device shown in the TTL Gate of Figure 3.19) takes up 20 times the space of a FET device. As a result, it is rare to see MOSFET based digital circuits containing diffused resistors and instead, MOSFET loads are used. Consider again the simplistic inverter circuit of Figure 3.17 (the Common-Emitter arrangement). The MOSFET equivalent would be a Common-Source circuit using one MOSFET as a driver (Q1) and another as a pull-up resistance (Q2) as shown in Figure 3.31. Another common technique is to use Complementary MOSFETs, or CMOS devices in order to create the driver and load from an n channel MOSFET (Q1) and a p-channel MOSFET (Q2) respectively. A CMOS inverter circuit is shown in Figure 3.32.

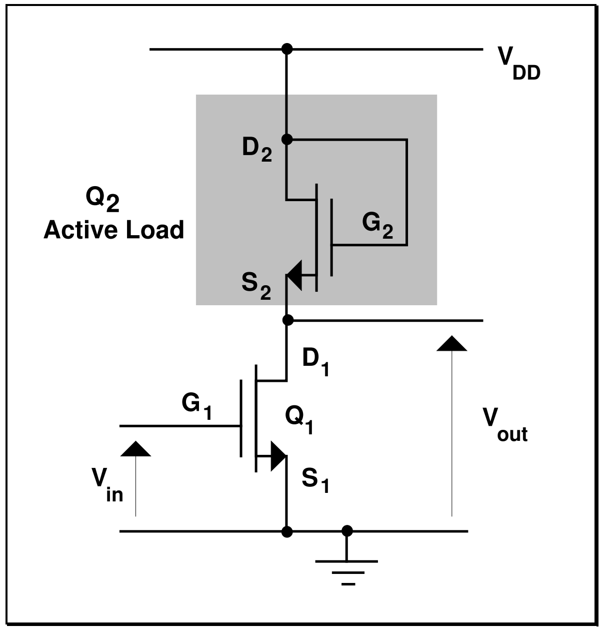


Figure 3.31 - Digital Inverter Gate Based Upon MOSFETs

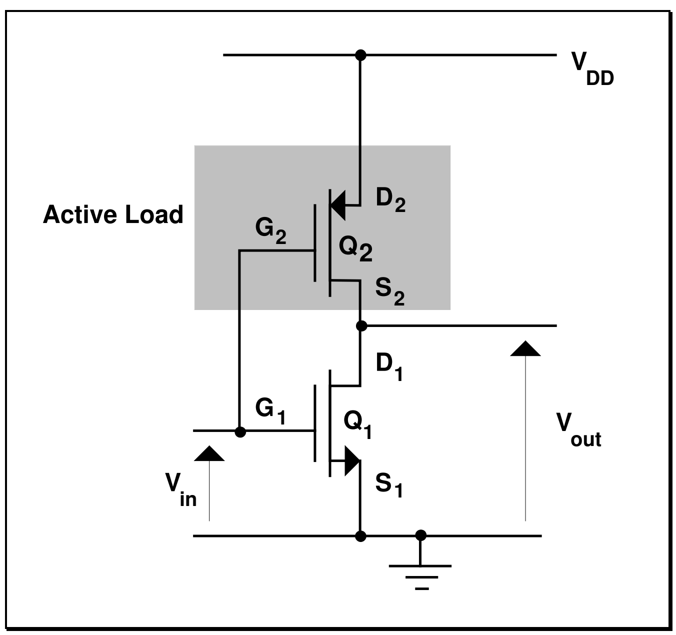


Figure 3.32 - CMOS Based Inverter Circuit with PMOS (Q2) Load

The sorts of MOSFET based digital circuits shown in Figures 3.31 and 3.32 provide a number of advantages over BJT based (TTL) circuits. The major advantage of course is that the size of FET gates is much smaller than BJT gates, so that a much higher chip density can ultimately be achieved in order to implement more complex digital functions. Secondly, however, MOSFET based circuits consume far less power than BJT circuits because they are essentially closer to the concept of an ideal switch (with very little current flowing when the devices are cut off and almost no resistance when short-circuited). The disadvantage of MOSFET based circuits is of course that they are slower than equivalent TTL devices, although the CMOS based circuits generally provide superior performance to the standard MOSFET devices.

MOSFETs are also referred to as "Insulated Gate FETs" or IGFETs, because of the Silicon Dioxide layer isolating the gate terminal from the semiconductor material and because of their high-input impedances. It is also possible to create hybrid devices with the high-input impedance characteristic of MOSFETs and the switching performance of BJTs. These devices are referred to as "Insulated Gate Bipolar Transistors" or IGBTs.

In general, the number of analog circuit applications is diminishing in respect to the number of digital circuit applications - primarily because it is possible to create digital switching circuits which:

- Emulate in digital form the functions of analog circuits (amplification, voltage conversion, etc.)
- Integrate with computer-based (microprocessor or Digital Signal Processor) controls
- Dissipate far less power than analog circuits because power is only supplied to devices when they are performing their required function.

As a result, the number of applications requiring MOSFET type devices is also proportionally increasing. However, JFET devices still have a role to play in low-noise, high-input impedance analog circuits, particularly amplifiers. To this end, we still need to configure JFETs in feedback circuits in order to create amplifier circuits whose gain is insensitive to ill-defined FET parameters. In particular, the pinch-off voltage level V_p and the drain current flowing with zero gate-source voltage (referred to as I_{DSS}) are both temperature dependent parameters whose values (although they can be calculated from manufacturer's data at any temperature) need to be isolated from the transfer characteristic of amplifier circuits.

The JFET analogy to the Emitter-Feedback circuit of Figure 3.20 is one of many circuits that can be used to stabilise the gain of the FET device as an amplifier. This circuit is shown in Figure 3.33 and the complete circuit, including a.c. signals and the Thévenin equivalent of the resistive ladder input stage, is shown in Figure 3.34.

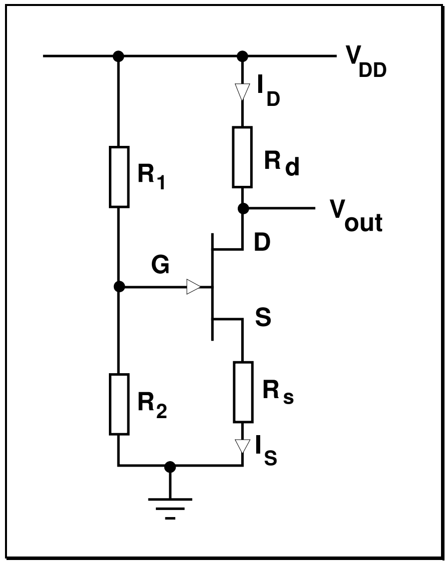


Figure 3.33 - Amplifier Feedback Arrangement for n-Channel JFET

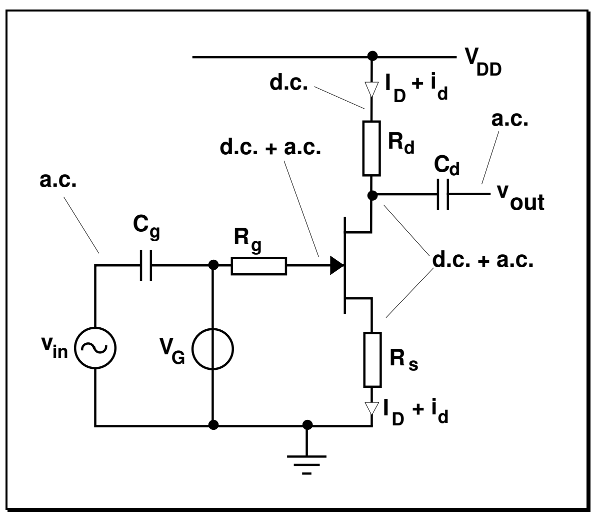


Figure 3.34 - Source-Feedback Amplifier Circuit of Figure 3.33 with a.c. Signals and Thévenin Equivalent of Input Stage

In analysing the circuit of Figure 3.34, we follow the same sort of procedure that we used with the BJT device. That is, we undertake a large-signal (quiescent, d.c.) analysis to establish the operating points of the system and then a small-signal (a.c.) analysis to examine the amplification of the circuit.

We first analyse the quiescent levels in the circuit so that we can select resistive components that will place the transistor into its normal mode for amplification - in the case of the FET this means the pinch-off or saturation region in the output characteristic. In doing this analysis, we set the a.c. voltages to zero and we calculate all the resistive values required to place the transistor into its required mode by using the output characteristic (I_D vs V_{DS}) of the transistor in question. With the FET device, we can assume that the quiescent gate current is negligible in comparison with the drain current.

For the second stage of the analysis, we restore the a.c. input and set all d.c. voltage sources to zero. We then replace the FET circuit symbol with the small-signal model for that device. The small-signal model for the FET is not unlike that for the BJT device. It is shown in Figure 3.35.

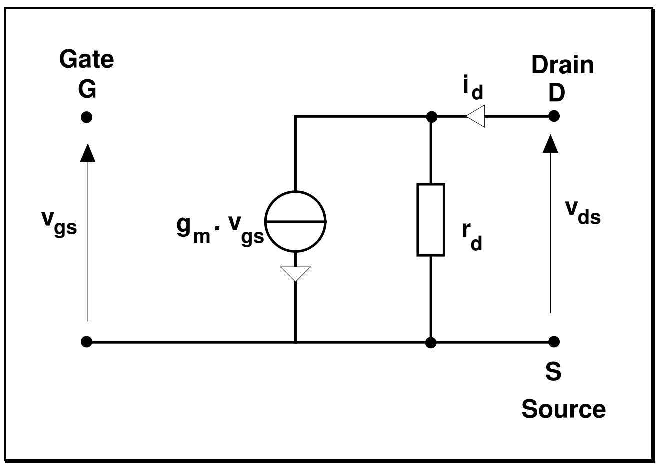


Figure 3.35 - Small Signal (Hybrid-π) Model for JFET or MOSFET

One interesting difference between the small-signal model for the FET and that of the BJT is that there is no (measurable) analogy to the base-current path for the FET - in other words, the gate and source terminals are open-circuit. The other parameters of the small signal model however are derived in an analogous manner to the BJT. For example, the small-signal transconductance of the FET is defined as the partial derivative of drain current with respect to gate-source voltage (in an analogous manner to the transconductance of the BJT). We will again simply cite the relationship between the transconductance and quiescent operating parameters for the transistor without proof, since this is outside the scope of this text. The transconductance relationship is as follows:

$$g_m = \frac{2 \cdot \sqrt{I_{DSS} \cdot I_D}}{|V_P|} \qquad \text{...(11)}$$

where:

- I_{DSS} is the drain current for zero gate-source voltage
- I_D is the quiescent operating drain current in the saturation region
- V_P is the "Pinch-Off" voltage for a depletion FET or the Threshold Voltage for an enhancement FET

All of the above parameters are either quoted in manufacturer's data for a given transistor or else can be deduced from provided or measured operating characteristics.

The dynamic drain resistance of the FET ("r_d" in the small-signal model) is obtained by taking the partial derivative of the drain-source voltage with respect to drain current, for a given value of gate-source voltage. This needs to be done in the saturation or pinch-off region of the transistor's operation and can be achieved by measuring the slope of the output characteristic (as in Figure 3.30) for a given value of gate-source voltage.

Once the small-signal model has been placed into the circuit, and all d.c. voltages have been set to zero, then an a.c. or small-signal analysis can take place in order to determine the transfer characteristic of the amplifier circuit in the same way in which the BJT circuits are analysed.

3.4 Analog and Digital Circuit External Characteristics

Regardless of whether circuits are analog or digital and regardless of whether they are BJT or FET based, there are a number of common characteristics which need to be determined in order to use them in conjunction with other circuits. In particular, we need to know what sort of an electrical load a particular circuit will place upon another circuit, or perhaps what sort of output characteristic a given circuit may have. For this reason, we commonly need to identify, measure or calculate the external characteristics of a given circuit. In order to this, we normally model a circuit, be it an amplifier or digital gate, as a black-box with some basic external parameters, as shown in Figure 3.36. We sometimes then treat systems as a cascade of black-box circuits as shown in Figure 3.37 (a).

