CHAPTER 1

BASIC ELECTRONICS

CONTENTS

1.1 Introduction to Electronics

1.2 Atomic structure

1.3 Energy bands

1.4 Classification of materials

1.5 Semiconductors

1.5.1 Types of semiconductors

1.5.2 p-n junction theory

1.6 pn junction diode

1.6.1 Diode biasing

1.6.2 Diode current

1.7 Special diodes

1.7.1 Zener diode

1.7.2 Tunnel diode

1.7.3 Varactor diode

1.7.4 PIN diode

1.7.5 Schottky diode

1.8 Bipolar Junction Transistors

1.8.1 Theory and operation

1.8.2 CB,CE and CC configurations

1.8.3 Biasing of transistor

1.9 Field Effect Transistor

1.9.1 Junction field effect transistor(JFET)

1.9.2 Metal Oxide Semiconductor FET

1.10 Introduction to Heat sink

1.11 Summary

Data sheets

Feedback questions

References

CHAPTER 1 BASIC ELECTRONICS

1.1 Introduction to electronics

The word *electronics* originates from electron mechanics. Electronics deals with the study of movement of electrons under the influence of externally applied electric or magnetic field. Electronics technology experienced a revolution in 1948 with the invention of *transistor*. Transistors control the flow of electrons through solid *semiconductor* substances rather than through vacuum, and so transistor technology is aptly referred to as *solid-state* electronics. The institute of electrical and electronics engineers(IEEE)has defined electronics as "that field of science and engineering which deals with the electron devices and their utilization".

An electronic device is one in which conduction takes place by movement of electrons through vacuum,gas or a semiconductor. The semiconductor is in general a single crystal material. The flexibility of the semiconductor material lies in the fact that the conductivity can be controlled to produce effects such as amplification, rectification, oscillation, signal mixing, and switching. The advancement in the field of electronics include the wireless communication, embedded systems digital computers, real time systems and the like.

1.2 Atomic structure

To understand the fundamental concepts of semiconductors, one must apply modern physics to solid materials, more specifically in semiconductor crystals. Crystals are solid materials consisting of atoms, which are placed in a highly ordered structure called a lattice. The electrical properties of a single crystal material are determined by the arrangement of atoms in the solid. Atoms

themselves are made up of even smaller components like protons, neutrons and electrons. The part of an atom that gives an element its identity is the nucleus. It has two kinds of particles, the proton and the neutron. Protons carry positive charge and Neutrons are electrically neutral. Electrons are the particles surrounding the nucleus of an atom which have opposite electric charge from that of the protons. The figure 1.1 shows the structure of an atom.

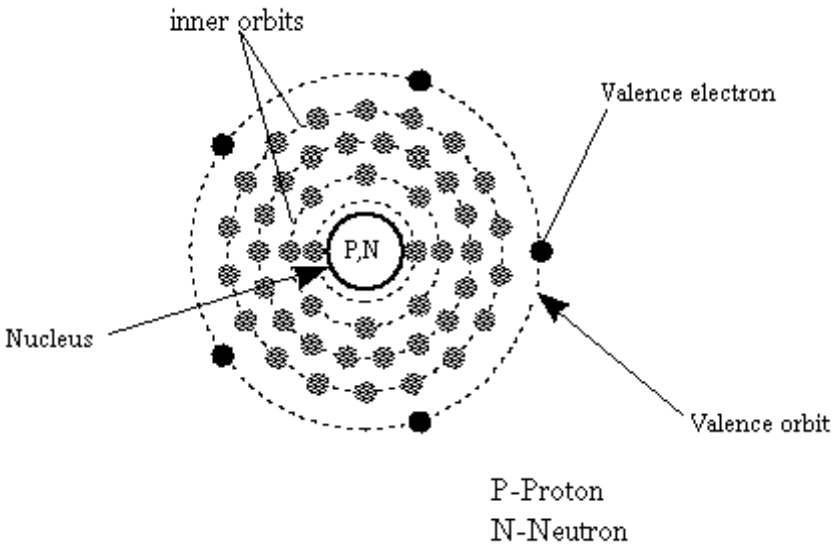


Figure 1.1 Structure of an atom

Generally, the number of electrons in an atom is the same as the number of protons. The negative charges therefore exactly cancel out the positive ones, and the atom is electrically neutral. But there can be an excess or shortage of electrons under some conditions like high levels of radiant energy, extreme heat or the presence of an electric field upsetting the balance. Atoms are the building blocks out of which all matter is constructed. Atoms bond with each other to form *substances*. Substances composed of just one type of atom are called *elements*. For example, copper, gold and silver are all elements; that is, each of them consists of only one type of atom. More complex substances are made up of more than one atom and are known as *compounds*. Water, which has both hydrogen and oxygen atoms, is such a compound. The smallest unit of a compound is a *molecule*. A water molecule, for example, contains two hydrogen atoms and one oxygen atom.

1.3 Energy bands

It has been found that most of the metals and semiconductors has crystalline structure. In the case of single crystal atoms are closely packed. In an isolated atomic structure(single atom) there are individual or discrete energy levels associated with each orbiting electron. The more distant electron from the nucleus, the higher the energy state, and any electron that has left its parent atom has a higher energy state than any electron in the atomic structure.[1]

As the atoms of the material are brought closer together to form the crystal lattice structure, there is an interaction between atoms that will result in the electrons in a particular orbit of one atom having slightly different energy levels from electrons in the same orbit of an adjoining atom. As a result of coupling between the outer shell electrons, the individual similar energy levels group together to form a *energy band*.

The range of energies possessed by valence electrons is known as ***valence band***. In semiconductors and insulators, the ***conduction band*** is the range of electron energy, higher than that of the valence band. Electrons in the conduction band are free to move within the crystal, thus the name free electrons. These free electrons constitute conduction band. The band gap energy between valance band and conduction band is called as *forbidden energy gap*. It is the minimum energy needed for the electrons to leave from the valence band and jump into the conduction band.

1The energy of an electron is measured in electron volts(eV). $1eV=1.6* 10^{-19}J$

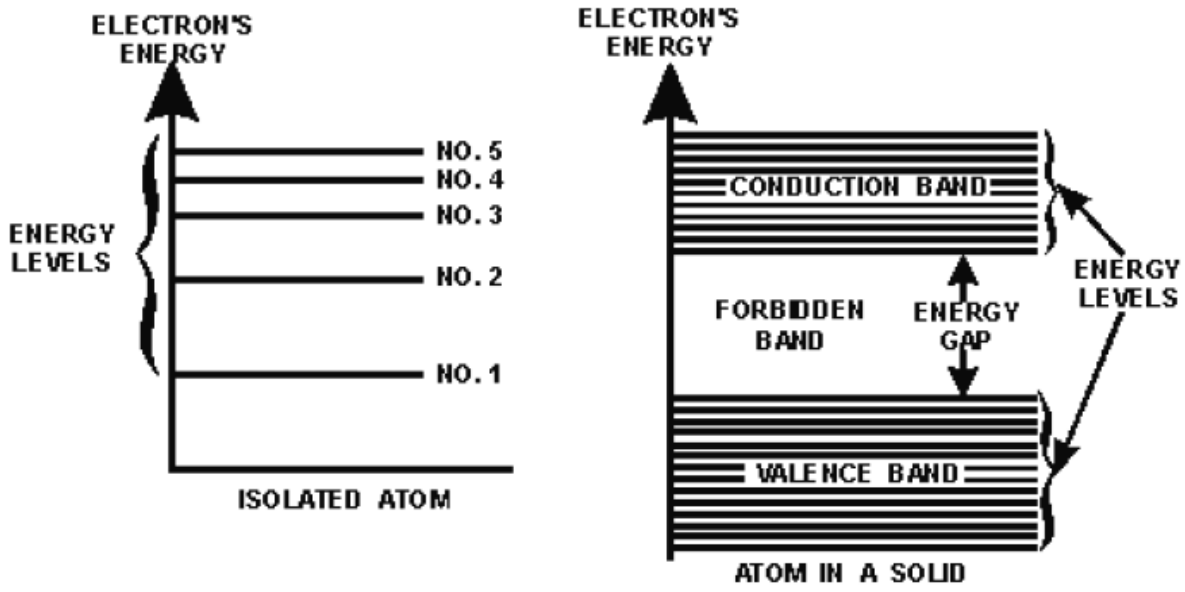


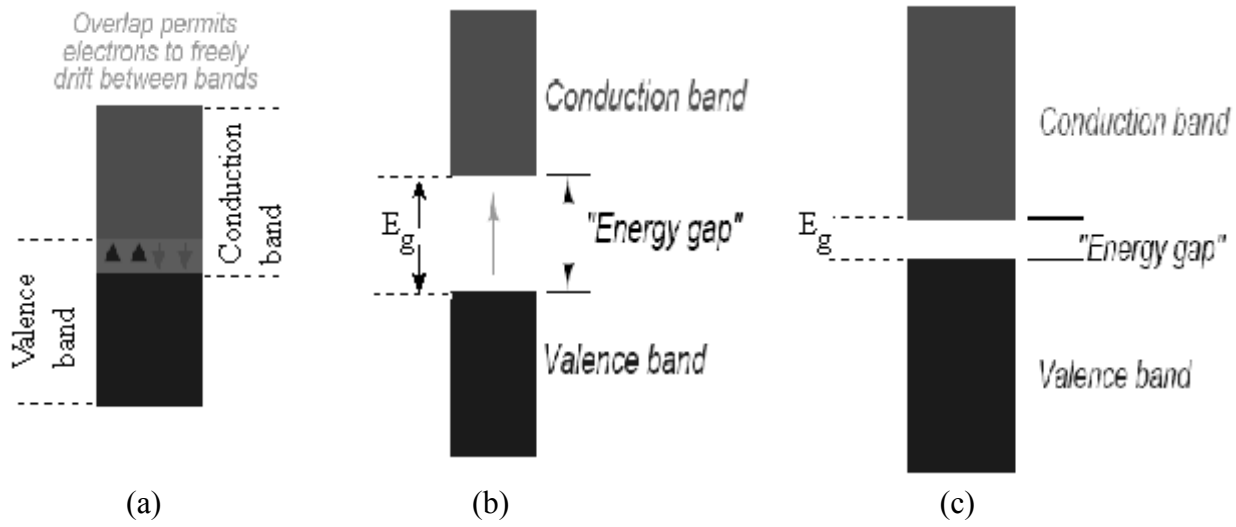
Figure1.2 Energy band formation

1.4 Classification of materials

Based on the energy gap the materials are classified as conductors, semiconductors and insulators.

Conductor is a substance in which plenty of free electrons are available for conduction without any external excitation at room temperature.[2] The outer electrons of the conductor atoms are loosely bound and free to move through the material. Metals are generally excellent electrical conductors. The energy band structure of a metal with over lapping valence and conduction bands is shown in figure 1.3(a). Examples of good metallic electrical conductors are silver,copper and aluminium.

2 Even at 0K(-273.15C) conductor has few number of free electrons.



(a) (b) (c)

Figure1.3 (a) Energy band structure of a metal,(b) Energy band structure of an insulator, (c)Energy band structure of semiconductor

An insulator is a material with extremely poor electrical conductivity. Figure1.3(b) gives the energy band structure of a typical insulator. The band gap energy E_g of an insulator is usually on the order of 3.5 to 6 electron volts(eV) or larger. At room temperature(300K,25C),these are essentially no electrons in the conduction band and the valence band remains completely full which means that electrons are so tightly bound. An insulating material is sometimes called a *dielectric.*

Semiconductor is a substance which has very few free electrons at room temperature. Its conductivity varies depending upon several factors and the range lies between those of conductors and insulators. A semiconductor has a small forbidden energy gap of about 1 electron volt. The figure 1.3(c) shows the energy band structure of semiconductor material.

The conductivity of the conductor is very high in the range of 10^3 to 10^8 and for insulator it is very low and in between 10^{-18} and10^{-8}. For the case of semiconductor the conductivity is in between insulator and conductor in the range of 10^{-8} to 10^3. It is illustrated in the figure 1.4.

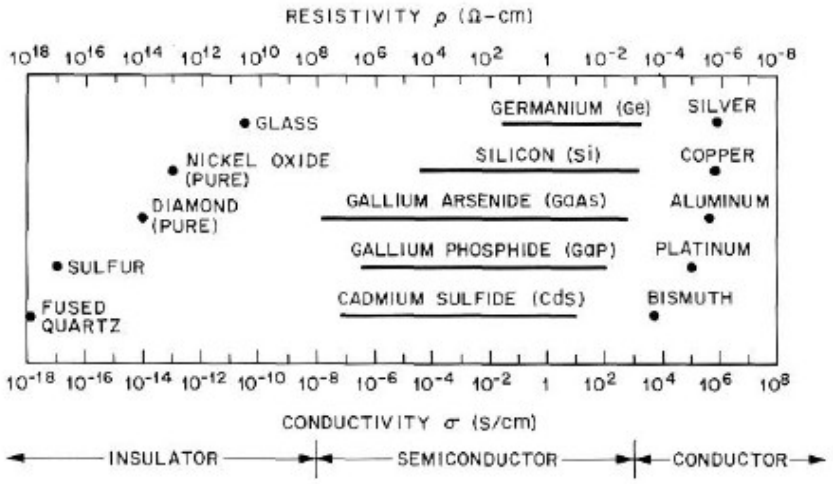


Figure 1.4 Classification of materials based on conductivity

1.5 Semiconductors

Two general classifications of semiconductor materials are the elemental semiconductor materials and the compound semiconductor materials.

Table 1.15 Elements: periodic table

Periods	*Groups*																	
	Ia	2a	3b	4b	5b	6b	7b	8b	8b	8b	1b	2b	3a	4a	5a	6a	7a	0
I	1 H				Metals									Non-metals				2 He
II	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
III	11 Na	12 Mg			Transitions								13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
IV	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
V	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
VI	55 Cs	56 Ba	57 La	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
VII	87 Fr	88 Ra	89 Ac															
		Rare earths			58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
		Actinides			90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr

Table 1.1 Periodic table

From the periodic table as shown in the table 1.1, the materials in the group IV are called the elemental semiconductor materials which are also known as tetravalent elements. The elemental materials are composed of identical atoms. The compound semiconductors are formed from special combinations of group III (trivalent) and group V(pentavalent) elements. The examples of compound semiconductor materials are aluminium phosphide(AlP), aluminium arsenide(AlAs),Gallium phosphide(GaP),Gallium arsenide(GaAs),Indium phosphide(InP) which are formed by combining group III and group V elements. Gallium arsenide is one of the most common of the compound semiconductors whose good optical properties make it useful in optical devices.

Silicon and germanium which have four* valence electrons are the widely used elemental semiconductor materials in diodes, transistors, and integrated circuits. The figures 1.5 (a) & (b) show the atomic structure of silicon and germanium.[3]

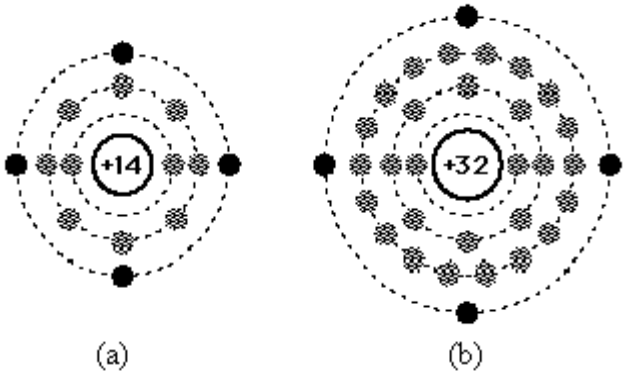


Figure 1.5 (a) & (b) atomic structure of silicon & germanium

1.5.1 Types of semiconductor

Intrinsic semiconductor

An intrinsic semiconductor is the semiconductor material in its extremely pure form. Example: germanium and silicon. When the valence electron gets sufficient external energy it enters conduction band leaving a hole in the valence band. Thus an electron-hole pair is created. In an intrinsic semiconductor even at room temperature a few electron hole pairs are created. Electron current is due to the directed movement of electrons in the conduction band whereas hole current is due to the directed movement of the holes jumping from one atom to another in the valence band. At room

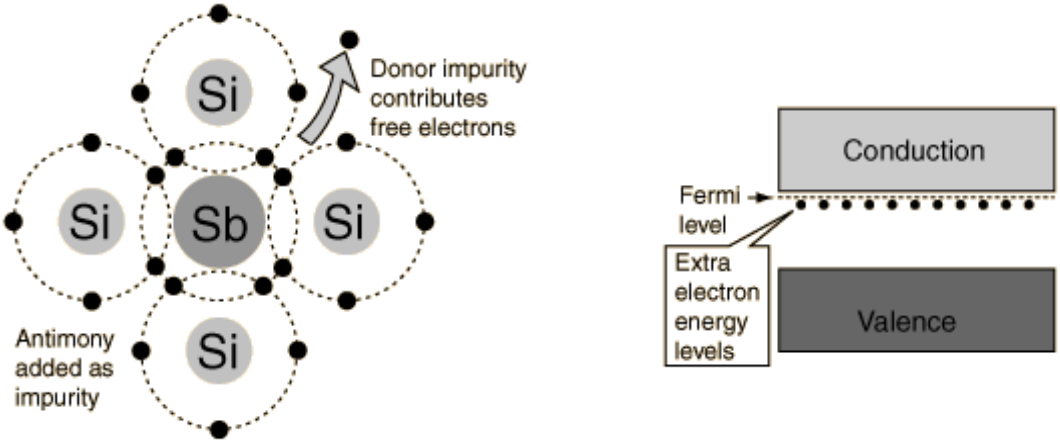
3 Elements that have less than 4 valence electrons behave as good conductors and more than 4 electrons behave as insulators.

temperature the net current flowing in the semiconductor is zero*. The net current in a semiconductor is always due to electron and hole currents.

Extrinsic semiconductor

In order to increase the electrical conductivity of the intrinsic semiconductor impurity atoms are added. This is called doping and the resultant semiconductor is known as extrinsic semiconductor. The usual dopants are pentavalent and trivalent atoms. The pentavalent dopant is known as *donor* and trivalent dopant is known as *acceptor* and the corresponding doped semiconductors are called as *n-type semiconductor* and p-*type semiconductor*. In n-type semiconductor,electrons are majority carriers and holes are minority carriers and for p-type holes are majority carriers and electrons are minority carriers.[4]

To understand the formation of n-type semiconductor, consider a pentavalent material, such as antimony, is added to pure silicon crystal. As shown in the figure 1.6(a), antimony atom forms covalent bonds with the surrounding four silicon atoms. The fifth valence electron of antimony finds no place in covalent bonds and hence it is free. Therefore,for each antimony atom added, one free electron will be added to the silicon crystal.



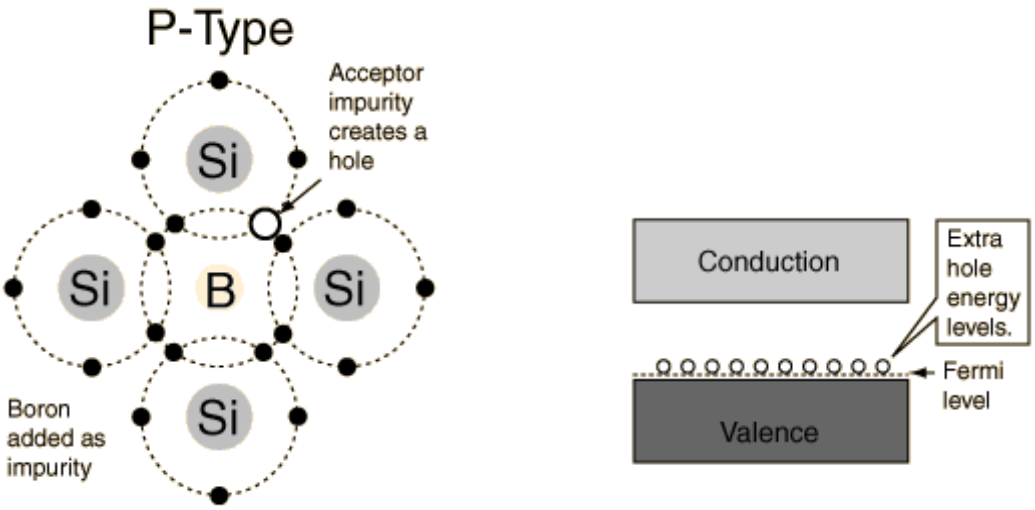
(a) (b)

Figure 1.6 n-type semiconductor

4 For silicon atom atomic number is 14. Hence it will have 14 orbital electrons and equal nos of protons in the nucleus. So it is electrically neutral. Similarly antimony whose atomic number is 51 is also electrically neutral. By doping silicon with antimony, resultant molecule will have equal number of electrons and protons thus remaining electrically neutral.

Since concentration of electrons in the conduction band is more than the concentration of holes in the valence band, *fermi level*[5] shifts nearer to the conduction band as shown in the figure1.6(b). conduction is by means of electron movement in the conduction band.

Consider a small amount of trivalent impurity, like boron, is added to silicon crystal as shown in the figure 1.7(a). Since boron has three valence electrons it forms only three covalent bonds with silicon .The fourth covalent bond needs one more electron to complete itself. This place where an electron is needed is called *hole*. Therefore, for each boron atom added, one hole is created.



(a) (b)

Figure 1.7 p-type semiconductor

Since concentration of holes in the valence band is more than the concentration of electrons in the conduction band, *fermi level* shifts nearer to the valence band as shown in the figure1.7(b). conduction is by means of hole movement at the top of valence band, acceptor level readily accepting electrons from the valence band.

5 Fermi level is defined as a measure of probability of all possible energy states being occupied under thermal equilibrium condition. In intrinsic semiconductor it lies at the centre of the forbidden energy gap.

1.5.2 p-n junction theory

When a p-type semiconductor is suitably joined to n-type semiconductor as shown in figure 1.8(a), a p-n junction is formed. The working of p-n junction forms the basis for all semiconductor devices. Note there is a high concentration of electrons in n-region than in p region. Similarly hole concentration in p-region is higher than n-region. This concentration difference establishes a flow of carriers from high concentration to low concentration near the junction even without any external excitation. This process is called as *diffusion*.

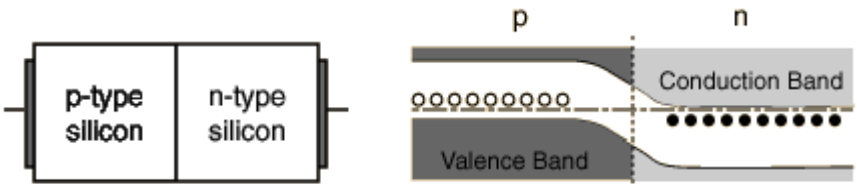


Figure 1.8 (a) p-n junction, (b) Energy band diagram

As the free electrons move across the junction from n-type to p-type, positive donor ions are left behind. Hence a positive charge is built on the n-side of the junction. At the same time, the holes cross the junction from p side to n side and uncover the negative acceptor ions. Therefore, a net negative charge is established on p-side of the junction as illustrated in the figure 1.9. The electrons and holes recombine while crossing the junction and disappear. This diffusion can not continue indefinitely because a *potential barrier* is set up against further movement of charge carriers. This potential barrier is 0.7v for silicon and 0.3v for germanium.

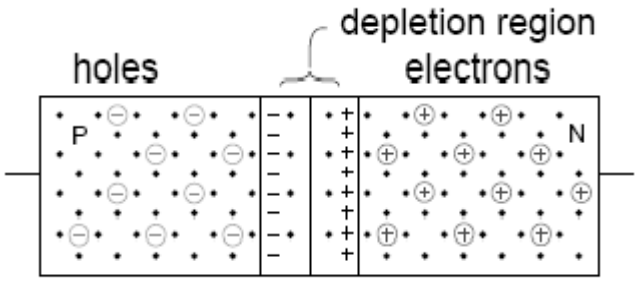


Figure 1.9 p-n junction diode

It should be noted that outside this barrier on each side of the junction, the material is still neutral. Only inside the barrier,there is positive charge on n-side and negative charge on p-side. This region is called *depletion region* or *depletion layer*. It is so named because this region is depleted of mobile carriers like electrons and holes. Its thickness is typically 1um.

1.6 p- n junction diode

The pn junction diode conducts current heavily in one direction and blocks it in the other. The terminal attached to p- type region is called as anode and that of n-type region is called as cathode. From the diode symbol shown in figure 1.11(a) the arrow head indicates the conventional direction of current flow when forward biased.

1.6.1 Diode biasing

Forward bias

Applying external voltage across the p-n junction diode in such a direction that it cancels the potential barrier, thus permitting current flow, is called *forward biasing*. To apply forward bias, connect positive terminal of the battery to p-type and negative terminal to n-type as shown in the figure 1.10.

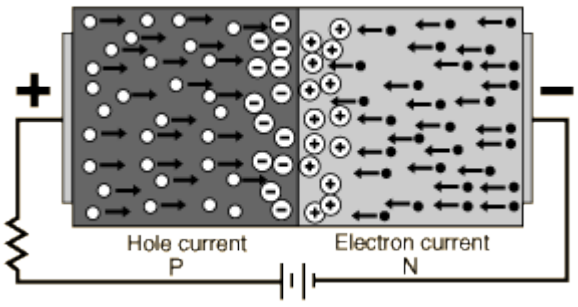


Figure 1.10 Forward biasing of p-n junction diode

Under the influence of forward voltage, both electrons and holes are driven towards the junction where they neutralise the ions thus reducing the barrier. Therefore junction resistance becomes almost zero and a low resistance path is established in the entire circuit which results in current flow in the circuit. This current is called *forward current*. As potential barrier voltage is very small a small forward voltage(0.3V for Ge and 0.7V for Si) is sufficient to completely eliminate the barrier. This voltage is known as *cut-in voltage or knee voltage.* It is illustrated in the V-I characteristic curve shown in the figure 1.11(b).