One of the most important parameters that we need to derive for a circuit is the input impedance (Z_{in}). The input impedance defines the amount of current that the circuit will draw from a preceding circuit stage, for a given output voltage from that stage. This is important because many circuits, particularly digital circuits, cannot supply very much current and so we need to know exactly how much current the subsequent circuit will drain. In addition, in digital systems, we normally connect a number of circuits, in parallel, to the output of one particular circuit. This is referred to as "fan-out" and is shown in Figure 3.37 (b). With "fanned-out" circuits, we often need to know how many subsequent circuits, in total, can be connected to the output of a given device.

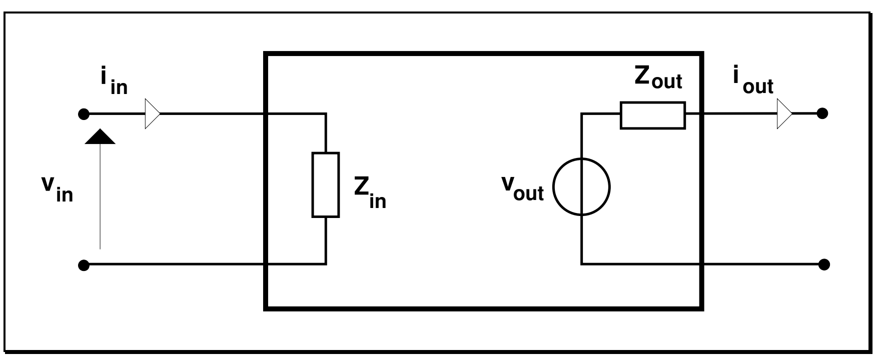


Figure 3.36 - Modelling Analog and Digital Circuits in Terms of External Characteristics

When we deal with families of circuits, such as with digital TTL or CMOS gates, the manufacturers specify the total fan-out for a device, assuming that other devices of the same family type will be connected to the output of the first stage. These characteristics are all defined on the basis of the input impedance of the family of devices in question.

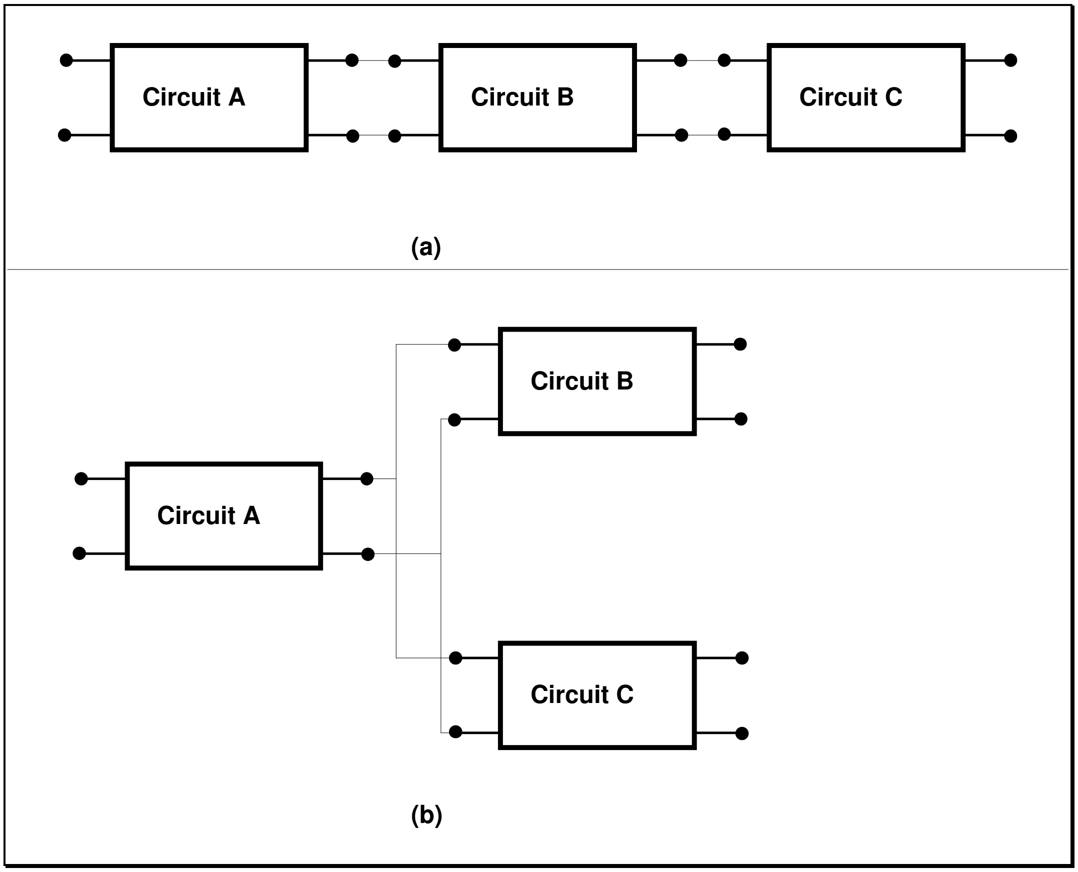


Figure 3.37 - (a) Cascaded Circuits
(b) Fan-out From One Circuit

The input impedance for a circuit can either be derived or measured, depending upon whether or not the circuit diagram for the device is actually known. In either event, in terms of Figure 3.36, the input impedance is defined to be:

$$Z_{in} = \frac{v_{in}}{i_{in}} \qquad ...(12)$$

If a known voltage is applied to the input terminals of the device and the input current measured, then the input impedance can readily be determined. If, for example, the circuit was known to be an analog amplifier, such as an emitter feedback circuit, then the input impedance could be algebraically determined from the complete small-signal model for the system, by deriving an expression for v_{in}/i_{in}.

In general, circuits are composed of resistors, capacitors, inductors and, in the case of small-signal transistor circuits, dependant current sources. As a result, the input impedance, to a.c. signals may be frequency dependent, due to the "jω" terms that arise in applying the phasor method to circuit solution. Moreover, the a.c. and d.c. input impedances will not generally be the same due to this frequency dependence. For this reason, if we are measuring input impedance, we will be measuring magnitude only and we need to be sure that we are measuring at a frequency commensurate with the frequency at which the circuit will normally be operating. When we derive expressions for input impedance, we can of course calculate the magnitude of the impedance at any given frequency. For analog circuits, we need to be aware of the frequency dependence of input impedance, whereas in digital circuits, we are primarily concerned with only the d.c. impedance to current flow.

Figures 3.37 (a) and (b) can be somewhat misleading in the sense that they may induce a novice to believe that any circuit can be divided up into isolated stages which can be analysed independently and then brought together at a later stage. This is only true if we always assume that there will be an output current and voltage from each stage of the circuit - that is, that every stage will be loaded. We cannot simply derive characteristics on the assumption that each stage will have no output current and then join these "independent" stages together. In general, the characteristics of a circuit, including its input impedance, depend upon the load applied on the output of that circuit, and hence, the subsequent stages of the system.

There may be some specific circuits, particularly amplifiers and digital circuits, whose input characteristics do not vary greatly with the load applied to the output. This is however a function of the isolating properties of transistors and should not be taken as a general rule. If you are still in doubt, take a circuit composed of resistors, capacitors and inductors and determine the input impedance for the entire circuit. Then, divide the circuit at an appropriate point, analysing each stage on the assumption that the other stage has no loading effect - note the difference.

In order to understand the concept of output impedance, it is necessary to review one of the most basic elements of network analysis - that is, the Thévenin and Norton equivalent circuits. Essentially, Thévenin and Norton showed that any circuit, regardless of how complex it is, can be represented as:

- A voltage source in series with a resistor (called a Thévenin equivalent circuit)
- A current source in parallel with a resistor (called a Norton equivalent Circuit).

The two equivalent circuits are shown in Figure 3.38 and are analytically derived as shown in Figure 3.39, using the following technique:

Take the original circuit and calculate the voltage output when the output is open-circuited (this is called v_{oc}). Take the original circuit and calculate the current that flows when the output terminals are short-circuited (this is called i_{sc}). The values for the Thévenin and Norton equivalent circuits are then:

$$v_{TH} = v_{OC}$$

$$Z_{TH} = \frac{v_{OC}}{i_{SC}}$$

$$i_N = i_{SC}$$

$$Z_N = \frac{v_{OC}}{i_{SC}}$$

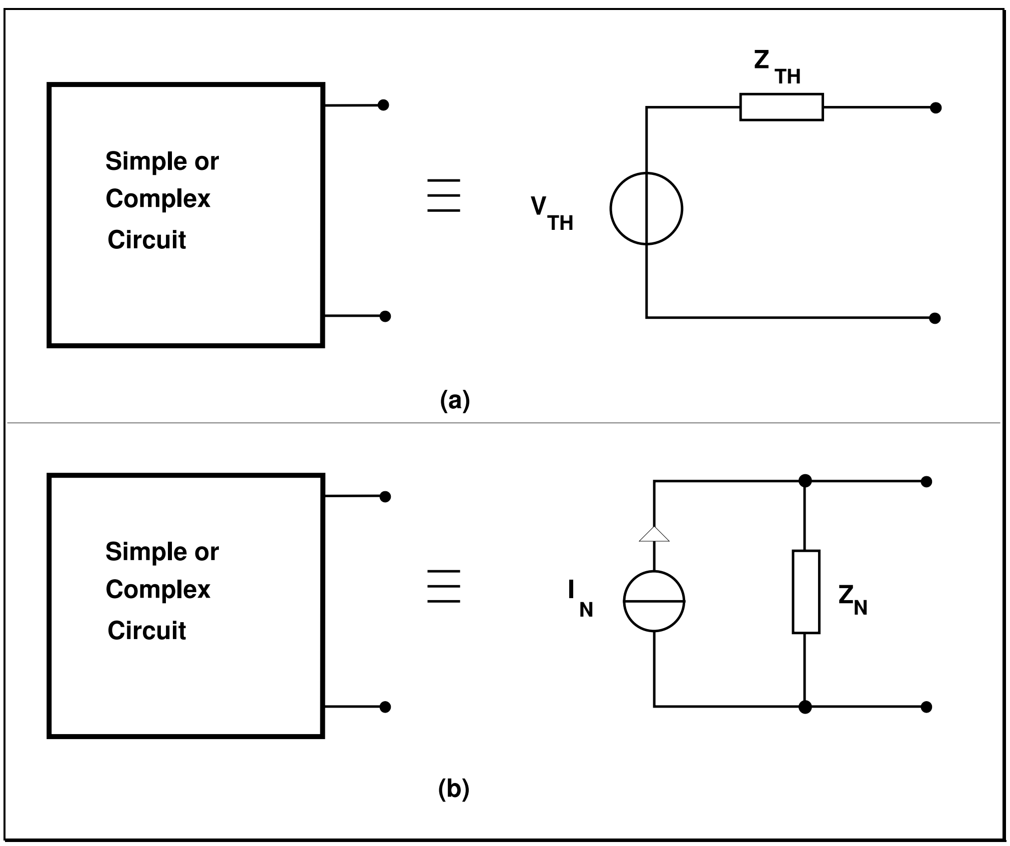


Figure 3.38 - (a) Thévenin Equivalent of a Circuit
(b) Norton Equivalent of a Circuit

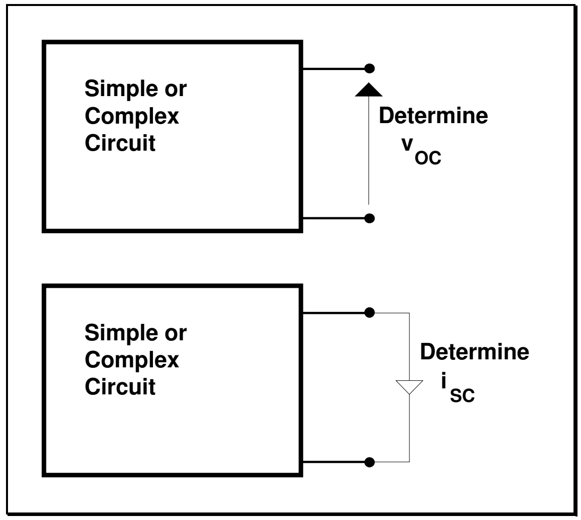


Figure 3.39 - Determining Thévenin and Norton Equivalent Circuits from an Existing Circuit

Having briefly re-examined the concept of the Thévenin and Norton equivalent circuits, we are in a position to understand the concept of the output impedance of a given circuit. Referring back to Figure 3.36, we can see that the output of our "black-box" circuit is in fact a Thévenin equivalent model. The question is, how can we determine the value of the output impedance. Analytically, we can do the open-circuit-voltage and short-circuit current test to calculate the value of the impedance, as described in terms of the Thévenin circuit model. However, in terms of measurement, a little more thought is required in order to determine the value.

Theoretically, if we measure the open-circuit voltage and short-circuit current of a circuit, then we should be able to determine the output impedance. The problem with this technique is that in endeavouring to short-circuit the output we may damage the device in question by causing an excessive current to flow. The solution therefore is to measure the open-circuit voltage and then to measure the current flow when a known impedance (R_L) is connected as a load (rather than a short-circuit). The value of the known impedance is chosen such that the maximum current flow in the circuit (ie: assuming the output impedance is zero) does not exceed the ratings of the circuit in question. The measured current flow in the circuit is then equal to:

$$\left|i_{out}\right| = \frac{\left|v_{out}\right|}{\left|Z_{out} + R_L\right|}$$

from which the magnitude of the output impedance can be calculated. In analog circuits, we must again keep in mind that the output impedance, like the input impedance may be frequency dependant.

3.5 Operational Amplifiers

In Section 3.3 of this chapter, we examined the basic characteristics of various transistors and how they could be used in order to create digital and analog circuits, with some emphasis on analog amplifiers. We also examined amplifiers that were created with a single transistor, together with a feedback circuit for gain stabilisation. There are many variations on single-transistor amplifier circuits and many more variations on multiple transistor amplifier circuits. In general, the detailed, small-signal analysis of multiple transistor circuits is quite complex without computer simulation and the parasitic effects (such as stray capacitances, etc.) of transistor circuits can cause them to oscillate or resonate with minor changes in feedback resistance values.

In the final analysis, the design of realistic amplifier circuits from first-principles of transistor theory is a specialised task, and not one which would be tackled by engineers involved in general-purpose systems design. For this reason, a number of commercially available amplifier packages have been designed as building blocks for systems design work. The generic name for these circuits is "Operational Amplifiers". Developments in power semiconductor technology have meant that circuits are now available for both low power and medium power applications. The objective of our discussions on operational amplifiers is not so much to delve into the intricacies of their design characteristics, but rather to understand their functionality and their typical applications in the process of interfacing microprocessor based control systems to industrial level signals.

Operational amplifiers are not new devices. One of the oldest types of operational amplifier is the 741 series, which was introduced in the mid-1960s and has been in widespread use since that time. Despite the fact that the 741 has been superseded by more modern devices, it tends to remain as a reference system upon which the more modern designs have been based. Even this rather old amplifier design is based upon some 24 bipolar junction transistors and some 11 diffused resistors and hence its analysis is outside the scope of this book. A simplified version of this amplifier circuit is shown in Figure 3.40, wherein a number of the transistor circuits have been replaced with their functional equivalents.

The operational amplifier is essentially a "difference" amplifier. It only amplifies the difference between the voltages at terminals A and B (as shown in Figure 3.40). Transistors Q_1 and Q_2 provide a high input impedance to the circuit (so that it only draws a very small current from preceding circuits) and Q_3 and Q_4 amplify the difference in the signals. Q_5 and Q_6 provide an active resistive load for the system. Ultimately, as the differential input is converted to a single-ended input and again amplified. The combined collection of transistors creates a device with:

- High-Input Impedance
- Low Sensitivity to Common-Mode Signals
- Low-Output Impedance.

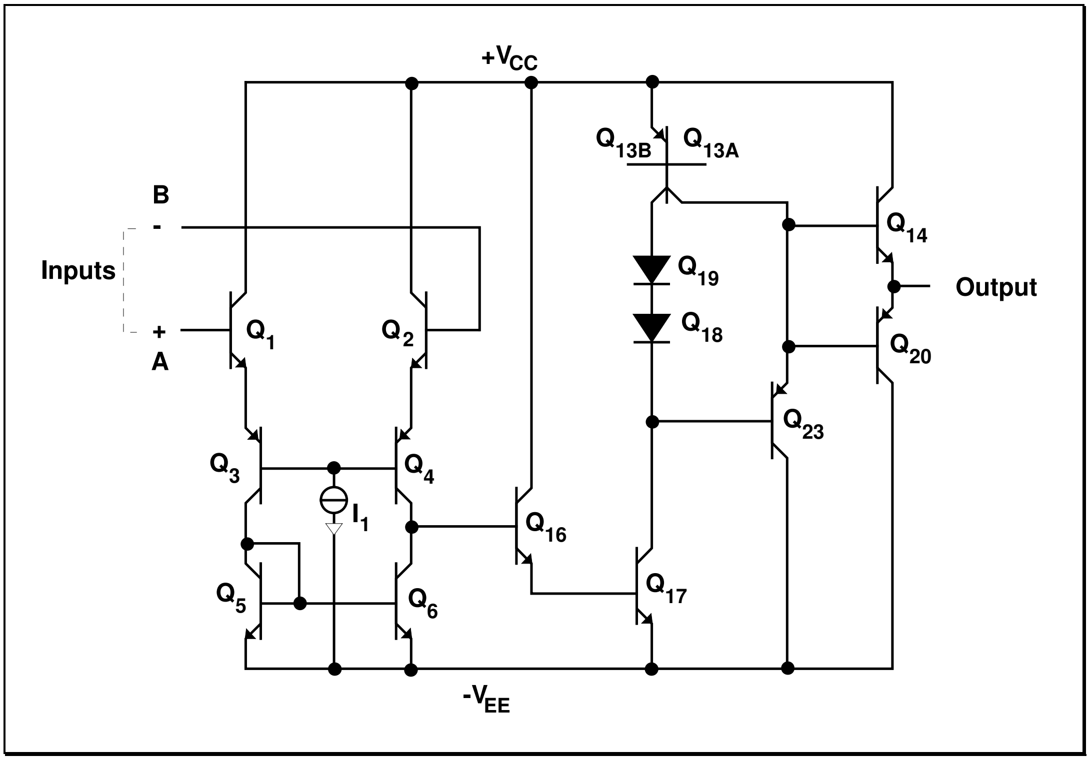


Figure 3.40 - Simplified Circuit Diagram for Single-Chip 741 Operational Amplifier

One of the key points to note about the 741 device shown in Figure 3.40 is the supply rail system. In this rather old device, a positive 15 volt d.c. supply needs to be connected to create the V_{CC} supply rail and a negative 15 volt d.c. supply needs to be connected to create the V_{EE} supply rail. As with all analog amplifiers, it is the supply rails which provide the additional power that is injected into the system in order to provide amplification. The input signals are, for the most part, just signals and their energy levels are generally very low in comparison to the output energy. In digital circuits, the same philosophy applies - supply rails provide the energy required by the transistors in order to switch from their open-circuit mode to their short-circuit mode, and again, the inputs are just signals and not generally providers of the switching energy. One of the problems with the 741 design is its requirement for positive and negative d.c. supply rails, which add to system cost, and hence more modern designs endeavour to achieve the same sort of functionality as the 741 device with only a single, positive d.c. supply rail.

The most common way to design systems using operational amplifiers is to treat the entire amplifier as a black-box device, normally represented with a circuit symbol as shown in Figure 3.41(a). The input marked with the "-" sign is referred to as the inverting terminal and the input with the "+" sign is referred to as the non-inverting terminal. The supply rails are not normally shown, because they vary from device to device and do not actually affect the functionality of the amplifier itself. However, it must always be remembered that it is these rails that make the operational amplifier an active device which can inject energy into a system. The idealised circuit model of the operational amplifier is shown in Figure 3.41 (b).

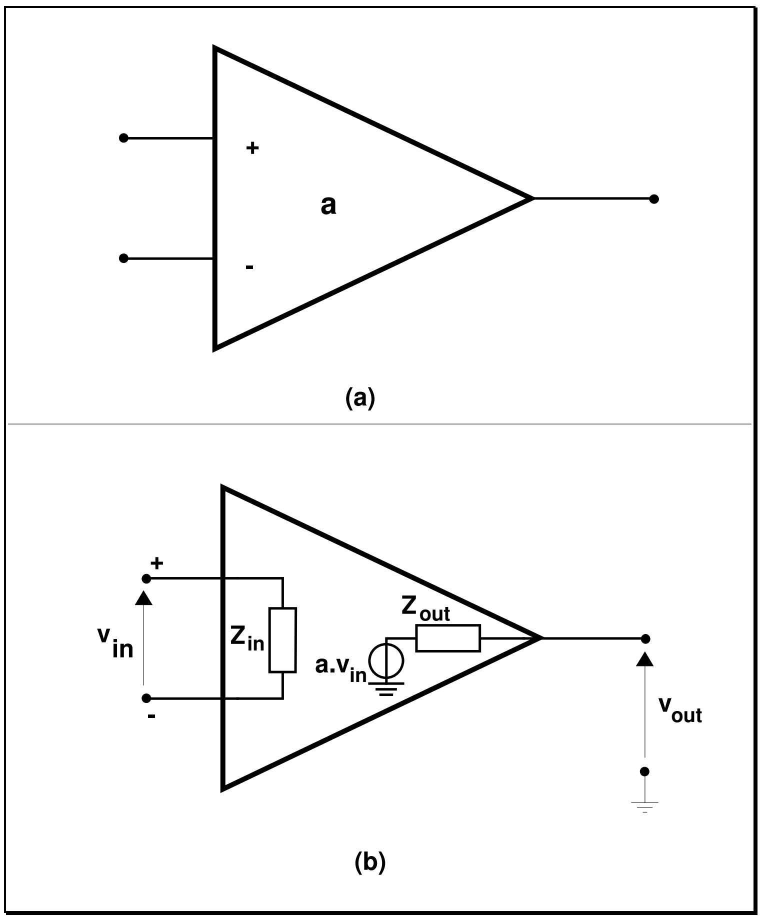


Figure 3.41 - (a) Circuit Symbol for Operational Amplifier (b) Idealised Model of Operational Amplifier Circuit

In most operational amplifier system designs, we often make assumptions to further simplify the idealised circuit of Figure 3.41 (b). For example, we assume that the amplification factor, or gain, of the amplifier "a" is very large and tending towards infinity - hence, for any realistic output voltage, v_{in} must always be approximately zero. For rough calculations we can also assume that the input impedance of the device is infinite and the output impedance is zero.

One may well ask what the practical significance of an amplifier with infinite gain would be in terms of electrical circuits. This clearly doesn't happen in the real world and, as it turns out, the so-called open-loop gain of an operational amplifier is actually a finite quantity albeit only nominally defined. Moreover, most operational amplifier based systems are effectively designed in a manner which is not specific to a particular type of amplifier. The net result is that operational amplifiers are not used as open-loop devices, but rather as elements in closed-loop systems, surrounded by feedback elements. A proper understanding of closed-loop amplifier systems requires an understanding of classical control theory, where the objective is to realise a stable system transfer characteristic, which is independent of all unstable or ill defined variables. Once we understand this concept, then we should appreciate the fact that an infinite open-loop gain does not necessarily lead to a system with infinite closed loop gain, but rather a gain which is dependent upon the feedback elements in the system (as shown in Figure 3.16 and described by equation (5)).