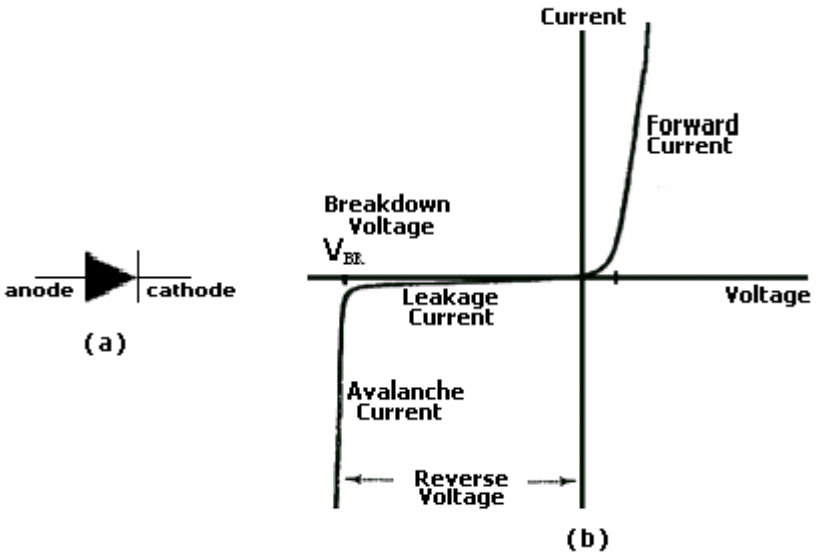


Figure 1.11 (a) Symbol, (b) V-I characteristics of diode

Reverse bias

Applying external voltage across the p-n junction diode in such a direction that potential barrier is increased is called *reverse biasing*. To apply reverse bias, connect negative terminal of the battery to p-type and positive terminal to n-type as shown in the figure 1.12.

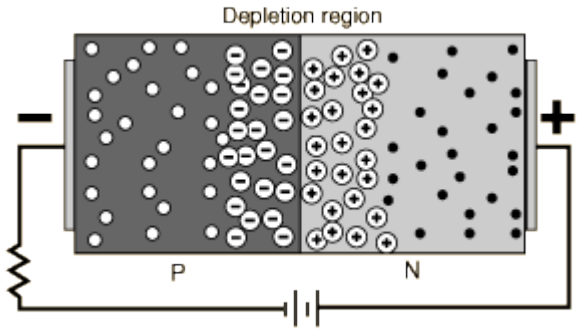


Figure 1.12 Reverse biasing of p-n junction diode

In reverse bias condition when battery voltage increases, the electrons are attracted by the positive terminal of the battery and holes are attracted by its negative terminal so that both electrons and holes move away from the junction. Therefore depletion region increases and its resistance becomes very high so that there is no current conduction. But a very small amount of current flows through the diode due to the minority carriers present in the p and n-type materials. This current is called as *reverse leakage current* and it is expressed in the equation xxxxx. It is of the order of 10^{-9} amperes for silicon and 10^{-6}amperes for germanium.

The maximum reverse-bias voltage that the diode can withstand is known as the *Peak Inverse Voltage(PIV)* also called *break down voltage* and it may be obtained from the data sheet . If the applied reverse-bias voltage becomes too high, the diode will *breakdown*, resulting in heavy current flow that damages the diode.

Diode current

A diode current equation describes the exact current through a diode, given the voltage drop across the junction, the temperature of the junction and several physical constants. This forward bias current is expressed as

$$I_f = I_0\left(e^{(qV_{appl}/nk_bT)} - 1\right)$$

where

I_o reverse saturation current in amps (typically 1 x 10^{-12} amps) and expressed as

$$I_0 = n_i^{\,2} qA\left(\frac{1}{N_A}\sqrt{\frac{D_p}{\tau_p}} + \frac{1}{N_D}\sqrt{\frac{D_n}{\tau_n}}\right)$$

A – junction area cm^2

D_p – Diffusivity of holes cm2 / sec
D_n – Diffusivity of electrons cm2 / sec
τ_p – hole life time sec
τ_n – electron life time sec

V_{appl} – Voltage applied across diode in volts

η – "Non ideality" or "emission" coefficient (typically between 1 and 2)

k_b – Boltzmann's constant (1.38 x 10^{-23}) joules/Kelvin

T – Junction temperature in Kelvins

q – charge of electron (1.6 x 10^{-19} coulombs)

The term kT/q in equation xxxx describes the voltage produced within the p-n junction due to the action of temperature, and is called the *thermal voltage(V_T)*of the junction.

Inductive flyback clamp circuit with reverse polarity protection

The term Relay generally refers to a device that provides an electrical connection between two or more points in response to the application of a control signal. The most common and widely used type of relay is the Electromechanical Relay or EMR. Relays are basically electrically operated switches that come in many shapes, sizes and power ratings suitable for all types of applications but in this section we are just concerned with the fundamental operating principles of "light duty" electromechanical relays. Such relays are used in general electrical and electronic control or switching circuits either mounted directly onto PCB boards or connected free standing and in which the load currents are normally fractions of an Ampere up to 20+ Amperes.

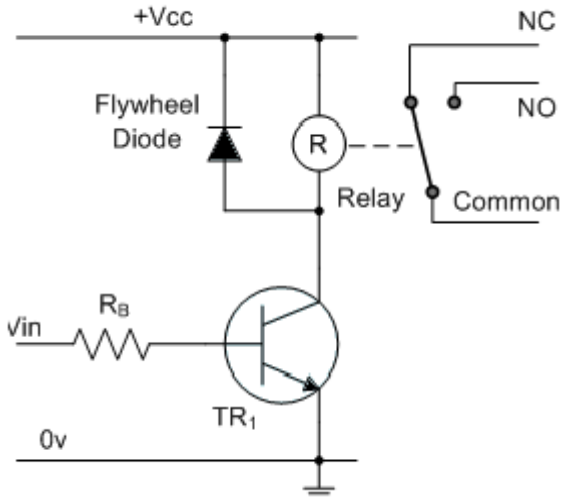


Figure 1.13 Inductive flyback clamp circuit using pn diode with reverse polarity protection

One of the more important parts of any relay is the coil. This converts electrical current into an electromagnetic flux which is used to operate the relays contacts. The main problem with relay coils is that they are "highly inductive loads" as they are made from coils of wire. Any coil of wire has an

impedance value made up of Resistance R and Inductance L in series (AC Circuit Theory). As the current flows through the coil a self induced magnetic field is generated around it. When the current in the coil is turned "OFF", a large back EMF (Electromotive Force) voltage is produced as the magnetic flux collapses within the coil (Transformer Theory). This induced reverse voltage value may be very high in comparison to the switching voltage, and may damage any semiconductor device such as a transistor, FET or microcontroller connected to the coil and used to control the relay.

One way of preventing damage to the transistor is to connect a reverse biased diode across the relay coil. When the current flowing through the coil is switched "OFF", an induced back EMF is generated as the magnetic flux collapses in the coil. This reverse voltage forward biases the diode which conducts and dissipates the stored energy preventing any damage to the semiconductor transistor.

When used in this type of application the diode is generally known as a "Flywheel Diode". Other types of inductive loads which require a flywheel diode for protection are solenoids and motors. Besides flywheel diodes, other devices used for protection include RC Snubber Networks, Metal Oxide Varistors(MOV) and Zener Diodes.

Diode as voltage regulator

The voltage dropped across a forward-biased diode changes little for large variations in diode current. Consider in a silicon diode approximately 0.7 V is maintained across the terminals while it is forward biased. This behavior makes the diode as voltage regulator.

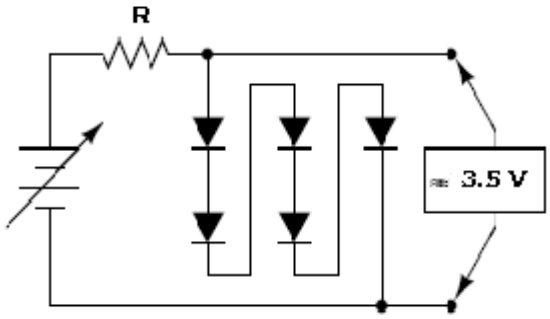


Figure 1.14 Voltage regulator circuit using pn diodes

By connecting multiple diodes in series, the voltage regulation point is increased. For example if we connect 5 diodes in series as shown in the figure 1.14, then the regulated voltage may be 5 times

0.7 or 3.5 V. So long as the battery voltage is maintained above 3.5 V, there would always be about 3.5 volts across the series of five diodes.

The following figure 1.15 shows a measurement setup to check a diode whether it is in good or bad condition before use it in a circuit by determining its forward and reverse resistance. Usually the forward resistance is in range of few ohms and the reverse resistance is in the range of mega ohms.

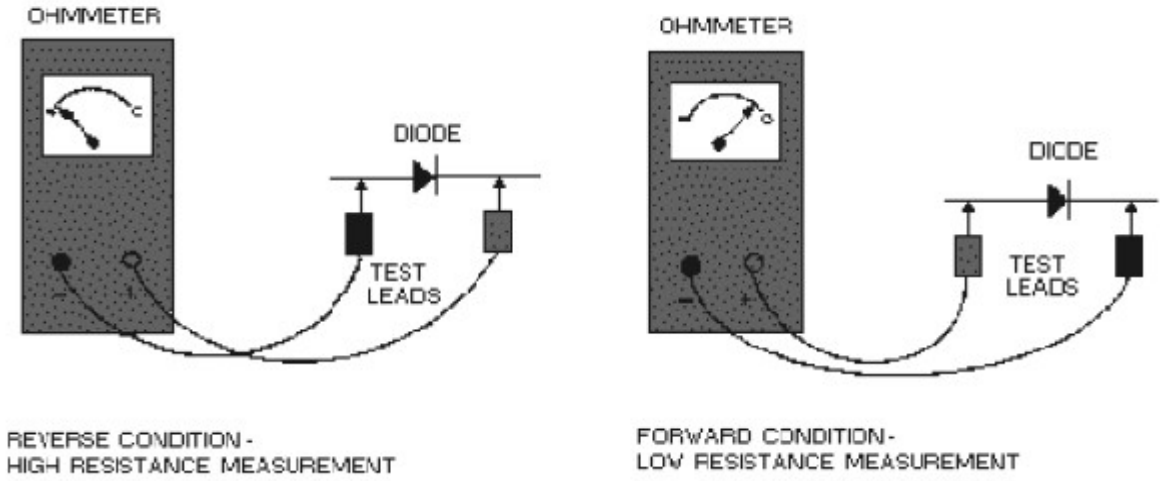


Figure 1.15 Forward and reverse resistance measurement of diode

1.7 Special diodes

1.7.1 Zener diode

A Zener Diode is a heavily doped pn junction diode which is operated in the breakdown region during reverse bias . When forward biased it is operated as normal pn junction diode. Under reverse biased condition when the voltage reaches certain value, the reverse current(Iz) through the diode increases sharply. This condition is known as zener breakdown and the voltage is known as *Zener break down voltage* or *Zener knee voltage* (Vz). After zener breakdown, the diode maintains a constant voltage (Vz) across its terminals over a specified range of current values. Because of this unique characteristic,zener diode can be typically used as a voltage regulator.

The zener voltage depends upon the amount of doping. If the diode is heavily doped ,depletion layer will be thin and consequently the breakdown of the junction will occur at a

lower reverse voltage. In other words for lightly doped diodes, zener breakdown voltage becomes high and breakdown is then predominantly by avalanche multiplication which permanently destroys the junction.

So long as the power dissipated by the reverse current(Iz) does not exceed the diode's thermal limits, the zener diode will not be harmed as illustrated in the figure 1.16(b). Zener diodes are manufactured with zener voltages ranging from 2.4V to 200V and power value ranging from 150mw to 50w.

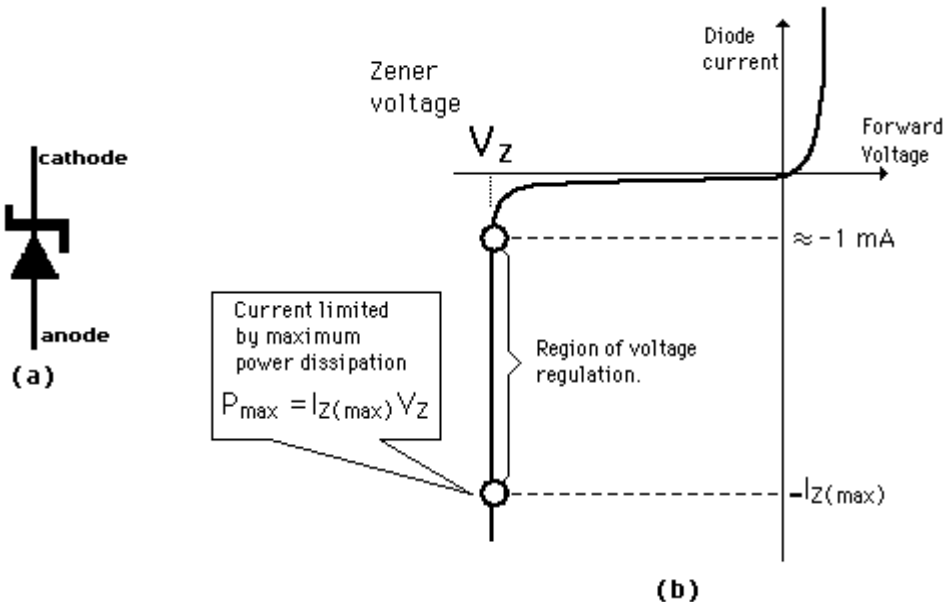


Figure 1.16 (a) Symbol, (b) V-I characteristics curve of zener diode

A simple voltage regulation circuit as shown in the figure 1.17 with the supply voltage is greater than the zener voltage. The constant reverse voltage of the zener diode makes the regulation of the output voltage against both variations in the input voltage from an unregulated power supply or variations in the load resistance.

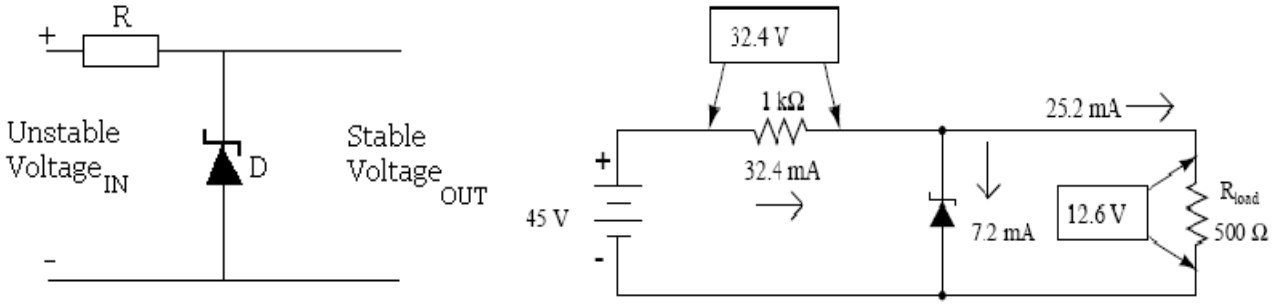


Figure 1.17 Zener Diode Voltage Regulator Circuit

The current through the zener will change to keep the voltage within the limits of the threshold of zener action and the maximum power it can dissipate.

Case:1

Assume the load current is constant. When the input voltage(unregulated) decreases the total current drawn from the supply will decrease and hence the zener current decreases. As a result the voltage across the resistor(R) decreases to maintain stable output voltage across the Rl. However the input voltage should never be lesser than the zener breakdown voltage. Similarly when the input voltage rises zener maintains a constant output voltage.

Case:2

When the input voltage is constant and greater than the zener voltage, the current flowing through the resistor R is also constant. When the load current increases the zener current decreases by the same amount.

1.7.2 Tunnel Diode

A tunnel diode is a *negative resistance* device which can be employed as an amplifier, an oscillator, or a switch. The tunnel diode is a pn junction in which both n and p regions are degenerately doped, ie the fermi energy level lies within the conduction band for n-type and within the valance band for p-type due to their high impurity concentrations. This concentration is high as compared with zener

diode and therefore even at 0K electrons will exist in the conduction band of the n-type material and holes will exist in the valence band of p-type material. Thus the tunnel diode requires smaller bias voltage than most other electronic devices.

Because of the heavy doping, a tunnel diode exhibits an unusual current-voltage characteristic curve as compared with that of an ordinary junction diode. The symbol and I-V characteristic curve is shown in the figure 1.18.

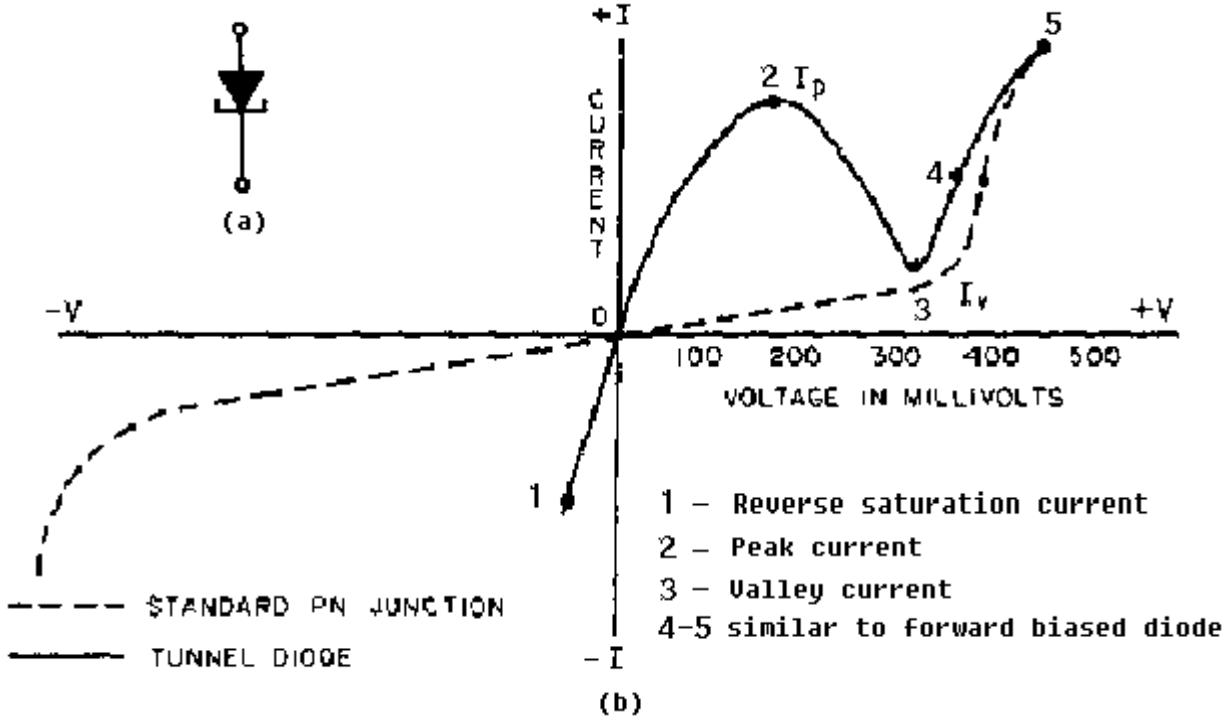


Figure 1.18 (a) Symbol, (b) V-I characteristics curve of tunnel diode

The forward current increases to a peak (I_p) with a small applied forward bias voltage V_P. Further increase in forward bias voltage causes current to decrease to a minimum value, called valley current (I_v) as illustrated in the characteristic curve. The region between I_p and I_V where current falls as voltage rises is called the *negative resistance region,* beyond which the current starts increasing again similar to an ordinary pn junction diode. This diode is normally used as a switch in high frequency applications because of its fast response due to the narrow depletion region.

1.7.3 Varactor diode

Varactor diode is a two-terminal electronic device designed to provide voltage-controlled capacitance when operated under reverse bias. Varactor diodes are also called as *varicap or VVC(Voltage Variable Capacitance) or tuning diodes.* The symbol of varactor diode is shown in the figure 1.19.

Figure 1.19 Symbol of varactor diode

When the junction reverse-bias is increased the depletion region widens. It increases the effective distance(depletion width) between the conducting p and n layers and so the effective capacitance decreases. The value of this junction or transition capacitance is determined using the following equation

Transition capacitance $C_T = \varepsilon_0 \varepsilon_r A / W_d$

Where ε_r is the relative permittivity of the semiconductor material, **A** is the p-n junction area and $\mathbf{W_d}$ is the depletion width.

By adjusting the doping gradient and applied reverse voltage, the junction width and eventually the capacitance range can be controlled. Performance specifications for varactor diodes include reverse voltage, reverse current, capacitance ratio, terminal capacitance, junction operating temperature, and power dissipation. The variable capacitance property of the varactor allows it to be used as parametric amplifiers that produce much lower internal noise levels at high frequencies than circuits that depend upon resistance properties. Since noise is of primary concern in receivers, circuits using varactors are an important development in the field of low-noise amplification.

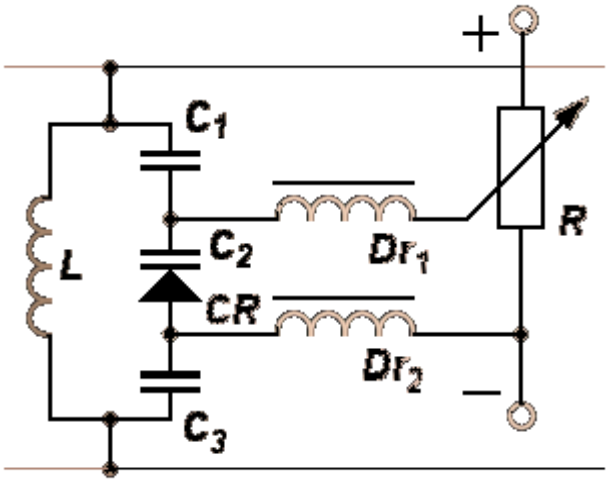


Figure 1.20 Varactor tuned resonant circuit

The figure 1.20 shows that a varactor tuned resonant circuit. Here the combination of L, C1, varactor diode(C2) and C3 form a tank circuit. Since L, C1 and C3 are fixed, the resonance frequency of oscillation can be changed by varying varactor diode capacitance C2. Its capacitance can be varied by changing the voltage applied across it through a variable resistor R, Dr1 and Dr2.

1.7.4. PIN diode

A PIN diode is a diode with a wide intrinsic semiconductor region between heavily doped p-type and n-type semiconductor regions. The figure 1.21 shows the structure of the PIN diode. When reverse biased, it acts like an almost constant capacitance and when forward biased it behaves as a variable resistor. Since its forward resistance can be changed by varying the bias, it can be used as a modulating device for AC signals. It is used in microwave switching applications. The same semiconductor material, usually silicon, is used for all three areas. Silicon is used most often for its power-handling capability and because it provides a highly resistive intrinsic (*i*) region. The *pin* diode acts as an ordinary diode at frequencies up to about 100 megahertz, but above this frequency the operational characteristics change.

The large intrinsic region increases the transit time of electrons crossing the region. Above 100 megahertz, electrons begin to accumulate in the intrinsic region. The carrier storage in the intrinsic

region causes the diode to stop acting as a rectifier and begin acting as a variable resistance. When the bias on a *pin* diode is varied, the microwave resistance changes from a typical value of 6 kilohms under negative bias to about 5 ohms when the bias is positive. Thus when the diode is mounted across a transmission line or waveguide, the loading effect is insignificant while the diode is reverse biased, and the diode presents no interference to power flow. When the diode is forward biased, the resistance drops to approximately 5 ohms and most power is reflected. In other words, the diode acts as a switch when mounted in parallel with a transmission line or waveguide. Several diodes in parallel can switch power in excess of 150 kilowatts peak.

Figure 1.21 Structure of PIN diode

In the PIN diode, the depletion region exists almost completely within the intrinsic region. This depletion region is much larger than in a pn diode, and almost constant-size, independent of the reverse bias applied to the diode. The intrinsic region offers relatively high resistance. The advantage of this high resistance region is to decrease the capacitance offered by the PIN structure because capacitance is inversely proportional to the separation of p and n regions. So that the diode has faster response time and it can be used at high frequencies(more than 300 MHz). The another advantage is the greater electric field between the p and n regions enhances the electron hole pair generation thereby enabling PIN diode to process even very weak input signals.

When reverse biased, it acts like an almost constant capacitance and when forward biased it behaves as a variable resistor. A PIN diode obeys the standard diode equation only for very slow(low frequency) signals.

1.7.5 Schottky diode

A Schottky diode is a metal semiconductor junction diode with no depletion region .It uses a metal-semiconductor junction as a Schottky barrier instead of a semiconductor-semiconductor junction as in conventional diodes. This Schottky barrier results in both very fast switching times and

low forward voltage drop. The structure of Schottky diode and its schematic symbol are shown in the figures 1.22 (a)&(b).

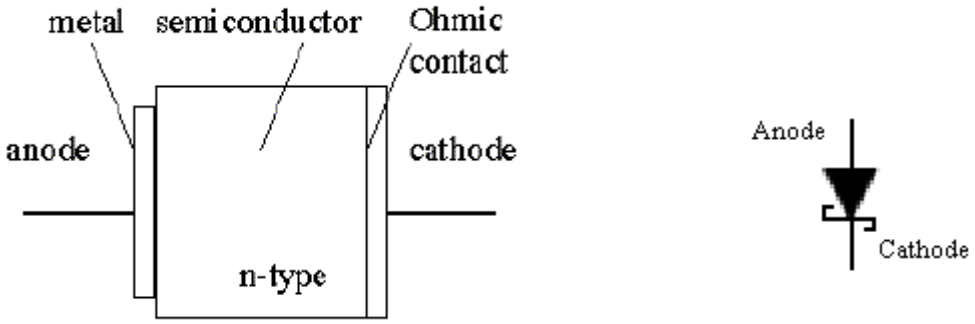


Figure 1.22 (a) Structure of schottky diode, (b) Symbol of schottky diode

When forward biased, a Schottky diode provides a low-resistance current path and, when reverse-biased, a high-resistance current path. A Schottky diode current is mainly due to flow of majority carriers. Schottky diodes often have faster switching times than p-n junction diodes and mainly used as a rectifier in the frequency range more than 300MHz.

The voltage-current characteristics of Schottky diodes are very similar to the voltage-current characteristics of p-n junction diodes as shown in the figure 1.23. During conduction, in normal pn diode the forward voltage drop between 0.3v and 1.7 v while a Schottky diode has a voltage drop between approximately 0.15v and 0.45v. This lower voltage drop causes the system efficiency to be high.

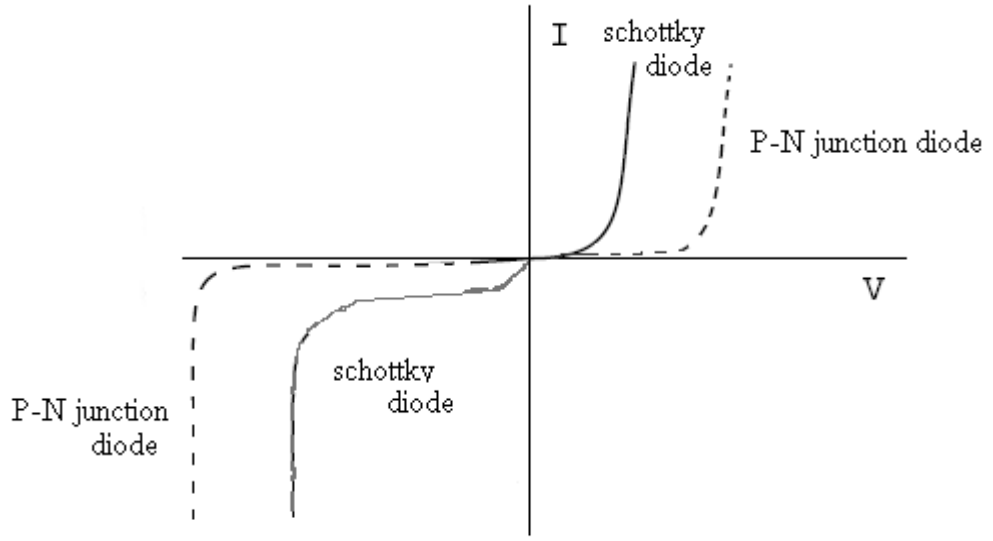


Figure 1.23 V-I curve comparison between Schottky and p-n junction diode

Schottky barrier diodes are widely used in integrated circuits in applications such as decoupling devices in digital circuits and as clamping devices to prevent heavy saturation of bipolar transistors. Also, Schottky diode is an important power device extensively used as output rectifiers in switching-mode power supplies and in other high-speed power switching applications.

1.8 Bipolar Transistors

Bipolar transistors have ability to control current flow by means of applied control signals which makes them essential elements in electrically controlled switching circuits, current and voltage-regulators, amplifiers, oscillators, and memory devices. Basically, a bipolar transistor consists of two back to back p-n junctions manufactured in a single piece of a semiconductor crystal. It has three regions named emitter, base and collector. The emitter region is heavily doped than any of the other regions because its main function is to supply majority carriers to the base and collector. The base region is very thin as compared to either the emitter or collector and is lightly doped. The collector region is to collect majority charge carriers coming from the emitter through the base. For proper working of a transistor, *emitter base junction* is always *forward biased* and *collector base junction* is always *reverse biased*. BJTs come in either *npn* or *pnp* configurations and their structures and symbols are shown in the figures 1.24(a) & 1.24(b) .

Bipolar transistors are called *bi*polar because both majority and minority carriers are involved in current flow through the transistor. Transistor functions as *current controlled device* by allowing a small base current to control a larger collector current. The amount of current allowed between collector and emitter is primarily determined by the amount of current moving between base and emitter.

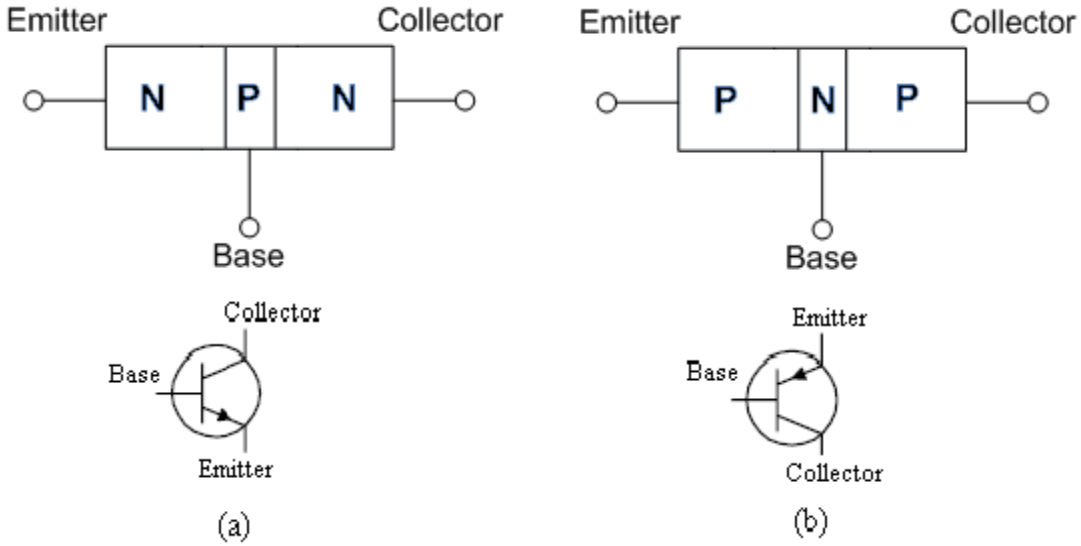


Figure 1.24 (a) npn transistor, (b) pnp transistor

1.8.1 Transistor operation

Operation of npn Transistor

When the base emitter pn junction is forward biased and base collector pn junction is reverse biased, the electrons from the emitter are injected across the B-E junction into the base as shown in the figure 1.25. These injected electrons create an excess concentration of minority carriers in the base. Since the base is very thin and lightly doped only few electrons combine with the holes and thus most of the electrons reach the collector region because of the B-C junction reverse biased.

Increasing the base voltage increases the emitter-to collector electron flow. It is to be noted that conventional current representation is in the opposite direction to the electron flow.

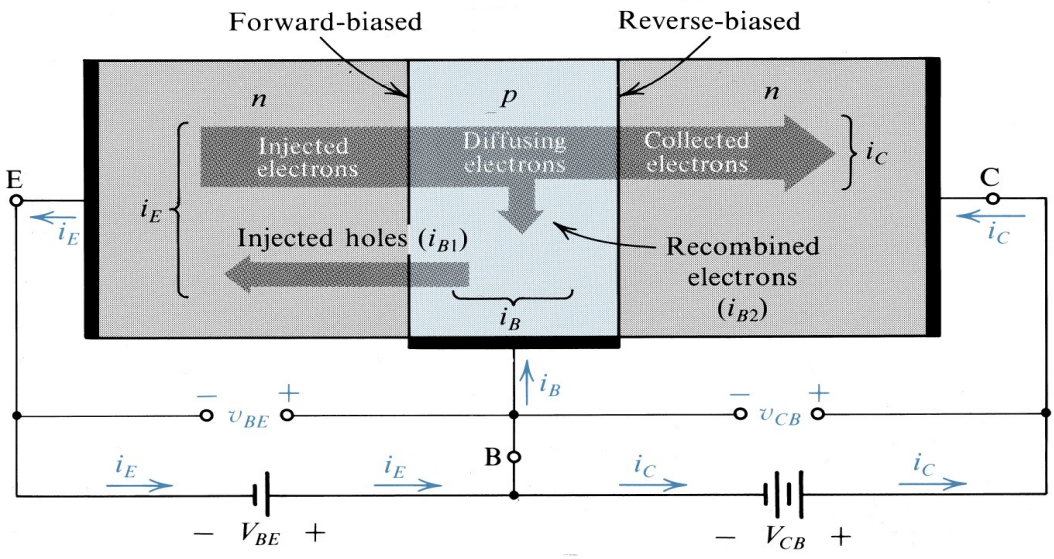


Figure 1.25 Illustration of current flow in npn Transistor

Where

I_B – Base current I_E – Emitter current I_C – Collector current

VBE – Base emitter voltage VCB – Collector base voltage

Collector Current has two parts:

(i) $I_C = \alpha . I_E$ (as $\alpha = I_C/I_E$); this is the part of emitter current reaching at the collector.

(ii) As the collector base junction is reverse biased some leakage current will also flow due to minority carriers, i.e. due to holes in case of npn and due to electrons in case of pnp transistor.

For an npn transistor, the voltage at the collector V_C must be greater than the voltage at the emitter V_E by at least a few tenths of a volt; otherwise, current will not flow through the collector-emitter junction, no matter what the applied voltage is at the base.

Operation of pnp Transistor

The pnp transistor behaves exactly the same as an npn device, with the exception that the majority charge carriers are holes. When the pnp transistor emitter base junction is forward biased and collector base junction is reverse biased, holes are emitted from the p-type emitter across the forward biased EB(Emitter Base) junction into the base. In the lightly doped n-type base(thin region), the holes find few electrons to recombine with. Some of the holes flow out through the base terminal, but most are drawn across the reverse biased junction in to the collector.

For pnp transistors, the emitter voltage must be greater than the collector voltage by at least a few tenths of a volt.

There are three transistor configurations based on keeping one transistor terminal as common.

· Common base (CB) configuration

· Common emitter (CE) configuration

· Common collector (CC) configuration

Common Base configuration

In CB configuration, the base connection is common to both the input voltage and the output voltage. The figure 1.26 shows the common base configuration of npn transistor.

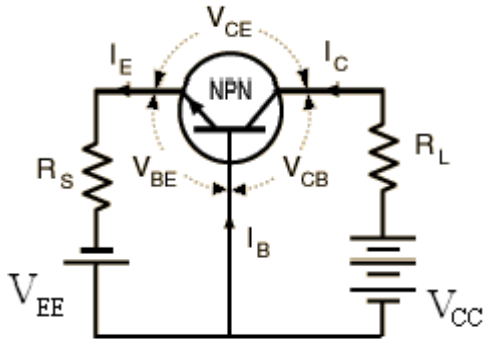
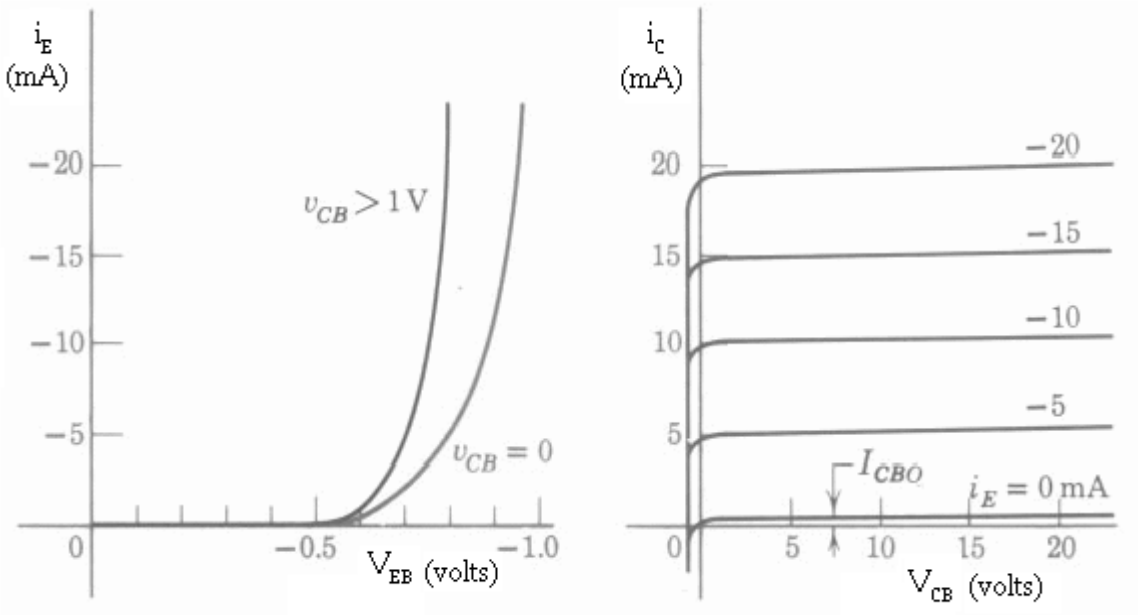


Figure 1.26 Common base configuration

The input characteristic relates the variation in the current I_E with respect to V_{BE} when

voltage V_{CB} is kept constant. Voltage V_{EE} is increased in a number of discrete steps and corresponding values of I_E are noted. Then the graph is plotted between I_E and V_{BE} as shown in the figure 1.27(a). This characteristic may be used to determine the input resistance of the transistor. The input resistance Rin is expressed as $Rin=\Delta V_{BE}/\Delta I_E$ (V_{CB} is constant)

From the graph it is seen that, the E-B junction of the transistor starts conducting when the voltage Vbe is >0.3V for Ge and 0.7>for Si.



(a) (b)

Figure 1.27 (a) Input characteristics, (b) Output characteristics

To obtain the output characteristics, V_{CB} is adjusted in convenient steps, and the corresponding levels of collector current I_C is measured by maintaining constant input current I_E. In this way, a table of values is obtained for various I_E levels and then output characteristic curve is drawn for those values as shown in the figure 1.27.(b).

Current Amplification Factor (α)

The ratio of output (I_C) to the input current (I_E) in the CB common base connection is known as current amplification factor. It is represented by **α**.

$$\alpha = I_C / I_E \text{ or } \Delta I_C / \Delta I_E$$

The input current flowing into the emitter is quite large as it is the sum of both the base current and collector current. Therefore, the current gain for CB circuit of less than 1 or in other words it attenuates the input signal current.

This type of amplifier configuration is a non-inverting voltage amplifier circuit, in that the signal voltages Vin and Vout are in-phase. This type of arrangement is not very common due to its unusually high voltage gain characteristics. Also this type of configuration has a low input resistance Rs and high output resistance(Rl) therefore the voltage gain (Av) is high and it is expressed as

$$A_V = \frac{I_c \times R_L}{I_e \times R_S} = \alpha \times \frac{R_L}{R_S}$$

The Common Base circuit is generally used in single stage amplifier circuits such as microphone pre-amplifier or RF radio amplifiers due to its very good high frequency response.

Common Emitter configuration The common-emitter configuration (CE) is the most frequently used configuration in practical amplifier circuits, since it provides high voltage and current gain. In this configuration emitter is common to input and output terminals as shown in the figure 1.28.

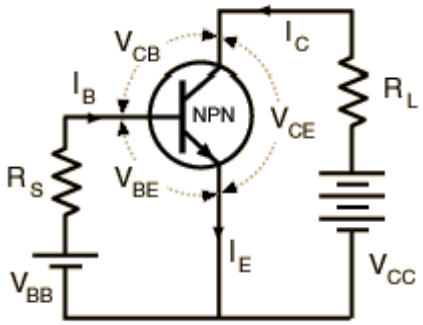
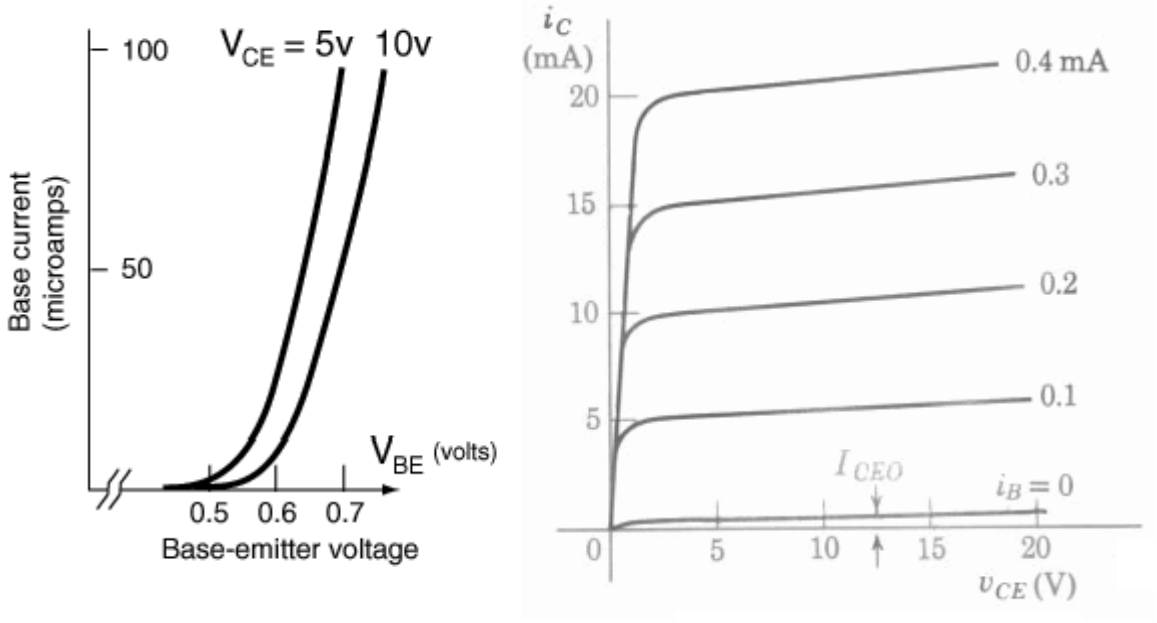


Figure 1.28 Common emitter configuration of npn transistor

Input characteristics

The input characteristics of common emitter configuration of N PN transistor shows how input current(I_B) varies with changes in V_{BE} when V_{CE} is held constant . To begin with, voltage V_{CE} is maintained constant and then V_{BE} increased in steps corresponding values of I_B are noted at each step. This procedure is then repeated by maintaining V_{CE} as constant in different values. A typical input characteristic curve is shown in the figure 1.29(a).



(a) (b)

Figure 1.29 (a) input characteristics, (b) output characteristics

From the curve, the input resistance of the transistor is obtained using the formula

$$R_{in} = \Delta V_{BE}/\Delta I_B$$

Output characteristics

To obtain the output characteristics, variation in I_C is noted by changing the values of V_{CE} for a constant I_B. This procedure is repeated for various values of I_B. From the output characteristics curve shown in the figure 1.29(b) *the current amplification factor(β)* is obtained as

$$\beta=\Delta I_C/\Delta I_B$$

In this type of configuration since I_B is constant any change in emitter current flowing into the transistor must be equal to the change in collector current flowing out of the transistor as the emitter current is given as $I_E = I_C + I_B$. The current gain(β) of the C E configuration is quite large as it is the ratio of I_C/I_B. Thus small changes in base current will control the current in the emitter/collector circuit.

By combining the expressions for both α and β the mathematical relationship between these parameters can be given as:

$$I_E = I_C + I_B$$

$$\alpha = \frac{I_C}{I_E} \quad \text{and} \quad \beta = \frac{I_C}{I_B}$$

$$\alpha = \frac{\beta}{\beta + 1} \qquad \beta = \frac{\alpha}{1 - \alpha}$$

Because of its high power gain CE configuration is used in middle stages of the multistage amplifier.

Common Collector configuration

In common collector configuration collector terminal is common to both base and emitter as shown in the figure 1.30. This configuration is primarily used for impedance matching purposes since it has high input impedance and low output impedance.

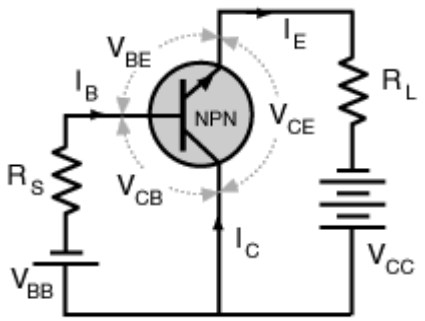
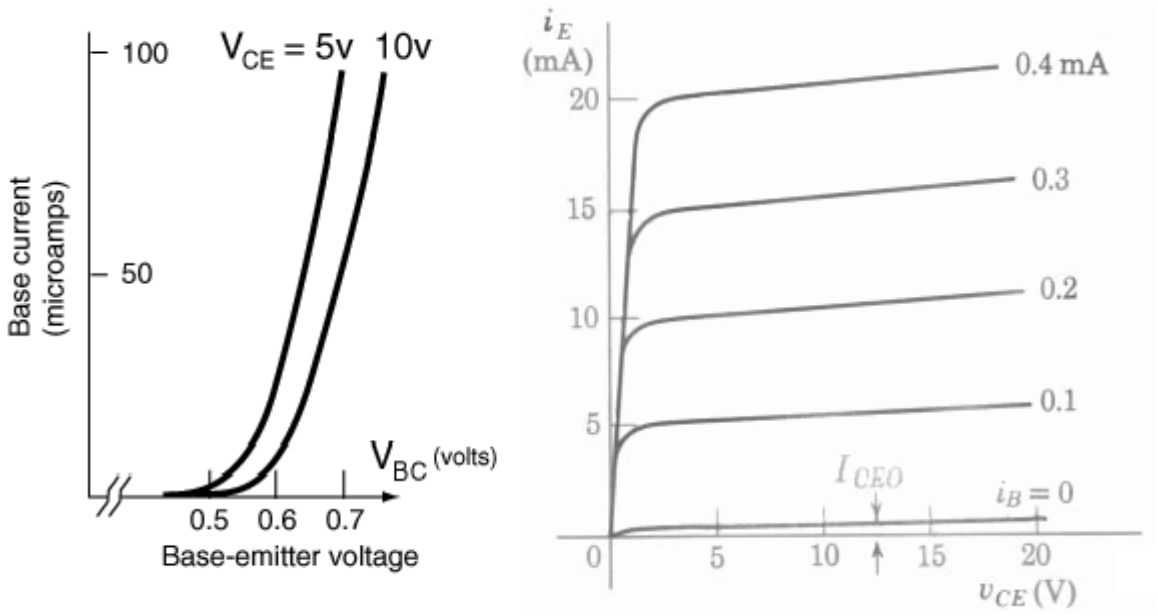


Figure 1.30 common collector configuration of npn transistor

The input characteristics of common collector configuration is plotted between I_B and V_{CB} for fixed values of V_{CE} and the output characteristics is plotted between I_E and V_{CE} for fixed values of I_B. The input and output characteristics of common collector configuration is same as the common emitter configuration and it is illustrated in the figure 1.31.



(a) (b)

Figure 1.31 (a) input characteristic, (b) output characteristic

The various parameters of CB,CE and CC configurations are compared and listed below.

AMPLIFIER TYPE	COMMON BASE	COMMON EMITTER (Emitter Resistor)	COMMON COLLECTOR (Emitter Follower)
INPUT/OUTPUT PHASE RELATIONSHIP	0°	180°	0°
VOLTAGE GAIN	HIGH	MODERATE	LOW
CURRENT GAIN	LOW α	MODERATE β	HIGH
POWER GAIN	LOW	HIGH	MODERATE
INPUT RESISTANCE	LOW	MODERATE	HIGH
OUTPUT RESISTANCE	HIGH	MODERATE	LOW

Table 1.2 Comparison of CB, CE and CC transistor configurations

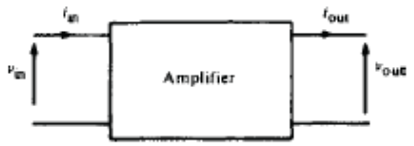
Transistor as an amplifier

The process of raising strength of a weak signal without any change in its general shape is known as *faithful amplification*. To achieve the faithful amplification the following conditions have to be met.

- Proper zero signal collector current
- Proper base emitter voltage at any instant
- Proper collector emitter voltage at any instant

Amplifier gain

Strictly the term gain of an amplifier should be restricted to refer only to the ratio of output signal power to input signal power while the more specific terms voltage gain and current gain clearly relate input and output signals, both defined as voltages or as currents. The three expressions for power, voltage and current gain are dimensionless ratios since input and output signals are in the same units.



$$\text{(Power) gain, } A = \frac{\text{Output signal power}}{\text{Input signal power}} = \frac{p_{out}}{p_{in}}$$

$$\text{Voltage gain, } A_v = \frac{\text{Output signal voltage}}{\text{Input signal voltage}} = \frac{v_{out}}{v_{in}}$$

$$\text{Current gain, } A_I = \frac{\text{Output signal current}}{\text{Input signal current}} = \frac{i_{out}}{i_{in}}$$

There are cases where the input and output signals of amplifiers are not in the same units; for example, the input may be in the form of a current (say, from a photodiode) and the amplifier generates a proportional output voltage. In this case the gain of the amplifier is the ratio of output voltage to input current, an expression which has dimensions of resistance and is defined as the transfer resistance (or transresistance), symbol r_T. Conversely the gain of an amplifier which converts voltage to current would be specified in terms of the ratio of output current to input voltage—a transfer conductance (or transconductance) g_T.

$$\text{Transfer resistance, } r_T = \frac{\text{Output signal voltage}}{\text{Input signal current}} = \frac{v_{out}}{i_{in}}\ \Omega$$

$$\text{Transfer conductance, } g_T = \frac{\text{Output signal current}}{\text{Input signal voltage}} = \frac{i_{out}}{v_{in}}\ \text{S}$$

For example consider a circuit shown in figure 1.32 having an input signal maximum amplitude of 50mV and the collector resistance R_L of 10KΩ. Base current I_B is 50 μA and current gain of the transistor β is 100. From the equation $\beta = I_C/I_B$.

$I_C = I_B * \beta$

$= 50*10^{-6}*100$

$=5mA$

$V_C = I_C * R_L$

$= 5*10^{-3}*1*10^{3}$

$V_C = 5V$

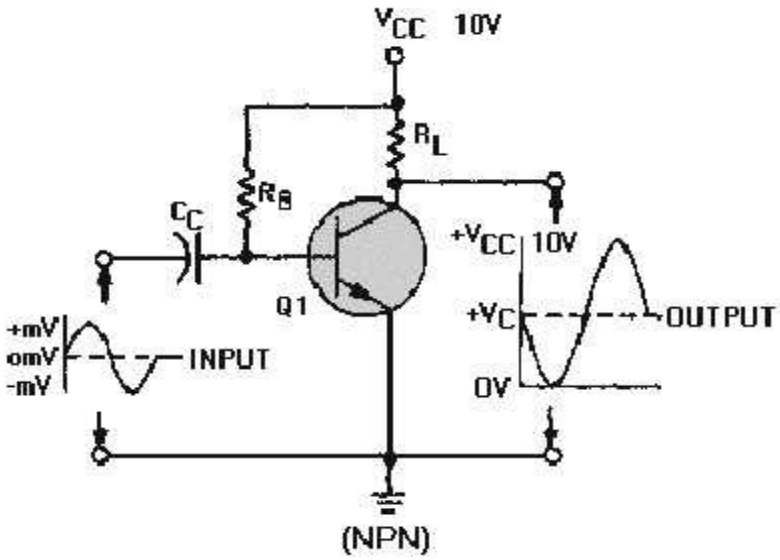


Figure 1.32 CE amplifier

Hence the 50 mV input signal is amplified to 5V. Similarly the output voltage amplitude can be increased by increasing the value of R_L. Note that the output signal is has the phase shift of 180° out of phase.

1.8.3 Biasing of transistor

Biasing is the method of establishing predetermined voltages and currents at various points of a circuit to set an appropriate operating point. The basic purpose of transistor biasing is to keep the base emitter junction properly forward biased and collector base junction properly reverse biased during the application of input signal.