In order to understand the purpose of a particular operational amplifier circuit, analysis is normally carried out using the approximate model in the following way:

(i) Assume that the voltage across the input terminals is approximately zero
(ii) Assume that the current flowing through the input terminals is approximately zero
(iii) Apply Kirchoff's Voltage and Current Laws in order to obtain an expression for output voltage or current in terms of input voltage or current

A more accurate analysis can be carried out by assuming a finite voltage gain "a" between the input and output terminals and by assuming an input impedance of Z_{in} between the input terminals. Values for these parameters can normally be obtained from relevant data sheets for the specific operational amplifier in question.

The first operational amplifier based circuit to be examined is called the d.c. voltage-follower. The circuit diagram for this is shown in Figure 3.42 (a). This is a rather peculiar circuit because the entire output is used as feedback into the inverting terminal of the amplifier.

Applying Kirchoff's Voltage Law to the d.c. voltage-follower circuit gives us the following relationship:

$$v_s - v_{in} = v_{out}$$

However, v_{in} is approximately zero and hence:

$$v_{out} = v_s$$

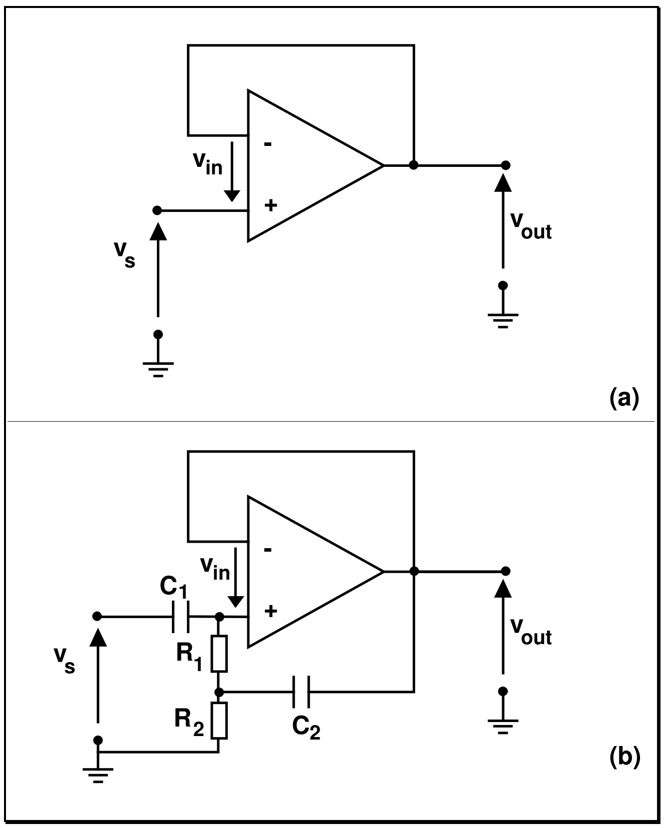


Figure 3.42 - (a) Operational Amplifier Based "d.c. Voltage Follower" Circuit (b) Operational Amplifier Based "a.c. Voltage Follower" Circuit

The voltage-follower circuit may appear to be somewhat unusual in the sense that it doesn't seem to do anything - there is no amplification or attenuation. However, the voltage-follower has a very important role to fulfil. It acts as a buffer or interface that provides a high input impedance circuit (ie: small load) to a preceding circuit. The voltage-follower replicates the voltage from the previous stage (unity gain) and is capable of driving a subsequent circuit that has a low input impedance (ie: high load).

In Figure 3.42 (b), the a.c. equivalent of the d.c. voltage follower circuit is shown. This circuit, like the one in Figure 3.42 (a) also acts as a high input impedance buffer between a primary circuit and a high-load circuit that would normally draw a current that is too large for the primary circuit to supply. The a.c. equivalent circuit has two capacitors, C_1 and C_2, whose sizes are selected in order to provide zero impedance at the a.c. operating frequencies of the system. R_1 and R_2 provide a path for d.c. current flow in the system.

A common role for operational amplifiers is acting as energy transducers in circuits. That is, to convert voltage into current (transconductance amplification) or current into voltage (transresistance amplification). Figure 3.43 (a) and (b) show two circuits suitable for transconductance amplification.

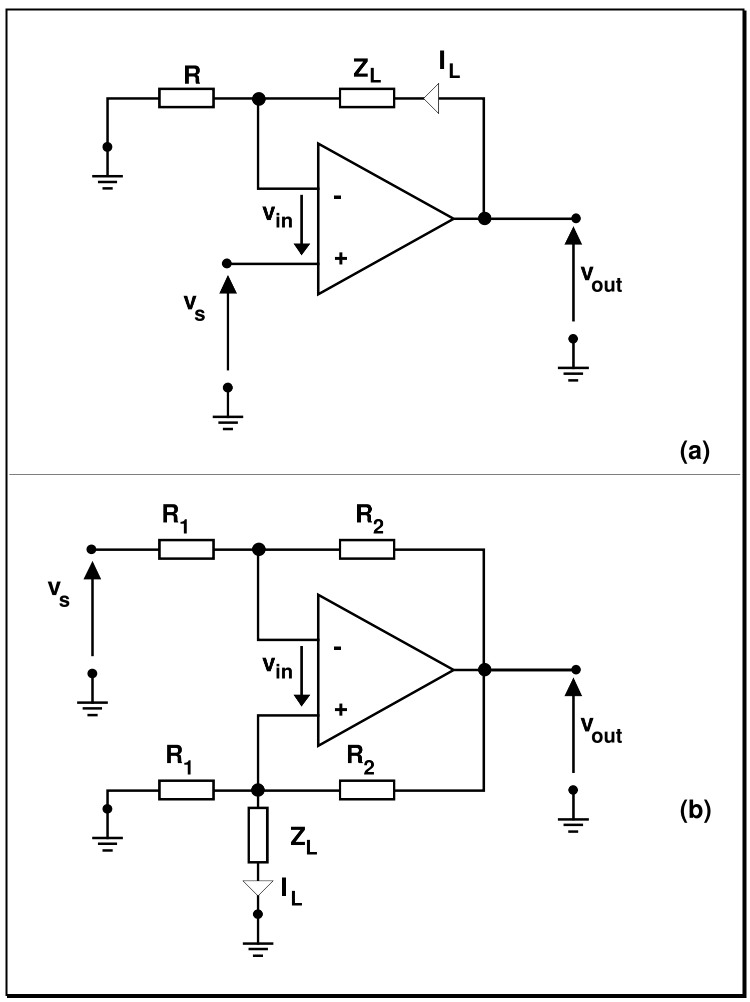


Figure 3.43 - (a) Transconductance Amplifier for Floating Load
(b) Transconductance Amplifier for Grounded Load

The circuit in Figure 3.43 (a) is the simpler alternative that is used when neither end of the load circuit (Z_L) is grounded. The more complex arrangement of Figure 3.43 (b) needs to be used whenever one end of the load is grounded. A simple analysis of the circuit (assuming infinite input impedance and amplifier gain) reveals that for the circuit of Figure 3.43 (a), the output load current is:

$$i_L = \frac{v_s}{R} \qquad ...(13)$$

For the circuit of Figure 3.43 (b), the output load current is:

$$i_L = -\frac{v_s}{R_1} \qquad ...(14)$$

The dual circuit of those shown in Figure 3.43 is the transresistance amplifier which converts current-based signals to voltage. This is shown in Figure 3.44.

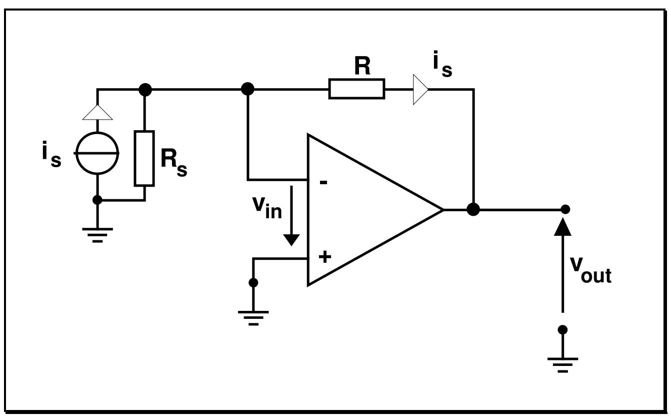


Figure 3.44 - Transresistance Amplifier - Current to Voltage Conversion

The transresistance amplifier is simple to analyse. The non-inverting terminal is connected to earth and hence the inverting terminal is also approximately at earth potential (virtual earth). The current through the source resistance R_s is also zero since one end is at earth potential and the other end is at virtual earth. The output voltage is therefore simply:

$$v_{out} = -i_s \cdot R \qquad ...(15)$$

One of the most common applications of operational amplifiers is in amplification and attenuation of signals in control systems. This can be readily accomplished with the inverting amplifier circuit of Figure 3.45 (a) or the non-inverting amplifier circuit of 3.46 (a). One circuit provides a negative amplification and the other a positive amplification.

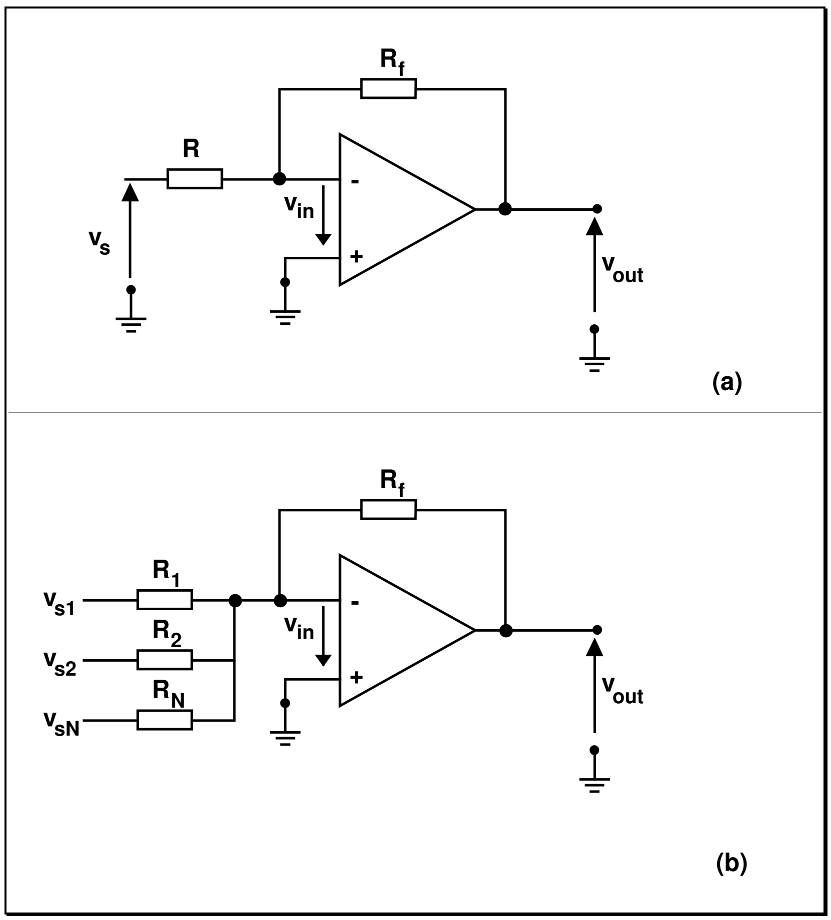


Figure 3.45 - (a) Inverting Amplifier Arrangement
(b) Inverting "Summing" Amplifier for Providing the Weighted Sum of "N" Input Voltages