Transistors are used as amplifiers,oscillators and switching circuits. Amplification is the process of increasing the magnitude of the given input signal. In an amplifier the transistor is required to work as a linear device in which the output voltage varies as a linear function of input voltage. Linear operation is obtained by selecting a proper zero signal operating point typically in the middle of active region. Thus selection of zero signal operating point is of great importance and it is done by using suitable biasing arrangement i.e. by applying proper dc voltages at emitter to base junction and collector to base junction. The operating point of a transistor also known as bias point or Quiescent point or Q-point is a DC voltage that causes it to operate in a certain desired region. It is illustrated in the figure 1.34.

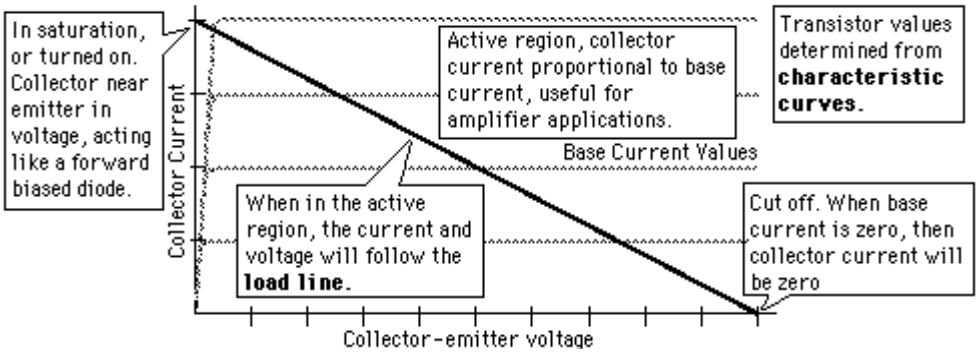


Figure 1.33 Various regions in a load line of a CE circuit

The figure 1.33 shows saturation, active and cutoff regions of a transistor. In saturation acts as a forward biased diode. In cutoff the transistor act as a reverse biased diode. In the active region the transistor will act as a properly controlled device.

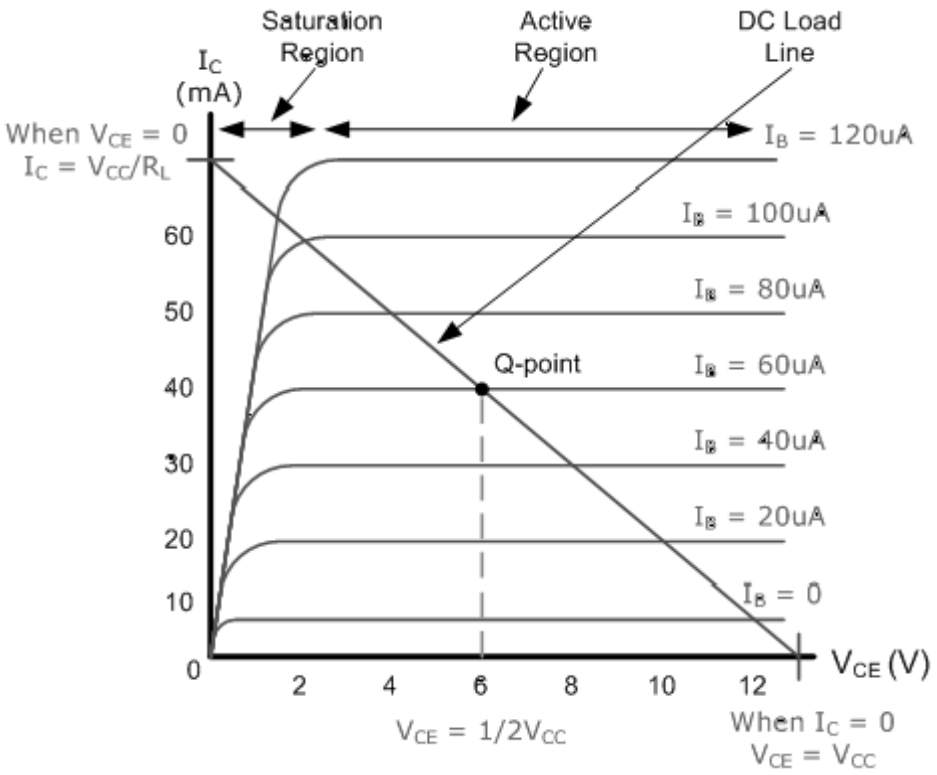


Figure 1.34 Load line of a CE circuit

The operation of transistor is linear when the operation is restricted to the active region of its characteristics. An improperly selected operating point results in output signal distortion. The operating point stability may be influenced by the *change of β on transistor replacement* and *thermal variation*. When temperature increases the collector reverse saturation current I_{CO} increases. This increase in I_{CO} causes collector current I_C to increase, which in turn increases the junction temperature. This may considerably shift the operating point.

Types of bias circuit

1. Fixed bias
2. Fixed bias with emitter resistor
3. Voltage divider bias

Fixed bias

The figure 1.35 shows the fixed bias arrangement for a transistor with common emitter configuration. Here single power source is connected to both collector and base of the transistor although separate sources can also be used.

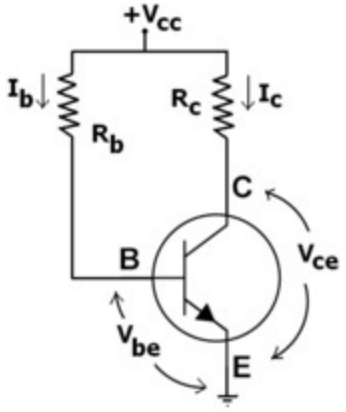


Figure 1.35 Fixed bias arrangement

Applying kirchhoff's voltage law for the base emitter loop

$V_{CC} - I_B R_B - V_{BE} = 0$

$V_{CC} = I_B R_B + V_{BE}$

Therefore,

$I_B = (V_{CC} - V_{BE})/R_B$

As V_{CC} is of fixed value, on selection of R_B, the base current I_B is fixed. Therefore this type of bias is called fixed biasing .

Also applying kirchhoff's voltage law for collector emitter loop

$V_{CC} - I_C R_C - V_{CE} = 0$

$V_{CC} = I_C R_C + V_{CE}$

Therefore,

$V_{CE} = V_{CC} - I_C R_C$

Since $I_C = \beta I_B$, we can obtain V_{CE} from the above equation. In this manner, operating point (V_{CE},I_C) can be set for a given transistor using the fixed bias method.

Demerits

1. The collector current does not remain constant against variation in temperature. Therefore the operating point is unstable in fixed bias.
2. When the transistor is replaced with another one, slight (if not large) change in the value of β will shift the operating point.

Usage

Due to the above inherent drawbacks, fixed bias is rarely used in linear circuits, ie. those circuits which use the transistor as a current source. Instead it is often used in circuits where transistor is used as a switch.

Fixed bias with emitter resistor

The fixed bias circuit is modified by connecting an external resistor with the emitter terminal as shown in the figure 1.36 to improve the stability level.

Applying KVL to input loop

$I_B = (V_{CC} – V_{BE}- I_E R_E)/R_B$

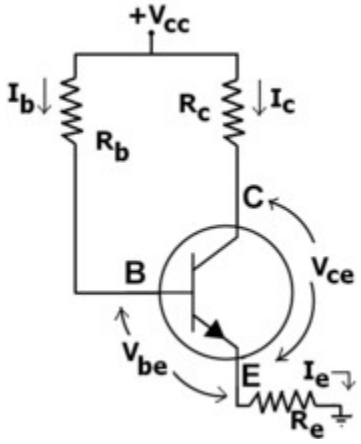


Figure 1.36 Fixed bias with emitter resistor

Applying KVL to output circuit,

$V_{CE} = V_{CC} - (R_C+R_E)I_C$ (since I_C roughly equals I_E as I_B is very small)

When the temperature increases, the leakage current increases. Therefore there is increase in I_C and I_E. This increases the emitter voltage, which reduces the voltage across the base resistor. This reduces the base current which results in less collector current. Thus collector current is not allowed to increase, and operating point is kept stable.

Similarly, if the transistor is replaced by another, there may be a change in I_C corresponding to change in β-value. By similar process as above, the change expected in I_C is negated and operating point kept stable.

Merits:

The circuit has the ability to stabilize operating point against changes in temperature and β-value.

Demerits

In this circuit, for proper functioning, the following condition must be met:

$R_E >> R_B/\beta$

As β-value is fixed for a given transistor, this relation can be satisfied either by keeping R_E very large, or making R_B very low.

- If R_E is of large value, high V_{CC} is necessary. This increases cost as well as precautions necessary while handling.
- If R_B is low, a separate low voltage supply should be used in the base circuit. Using two supplies of different voltages is impractical.
- R_E causes ac feedback which reduces the voltage gain of the amplifier.

Usage:

Due to the above disadvantages, this type of biasing circuit is generally not used.

Voltage divider bias

It is also called as potential divider bias and this method is widely used in electronic circuits. For a single stage amplifier this circuit is more flexible against changes in temperature and device characteristics. The voltage divider bias circuit is shown in the figure1.37.

Here R1 and R2 form a potential divider, which will fix the base potential of the transistor. By proper selection of resistors R_1 and R_2, the operating point of the transistor can be made independent of β. The current(I_1) through this divider circuit is usually set at 10 times greater than the base current required by the transistor. The base emitter voltage drop of the silicon transistor is approximated as 0.6 volt. There will also be a voltage drop across the emitter resistor, R_E, this is generally set to about 10% of the supply voltage. The inclusion of this resistor helps to stabilize the bias for temperature variation.

The equations for potential divider bias design are as follows:

$R_C = V_C / I_C$

$I_E = I_B + I_C$ as $I_C >> I_B$ then $I_E \sim I_C$

$V_E = 10\% * V_{CC}$

$R_E = V_E / I_E$

$V_B = V_E + 0.6$

$V_B = V_{CC} * R_2 / (R_1 + R_2)$

$R2 = V_B / 10 * I_B$

$R1 = V_{CC} - V_B / 10 * I_B$

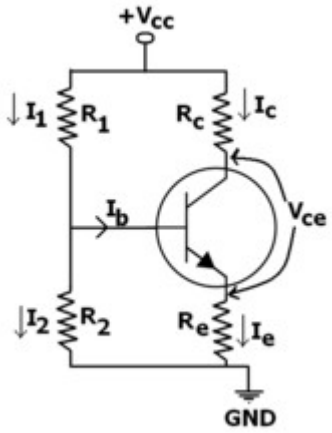


Figure 1.37 Voltage divider bias

1.9 Field Effect Transistor

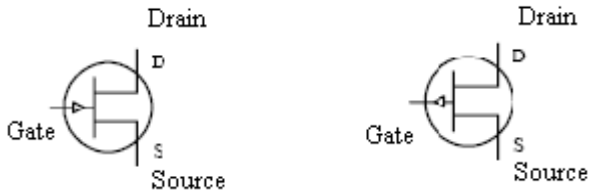
Field Effect Transistor(FET) is a *voltage-controlled device* which means that the output current of the FET is controlled by input voltage, unlike a bipolar transistor which requires input current to control the output current. FET is a *unipolar device* because the output current is mainly due to the conduction of majority carriers only.

There are two basic types of FET:

- Junction field effect transistor (JFET)
- Metal oxide semiconductor FET (MOSFET)

1.9.1 Junction Field-Effect Transistors

Junction field-effect transistors (JFETs) are used as electrically controlled switches, amplifiers and voltage-controlled resistors. JFET has three terminals such as drain,source and gate. JFETs are classified as *n-channel* or *p-channel* configurations. The symbols for both types of JFETs are shown in the figure 1.38(a) & (b).



(a) (b)

Figure 1.38 (a) n-channel JFET, (b) p-channel JFET

n-channel JFET

An n-channel JFET consists of a long channel of n-type semiconductor which contains two heavily doped small p-type regions as shown in the figure 1.39. The two p-type materials form two p-n junction with the n-channel. The one end of the channel is connected through an ohmic contact to a terminal referred to as *drain(D)*, while the another end of the channel is connected through an ohmic contact to a terminal referred as *source(S)*. The two p-regions are internally connected and a single lead is brought out which is called *gate(G)*. Since the channel is made up of n type semiconductor only majority carriers electrons flowing from source to drain contribute to current.

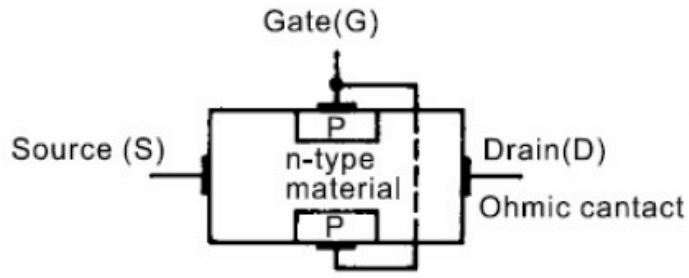


Figure 1.39 Structure of n-channel JFET

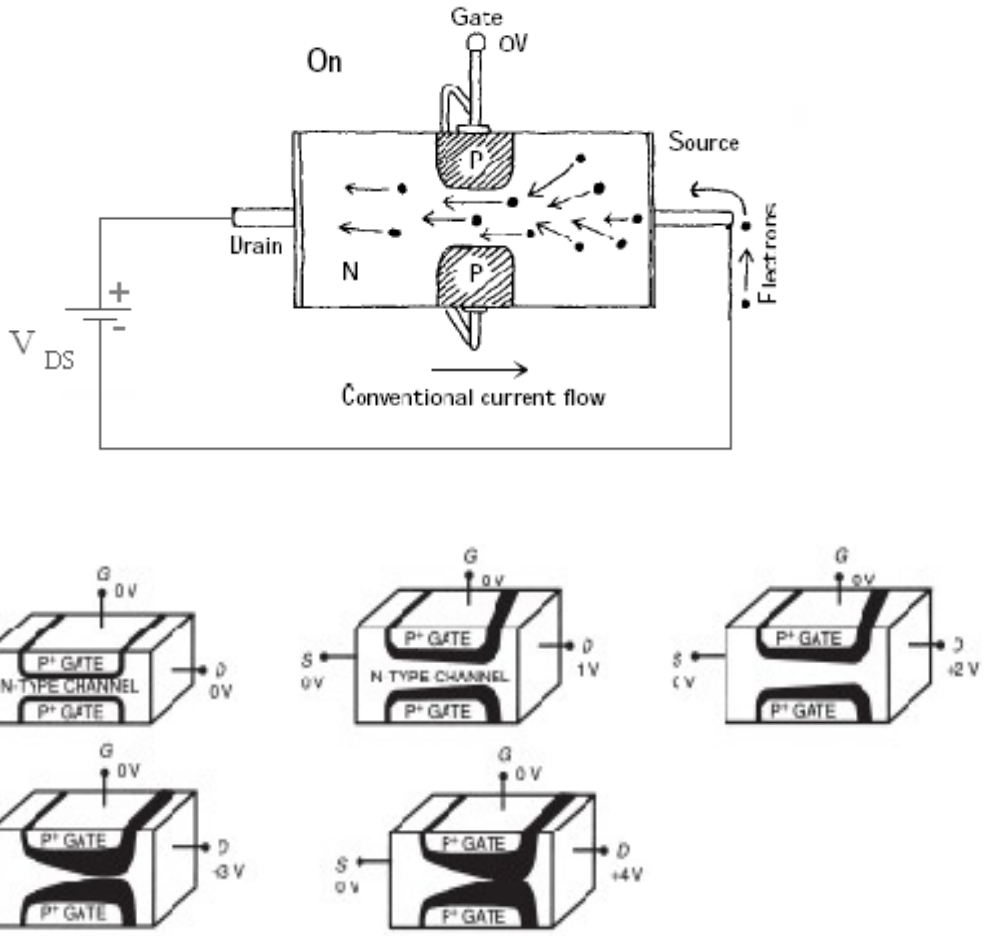


Figure 1.40 n-channel JFET with V_{GS}=0 and various V_{DS}

At V_{GS}=0 if a positive voltage(V_{DS}) is applied across the drain and source terminals, a drain current I_D flows through the channel from drain to source as shown in the figure 1.40. By applying a reverse bias voltage between gate and source, the depletion region width is increased in both the junctions and the width of the conducting channel is reduced. This means that the channel resistance is increased. Consequently the current from drain to source is decreased. Therefore the application of gate voltage(V_{GS}) controls the current flow from drain to source(I_D).

Static drain characteristics

The *drain characteristics* shown in figure 1.41 describes how the drain current (I_D) is influenced by the gate-to-source voltage (V_{GS}) and the drain-to-source voltage (V_{DS}). From the characteristics curve, when V_{DS} increases I_D increases linearly following ohm's law and JFET behaves like a ordinary *resistor.* This region is known as the *ohmic region.* When V_{DS} increases further to a certain value called the *pinch off voltage(V_P)* ,the JFET enters in to the *saturation region* also called as *pinch off region*

where the drain current becomes constant .In the saturation region JFET operates as a constant current device because when V_{DS} increases, channel resistance also increases proportionally so that I_D is maintained constant. If V_{DS} reaches the breakdown voltage value, the device enters into breakdown region where I_D increases to an excessive value.

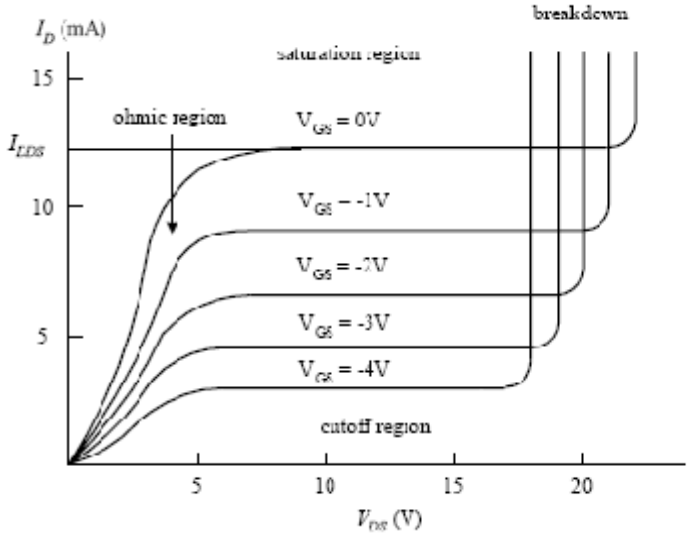


Figure 1.41 Drain characteristics

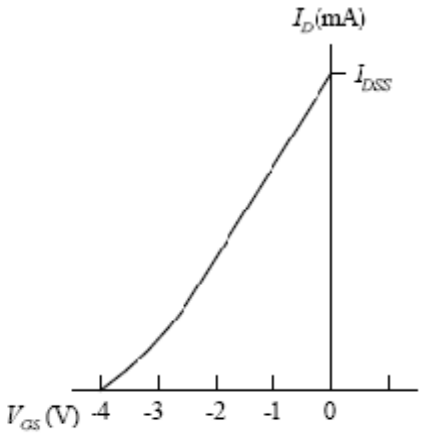


Figure 1.42 Transfer characteristics

Transfer characteristics

The figure 1.42 shows the *transfer characteristics* which gives relationship between I_D and V_{GS} for a constant V_{DS}. When V_{GS}=0V the maximum current flows in to the device and this current is known as *shorted gate drain current* (I_{DSS}) because gate is shorted with the source to apply 0V. Further increasing of negative V_{GS} ,I_D will decrease and becomes zero when V_{GS}= - V_P.

JFET Parameters

The various parameters of a JFET can be obtained from its drain and transfer characteristics. The parameters are ac drain resistance(r_d), transconductance(g_m), amplification factor(μ) and DC drain resistance(R_{DS}).

When JFET is operating in the saturation region, the ac resistance between the drain and source is called as ac drain resistance (r_d) and it is expressed as

$$r_d = \frac{\Delta V_{DS}}{\Delta I_D} \quad \text{at constant } V_{GS}$$

$$\text{Transconductance } (g_m) = \frac{\Delta I_D}{\Delta V_{GS}} \quad \text{at constant } V_{DS}$$

$$\text{Amplification factor}(\mu) = \frac{\Delta V_{DS}}{\Delta V_{GS}} \quad \text{at constant } I_D$$

dc drain resistance or Static resistance (R_{DS}) = V_{DS} / I_D

p-channel JFET

In the p-channel JFET, the channel is p-type material and the gate regions are n-type materials. The figure 1.43 shows the structure of p-channel JFET. Since the channel is p-type semiconductor only majority carriers holes contribute the drain current.

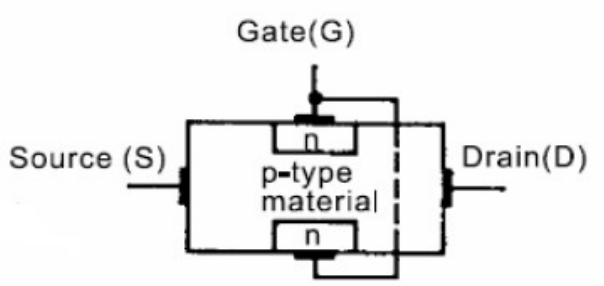


Figure 1.43 Structure of p-channel JFET

When no voltage is applied to the gate of an p-channel JFET, if the positive voltage is applied to the source with respect to drain, then maximum drain current Idss will flow through the channel. The drain characteristic of p-channel FET is illustrated in the figure 1.44.

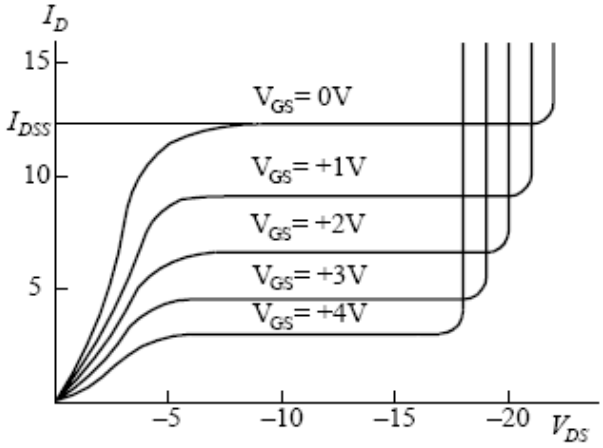


Figure 1.44 Drain characteristics

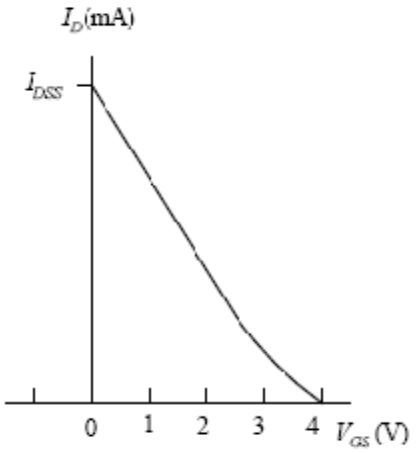


Figure 1.45 Transfer characteristics

Next for a fixed value of V_{DS} when increasing V_{GS} to a positive value the depletion region of gate-channel junction will be increased so that the width of the channel get reduced and the current(I_D) through the channel also reduced. The transfer characteristic curve shown in the figure 1.45 shows that,increase in V_{GS} to positive values cause reduction in current(I_D).

1.9.2 Metal Oxide Semiconductor Field Effect Transistor(MOSFET)

MOSFETs are used in high-speed switching applications due to their low power consumption and high reliability. The basic principle of the MOSFET is that the source to drain current is controlled by the gate voltage similar to JFET.

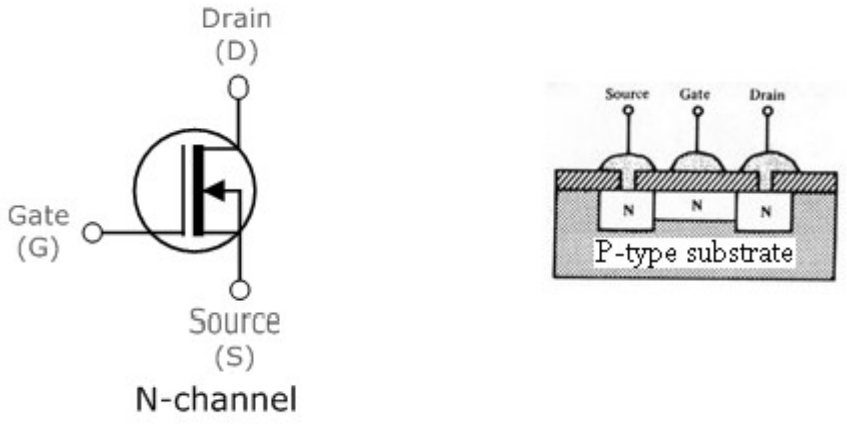
The input impedance of a MOSFET is *much greater* than that of a JFET because the gate terminal is insulated from the main current carrying channel by a thin gate oxide layer so that no current flows into the gate. Hence MOSFET is also known as *insulated gate field effect transistor*(IGFET or IGT). The two types of MOSFETs viz depletion type and enhancement type are available and each has both n-channel and p-channel versions. The depletion type is normally ON, and operates as a JFET. The enhancement type is normally OFF, which means that the drain to source current increases as the voltage at the gate increases. No current flows when no voltage is supplied at the gate.

The main difference between the depletion and the enhancement mode MOSFET is that a physical channel is present in the depletion mode MOSFET whereas a channel does not exist in enhancement mode MOSFET but will be created during its operation.

Depletion type MOSFET

n-channel depletion type MOSFET

The symbol and basic structure of n-channel depletion MOSFET are shown in the figures 1.46 (a) & (b). In the p-type substrate two blocks of heavily doped n+ material are diffused and a lightly doped n-type channel is introduced between the two n+ regions. The drain and source terminals are taken from each n-type blocks through metallic contacts. The gate is also connected to a metal contact surface but remains insulated from the n-channel by a thin layer of silicon dioxide(SiO_2) which makes no direct electrical connection between the gate and channel. In some devices the substrate is internally connected with source, in many devices a separate terminal named *ss* is taken out and externally connected with source.



(a) (b)

Figure 1.46 (a) Symbol of n-channel depletion MOSFET
(b) Structure of n-channel depletion MOSFET

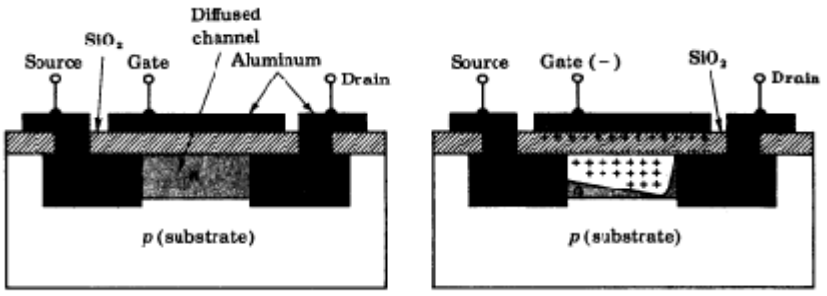
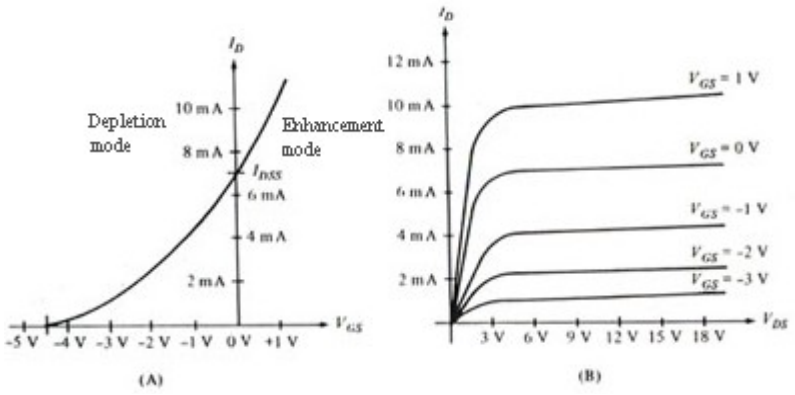


Figure 1.47 Illustration of the effect in the channel due to V_{GS}

Theory and operation

When applying 0V to V_{GS} and the drain is made positive with respect to source(V_{DS}), electrons can move freely from source to drain through the conducting channel which exists between the two n-type regions. When negative voltage is given to V_{GS}, some of the negative charge carriers are repelled from the gate and driven out of the n-type channel as shown in the figure 1.47 and increases the resistance of the channel. Due to the increase in channel resistance drain current will be reduced similar to the effect in n-channel JFET. Since the action of the negative voltage on the gate is to deplete the channel of free n-type charge carriers, the device is referred to as a *depletion mode* MOSFET. The figure 1.48 (a) shows the transfer characteristics and 1.48 (b) shows the drain characteristics.



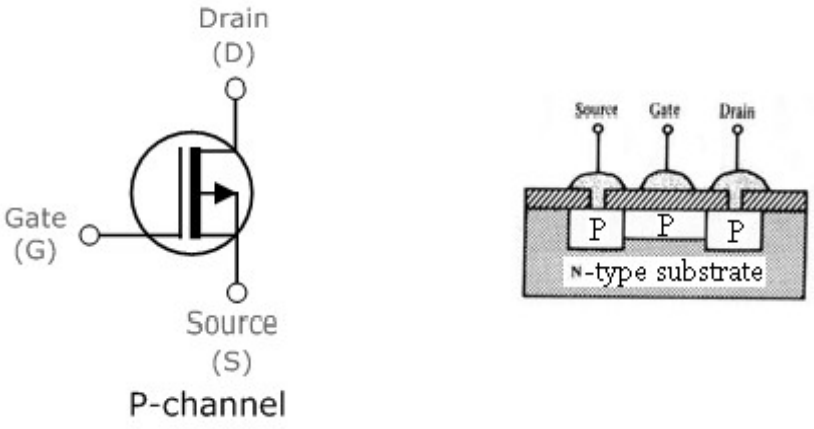
(a) (b)

Figure 1.48 (a) Transfer characteristics, (b) Drain characteristics

If positive voltage is applied to the gate terminal with respect to the substrate, additional n-type charge carriers are attracted from the p-type substrate into the channel, so the channel resistance decreases and the drain current will increase at a rapid rate. For this reason, this operating region is referred as *enhancement region.* But when increasing positive voltage, there is a chance of drain current exceeding maximum current rating.

p-channel depletion type MOSFET

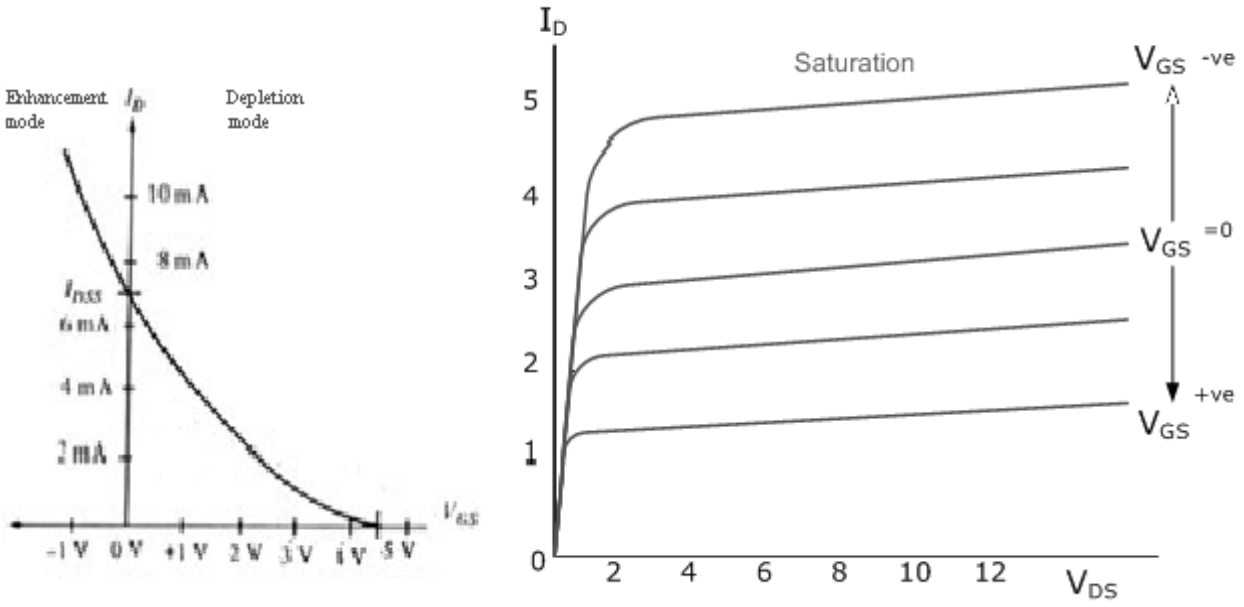
The p- channel depletion type MOSFET is similar in structural arrangement as the n-channel depletion MOSFET except that the regions are inversed as shown in figure 1.49(b). The symbol and basic structure of p-channel depletion MOSFET are shown in the figure 1.49 (a) & (b).



(a) (b)

Figure 1.49 (a) Symbol of p-channel depletion MOSFET
(b) Structure of p-channel depletion MOSFET

In p-channel depletion MOSFET, the voltage polarities and current directions are opposite to n-channel depletion MOSFET. When applying positive voltage to the gate, the width of the conducting channel is reduced and so the current through the channel is reduced. If negative voltage is applied to gate the current will increase. The transfer and drain characteristic of a p-channel depletion MOSFET are shown in the figures 1.50 (a)&(b) .

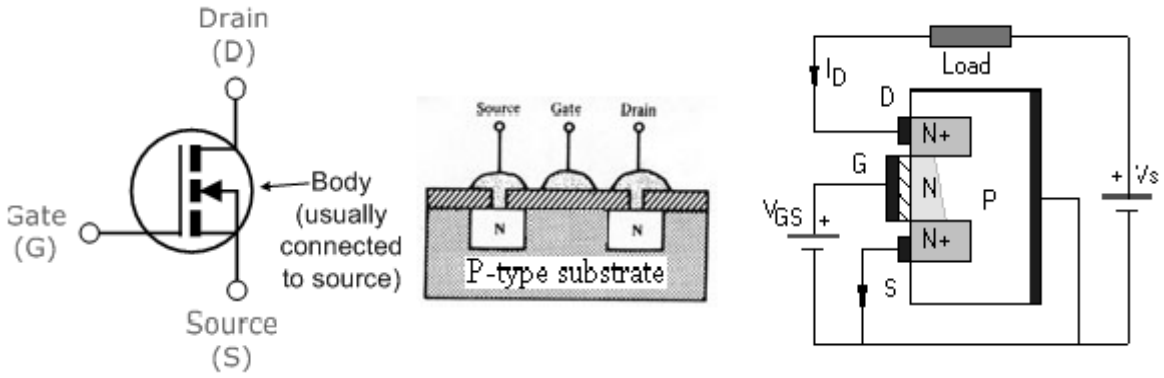


(a) (b)

Figure 1.50 (a) Transfer characteristics, (b) Drain characteristics

n-channel Enhancement MOSFET

The symbol and basic structure of n-channel enhancement MOSFET are shown in the figure 1.51 (a) & (b). The n-channel enhancement MOSFET consists of two heavily-doped n-type regions which are diffused into the p-type substrate. The source and drain terminals are taken from the n-type materials through the metallic contacts. The silicon dioxide layer is present on the surface of the p-type substrate to isolate the gate metallic platform from the region between drain and source as similar to the n-channel depletion MOSFET. The only difference between the depletion MOSFET and the enhancement MOSFET is absence of the channel as the constructed component of the device in the enhancement MOSFET. The figure 1.51 (c) shows the n- channel enhancement MOSFET with drain and gate voltage applied.



(a) (b) (c)

Figure 1.51 (a) symbol, (b) structure, (c) Cross sectional view of n- channel enhancement MOSFET operating in the linear region

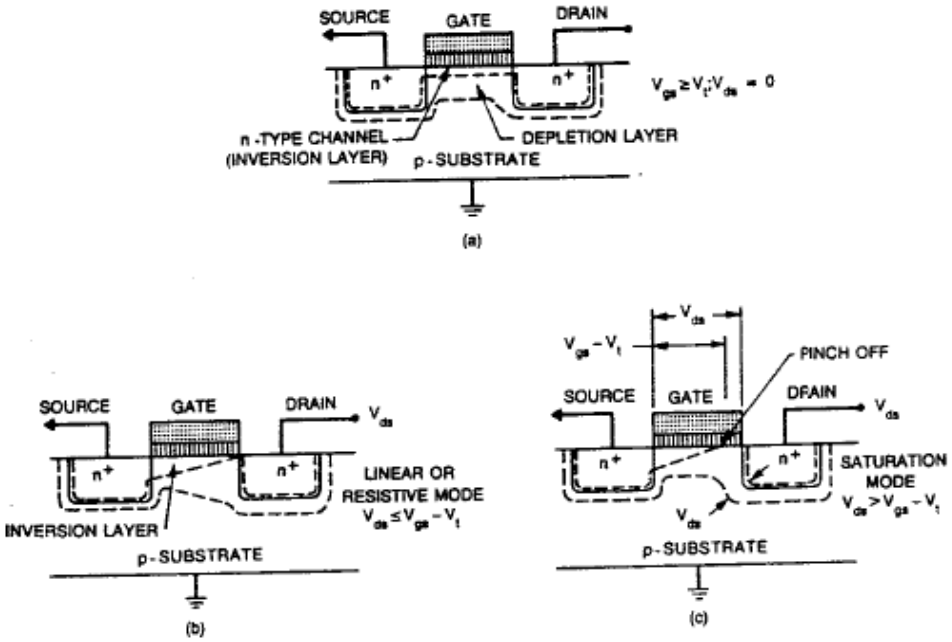


Figure 1.52 Operation of n- channel enhancement MOSFET with various conditions of V_{DS}

When applying 0V to V_{GS} and the drain is made positive with respect to source effectively no current flow through the device due to the absence of n-channel between the two n-type materials. If positive voltage is applied to the gate terminal, the holes in the p-substrate along the edge of the SiO_2 layer are pressured to leave the area and it causes formation of depletion region near the SiO_2 layer.

As gate voltage increases the electrons in the p-type substrate(minority carriers) are attracted and get accumulated near the surface of the SiO_2 layer. Further increase in V_{GS} causes the concentration of electrons to increase eventually to form the induced n-type channel known as inversion layer.

The level of V_{GS} that requires to create a channel is called the *threshold voltage*(V_{T0}). As V_{GS} increases beyond the threshold level, the density of free carriers in the induced channel will increase, resulting in an increased level of drain current as shown in the drain characteristics curve(figure 1.53). From the figure 1.52 as the drain voltage is increased the inversion layer charge and the channel depth at the drain end start to decrease. Eventually for V_{DS} =Vdsat=V_{GS}-V_{T0} the inversion charge at the drain is reduced to zero which is called the pinch-off point and the voltage is known as *pinch-off voltage or saturation voltage*. Beyond the pinch-off point i.e. for V_{DS}>Vdsat the operating mode is known as saturation mode. In this mode the effective channel length is reduced as the inversion layer near the drain disappears while the channel end voltage remains essentially constant and becomes equal to the Vdsat. Note that the pinched-off section of the channel absorbs most of the excess voltage drop V_{DS}-Vdsat and a high field region forms between the channel end and the drain boundary. Electrons arriving from the source to the channel end are injected in to the drain depletion region and are accelerated toward the drain in this high electric field and eventually constant current flows between source and drain even after pinched-off.

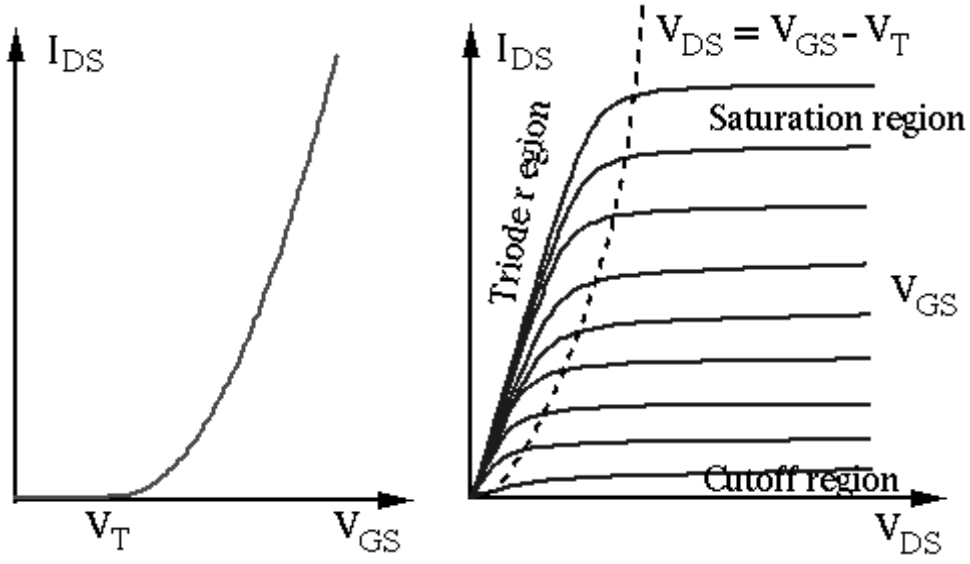
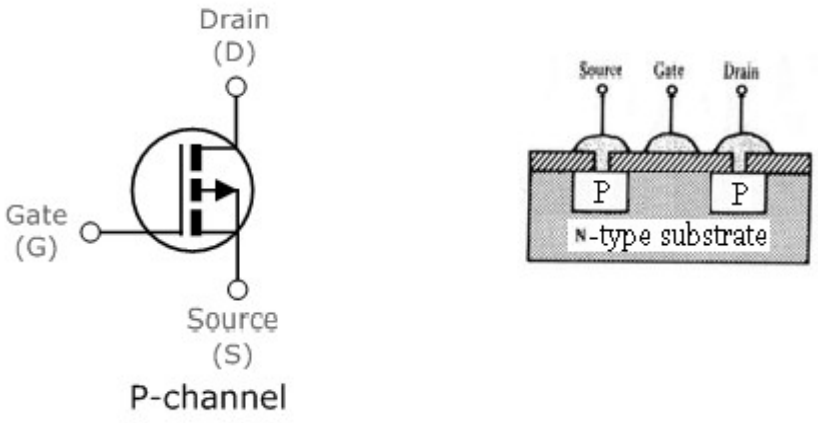


Figure 1.53 (a) Transfer characteristic, (b) Drain characteristics

p-channel enhancement type MOSFET

The structure of a p-channel enhancement MOSFET is exactly opposite to the n-channel enhancement MOSFET. The basic structure and its symbol are shown in the figure 1.54 (a) & (b).



(a) (b)

Figure 1.54 (a) symbol, (b) structure

The figure 1.55 shows the drain characteristics of the p-channel enhancement MOSFET. When increasing negative values of V_{GS} beyond threshold voltage value(V_{T0}) results increase in drain current. The threshold voltage for pmos is negative unlike it is positive for nmos.

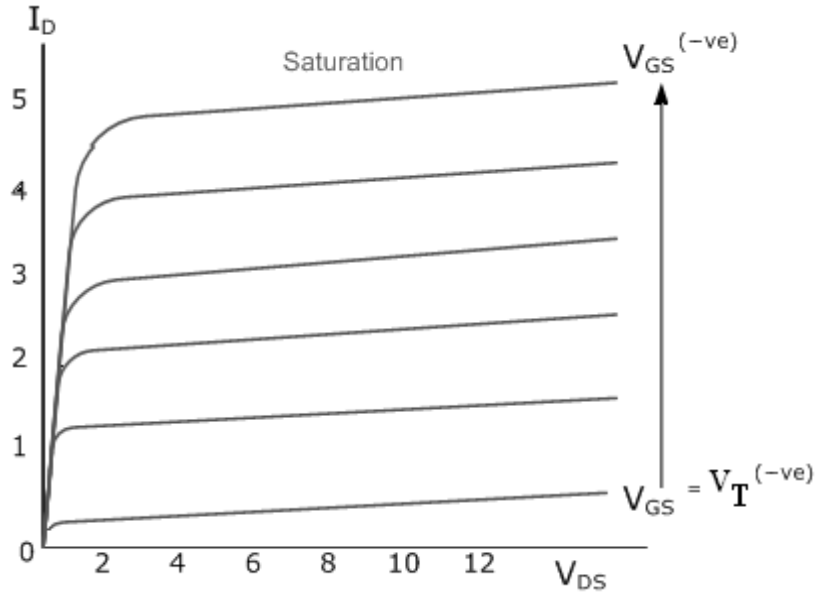


Figure 1.55 Drain characteristics

1.10 Introduction to Heat sink

All semiconductor devices have some electrical resistance, just like resistors and coils, and the like. This means that when power diodes, power transistors and power MOSFETs are switching or otherwise controlling reasonable currents, they dissipate power as heat energy. This heat must be removed from inside the device (usually the collector-base junction for a bipolar transistor, or the drain-source channel in a MOSFET) at a fast enough rate to prevent excessive temperature rise. The most common way to do this is by using a *heat sink*.

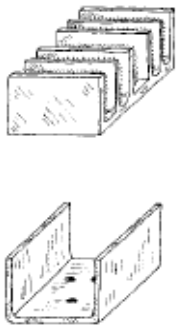


Figure 1.56 Heat sinks

Heat sinks are metal devices with large surface areas designed to reduce the temperature of an electronic device by dissipating heat into the surrounding air. The figure 1.56 shows the heat sinks in different styles. A heat sink is usually connected to a component by means of a screw and washer fastener. Silicon grease or heat-conducting paste placed between the washer and heat sink is often used to enhance the thermal conductivity between the electrical component and heat sink. Heat sinks are generally made of an aluminum alloy.

In some modern electronic systems heat sinks are used with fans. For example CPU require a heat sink and some also require a fan. A heat sink without a fan is called a *passive heat sink;* a heat sink with a fan is called an *active heat sink.*

Heat sink ratings

Heat sinks are rated by their thermal resistance (Rth) in °C/W. For example 3°C/W means the heat sink will be 3°C hotter than the surrounding air for every 1W of heat it is dissipating. Note that a lower thermal resistance means a better heat sink.

The required heat sink rating for a transistor can be chosen as follows.

Work out thermal power to be dissipated, $P = I_C \times V_{CE}$

where I_C is the maximum collector current and V_{CE} is the half of the supply voltage.

For example if a power transistor is passing 1A and connected to a 12V supply,

the power P is

$P = I_C \times V_{CE}$

$=1 \times 6 = 6W.$ $(V_{CE}=1/2\ V_{CC})$

Find the maximum operating temperature (Tmax) for the transistor (assume Tmax = 125°C.)

Estimate the maximum ambient (surrounding air) temperature (Tair). If the heat sink is going to be outside the case Tair = 25°C is reasonable, but inside it will be higher (perhaps 40°C) allowing for everything to warm up in operation.

Work out the maximum thermal resistance (Rth) for the heat sink using: Rth = (Tmax - Tair) / P

With the example values given above: Rth = (125-25)/6 = 16.67°C/W.

Choose a heat sink with a thermal resistance which is less than the value calculated above (remember lower value means better heat sinking) for example 7°C/W would be a sensible choice to allow a safety margin. A 7°C/W heat sink dissipating 6W will have a temperature difference of 7 × 6 = 42°C so the transistor temperature will rise to 25 + 42 = 67°C (safely less than the 125°C maximum).

All the above assumes the transistor is at the same temperature as the heat sink. This is a reasonable assumption if they are firmly bolted or clipped together. However, if a mica sheet or similar is placed between them to provide electrical insulation, then the transistor will be hotter than the heat sink and the calculation becomes more difficult. For typical mica sheets you should subtract 2°C/W from the thermal resistance (Rth) value calculated above.

Chapter 2
Circuit theory

Contents

2.1 Introduction

2.2 Voltage and Current sources

2.3 Circuit elements

2 .4 Voltage and Current laws

2.4.1 Ohm's law
2.4.2 Kirchhoff's laws
2.4.3 Voltage divider rule
2.4.4 Current divider rule

2.5 DC Circuits

2.5.1 Simplification of resistances
2.5.2 Simplification of inductances
2.5.3 Simplification of capacitances
2.5.4 Star-delta and delta-star conversion

2.6 Network Theorems

2.6.1 Thevenin's Theorem
2.6.2 Norton's theorem
2.6.3 Maximum power transfer theorem
2.6.4 Superposition theorem
2.6.5 Millman's theorem
2.6.6 Telegan's theorem

2.7 Transient circuits

2.7.1 DC response of R-L circuit
2.7.2 DC response of R-C circuit
2.7.3 DC response of R-L-C circuit

2.8 Summary

2.1 Introduction

Circuit theory provides the basic principles required to understand the behaviour of electrical and electronic devices,circuits and systems .The electrical behaviour is analysed by a few basic experimental laws like Ohm's law, Kirchoff's Voltage law Kirchoff's current law. The principles,concepts,mathematical relationships and methods of analysis together constitute *circuit theory.*

A circuit or network is an interconnection of circuit elements and sources. The current flowing in a circuit transfers and transforms energy from one part of the circuit to another .There are several *techniques* available to analyze a given circuit. Depending on the nature and complexity of the circuit or network a particular technique of circuit analysis is usually adopted. Circuits are classified in to linear and non-linear circuits based on ohm's law. This chapter deals with linear circuits which consists of independent sources,linear dependent sources and linear elements .A Linear element is a passive element that has a linear voltage-current relationship .Example for passive elements are resistance,inductance and capacitance .A circuit is said to be bilateral if it has identical properties when the source and load are interchanged. This chapter deals with the laws, theorems, simplification and transients pertaining to dc circuits.

2.2 Voltage and Current sources

An *Ideal voltage source* will supply a constant terminal voltage independent of the current delivered .It has zero internal series impedance .Hence it is not practical to have an ideal voltage source as it implies that infinite amount of energy can be drawn from it .

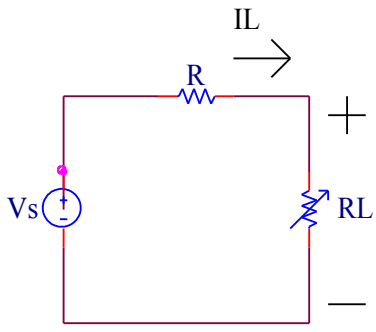


Figure2.1 Ideal voltage source

In the figure 2.1 ,Vs is ideal voltage source, R_s internal series impedance that tends to zero,I_t the terminal current , V_t the terminal voltage. and R_L the load.

Practical voltage sources will have negligible internal resistance .For good voltage sources the variation in the output will be negligible for specified load conditions

eg. battery,power supplies

An ideal current source is one which delivers a constant current through the load connected to it, independent of the voltage across it. An ideal current source has an infinite parallel impedance. The voltage across the current source is determined by the value of the load impedance and is zero for open circuit The ideal voltage and current sources are also called *independent sources.*

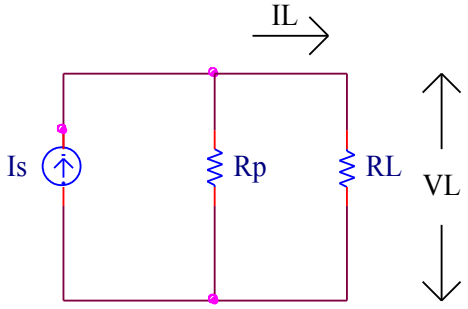


Figure2.2 Ideal current source

In the figure above I_s denotes the ideal current source,R_p the infinite parallel impedance,I_L the terminal current,V_L voltage across the terminals and R_L the load.

Dependent or Controlled sources

The sources whose voltage or current is dependent on the internal voltage or current existing at some other location in the remaining part of the circuit are called dependent sources. These types of sources mainly occur in the analysis of equivalent circuits of transistors.

The types of dependent sources are as follows

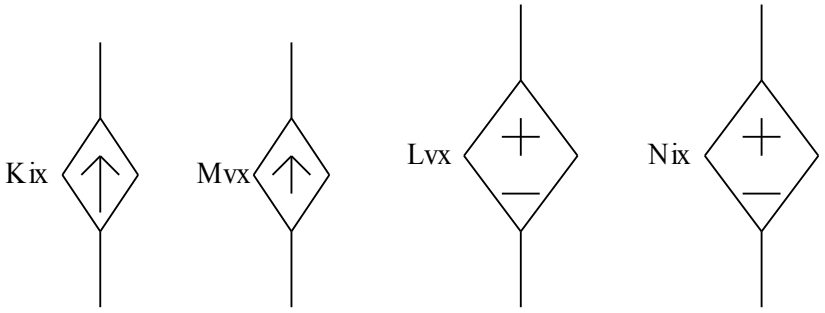
1.Current controlled current source

2.Current controlled voltage source

3.Voltage controlled current source and

4.Voltage controlled voltage source

To distinguish between dependent and independent sources we use the diamond symbol. K,L are dimensionless scaling constants, gM is a scaling factor with units of A/V, N is a scaling factor with units of V/A.ix is the controlling current and vx the controlling voltage.