For the circuit of Figure 3.45 (a), the gain of the system is defined by the feedback resistors and the output signal is the negative of the input signal:

$$v_{out} = -\frac{R_f}{R} \cdot v_s \qquad \text{...(16)}$$

For the circuit of Figure 3.45 (b), a number of weighted inputs can be summed together, amplified and inverted. For this amplifier:

$$v_{out} = -\left(\frac{R_f}{R_1} \cdot v_{s1} + \frac{R_f}{R_2} \cdot v_{s2} + .. + \frac{R_f}{R_N} \cdot v_{sN}\right) \qquad ...(17)$$

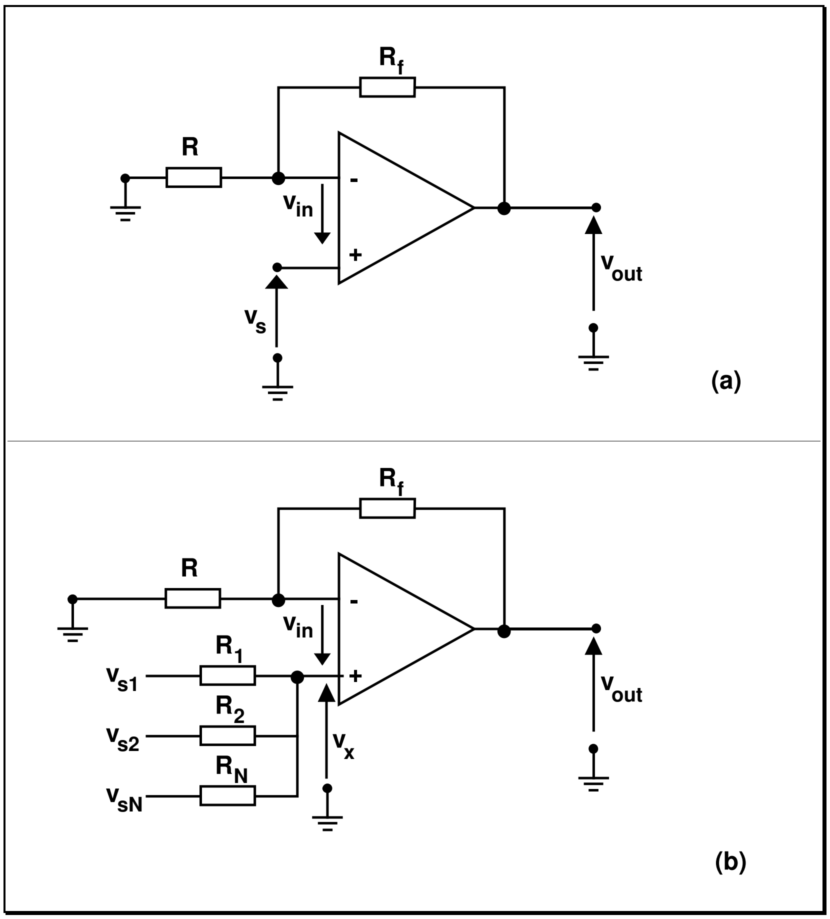


Figure 3.46 - (a) Non-Inverting Amplifier
(b) Non-Inverting Summing Amplifier for Providing the Weighted Sum of "N" Input Voltages

For the circuit of Figure 3.46 (a), the output waveform is a scaled, positive version of the input and is defined as follows:

$$v_{out} = \left(1 + \frac{R_f}{R}\right) v_s \qquad ...(18)$$

For the circuit of Figure 3.46 (b), it is somewhat more difficult to calculate the transfer function, which ultimately depends upon the number of inputs. We know however that the transfer function can be derived from single input circuit equation (18) and will be:

$$v_{out} = \left(1 + \frac{R_f}{R}\right) v_x \qquad \text{...(19)}$$

where v_x is the voltage at the non-inverting terminal of the amplifier. The method for calculating v_x is to take the Thévenin equivalent circuit of all the inputs at the non-inverting terminal, assuming that no current flows into that terminal. The result is then inserted into equation (19).

A number of energy transducers provide a pair of outputs and it is often the difference between the two that needs to be amplified or attenuated. This can be achieved by using a differential amplifier configuration, as shown in Figure 3.47.

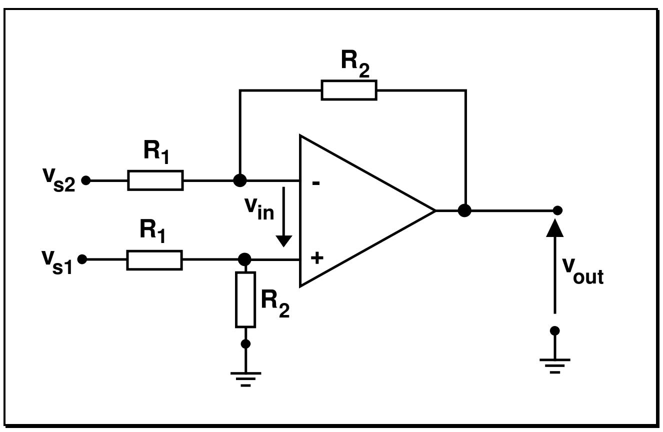


Figure 3.47 - Differential Amplifier Configuration

The output can be determined in terms of the two inputs by using the principle of superposition (ie: calculating the output voltage for each of the inputs individually and then adding the result). The relationship is as follows:

$$v_{out} = \frac{R_2}{R_1} \cdot \left(v_{s1} - v_{s2}\right) \qquad \text{...(20)}$$

The relationship is not altogether surprising since the differential configuration is a hybrid of the inverting and non-inverting amplifier configurations.

The final two operational amplifier circuits to be examined are the integrator and the differentiator. These are shown in Figure 3.48 (a) and 3.48 (b) respectively. The purpose of these circuits, as their names suggest, is to provide an output voltage proportional to the integral or differential of the input voltage waveform. This sort of functionality is achieved by using a capacitor, whose current is proportional to the derivative of the voltage.

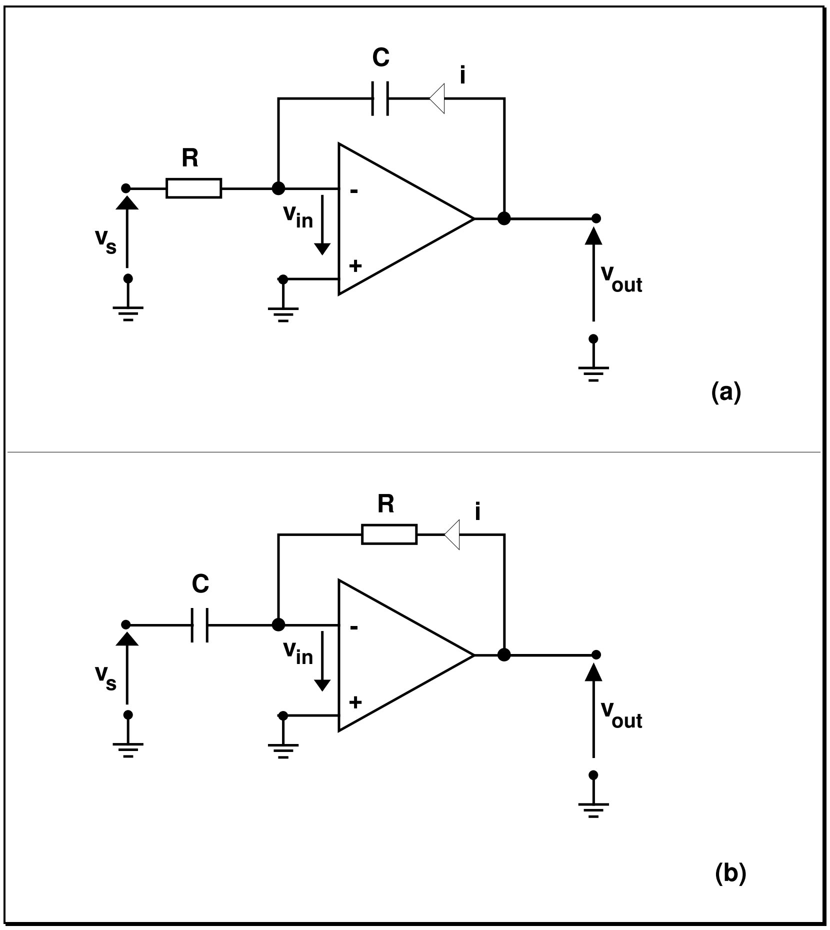


Figure 3.48 - (a) Integrator Circuit
(b) Differentiator Circuit

For the integrator circuit of Figure 3.48 (a), the inverting terminal is approximately at zero voltage, so the current through the capacitor is proportional to the derivative of the output voltage, and is equal to the current flowing through the resistor. The following relationships apply:

$$i = C \cdot \frac{dv_{out}}{dt} = -\frac{v_s}{R}$$

$$\Rightarrow v_{out} = -\frac{1}{RC} \cdot \int v_s \, dt \qquad ...(21)$$

For the differentiator circuit of Figure 3.48 (b), the inverting terminal is also approximately at zero voltage and we again equate the current through the resistor to the current through the capacitor (ie: assume infinite amplifier impedance) in order to ascertain the relationship between output and input, as follows:

$$i = \frac{v_{out}}{R} = -C \cdot \frac{dv_s}{dt}$$

$$\Rightarrow v_{out} = -R \cdot C \cdot \frac{dv_s}{dt} \qquad ...(22)$$

The integrator and differentiator both featured prominently in analog computers and controllers because they provided a very effective means of providing integrals and derivatives in real-time. These were particularly useful in classical control theory where Proportional Integral Derivative (PID) control techniques are used as a control strategy. However, despite the increase in availability of low-cost, sophisticated digital processing technology, these devices still retain their usefulness in control applications because they free the controlling processor from the task of carrying out integration and differentiation. This not only minimises the performance drain on the processor but additionally simplifies programming because many low-level controllers are often programmed in a machine or assembly language, where the software implementation of a digital integration or differentiation can substantially complicate the programming of a control algorithm.

3.6 Linearity of Circuits - Accuracy and Frequency Response

Analog circuits are commonly used to interface incompatible devices by:

- Scaling
- Converting from voltage to current
- Converting from current to voltage
- Injecting energy into systems from external supply rails.

All of these analog functions require circuits with a high degree of accuracy. In digital circuits, on the other hand, we only ever deal with voltage levels that correspond to the binary numbers one or zero. Therefore, the absolute values of voltage in digital circuits are not critical, provided that they are within the appropriate ranges used to describe those binary numbers. Analog circuits therefore provide a greater challenge to system designers than digital circuits because information is always explicitly contained within analog voltages.

There are three main factors which need to be considered when using analog devices such as energy transducers (of which, circuits such as the basic transistor amplifiers and the operational amplifiers are a subset). These are:

(i) Accuracy
(ii) Frequency response
(iii) Linearity.

The overall effect of these factors is shown schematically in Figure 3.49. There are an enormous variety of energy transducers currently applied to engineering designs. Rather than attempt to cover the entire spectrum of devices, we will only discuss the performance factors in terms of the sorts of devices that we have examined in this chapter, such as transformers, basic transistor amplifiers and operational amplifiers. Most other energy transducers have analogous limitations to those exemplified by these devices.

In order to begin our examination of deviations from ideal behaviour, let us recall the work we did in Section 3.3.2, where we examined the basic attributes of Bipolar Junction Transistors and noted that a number of BJT parameters were ill-defined and subject to temperature variation or variation with operating current. Our objective then was to design circuits that would make the transfer functions of these circuits independent of these parameters. This was achieved by the traditional method of feedback circuits, such as the one shown in Figure 3.20 for the BJT or those shown in Figures 3.42 - 3.48 for the Operational Amplifier. In all these cases, provided that the ill-defined open-loop gains of the systems are sufficiently large, then their transfer functions are only dependent upon the feedback circuits.