Source transformation

A practical voltage source has a negligible resistance in series with it which is so small in comparison with other circuit resistances that it can be effectively ignored when determining the operation of the circuit . Similarly a practical current source has a high internal resistance across it which is so large in comparison with the other circuit resistances that the internal resistance of the source can be ignored .The process of transforming voltage source to current source and vice versa is known as *source transformation*. Source transformation is helpful when analyzing networks with a view to reduce the complexity of the networks. The source of one type can easily be converted to the other type as follows.

To convert a voltage source Vs (in series with an internal resistance Rs) in to an equivalent current source.

1.Find Vs/R_s.This is the equivalent current.

2. Replace the voltage source with a current source of value(Vs/R_s)

3.Place the resistance R_S as R_p in parallel as shown in figure 2.4(b)

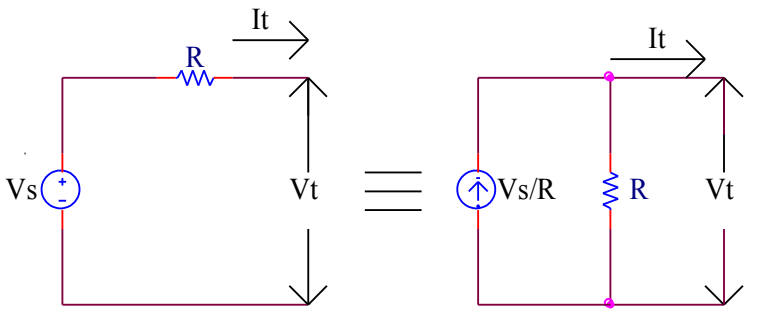


Figure 2.4(a)Voltage source (b)transformed current source

To convert a current source Is (in parallel with an internal resistance R_p) in to an equivalent voltage source.

1. Find $Is*R_p$. This is the equivalent voltage.

2. Replace the current source with a voltage source of value (IsR_p)

3.Place the resistance R_P as R_s in series as shown in the figure2.5(b)

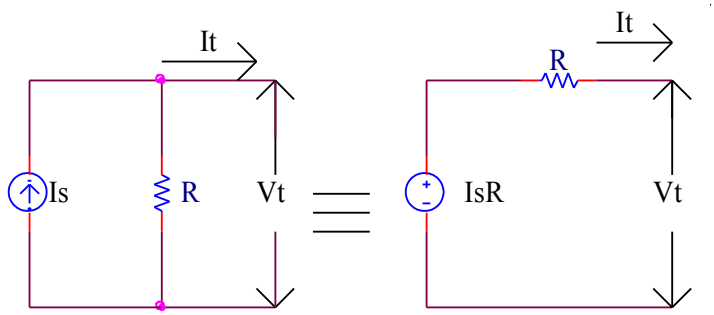


Figure 2.5(a)Current source (b)transformed voltage source

Proof

Consider the circuit shown in figure 2.6.The voltage across the load resistor is given by

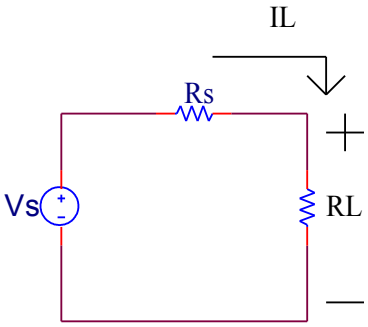


Figure 2.6

$$V_L=I_{L*}R_L \qquad \text{.....2.1}$$

The current through the resistor R_L is given as

$$I_L=Vs/(R_L+R_S) \qquad \text{.....2.2}$$

Now consider an equivalent source connected to the same load as shown in figure2.7.The current through the resistor R_L is given by

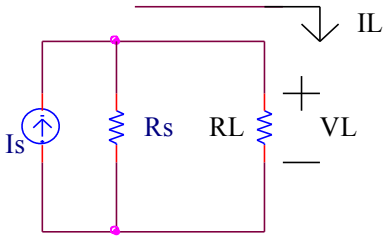


Figure 2.7

$$I_L=(R_S/(R_L+R_S))\ Is \qquad \text{.....2.3}$$

We know that $Is=Vs/Rs$ 2.4

substituting 2.4 into 2.3

$I_L=(R_S/(R_L+R_S))$ x (Vs/Rs)

$I_L=Vs/(R_L+R_S)$ This result is equivalent to the current obtained in equation 2.2

similarly from the figure2.7 the voltage across the resistor is given as

$$V_L=I_L R_L \qquad .....2.5$$

It is concluded that the load current and voltage are the same whether the source is a voltage source or an equivalent current source.

2.3 Circuit elements

The circuit elements are classified into active and passive. The basic passive elements are resistor , inductor and capacitor which are dependent upon external power source for their normal working .They consume power rather than producing it .The active elements are capable of generating power during their normal working . Examples are diodes , transistors etc. Passive components have limited frequency ranges, and operation of the part outside of that range can have some very unexpected results

Resistor

Resistor opposes the flow of current through it. Resistance has units of ohms(Ω).Resistance is given by

$$R= \rho l/A \qquad .....2.6$$

ρ- is the resistivity of the material.

l-is the length of the material

A-is the cross sectional area.

The voltage across the resistor is expressed as

$$v(t) = R\, i(t) \qquad .....2.7$$

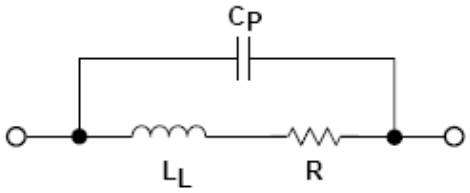
Sometimes, the v-i relation for the resistor is written i(t) = Gv(t), with G, the conductance, equal to 1/R. Conductance has units of Siemens (S).

The power dissipated by a resistor is given by the voltage across the resistor multiplied by the current through the resistor and is expressed as follows

$P(t) = Ri^2(t)$ (or) $(1/R)*v2(t)$ (or) $v(t)i(t)$ 2.8

Frequency Characteristics of Resistor

It is important that the analog circuit design starts with choosing the right passive components. This design process should consider the high frequency characteristics of passive components and putting the correct part outline from the start. The high frequencies that are radiated or conducted into a low-speed circuit will affect passive components. For example: a simple op amp low-pass filter may well turn into a high-pass filter at radio frequencies. . High-frequency performance of resistors is approximated by the schematic shown in Figurexxx



Resistors are typically of three types such as wire-wound, carbon composition, and film. Wirewound resistors can become inductive as they are coils of resistive wire. Similarly film resistors are also coils of thin metallic film. Therefore, film resistors are also inductive at high frequencies. The inductance of film resistors is lower and values under 2 kΩ are usually suitable for high frequencies. The end caps of resistors are parallel and they form capacitance along with the dielectric present in the middle. Usually, the resistance will make the parasitic capacitor so "leaky" that the capacitance does not matter. For very high resistances, the capacitance will appear in parallel with the resistance, lowering its impedance at high frequencies.

Capacitors:

A capacitor is a passive element that stores energy in the form of an electrostatic field. In its simplest form, a capacitor consists of two conducting plates separated by an insulating material called the dielectric .The process of storing energy in the capacitor is known as "charging", and involves electric charges of equal magnitude, but opposite polarity, building up on each plate. A capacitor's ability to store charge is measured by its capacitance . The relationship between the charge stored and the resultant voltage is

$$Q = C\,V \qquad \text{.....2.9}$$

In a capacitor, as the current is proportional to the rate of change of charge, the V-I relationship can be expressed in differential or integral form as

$$i(t) = Cd/dt(v(t)) \text{ or } v(t) = 1/c \int_{-\infty}^{t} i(t)dt \qquad .....2.10$$

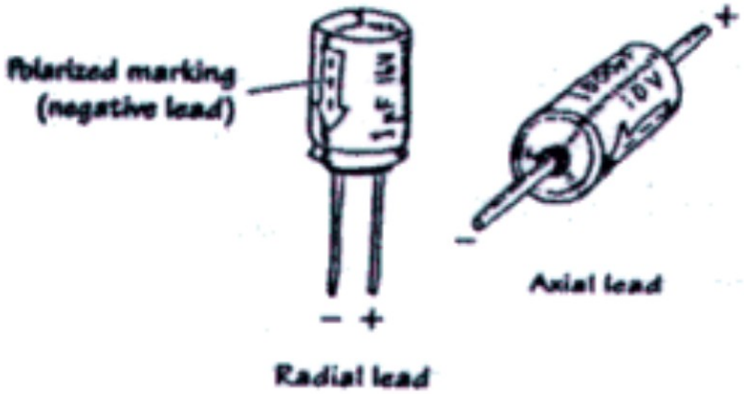
If the voltage across a capacitor is constant, then the current flowing into it equals zero. In this situation,the capacitor is equivalent to an open circuit. The power p(t) consumed or produced by a voltage applied to a capacitor depends on the product of the voltage and its derivative.

$$p(t) = v(t).\ C.\ d/dt(v(t)) \qquad .....2.11$$

Thus the total energy dissipated up to time t is given by

$$E(t) = \int p(t)dt = \int v(t).\ C.\ d/dt(v(t))dt = 1/2Cv^2$$

Capacitors are often used in electric and electronic circuits as energy storage devices. The capacitor also functions as a filter, passing alternating current (AC), and blocking direct current (DC) .Capacitive coupling is the transfer of energy within an electrical network by means of the capacitance between circuit nodes. Capacitive coupling is also known as ac *coupling* and the capacitor used for the purpose is known as ac *coupling or dc blocking capacitor*. Practically capacitors are available as through hole(leaded) or SMD components. Surface mount capacitors are extensively used compared to normal leaded capacitors, as they are much smaller and have no leads, due to which the



signal path becomes significantly shorter .As SM components are physically small, the layout can be very compact. This helps in minimizing both stray inductance and capacitance on the PCB or in the connecting wires and very suitable as power supply bypass capacitors for active components like transistors and IC's . The SM capacitors can be placed very close to the active device's power supply input, in order to keep the connecting PCB trace inductance at a minimum

Types of capacitors

Electrolytic capacitors

These capacitors include both aluminum and tantalum electrolytics. They are manufactured by an electrochemical formation of an oxide film onto a metal (aluminum or tantalum) surface. The metal on which the oxide film is formed serves as the anode or positive terminal, the oxide film acts as the dielectric, and a conducting liquid or gel acts as the cathode or negative terminal. Tantalum electrolytic capacitors have larger capacitance per volume ratios when compared with aluminum electrolytic. A majority of electrolytic capacitors are polarized. Electrolytic capacitors, when compared with nonelectrolytic capacitors, typically have greater capacitances but have poor tolerances (as large as +/- 100 percent for aluminum and about +/- 5 to +/-20 percent for tantalum), bad temperature stability, high leakage, and short lives. Capacitances range from about 1 uF to 1 F for aluminum and 0.001 to 1000 uF for tantalum, with maximum voltage ratings from 6 to 450 V.

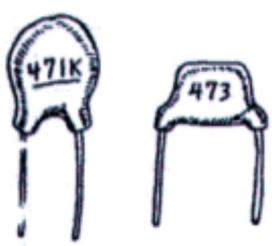
Tantalum capacitor

Tantalum capacitors are a form of electrolytic capacitor .They use tantalum powder, pressed into a pellet shape, as one "plate" of the capacitor with the oxide as the dielectric, and an electrolytic solution or conductive solid as the other "plate". Because the dielectric layer can be very thin (thinner than the similar layer in, for instance, an aluminium electrolytic capacitor), a high capacitance can be achieved in a small volume. They usually have longer life, especially at moderately elevated temperatures. For surface mount devices, they are generally smaller.

A tantalum capacitor is comprised of a permeable tantalum center section surrounded by tantalum pentoxide. A tantalum wire is inserted into the center section and then extends axially from the component. The tantalum pentoxide layer is coated with manganese dioxide, graphite, a silver conductive coating and (finally) solder.

- Tantalum capacitors have a higher volumetric efficiency
- Superior frequency characteristics
- ESR 10 times better than an aluminium electrolytic capacitor
- Highly reliable – electrical performance qualities do not degrade over time
- Excellent wide operating range from -55 degrees centigrade to +125 degree centigrade.

Ceramic



This type is very popular nonpolarized capacitor that is small and inexpensive but has poor temeperature stability and poor accuracy. It contains a ceramic dielectric and a phenolic coating. It is often used for bypass and coupling applications. Tolerances range from +/-5 to +/-100 percent, while capacitances range from 1 pF to 2.2 uF, with maximum voltage rating from 3 V to 6 kV.Figurexx shows the through hole and SMT type of ceramic capacitors.

Mylar

This type is a very popular nonpolarized capacitor that is reliable, inexpensive, and has low leakage current but poor temperature stability. Capacitances range from 0.001 to 10 uF, with

voltage ratings fr from 50 to 600 V.

Mica

This type is an extremely accurate device with very low leakage currents. It is constructed with alternate layers of metal foil and mica insulation, stacked and encapsulated. These capacitors have small capacitances and are often used in high frequency circuits (eg. : RF circuits). They are very stable under variable voltage and temperature conditions. Tolerances range from +/-0.25 to +/-5 percent. Capacitances range from 1 pF to 0.01 uF, with maximum voltage ratings from 100 V to 2.5 kV.

Multilayer capacitors

Features

1.High Q

2.Low ESR/ESL

3.Low noise

4.Extended WVDC upto 250 VDC

5.Capacitance range 0.1pF to 100pF

6.Ultra stable performance

7.High self-Resonance

8.Established reliability

RF Power Capacitors

1.High voltage

2.High current

3.High capacitance values available

4.Tighter tolerances

5.High reliability

6.High Q

7.Ultra-low ESR

BMC BroadBand Microwave Capacitor

1.Operating frequency-16Khz to 18Ghz

2.Low insertion loss:1dbmax

3.Solderable SMT terminations

4.Broad band performance

5.Flat frequency response

6.Excellent return loss through 18Ghz

7.Unit-to-unit performance Repeatability

8.Rugged ceramic construction

Single layer capacitors

1.Operating frequency upto 100Ghz

2.Rugged construction

3.Ultrahigh Q

4.Standard capacitance range 0.04 to 10,000pF

5.Dielectric constants from 14 to 25000

6.Voltage ratings upto 100WDC

7.Low cost

Ceramic trimmer capacitors

Features

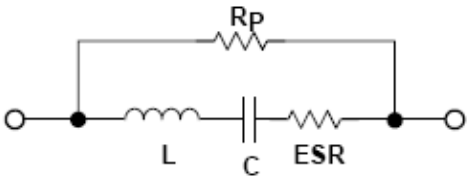
1. Ultra-small and thin

2. Unique construction with no plastic material provides superior soldering heat resistance to maintain excellent characteristic performance after reflow soldering.

3.Suitable for high frequency circuit due to high self resonant frequency

Frequency characteristics of Capacitors

At high frequencies a capacitor exhibits the characteristics of inductor and resistor and it is approximated as shown in the figure

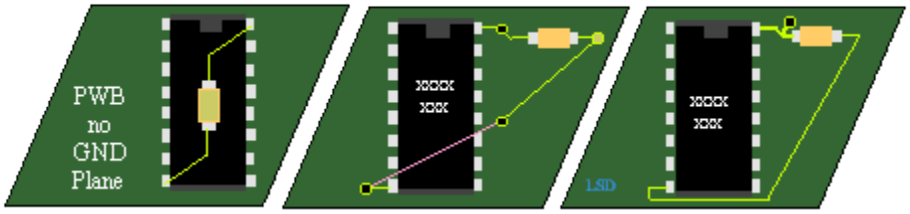


Film and electrolytic capacitors have layers of material wound around each other, which creates a parasitic inductance. Self-inductance effects of ceramic capacitors are much smaller, giving them a higher operating frequency. There is also some leakage current from plate to plate, which appears as a resistance in parallel (Rp) with the capacitor, as well as resistance within the plates themselves, which add a parasitic series resistance(ESR). The electrolyte in electrolytic capacitors is not perfectly conductive so as to reduce leakage current. These resistances combine to create the equivalent series resistance (ESR)

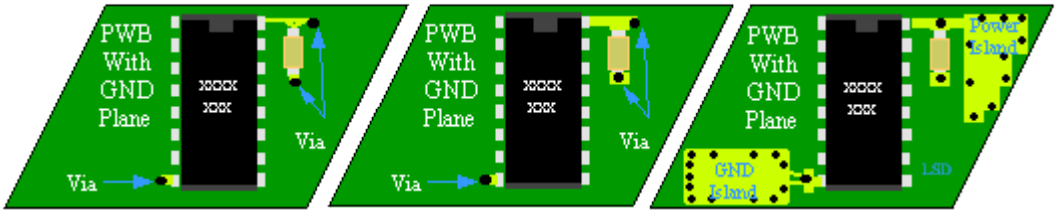
The capacitors used for decoupling should be of low ESR types, as any series resistance limits the effectiveness of the capacitor for ripple and noise rejection. Elevated temperatures also severely increase ESR and can be permanently destructive to capacitors. Therefore, if an aluminum electrolytic will be subjected to high temperatures, use the high-temperature grade(105°C), not the low temperature grade (85°C). For leaded parts, the leads themselves also add a parasitic inductance. For small values,if electrolytic capacitors are used in a design, make sure that the polarity is correctly observed. The positive terminal of the capacitor must be connected to the more positive of two dc potentials. If there is any doubt whatsoever which polarity is correct, design calculations must continue until it is known, or a prototype must be built. Incorrect polarity of electrolytic capacitors will cause them to conduct dc current, in most cases destroying the part — and probably the rest of the circuit as well. If there is a rare case in which there will be both polarities present, use a nonpolarized electrolytic (which is constructed by connecting two polarized electrolytic capacitors in series). Of course, one can always connect two capacitors in series on the PCB, keeping in mind that the effective capacitance will be cut in half for equal values of capacitor.

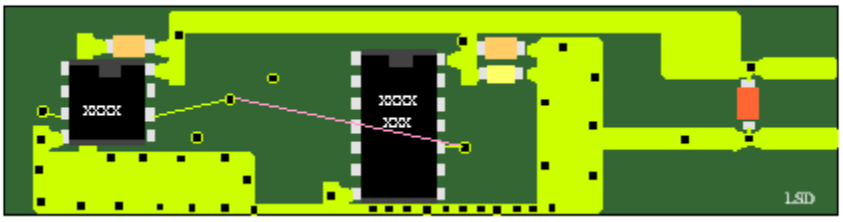
Capacitor placement in PCB

A design of a PCB to a circuit , involves in use of various types of capacitors whose placement in a layout design is one of an important consideration.The Decoupling capacitor(By-pass) is used to decouple the IC from the power source. The By-pass capacitor reduces the trace length from many inches to less then an inch. A reduction of 20nH an inch in trace inductance is achieved by introducing a decoupling capacitor in between the IC and its power source. Assuming the power source is a regulator 12 inches away, and the by-pass capacitor is placed within an inch of the IC, the trace inductance is reduced from 240nH to 20nH. The 20nH vaule is just an example



Running a long trace between the by-pass capacitor and the IC defeats the purpose of the by-pass capacitor. As the trace length increases, so does the trace inductance. For a PCB without a ground plane, the capacitor is usually placed in a equal distance between the supply and ground pins else it is placed near the supply pin .The remaining line should be directly routed to the ground trace so that it does not pass into an another layer with a via. A via in a PCB creates a low pass filter effect so its number should be limited.





In a proper designing of PCB low value capacitors are placed as close as to the device, leaving the shortest and widest possible trace. Additional larger value capacitors such as tantalum may be placed even at reasonable distance from the device. The tantalum capacitor helps to reduce the over-all

impedance of the power plane. .

Trace to Plane Capacitors

PCB traces, being composed of foil, form capacitance with other traces that they cross on other layers. Two traces crossing each other on adjacent planes seldom cause a problem. Coincident traces (those that occupy the same routing on different layers), form a long, skinny capacitor. Fortunately, these capacitances are usually small, only affecting high frequency designs. It is important, however, to minimize capacitance at the inverting input to an op-amp in high-speed designs. Otherwise, oscillation may occur. Capacitance can be reduced by shortening the PCB traces and reducing trace width. If oscillation still occurs, the resistors can be lowered a decade or two to damp it out. This will also result in lower circuit noise, but at the penalty of increased power consumption

Power Supply Bypass /Decoupling Capacitance

The power supplies are usually assumed to have an impedance low enough at the frequencies of interest to permit return currents to flow as desired. In fact, the connection from an integrated circuit to the power system has substantial inductance, which blocks all but the lowest frequency return currents from flowing through the power sources. Power supply decoupling provides three benefits to a design: (1) improves power supply integrity, (2) improves signal integrity, and (3) helps control EMI. Generally, designs with good power supply and signal integrity have the fewest EMI problems, and designs with good power supply integrity provide the best electrical environment for signal transmission. Proper decoupling —the placement of capacitance between power and the return (usually ground)—can make the difference between a design working marginally and one that is reliable. It's usually difficult to retrofit decoupling capacitance into a marginal design to turn it into one that works reliably. It's always better to properly design in the power supply and return path decoupling.

Parasitic Capacitance on Signal Traces

Parasitic capacitance can cause amplifiers to peak and in extreme cases oscillate. Fractions of a picofarad can control details of gain flatness and bandwidth. Whole picofarads can cause more gross effects. In general terms, capacitance is minimized by increasing distance between wires (reducing trace to trace capacitance) and minimizing trace length (reducing trace to ground capacitance). One of the most critical nodes to minimize parasitic capacitance on is the inverting input of an op amp. This is due to the interaction of parasitic capacitance with the feedback and gain set resistors, Rf and Rg. Many amplifiers are used at low gain, with relatively high Rf and Rg. Capacitance from the inverting input to ground is in parallel with Rg, and directly impacts the op amp closed loop transfer function. In addition, this capacitance can add a pole to the loop transmission, making the

feedback unstable.

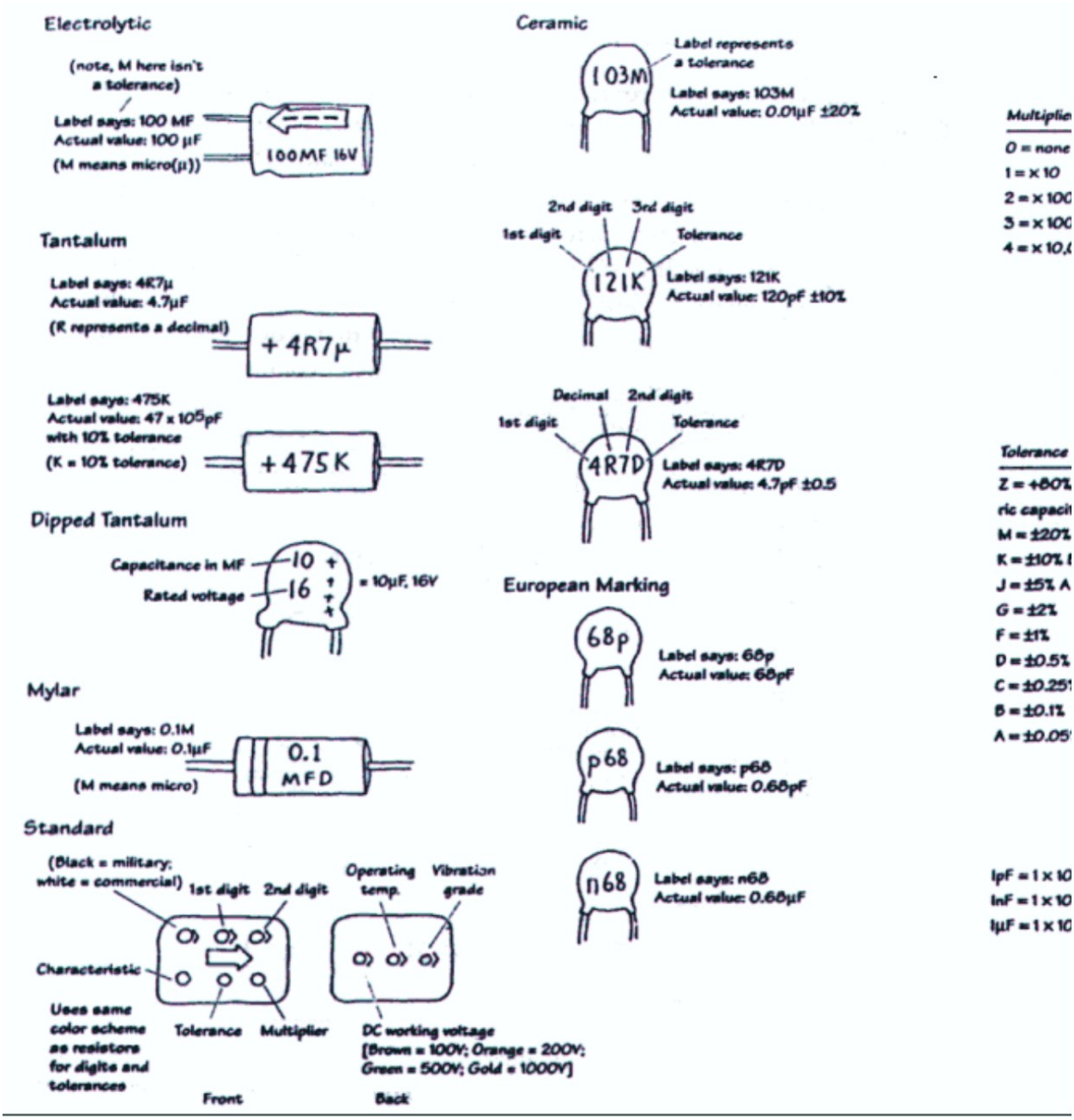
Parasitic Issues in Grounding

Ground and power planes should be removed from the vicinity of the inverting input pin. For very high speed amplifiers (>50MHz), removing ground plane from around the output pin is also a good idea. There are other issues to consider in component grounding. Ground planes are not ideal conductors; they are made up of a mesh of parasitic resistance and inductance. A current in the ground plane will flow in multiple paths, distributed relative to the conductance of each path. As the current flows through the ground plane impedance, it causes varying signal voltages to appear throughout the plane. The signal levels are quite small, but can have an impact on sensitive circuits.

Harmonic distortion mostly the even harmonics can be very sensitive to ground current routing, especially at high gains. Current leaving the output of an amplifier flows to the load and returns to the power supply bypass capacitors according to polarity. Since multiple supply pins (+VS and -VS) are not usually located next to each other, the ground current return path varies depending on polarity. Some of this current will route past the input grounds, with one supply having a stronger effect due to ground current routing. This causes one polarity of the input signal to be altered, but not the other polarity, leading to additional even harmonic distortion (2nd, 4th, 6th, etc.). Minimize this by rotating the bypass capacitors so that their ground connections are towards the output rather than the inputs. A very good approach for triple and quad bypassing is to ground the positive and negative bypass capacitors at a common point, forcing all polarities of return currents to flow by the same paths.