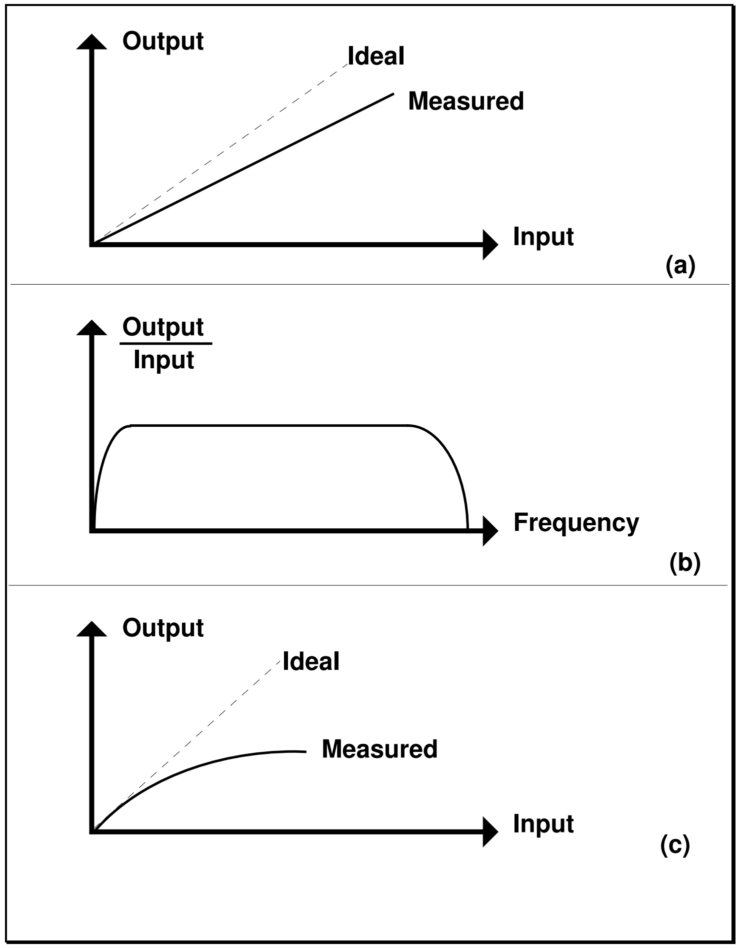


Figure 3.49 - Imperfections of Realistic Energy Transducers and Circuits (a) Accuracy Problems; (b) Limited Frequency Response; (c) Non-Linearity

The inverting amplifier of Figure 3.45 (a) is a classic example of a circuit whose gain is stabilised by the feedback of output to input. In this circuit, the transfer ratio of output voltage to input voltage is only dependent upon the ratio of circuit resistors as shown in equation (16). However, this analysis assumed that the open-loop gain and the input impedance of the amplifier were both infinite. Let us therefore examine the factors that cause the inverting amplifier circuit to deviate from ideal behaviour:

(i) ***Accuracy***

In a practical sense, we know that resistance values provided by resistor manufacturers are only nominal and subject to accuracy tolerances. The accuracy of the inverting amplifier transfer ratio (output to input) is therefore dependent upon the actual resistance values, rather than the specified values. Actual resistances can vary due to manufacturing tolerances and also due to the effects of age and operating temperature. Another factor affecting the accuracy of the circuit is the open-loop gain of the operational amplifier. If the open-loop gain of the operational amplifier is small, then the relationship in equation (16) no longer holds and so the circuit deviates from the expected behaviour.

(ii) ***Frequency Response***

All physical devices have performance characteristics that change with operating frequency. In the case of the operational amplifier, we have a system made up from transistor components, fabricated onto a single chip. When we examined the small-signal (hybrid-π) model of the BJT and the FET, we did not include the parasitic capacitances that exist between various points within the devices themselves because at normal operating frequencies these elements were insignificant. However, since the impedance of a capacitor is inversely proportional to frequency (in an a.c. system), we know that the impedance of these parasitic capacitances to high-frequency voltage components becomes zero (short-circuit). As a result, the parasitic capacitance between the base and emitter of a BJT (or gate and source of a FET) can ultimately form a short-circuit between these terminals, thereby lowering the gain of the transistor.

As a result of parasitic capacitances within the semiconductor structure, even the transistor has a limited operating frequency range. This limited range obviously applies to the combination of transistors within an operational amplifier. The ratio of output signal to input signal begins to attenuate outside the normal operating frequency range of the transistor devices until eventually, the output attenuates to zero at high frequencies. The frequency response of the operational amplifier is not quite like the one shown in Figure 3.49 (b) because the amplifier does provide gain even with zero frequency signals (d.c.). However, in the case of the inverting amplifier circuit, a reduction in the open-loop gain of the operational amplifier (at high frequencies) leads to a deviation from the assumption of "infinite gain" and hence a deviation from the transfer relationship of equation (16).

Another frequency issue with operational amplifiers is the so-called "slew-rate" or maximum rate of rise of output for a given input change. This specifies how quickly the outputs of an amplifier can respond to a change in input level.

(iii) Linearity

When we examined the use of the BJT in the emitter-feedback circuit, we came up with a total solution for the voltage at each node. This was shown in Figure 3.25. We noted then that the signal at the collector node of the transistor could never exceed the supply rail voltage and that as a result of increasing the magnitude of the input signal, the output signal would eventually distort (clip). This phenomenon occurs in all amplifier circuits, including the operational amplifier and ultimately leads to distortion, because the input and output waveforms are not linearly related. So, in the case of the inverting amplifier circuit, the system is only linear while the transistors within the operational amplifier have signals below the rail voltages.

Another familiar circuit that we can use in order to understand the three major limitations of circuits and transducers is the transformer:

(i) Accuracy

At normal load currents, realistic transformers do not provide a voltage transfer ratio equivalent to the turns ratio. This is because of the resistance of the windings and because of flux leakage in the core (as represented by circuit elements shown in Figure 3.7 (b)). At no load (zero output current), very little current flows in the windings and hence the voltage is almost identical to the turns ratio - however, as the load current increases, the voltage transfer ratio is lower than that specified by the turns ratio. This deviation from ideal is referred to as the "regulation" effect of the transformer and affects the accuracy of the circuits using the transformer as an energy transducer.

(ii) Frequency Response

Unlike operational amplifiers, transformers do not operate at zero frequency (d.c.). They do however, provide a relatively stable output/input ratio from low frequencies up into the kilohertz ranges and then gradually begin to attenuate the output voltage until at very high input voltage frequencies, no output voltage is obtained. The low-frequency limitation is due to Faraday's Law of Electromagnetic Induction and the high-frequency limitation occurs because of the magnetisation/demagnetisation of the core that has to take place with each voltage cycle - there are limits to the speed with which magnetic domains within a ferromagnetic material can be rearranged.

Every physical device has analogous frequency limitations caused by the inability of one energy form to be converted into another form at infinite speed. Even the resistance of a simple piece of wire increases with frequency, because of "skin-effect", where current tends to flow on the outside of a conductor at high frequencies. This tends to accentuate the attenuation of output voltage in transformers at high frequencies

(iii) Linearity

The flux density (***B***) in the core of the transformer is non-linearly related to the magnetic field intensity (***H***) applied to the core by what is known as the magnetisation curve for the core. At low values of ***H***, ***B*** and ***H*** are linearly related. However, as ***H*** is increased, the core saturates and the value of flux density in the core becomes almost independent of the magnetic field applied. ***H*** is dependent upon the current (hence voltage) applied to the primary windings. Increasing the primary voltage to a high level causes the transformer to saturate and creates non-linearity between the input and output sides of the device.

The above two examples illustrate the sort of reasoning that should be applied to any energy transducer before it is placed into a circuit, so that either the device can be operated within a limited range or else so that the deviations from ideal behaviour can be accounted for by other means. Manufacturers normally provide specifications for the limitations of their particular energy conversion device. For example, the specifications for an operational amplifier include frequency response characteristic curves, slew-rate, data for calculating (interpolating or extrapolating) parameters affected by temperature or operating current, etc. However, not all transducer manufacturers are as forthcoming with this sort of data, and in the absence of this sort of information, a reasoned system analysis cannot be carried out without a background investigation into the physics and design of the device in question.

3.7 Thyristors

3.7.1 Introduction

Once you begin to read this section, you may feel that it is rather peculiar that we have chosen to discuss diodes then transistors and amplifiers and have now reverted back to discussing diode-like devices again. In fact the reason we have done so is because, in semiconductor terms, the modern thyristor is more akin to a pair of transistors than it is to a single diode. However, the main purpose of thyristors is to provide a current path that is turned on and off whenever a triggering voltage exceeds a certain level. Thyristors were originally divided into a number of sub-groups that included the Silicon Unilateral Switch or "SUS" (which has now been discontinued) and the Silicon Controlled Rectifier or "SCR". The SCR is now by far the most common form of thyristor, although it is also possible to purchase bi-directional thyristors such as Diacs and Triacs, which we shall examine in section 3.7.3 as part of our overview of thyristor devices.

Thyristors can ideally be used with sinusoidal voltages in order to provide phase control, and by controlling conduction, they can reduce the average value of an a.c. input voltage in order to provide motor speed control. Thyristor-based circuits can also be used in light dimmer circuits in order to reduce the voltage supplied to light fittings. More importantly, in large-scale power circuits, thyristors can be used to replace traditional diodes in rectifier bridges. The advantage of thyristors over diodes in rectifier applications is that it is possible to use bridges for d.c. to a.c. conversion. This is the reverse procedure to rectification and is referred to as "inversion". The circuits that are used to accomplish this are called inverters.

3.7.2 Silicon Controlled Rectifiers

The semiconductor structure for an SCR is shown schematically in Figure 3.50 (a). Figure 3.50 (b) shows the circuit equivalent for the SCR, which is essentially a pnp and npn transistor pair, joined between the base of the pnp and the collector of the npn. Figure 3.50 (c) shows the common circuit symbol for the SCR, which is effectively that of a diode with a controlling gate terminal. The voltage-current characteristic for the SCR appears to be most unusual on a first glance and to some extent misleads one as to the true functionality of the device. The typical voltage-current characteristic for the SCR operating in the forward mode (anode voltage greater than cathode voltage) is shown in Figure 3.51, for a range of different gate currents.

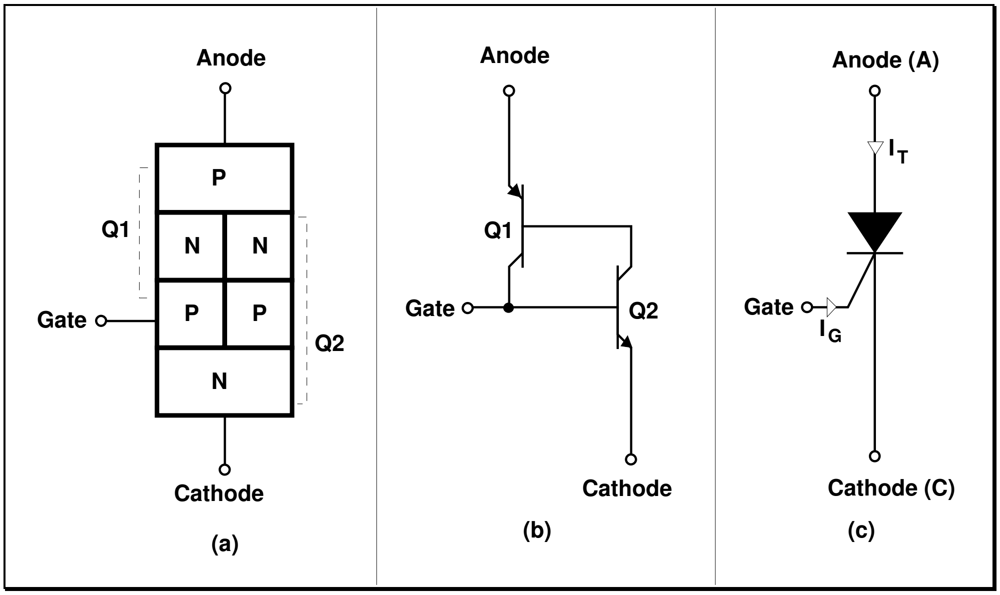


Figure 3.50 - Silicon Controlled Rectifier
(a) Schematic of Semiconductor Structure
(b) Equivalent Two-Transistor Circuit
(c) Circuit Symbol