Connecting to Decoupling Capacitors

The way in which a decoupling capacitor is connected into the power system determines its parasitic lead inductance and thus is a factor in determining the capacitor's resonant frequency. Often an integrated circuit's power and ground connections are made by connecting the pins by way of vias into power and ground planes running underneath the device. In this scheme, decoupling capacitors are located close by the integrated circuit and connect into these planes with their own vias. An alternative method is to connect the capacitors to the integrated circuit's pins with trace and then use vias to make the connection into the planes below .These schemes are illustrated in Figure 6.22. The total loop inductance is twice the sum of the via's inductance plus the inductance due to any trace connecting the capacitors' mounting pads to the via. The inductance is two times this value in order to account for the inductance in both capacitors' leads. As shown in Figure 6.22(a), the trace inductance depends on the height h above the plane and its width w. A 5-mil-wide trace 5 mils above a return plane will have an inductance of about 10 pH/mil length.



Reading capacitors labels is tricky business. Each family of capacitors uses its own unique labeling system as shown in the figurexxx above. The best way to configure out what a capacitors label means is to first figure out what family the capacitor belongs to. After that, try seeing if the capacitor label follows one of the convention like these .

Inductors

An *inductor* is a passive component which resists the changes in current through it. Inductors store electrical energy in the form of electromagnetic field around their coils .A current carrying conductor produce a magnetic field .This magnetic field is linearly related to the current which produced it. Similarly a changing magnetic field could induce a voltage in the neighbouring circuit .The voltage is proportional to the time rate of change of the current and therefore the VI relationship is expressed as

$$v(t) = L\,di(t)/dt \quad \text{or} \quad i(t) = 1/L \int v(\alpha)\,d\alpha \qquad .....2.12$$

where L is the propotionality constant known as inductance. Inductance has units of henries (H). From the above relation, the larger valued inductors are capable of storing more flux.

The power consumed by an inductor depends on the product of the inductor current and its derivative,

$$P(t) = L\,i(t)\,di(t)/dt \qquad .....2.13$$

and its total energy dissipated up to time t is given by

$$E(t) = 1/2Li^2(t) \qquad .....$$

From a field's perspective, inductance relates the number of magnetic flux lines in given enclosed region to the current (*I*) required to produce the field lines in that region . The general definition of inductance *L*:

$$L = \frac{\Phi}{I} = \frac{BA}{I}$$

The above relation shows inductance as a property of current flow and the magnetic field contained within a region.

Types of Inductors

Inductors are classified by the type of core and the winding. Kinds of inductors are coupled inductors, multi-layer inductors, wire wound ceramic core inductors, molded inductors, power inductors, shielded power inductors, molded inductors, conformal coated inductors and wide band chokes.

Coupled inductors: In these types of inductors, the magnetic flux of one conductor is linked to another conductor. These are used in special applications where mutual inductance is required like in transformers.

Multi-layer inductors: In this type, the coil is wound in multiple layers with insulation between each layer. This provides very high inductance.

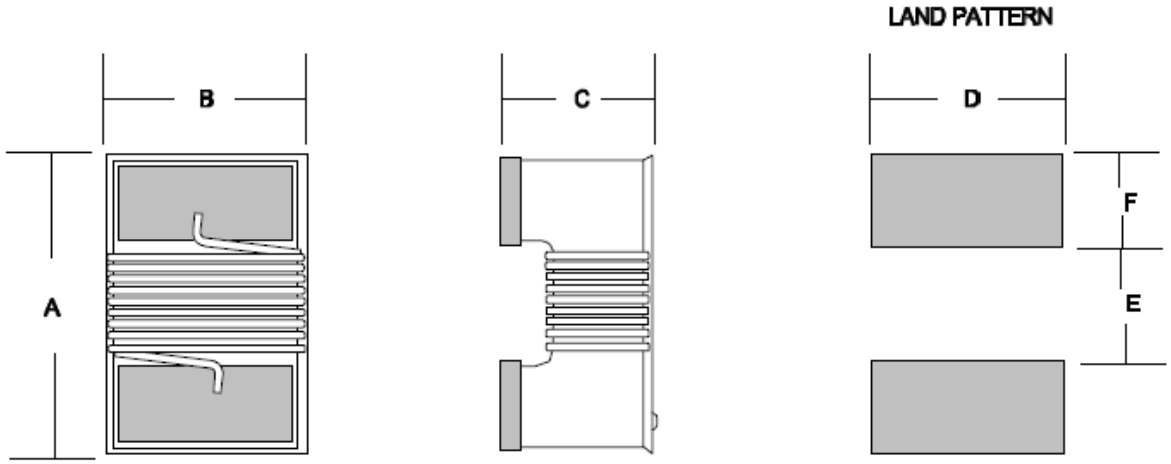
Ceramic core inductors: The core is made of a ceramic, which is a dielectric material and this type of inductor has high linearity, low hystersis and low distortion.

Molded inductors: These are low value inductors used in printed circuit boards. Usually bar or cylindrical in shape, they have windings on a core and the whole assembly is molded in plastic or ceramic insulation. Terminations are given at the ends and they come in two types : Through mount and Surface mount.

Surface Mount Inductors

Chip inductor has been designed to meet the requirements of the telecommunications market of high production capability and inexpensive cost. High self resonant frequencies and 'Q' are ensured by this optimum coil design under a zero defects quality program. The gold plated terminations allow excellent solderability by reflow and wave soldering processes as well as with many conductive adhesives.

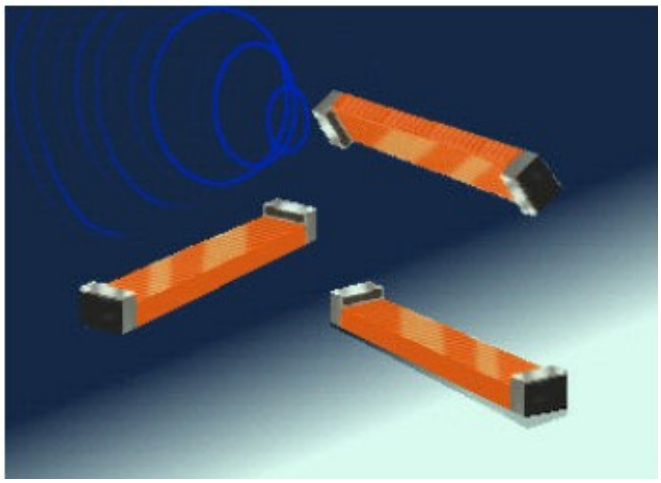
COMPONENT OUTLINE



ECM Chip Transponder Inductors

1. High Shock Resistance
2.High Temperature Tolerance
3.Ferrite Core
4. 2 Standard Sizes
5. Taped and Reeled
6.Typical Reel Size 3000pcs

The TR series of surface mountable wound inductors was the very first of its type specificallydesigned for transponder applications. Continuous product developments mean the TR range continues to be the most effective antenna coil available for a wide choice of applications. Its length and cross sectional area is optimised to achieve maximum sensitivity on the coils axis.For designs such as industrial and automotive applications, that need to withstand high shock andvibration characteristics, high shock versions are also available.



High Current SMD Inductors

FEATURES :

 Low profile (2.40mm max. height) and 7.3 mm max. square.

 Magnetically shielded and low DC resistance .

 Suitable for large current .

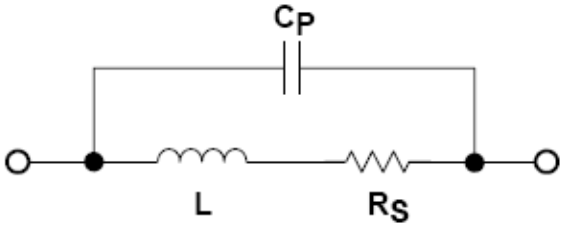
Ideal for DC – DC converter inductor application in hand help personal Computer ,etc .

 Frequency range up to 5MHZ .

 Large current handling capability

Frequency characteristics of inductors

High frequency performance of inductors is approximated by the schematic shown in FigureParasitic resistances are easy to understand — the inductor is constructed of wire, which has a given resistance per unit length. Parasitic capacitance is harder to visualize, unless one considers the fact thateach turn of wire in the inductor is located next to adjacent turns, forming a capacitor. This parasitic capacitance limits the upper frequency of this inductor to under 1 MHz. Even small wire-wound inductors start to become ineffective in the 10 MHz to 100 MHz range.



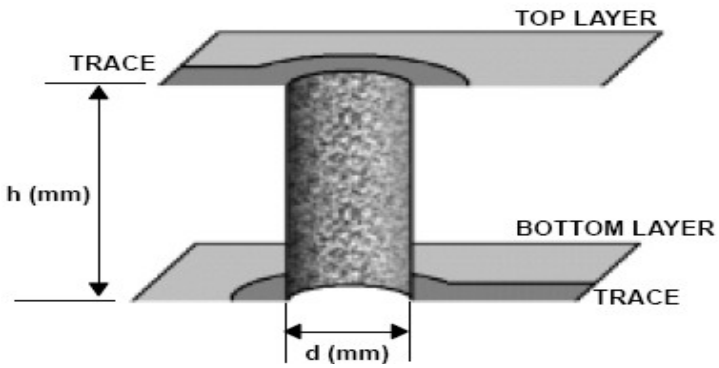
When current flows through conductors, there is inductance. Inductance is like the anti-current in that it is the opposition to any change in the flow of current. This means if there is no current on the trace,

inductance acts to briefly prevent it by building up magnetic fields first, before the energy can flow. Conversely, if there is current on the trace, inductance slows the dissipation of it by using up the energy in the magnetic fields before the current is completely dispelled. Since inductance always exists on a board, the designer must know how to control it. The magnetic fields are most problematic when they are changing. Signals that change often or change rapidly cause more interference to other signals. If the magnetic fields from one signal cut through a neighboring trace, they can induce a voltage in that trace when none should be there, causing problems on that signal. Extra spacing between traces would help in this instance.The inductive loop area should always be kept as small as possible. In order to keep the magnetic fields to a minimum, planes should be placed as close as possible to every signal layer. This reduces the loop area of the return current and allows the magnetic fields to partially cancel each other, reducing interference.Another way to control inductance is shorter traces because they create a smaller loop area and possibly keep the signal from reaching its maximum interference potential. Inductance also decreases very slightly if the trace is wider or thicker, but the percentage of decrease is so small it is insignificant (1 to 2%).

Inductive Vias

Whenever routing constraints force a via (connection between layers of a PCB, Figurexxx.a parasitic inductor is also formed. At a given diameter (d) the approximate inductance (L) of an via at a height of (h) may be calculated as follows:

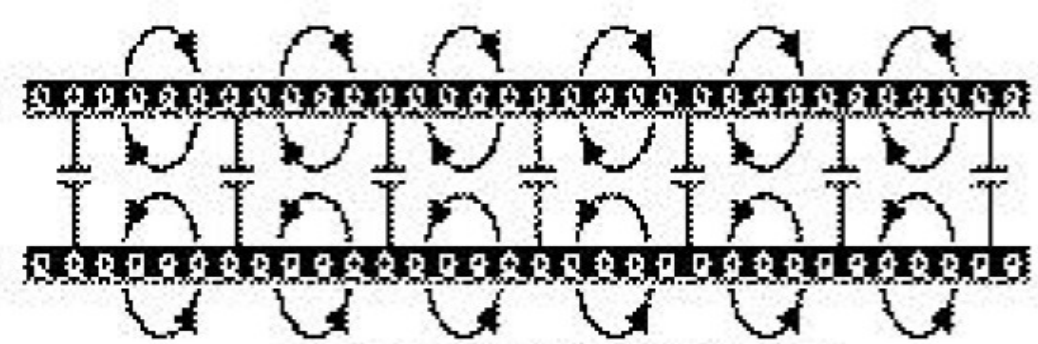
$$L = \approx \frac{h}{5} \times \left(1 + \ln\left(\frac{4h}{d}\right)\right) nH$$



For example, a 0.4-mm diameter via through a 1.5-mm thick PCB has an inductance of 1.1 nH. It must be kept in mind that inductive vias combined with parasitic capacitance can form resonant circuits. The self-inductance of a via is small enough that these resonances are in the GHz range, but inductors add in series, lowering the resonant frequency. Do not put several vias on a critical trace of a high-speed analog circuit! Another concern is that the vias put holes in ground planes, potentially creating ground

loops. They should be avoided — the best analog layout is one that routes all signal traces on the top layer of the PCB**.**

Trace to Trace Capacitors and Inductors

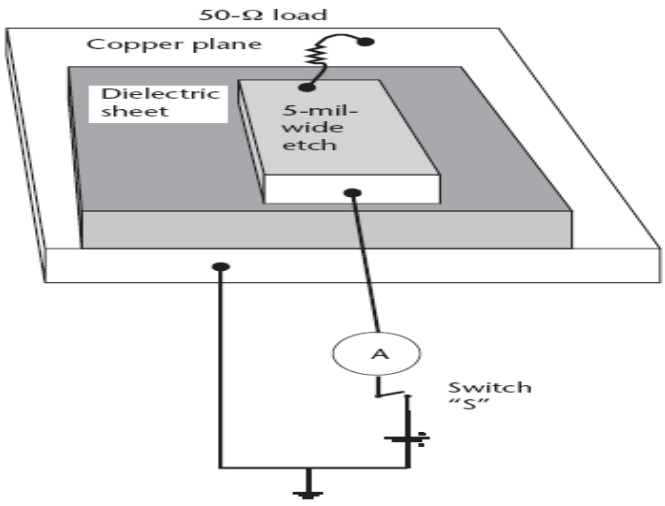


TRUE BEHAVIOR OF PARALLEL PCB TRACES

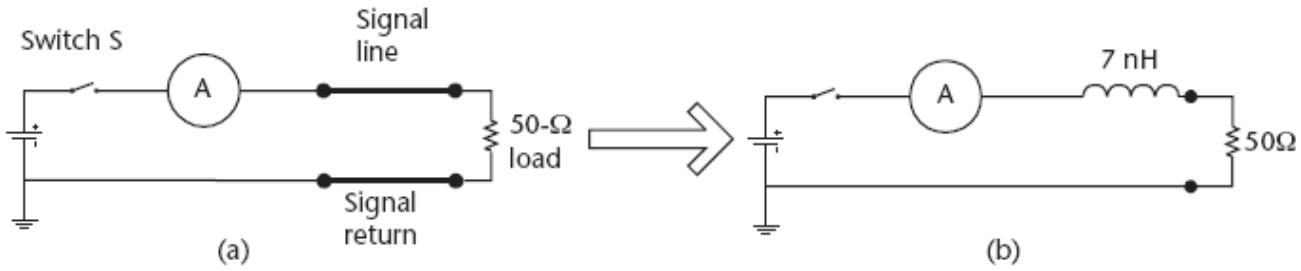
PCB traces are not infinitely thin. They have some finite thickness, as defined by the "ounce" parameter of the copper clad foil. The higher the number of ounces, the thicker the copper. If two traces run side-by-side, then there will be capacitance and inductive coupling between them. The formulas for these parasitic effects can be found in transmission line and / or microstrip references, but are too complex for inclusion here. Signal lines should not be routed parallel to each other, unless transmission line or microstrip effects are desired. Otherwise, a gap of at least three times the signal trace width should be maintained (figure xxx)

Circuit Behavior of Inductance

To illustrate inductive effects, consider a 5-mil-wide, 1-in-long, 50-Ω microstrip terminated in a50-Ω load, as shown in Figure 4.2. For purposes of this discussion, the trace will have no capacitance or resistance. A circuit schematic of the Figure 4.2 setup appears in Figure 4.3. From Ohm's law lthe ammeter should show a 20-mA step response when the switch is closed.

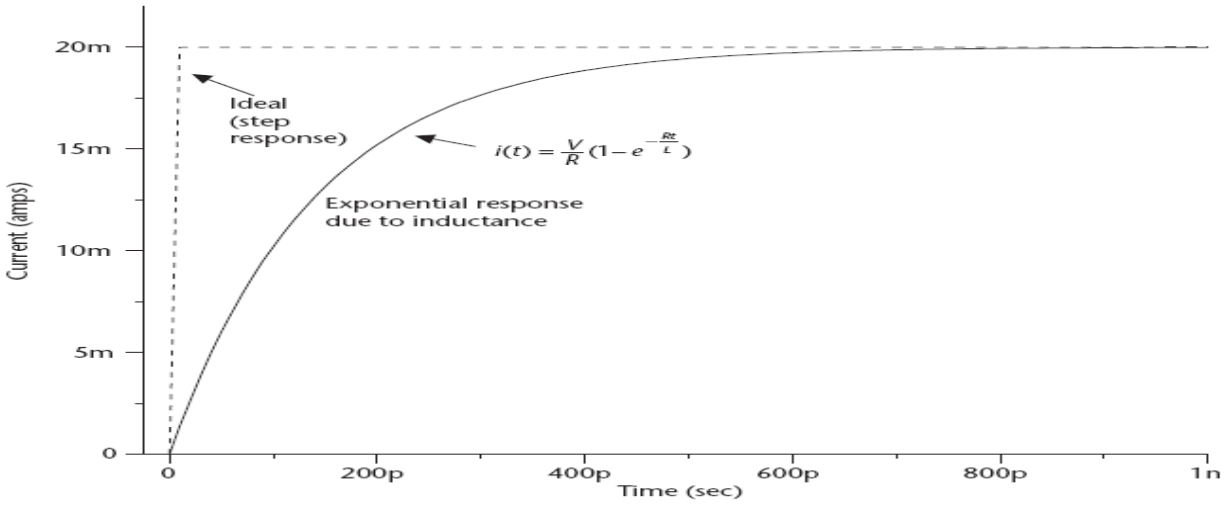


4.2 Microstrip connecting load to source



Schematic of Figure 4.2 topology, showing signal and return inductance may be represented by one 7-nH inductor, provided the proper reference point is preserved. In (a) the load is connected to S by inductance in the signal and return lines. These have been combined in (b) to a single inductor.Instead, the ammeter shows the current rising exponentially, requiring nearly 1 ns to fully reach 20 mA. This is shown in Figure 4.4. The reluctance of the current to change suddenly is due to the inductance of the path connecting *S* to the load. It's often said that a trace or wire has a certain amount of inductance, but it's actually the source/return loop that possesses this property. It's not possible to determine the signal's inductance unless the return path is known. In fact, the waveform in Figure 4.4 is only correct with respect to the reference connection shown. Inductance is therefore a property of the physical relationship between a signal and its return. As shown in Figure 4.3(b), on schematics it's often convenient to lump the inductance totally in the signal lead. If done carefully, this gives the proper circuit response. However, when doing so it's important not to lose sight that inductance is a property of the current flowing in an area and the relationship to the return path is key.

Figure 4.4 Ideal and actual current response of the inductive circuit illustrated in Figure 4.3.



Inductive Reactance

With some simple algebra, (4.4)Faraday's Law(e=-L di/dt) can be manipulated to determine *Xl*, the inductor's

reactance (4.6), assuming current varies sinusoidally

$$X_l = 2\pi fL$$

The reactance (*Xl*) has ohms as units, and the relationship between the current

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

through the inductor and its terminal voltage becomes (4.7):Tactically assumed in (4.7) is that the current lags the voltage by 90°. This is just the opposite of a capacitor, where the current leads the voltage by 90°.

2.4 Voltage and Current laws

2.4.1 ohm's law

Ohm's law states that at constant temperature current through a conductor is directly propotional to the potential difference across the conductor .The proportionality constant is known as resistance . The ohm's law is expressed mathematically as

I=V/R or V=IR 2.15

where *I* is the current in amperes *V* is the potential difference in volts,and *R* is a constant parameter called the *resistance* measured in ohms.

2.4.2 Kirchhoff's laws

Kirchhoff's current law and Kirchhoff's voltage law constitute Kirchhoff's laws. Kirchhoff's current law is applicable to nodes or junctions in a network .A node is a common point in a network to which more than two circuit elements are connected .Whereas Kirchhoff's voltage law is applicable to closed loops or meshes in the network. A closed loop is defined as any path which originates at a point travels around a circuit and returns to the original point without retracing any segments.

Kirchoff's current law(KCL)

In a circuit, at any node,the total current entering in to the node is equal to the total current leaving the node.Alternatively,the algebraic sum of all currents at any node is zero.

In mathematical form,Kirchoff's current law is stated as follows

$$\sum I_{enteringnode} = \sum I_{leavingnode} \quad \text{(or)} \quad \sum I_{enteringnode} - \sum I_{leavingnode} = 0 \qquad .....2.16$$

Kirchoff's voltage law(KVL)

Kirchoff's voltage law states that the algebraic sum of the voltages around any closed path or loop in a circuit is always zero.The sum of voltage rises and voltage drops around a closed loop is equal to zero,symbolically this is described mathematically as follows.

$$\sum V = 0 \text{ for a closed loop} \qquad \text{(or)} \qquad .....2.17$$

$$\sum V_{rises} = \sum V_{drops} \text{ for a closed loop} \qquad .....2.18$$

2.4.3 voltage divider rule

In a series circuit since the same current flows through each resistor the voltage drops across them are proportional to the values of resistors. This principle forms the basis of Voltage divider rule.

The general form of the voltage divider rule(VDR) is

$$V_{RX} = (R_X / R_T) \times V \qquad .....2.19$$

where VR_x is the voltage drop across the resistor R_x , R_T is the total resistance of the series circuit ,R_X is the resistance across which the voltage drop is calculated and V is the total applied voltage.

Consider the circuit shown in figure 2.8

By applying KVL, $V=V1+V2+V3= IR1+IR2+IR3$

$$=I(R1+R2+R3)$$

$$=I R_T \qquad .....2.20$$

Hence $V1/V=IR1/IR_T=R1/R_T$ or $V1=V(R1/R_T)$ 2.21

Similarly $V2=V(R2/R_T)$;

$V3=V(R3/R_T)$ are obtained 2.22

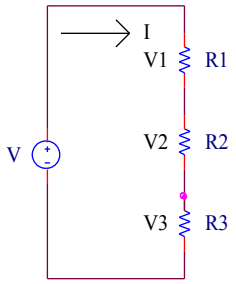


Figure 2.8

2.4.4 Current divider rule

In a parallel circuit the voltage remains same whereas the total current divides among the branches according to the resistance values .The branch having higher resistance allows less current and the branch with lower resistance allows more current. This forms the basis of current divider rule. The general form of the current divider rule (CDR) is

$$I_{RX} = (R_T/R_X) \times I_T \qquad .....2.23$$

where R_T is the equivalent resistance of the given parallel network .Irx current across the resistor Rx R_X is the resistance across which the current is calculated

Consider three resistors R1,R2,R3 connected in parallel as shown in figure

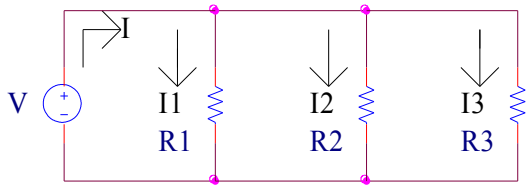


Figure 2.9

By applying KCL, the total current I is

$$I_T = I1+I2+I3 \qquad .....2.24$$

$$= V/R1+V/R2+V/R3 = V/R_T \qquad \text{where} \quad 1/R1+1/R2+1/R3 = 1/R_T$$

Now $I1/I = (V/R1)/(V/R_T) = R_T/R1$

or $I1 = I(R_T/R1)$ 2.25

Similarly $I2 = I(R_T/R2)$ and $I3 = I(R_T/R3)$ 2.26

In particular,if there are only two resistors R1 and R2 in parallel , then

$$R_T = (R1R2)/(R1+R2) \qquad .....2.27$$

and therefore

$$I1=IR2/(R1+R2) \text{ and } I2=IR1/(R1+R2) \quad .....2.28$$

2.5 DC circuits

2.5.1 Simplification of resistances

In real applications a circuit is complex and needs to be simplified for further analysis. For this purpse the circuit elements connected either in series or parallel can be replaced by their equivalent. Consider three resistors connected in series as shown in figure xxx which is excited by a constant dc source of V volts resulting in current of I Ampere.

By applying KVL,

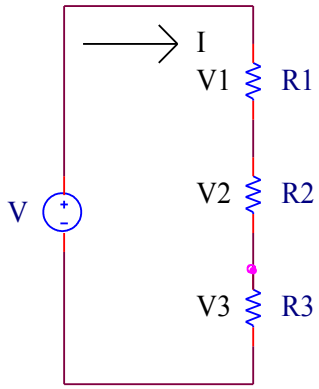


Figure 2.10

$$V=V1+V2+V3 \quad .....2.29$$

by applying Ohm's law,

$$= R1I+R2I+R3I$$

$$=I(R1+R2+R3)$$

$$=I R_s \quad .....2.30$$

$$R_s=R1+R2+R3 \quad .....2.31$$

Therefore for a circuit having n series resistors can be replaced by its single equivalent resistor R_s

$$R_s=R1+R2+.....+Rn=\sum Rx \quad .....2.32$$

Parallel connection

Consider three resistors R_1, R_2 and R_3 connected in parallel across a constant dc source of V volts .Let I be the total current and I_1,$I_2$and I_3 be the branch currents. Here the total current is given by

$$I=V(1/R_1+1/R_2+1/ R_3) \quad .....2.33$$

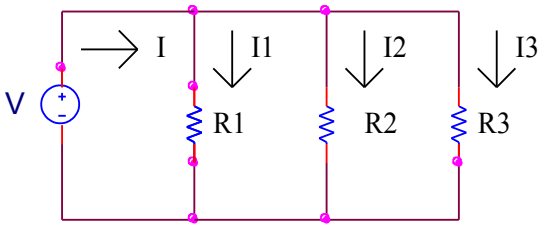


Figure 2.11

Let R_T denotes the single resistor which can replace the three resistors by

$1/R_p = 1/R_1+1/R_2+1/R_3$ 2.34

In general for a circuit of n resistors connected in parallel,The equivalent resistance is expressed as

$$1/R_p=1/R_1+1/R_2+1/R_3+.....+1/Rn = \sum(1/Rx)$$ 2.35

by applying ohm's law we get

$$I=V/R_T$$ 2.36

Let G_1,G_2,G_3 be the branch conductances(reciprocal of the resistance) and G_T is the equivalent conductance ,then

$$G_1= 1/R_1, \quad G_2= 1/R_2, \quad G_3=1/R_3.$$ 2.37

$$G_T= G_1+G_2+G_3$$

In general $G_T= G_1+G_2+G_3+.....+Gn = \sum Gx$ 2.38

2.5.2 Simplification of inductances

Inductance in series

Consider three inductors joined in series .Let a time-varying voltage v(t) be applied across them .It is evident that a time varying current i(t) results,and voltages v_1,v_2 and v_3 develop across the inductors,with instantaneous polarities.