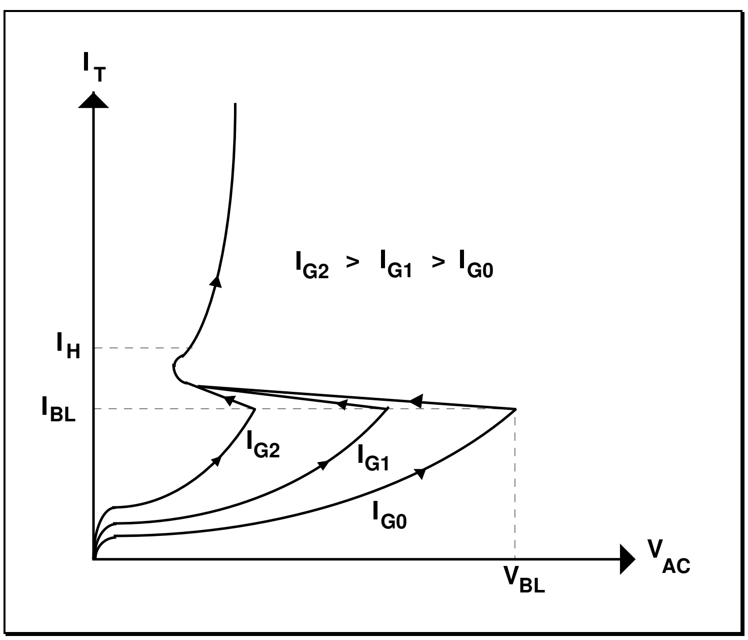


Figure 3.51 - SCR Voltage-Current Characteristic

In order to understand the SCR characteristic of Figure 3.51, first consider the curve for I_{G0}, which we can assume to be the characteristic for the situation where no current flows into the SCR gate terminal. In this curve, we can see that simply applying a forward bias from the anode to cathode does not cause a significant amount of conduction until the threshold blocking voltage, V_{BL} is reached. At this level of voltage, both transistors within the SCR become saturated and almost provide a short-circuit between the anode and cathode, whereupon the voltage across the SCR rapidly drops. The rectifier remains "on" until the current flowing through the device drops below a threshold holding level I_H, after which the device reverts back to its "off" mode and cannot be reactivated until the voltage again exceeds its threshold level.

The next stage in the understanding process is to examine what happens for increasing levels of gate current. Note from Figure 3.51 that the threshold voltage level at which the SCR begins to conduct, decreases for increasing levels of gate current. If the gate current is sufficiently high, then the SCR characteristic will resemble that of a normal diode. The net effect is that the rectifier is a diode which can be caused to conduct by either applying a sufficiently large voltage across its terminals, or by supplying a sufficient gate current and applying a low voltage across its terminals. When we trigger the operation of the SCR with a gate current, we say that the SCR is being "fired". Once the SCR has fired into conduction, it continues to do so until the current flowing through it falls below the threshold level and then reverts back to its low conduction state.

The reverse bias characteristic of the SCR has not been shown in Figure 3.51, because SCRs are normally designed to operate in their forward bias region and for all intents and purposes, the SCR is a unidirectional device. The reverse characteristic is not unlike that of any other diode, except that the reverse breakdown voltage is of a similar magnitude to the forward blocking voltage.

The actual mechanism for triggering an SCR is not directly through the application of current to the gate, but rather by the application of a rectangular voltage pulse to the gate terminal (which then causes the Q2 transistor in the SCR to saturate and the SCR to fire). A typical data sheet for an SCR device would specify the trigger level voltage ranges allowable at the gate terminal and the blocking voltage and blocking current.

In most circuits involving SCRs, the objective is to create a triggering pulse which will cause the SCR to conduct. This makes the SCR a useful device that can be controlled with digital circuits and microprocessor based control systems. However, it is interesting to note that a common parasitic problem with the SCR is that applying a rapidly changing anode to cathode voltage (V_{AC}) can also cause the SCR to fire at inappropriate times. This is referred to as dv/dt breakdown and is generally an undesirable phenomenon which is commonly overcome through a so-called "snubber" circuit.

A snubber circuit is simply a high frequency filter that consists of a resistor and capacitor in series. The snubber is connected between the anode and cathode terminals of the SCR. Since a capacitor has a very low impedance to high-frequency signals, the snubber circuit will draw current away from the SCR, thus preventing it from turning on. The series resistor is used to limit the current through the capacitor.

Most applications involving SCRs are generally concerned with switching and hence there is always a potential for unwanted voltage spikes to occur or be induced in circuits. While the snubber circuit protects the SCR suffering from dv/dt breakdown, the gate terminal also needs to be protected from unwanted noise accidentally firing the SCR. The most common approach is to simply connect a capacitor between the gate and the cathode. The low impedance of the capacitor to high frequency signals effectively means that any short-duration spikes will effectively cause a short circuit between the gate and the cathode, thereby preventing the SCR from switching on.

There are many applications to which SCRs are suited and our objective is only to examine a few that are of relevance. One major use of SCRs is in the inversion of voltages from d.c. to a.c. In Figure 3.11, we briefly examined the three-phase bridge rectifier circuit. However, we can actually replace all the standard diodes in this circuit with SCRs. If we set up the gate triggering such that the SCRs will fire whenever the anode voltage is greater than the cathode voltage, then the bridge will act as a normal three-phase rectifier. However, consider what happens if we reverse the operation of the three-phase bridge and supply a d.c. input across the terminals of the bridge and then selectively fire the SCRs. The net result is a conversion from d.c. to three-phase a.c. In other words, by having two three-phase bridges made up from SCR devices and connected via a pair of cables, we can have a d.c. transmission line system. This is shown in Figure 3.52.

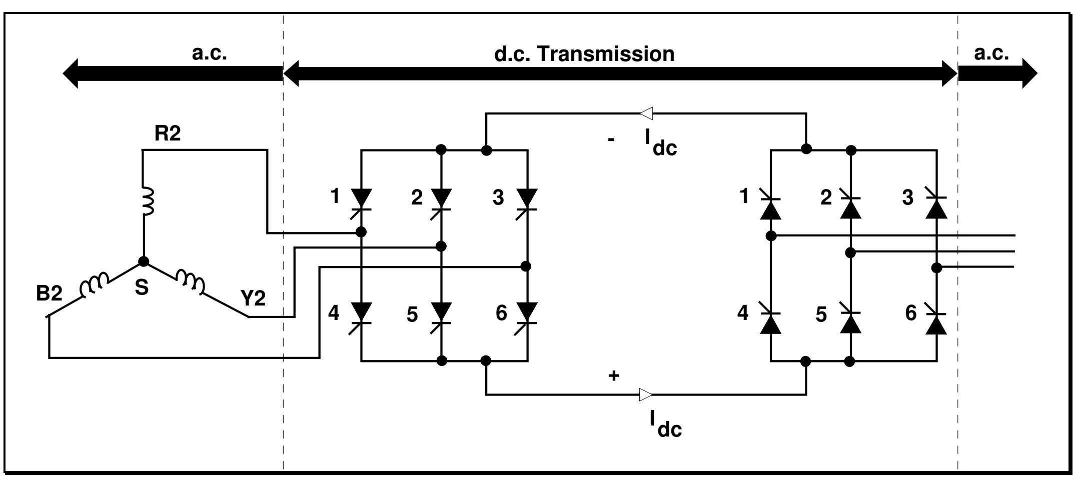


Figure 3.52 - SCRs for Inversion and d.c. Transmission Systems

One may well ask why there is a need for transmission of d.c. over long distances, but in fact, there are a number of potential benefits in terms of minimising the number of conductors and minimising transmission losses. In addition there are many systems that are based upon d.c. machines, such as tramway / trolley bus power systems and railway power systems. Since most power generation is in a.c. form, there are many benefits to be realised by transmitting at least some portion of the total energy requirements of a large city via d.c. systems such as those established with thyristors.

At a lower level, SCRs can be used in conjunction with Zener diodes in order to protect circuits from voltage spikes. This is achieved via a circuit known as a "SCR crowbar". The circuit diagram for the SCR crowbar is shown in Figure 3.53.

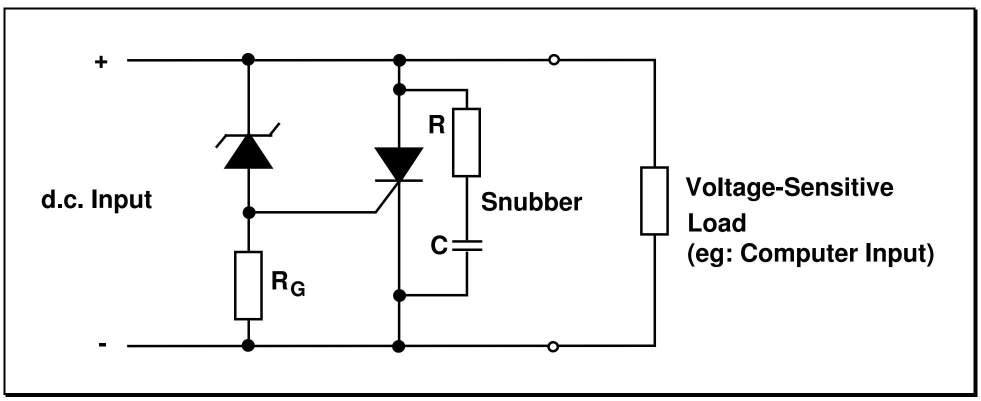


Figure 3.53 - SCR Crowbar Circuit

The operation of the circuit is really an extension of the voltage regulator design discussed in Section 3.2.3. Under normal operation, neither the Zener diode nor the SCR in Figure 3.53 are conducting and all current flows through the load. If however, a voltage spike occurs on the input side then the Zener diode goes into breakdown mode and conducts. This causes a current to flow in R_G and hence a voltage to be developed across it (thereby creating a voltage at the gate of the SCR). The SCR then turns on and acts as a by-pass for the current and reduces the voltage across the load. The SCR also has the effect of short-circuiting the device that created the voltage spike. If the device is protected by a fuse, then the fuse should blow. Another alternative is to place a "normally-closed" (ie: normally short-circuited) relay coil in series with the SCR and the switch terminals of the relay in series with the d.c. supply. When current flows in the SCR, the relay is activated and the d.c. supply is temporarily isolated from the load.

The SCR in the crowbar circuit essentially duplicates the role of the Zener diode. It is there because it is capable of withstanding a much higher current than the diode can withstand. The SCR in such circuits is expendable and really only needs to survive long enough to blow out a series fuse.

Another common application for the SCR is in controlling the average value of the voltage applied to a load. This is done with an SCR phase controller circuit such as the one shown in Figure 3.54. The load could be a simple light-globe or it could be the armature terminals of a universal motor. The input waveform is a sinusoidal a.c. voltage and the output is a d.c. waveform whose shape is determined by the firing of the SCR.

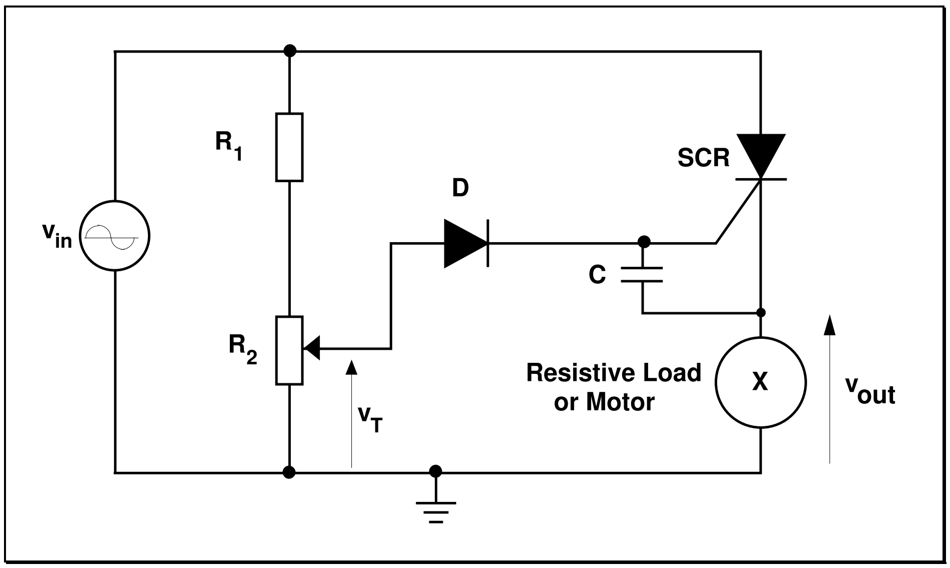


Figure 3.54 - SCR Phase Controller for Resistive Loads and Motors

It is evident, from Figure 3.54, that whenever the SCR is conducting then the output voltage will approximately equal the input voltage. When the SCR is not conducting, then the output voltage will be zero. The capacitor between the gate and the cathode of the SCR is present to prevent unwanted triggering. The SCR will only conduct when the anode to cathode voltage is positive and when the gate is triggered and so the output voltage can only be greater than zero when the input waveform is on the positive half of its cycle.

The SCR in Figure 3.54 is triggered from a voltage level derived from the sinusoidal input waveform. The voltage level at which triggering occurs can be changed by adjusting the variable resistor, R_2. This has the effect of changing v_T and hence the SCR gate voltage (which is v_T minus the diode voltage drop of 0.7 volts). If we adjust v_T (by adjusting R_2)so that the SCR is always triggered, then the output waveform will be a half-wave rectified version of the input (remember that the SCR does not allow for negative conduction). This provides the maximum possible average output voltage. Other values of R_2 provide a lower average value of output voltage. The effect is shown in Figure 3.55.

The purpose of the diode in the circuit of Figure 3.54 is to prevent the SCR from breaking down when the input voltage waveform is in the negative half of its cycle. Whenever the input voltage moves into the negative half of its cycle, the diode, D, becomes open circuit and so no voltage is applied to the gate.

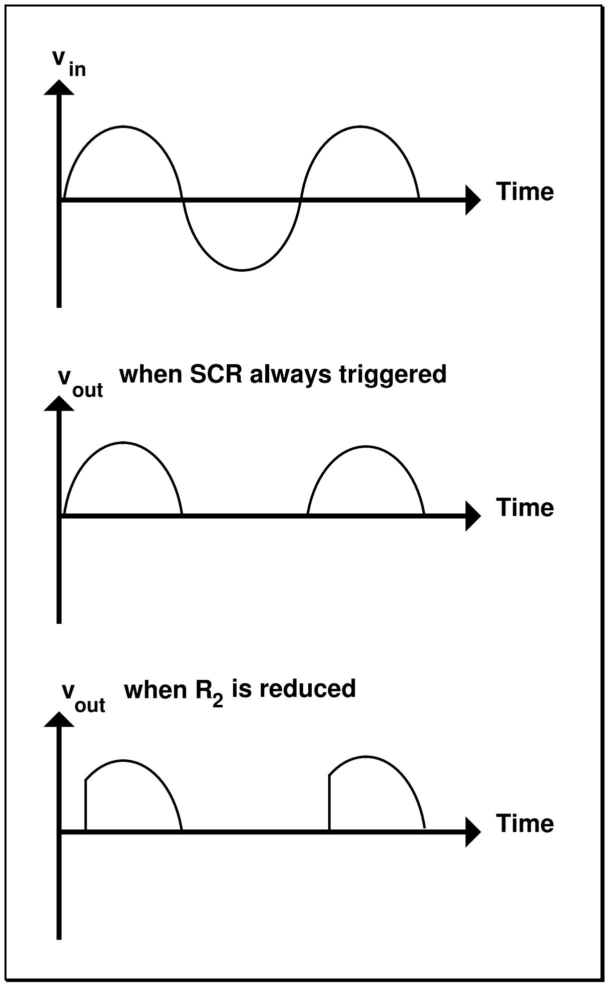


Figure 3.55 - Using a Variable Resistance to Adjust the Average Output From an SCR Phase Controller

The most common application of the circuit in Figure 3.54 is in light dimmer circuits, where the phase controller reduces the average voltage applied to the light. One may well ask why the same function cannot be achieved with a simple series resistor. The answer, of course, is that the series resistor would dissipate a great deal of energy and generate a lot of unwanted heat. The resistance values in the phase controller circuit however, can be made large, thereby conducting very little current and dissipating very little energy. The result is a much smaller light dimmer switch than can be achieved via a variable resistor.

A universal motor is essentially a d.c. motor whose armature and field terminals are both supplied from a common source. The end result is that the d.c. motor can function on either a.c. or d.c. voltages. In addition, the armature and field inductances act as a choke so that even when a time-varying voltage is applied, the currents flowing in the field and armature are relatively smooth. The SCR phase controller is therefore an ideal way of controlling the flow of energy to the motor without wasting heat in additional resistors. The SCR phase controller arrangement is used in domestic motors such those found in vacuum cleaners, power tools, etc.

3.7.3 Diacs and Triacs

In Section 3.7.2 we noted that the traditional thyristors (Silicon Unilateral Switches and Silicon Controlled Rectifiers) were unidirectional devices. That is, when triggered by either a large anode-cathode voltage or by a gate pulse, the devices would conduct current in one direction. Diacs and Triacs are, in effect, bi-directional versions of the SUS and the SCR, respectively.

A Diac is basically an npn or pnp transistor, except that unlike the BJT, the doping density in each region is identical and hence the device is bi-directional. The Diac has no gate terminal and hence it can only be triggered by applying a forward or reverse bias greater than the blocking voltage. Its forward and reverse characteristics are identical and the forward region of a Diac is similar to an SCR with zero gate current.

The Triac is a considerably more complex device than the Diac, with six doped semiconductor regions. Operationally, the Triac behaves like two complementary SCRs connected in parallel and is capable of providing conduction in both directions provided that either the magnitude of the voltage across the device is greater than the blocking voltage or the magnitude of the gate pulse voltage is greater than the triggering level. The circuit symbols for the Diac and Triac are shown in Figure 3.56.

The Diac is both physically and operationally a symmetrical device. The Triac, on the other hand, is not. In terms of operation it is best to view the Triac's Main Terminal 2 (MT 2) as being equivalent to the anode on the SCR and Main Terminal 1 (MT 1) as being equivalent to the cathode on the SCR.

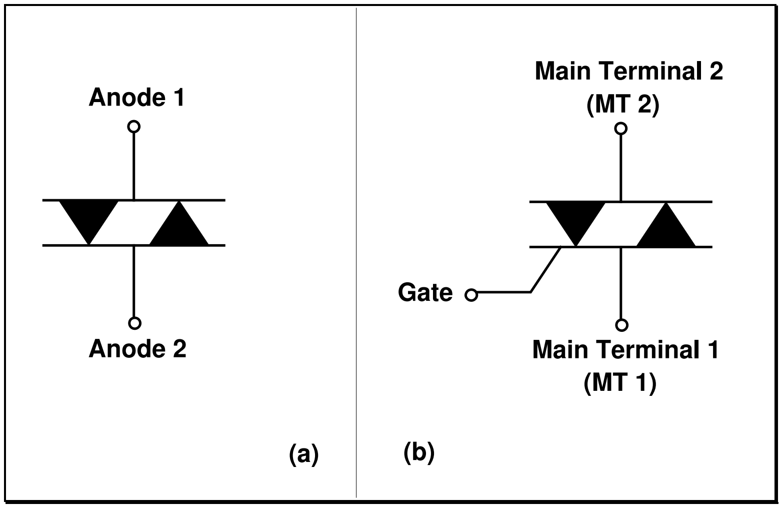


Figure 3.56 - (a) Circuit Symbol for Diac
(b) Circuit Symbol for Triac

In order to operate the Triac in its forward region (that is, identical to the SCR), we apply a voltage at MT 2 greater than MT 1 and apply a positive gate pulse. In order to operate the device in the reverse mode, we not only have to make MT 1 greater than MT 2, but we also have to apply a negative pulse to the gate terminal.

The Triac can be used for phase control of a.c. waveforms just like the SCR, except that it provides the potential for full-wave phase control, rather than the half-wave control discussed in 3.7.2.

3.7.4 Unijunction Transistors (UJTs)

The Unijunction Transistor or UJT is a device specially designed for use as a triggering mechanism for SCRs and Triacs. The construction of the UJT is not unlike the JFET and its circuit symbol is also similar. The circuit symbol for the UJT and the equivalent circuit are shown in Figure 3.57 (a) and Figure 3.57 (b) respectively.

The method of UJT operation is relatively straightforward and can be understood by examining Figure 3.57 (b). Base 1 (B1) is the output of the device and the Emitter is the input. When the voltage between the emitter and B1 is sufficiently high, the diode conducts and B1 is approximately equal to the Emitter voltage. However, by varying the voltage between B2 and B1, we increase the cathode voltage of the diode and hence the emitter voltage required to cause the diode to conduct. The base to base voltage therefore determines the threshold level at which the input is passed through to the output.

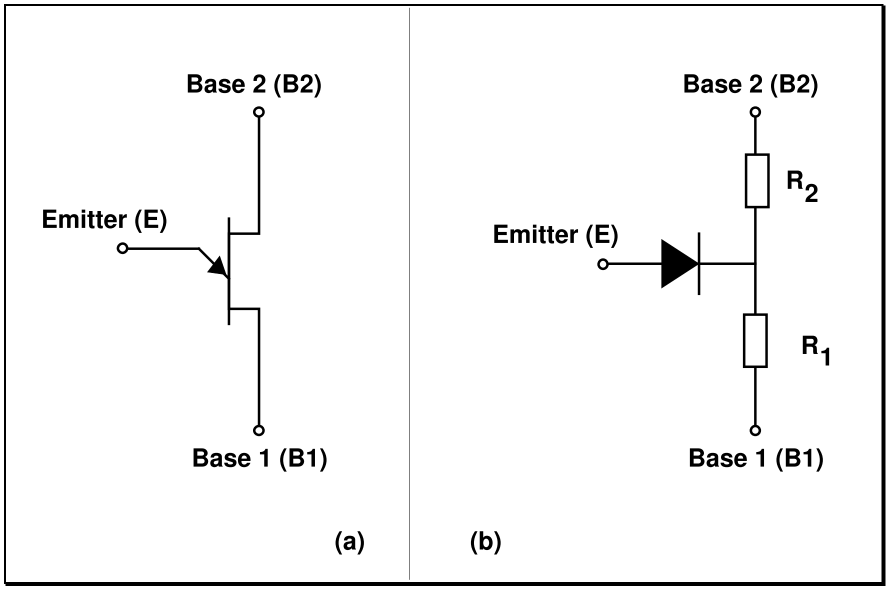


Figure 3.57 - Unijunction Transistor
(a) Circuit Symbol
(b) Equivalent Circuit

The voltage at the cathode of the diode can be determined by using voltage division:

$$V_C = V_{B2B1} \cdot \frac{R_1}{R_1 + R_2} \qquad ...(23)$$

where:

V_C is the voltage at the cathode of the diode
V_{B2B1} is the voltage between B2 and B1.

The resistance ratio, derived from voltage division in equation (23) is referred to as the intrinsic stand-off ratio "η" for the UJT and is typically in the order of 0.8.

In a typical application for the UJT, the B1 terminal of the device would be connected to the gate of the SCR. B2 would be connected to a supply rail and the Emitter would be fed with an input which is some scaled version of the supply rail voltage.

The UJT is really only designed as a mechanism for firing SCRs and Triacs and is seldom used for other purposes. Although the UJT is often associated with SCRs and Triacs, it is not in itself a thyristor.