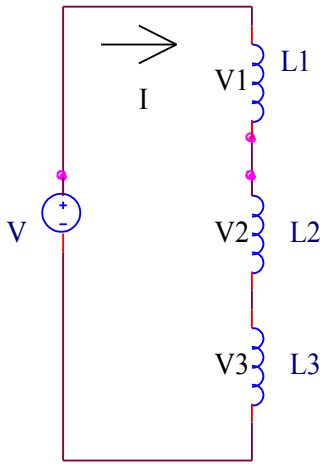


Figure 2.12

We have , $v_1=L_1\, di/dt$, $v_2=L_2\, di/dt$ and $v_3=L_3\, di/dt$ 2.41

Let the three inductors be replaced by an equivalent inductance L_s

Then,$v(t)=L_s di/dt$ 2.42

by KVL $v(t)=v1+v2+v3$

Substituting the values of v1,v2 and v3 into the above equation and simplifying

we get $L_s= L_1+L_2+L_3$. 2.43

Inductance in parallel

Let us assume three inductors connected in parallel as shown in the figurexxxxxxxxx

by applying KCL $i(t)=i_1+i_2+i_3$ 2.44

$i_1=1/L_1 \int vdt$ 2.45

$i_2=1/L_2 \int vdt$ 2.46

and $i_3=1/L_3 \int vdt$ 2.47

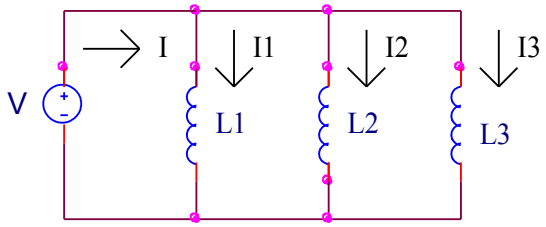


Figure 2.13

Let $L_p$be the Equivalent inductance

Hence $1/L_p=1/L_1+1/L_2+1/L_3$ 2.48

2.5.3 Simplification of capacitances

The following relationships holds true for a capacitor

$Q=CV$ 2.49

$i= C\, dv/dt$ 2.50

From the above expression we get

$v =1/C \int i dt$ 2.51

Capacitance in series

Consider three capacitors connected in series .The same charge Q flows through all the capacitors,when a steady voltage V is applied .As the capacitances are different, different voltages develop across the capacitors.

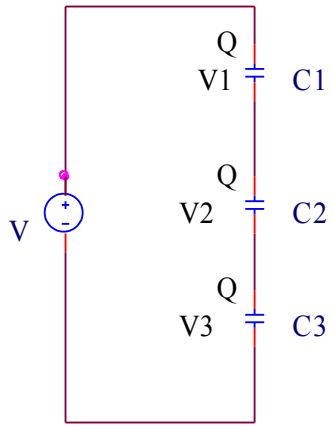


Figure 2.14

$$V_1=Q/C_1 \quad \text{.....2.52}$$

$$V_2=Q/C_2 \quad \text{.....2.53}$$

$$V_3=Q/C_3 \quad \text{.....2.54}$$

$$\text{Thus } V=Q(1/C_1+ 1/C_2+ 1/C_3) \ . \quad \text{.....2.55}$$

If these capacitors are replaced by an equivalent capacitor C_s we have $V=Q/C_s$

$$1/Cs= 1/C_1 + 1/C_2 + 1/C_3 \quad \text{.....2.56}$$

Capacitance in parallel

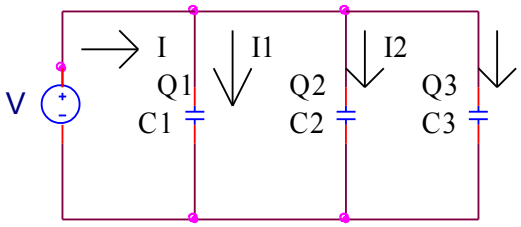


Figure 2.15

Consider a parallel combination of capacitors C_1,C_2 and C_3

we have $Q=VC$ in general 2.57

Total charge on all capacitors is given by

$$Q=V(C_1+C_2+C_3) \quad \text{.....2.58}$$

If the capacitors are replaced by a single equivalent capacitor C_p, we have

$$C_p= C_1+C_2+C_3 \quad \text{.....2.59}$$

$$\text{Thus } 1/ C_s=\sum 1/C \quad \text{for capacitors in series} \quad \text{.....2.60}$$

$$C_p=\sum C \quad \text{for capacitors in parallel} \quad \text{.....2.61}$$

2.5.4 Star-delta and delta-star conversion

The star delta transformation is useful in solving complex networks.The circuit is said to be in star connection,if three elements are connected as shown in figure,when it appears like a star.Similarly,the circuit is said to be in delta connection,if three elements are connected as shown in the figure,when it appears like a delta (Δ) .In some complex networks,before applying KCL or KVL or any other technique,it may be necessaryto modify the network so that the different branch impedence are arranged in a pattern which can easily be recogonized as series combination or parallel combination .This modification is usually brought about by the technique of Star-delta or delta-star transformation.

r1=(R1R2)/(R1+R2+R3)

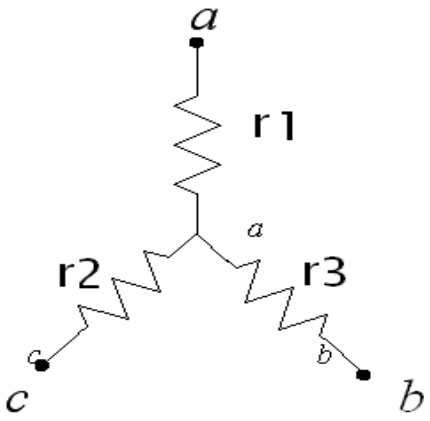
r2=R1R3/(R1+R2+R3)

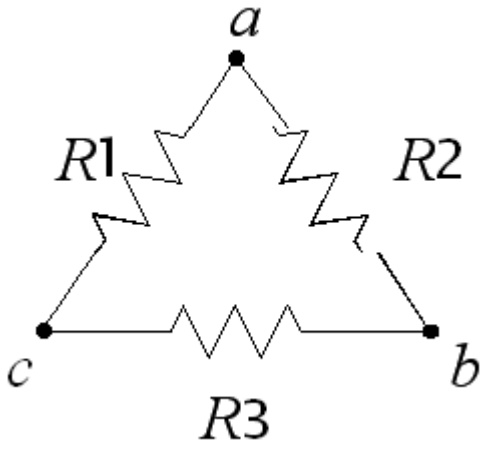
r3=R2R3)/(R1+R2+R3)

R1=(r1r2+r2r3+r3r1)/(r3)

R2=(r1r2+r2r3+r3r1)/(r2)

R3=(r1r2+r2r3+r3r1)/(r3)





2.6 Network Theorems

Network theorems are used to simplify a complex network into simpler one for easy analysis.When any part of a network is under analysis the remainder of the network may be replaced by a simple equivalent network determined by using network theorems. **The prominently used theorems are discussed here.**

2.6.1Thevenin's Theorem

Thevenin's theorem states that any active linear bilateral network with two terminals can be replaced by an equivalent circuit with a voltage source in series with a resistance..The voltage is calculated by open-circuiting the terminals and is called Thevenin's voltage .The resistance called thevenin's resistance is calculated across the open circuit terminals with all voltage sources in the network replaced by their internal resistances.

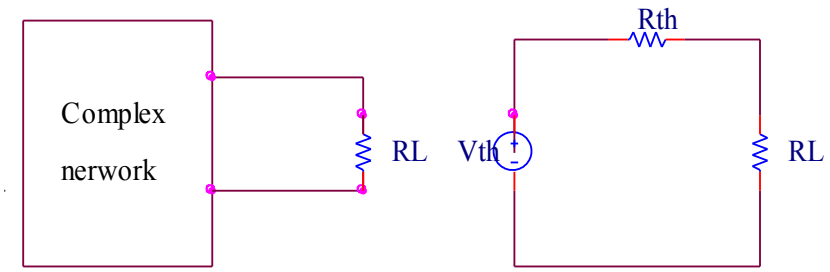


Figure 2.17 Thevenin's circuit

The basic procedure for solving Thevenins Analysis equations is as follows:

1. Remove the load resistor R_L or component concerned.
2. Find R_{th} by shorting all voltage sources or by open circuiting all the current sources.
3. Find V_{th} by the usual circuit analysis methods.
4. Find the current flowing through the load resistor R_L.

The Thevenin's theorem is applicable to circuits containing several independent and dependent voltage and current sources.

Note:Practical limitations of Thevenin's theorem.

- If the circuits are non linear the Thévenin equivalent is not valid. However the Thevenin's equivalent is valid for the linear range of operations .
- The Thevenin circuit has an equivalent I-V characteristic with respect to the specified terminals only.
- Power is not linearly dependent on voltage or current, the power dissipation of the Thévenin

equivalent is not identical to the power dissipation of the real system

2.6.2 Norton's theorem

This theorem states that any active,linear bilateral network with two terminals can be replaced by an equivalent circuit comprising of a current source and an impedance in parallel with it. The value of the current source being equal to the current passing through the short-circuit output terminals and the impedance being equal to the impedance between the open circuited output terminals,with all the sources in the network set equal to zero .

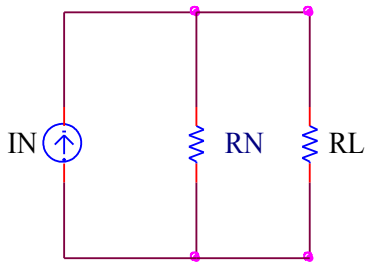


Figure 2.18 Norton's circuit

1.Find the Norton current IN. Calculate the output current

2.Find the Norton resistance RN. When there are no dependent sources (i.e., all current and voltage sources are independent), there are two methods of determining the Norton impedance .

3.Calculate the output voltage when in open circuit condition (i.e., no load resistor — meaning infinite load resistance). RN equals this output voltage divided by IN.

or

4.Replace independent voltage sources with short circuits and independent current sources with open circuits. The total resistance across the output port is the Norton impedance RN

Conversion to a Thevenin's equivalent

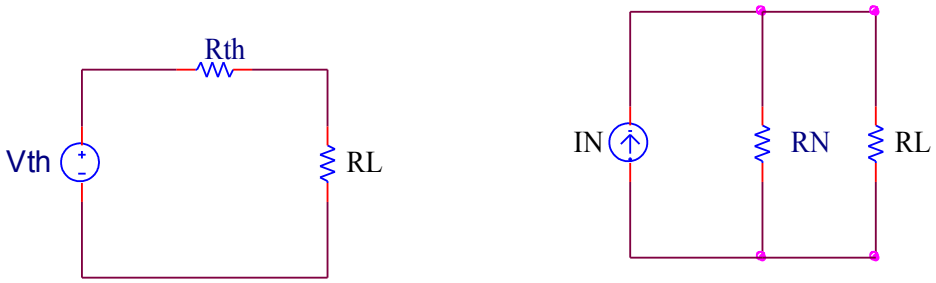


Figure 2.19

To convert to a Thevenin equivalent circuit the following equation is used

$$R_{th}=R_N, \quad V_{th}=I_N R_N \quad \text{.....2.68}$$

2.6.2 Maximum power transfer theorem

A network delivers the maximum power to a load resistance R_L when R_L is equal to the Thevenin equivalent resistance($R_L=R_{th}$) of the network.

Note:Therefore for maximum power transfer,the necessary condition to be satisfied is

Load resistance = network impedance,in the case of an a.c network.

Load resistance = network resistance,in the case of an d.c network.

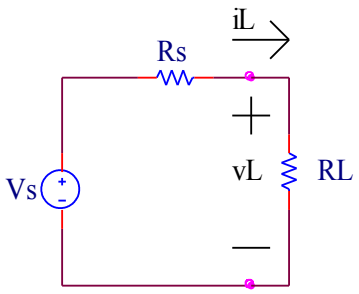


Figure 2.20 A practical voltage source connected to a load resistor R_L

Proof

For a practical voltage source (figure 2.20),the power delivered to the load R_L is

$$p_L=i^2_L R_L=(v^2_s R_L)/(R_s+R_L) \quad \text{.....2.69}$$

To find the value of R_L that absorbs a maximum power from the given practical source,differentiate the above equation with respect to R_L

$$d\, p_L/d\, R_L=((R_s+R_L)^2 v^2_s - v^2_s R_L(2)\,(R_s+R_L))/(R_s+R_L)^4 \quad \text{.....2.70}$$

equate the derivative to zero,that obtains

$$2R_L(R_s+R_L)=(R_s+R_L)^2 \quad \text{.....2.71}$$

or

$$R_s=R_L \quad \text{.....2.72}$$

2.6.3 Superposition Theorem

The Superposition theorem states that in any linear bilateral network containing two or more sources,the response in any element is equal to the algebraic sum of the responses caused by individual sources acting alone,while the other sources are nonoperative. that is,while considering the effect of individual sources,other voltage sources and current sources in the network are replaced by their internal impedances.

2.6.4 Tellegan's theorem.

Tellegan's theorem states that in an arbitrary linear or non-linear network, the algebraic sum of the powers in all branches at any instant is zero. All branch currents and voltages in that network must satisfy Kirchhoff's laws. Alternately, the algebraic sum of the powers delivered by all sources is equal to the algebraic sum of the powers absorbed by all elements. Resistors present in the network absorb and dissipate power,whereas capacitors and inductors store power.

2.6.5 Millman's theorem

Millman's theorem states that if in any network if the voltage sources V1,V2.....Vn in series with internal resistances(impedances) R1(Z1),R2(Z2) ..... Rn(Zn),respectively are in parallel,then these

sources may be replaced by a single voltage source V' in series with R'Z' as shown in figure

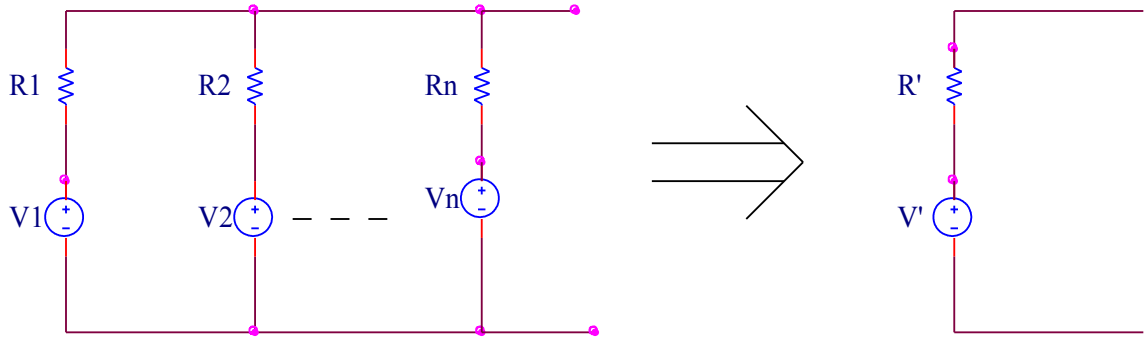


Figure 2.21

where V'=(V1G1+V1G1+.....VnGn)/(G1+G2+.....Gn) 2.73

Here Gn(Yn) is the conductance(admittance) of the n^{th} branch

and R'=1/(G1+G2+.....Gn) 2.74

Similarly for n current sources having internal conductances(admittances) which can be replaced by a single current source I' in parallel with an equivalent condutance (admittance).

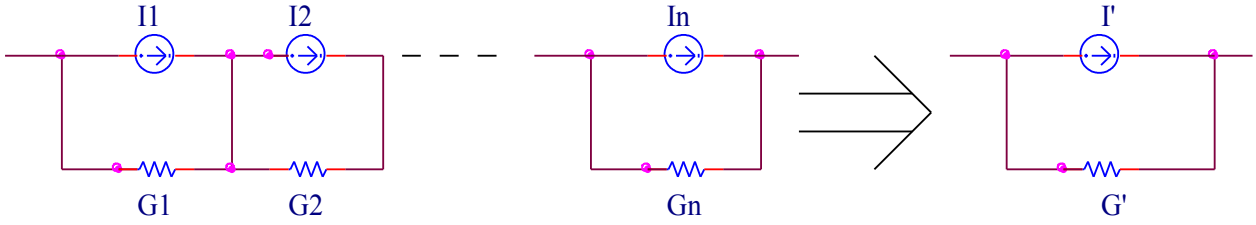


Figure 2.22

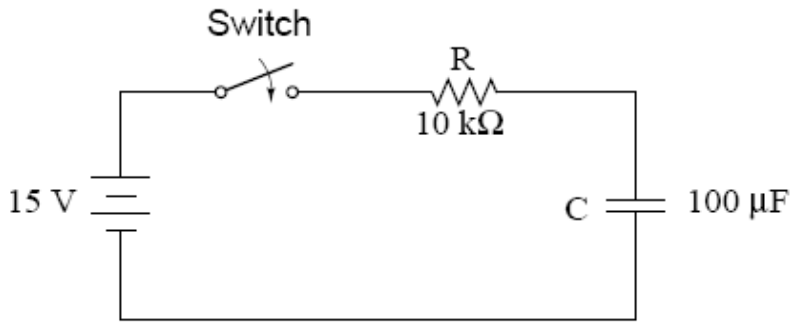
where I'= (I1G1+I1G1+.....InGn)/(R1+R2+.....Rn) 2.75

and G'=1/(R1+R2+.....Rn) 2.76

2.7 Transient circuits

A circuit having constant sources is said to be in steady state if the currents and voltages do not change with time. In a circuit when a d.c. voltage is applied to a capacitor C(or inductor L), and resistor R connected in series, there is a short period of time immediately after the voltage is connected, during which the current flowing in the circuit and voltages across C and R are changing .In a network containing energy storage elements like capacitors and inductors with excitation,the currents and voltages change from one state to other state .These changing values are called transients .The time from the moment the excitation is applied till the steady state is obtained is called the transient time .For a circuit containing storage elements which are independent of sources,the response depends upon the nature of the circuit and is called the *natural response*. The storage elements deliver their energy to the resistances .The resistance by dissipating the energy delivered by the storage elements gradually diminishes the change in response over time till it becomes steady. This is referred to as *transient* response.

2.7.1 DC response of an R-C Circuit



Let us assume the capacitor in the circuit is initially uncharged . When the switch is first closed, the voltage across the capacitor is zero volts; thus, it behaves as a short-circuit drawing maximum initial current . The charging current through the circuit is determined by the difference in voltage between the source and the capacitor, divided by the resistance R .At steady state, the capacitor voltage will rise to equal source voltage, ending in a condition where the capacitor behaves as an open-circuit.

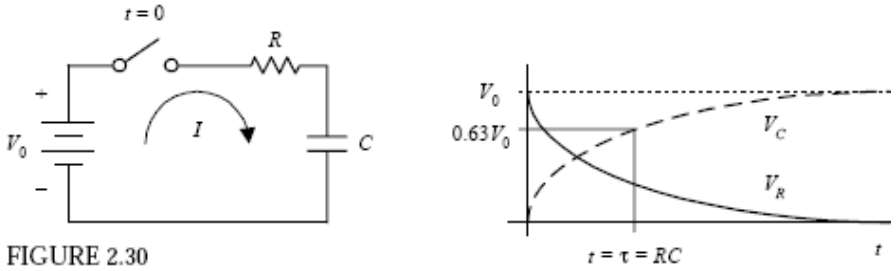


FIGURE 2.30

By applying Kirchhoff's voltage law to the circuit, the following expression is used to figure out how the current will behave after the switch is closed:

$$V_0=IR+1/C\int I\,dt \qquad .....2.77$$

To get rid of the integral, differentiate each term:

$$0=R\,dI/dt+1/C \quad \text{or} \quad dI/dt+(1/RC)I=0 \qquad .....2.78$$

This expression is a linear, first-order, homogeneous differential equation that has the following solution:

$$I=I_0e^{-t/RC} \qquad .....2.79$$

The *RC* term in the equation is called the *RC time constant* ($\tau = RC$).
With the solution for the current now known, the voltage across both the resistor and the capacitor can be found by substituting it into Ohm's law to find V_R and likewise substituting it into the general voltage expression of a capacitor to find V_C:

$$V_R=IR=V_0e^{-t/RC} \qquad .....2.80$$

$$V_c=1/C\int_0^t Idt=V_0(1-e^{-t/RC}) \qquad .....2.81$$

The graph of *VR* and *VC* versus time is shown in figure 2.30

.

2.7.2 DC Response of an R-L Circuit

When the switch is first closed, the voltage across the inductor will immediately rise to source voltage acting as though it were an open-circuit and decay down to zero over time eventually

acting as though it were a short-circuit. Voltage across the inductor is determined by calculating the voltage drop across R, given the current through the inductor, and subtracting that voltage value from the source voltage .

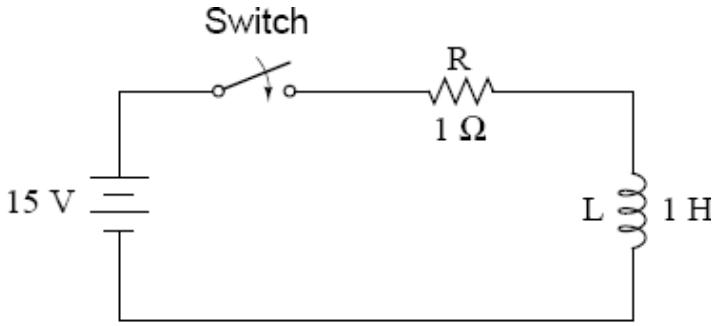
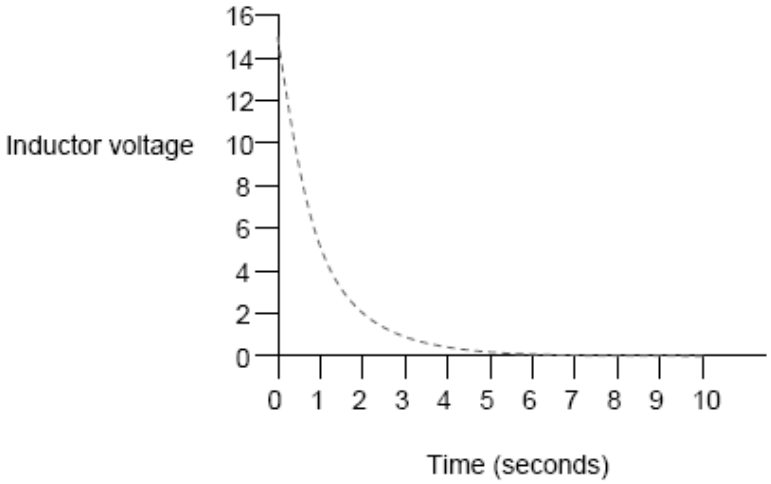
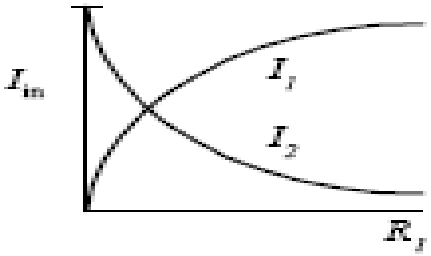
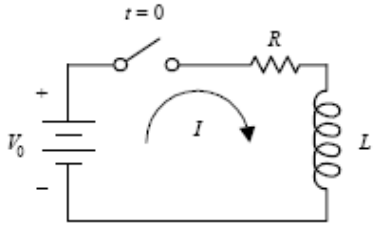


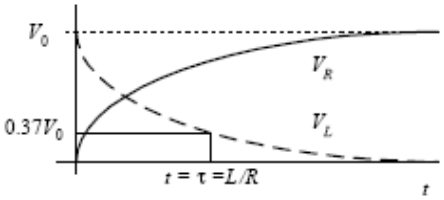
Figure 2.27





When the switch S is closed at *t=0,t*he complete solution for the current is determined





$$V_0=IR + L\,(dI/dt) \quad \text{or} \quad dI/dt +(R/L)I=V_0/L \qquad .....2.82$$

By applying Kirchhoff's voltage law to the circuit, the following expression is used to figure out how the current will behave after the switch is closed:

This expression is a linear, first-order, nonhomogeneous differential equation that has the following solution:

The L/R term is called the RL time constant ($\tau = L/R$).

With the solution for the current now known, the voltage across both the resistor and inductor can be found by substituting it into Ohm's law to find V_R and likewise substituting it into the general voltage expression of an inductor to find V_L:

$$V_R=IR=V_0(1-e^{-Rt/L}) \qquad .....2.83$$

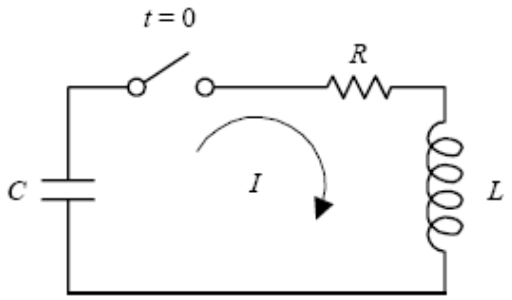
$$V_L=L\,dI/dt= V_0e^{-RtL} \qquad .....2.84$$

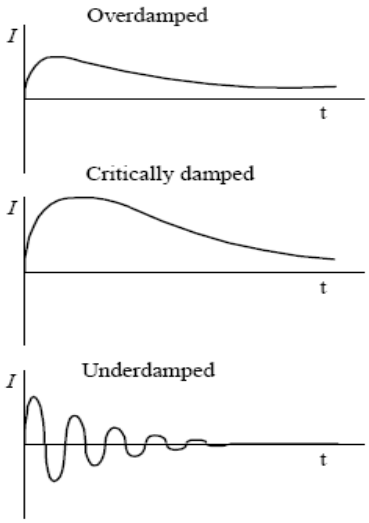
The graph of VR and VL versus time is shown above. Note that when t τ L/R, VL reaches 0.37 V0 or 37% maximum voltage.

2.7.3 RLC Circuit

In the following *RLC* circuit, assume that the capacitor is initially charged and the switch is initially open.

At t=0 the switch is closed .Applying Kirchhoff's voltage law , an expression for the current is





determined as

$$1/C\int I\,dt + IR + L\,dI/dt=0 \quad \text{or} \quad d^2I/dt^2 + R/L\,dI/dt + 1(I)/LC=0 \qquad .....2.85$$

This expression is a linear, second-order, homogeneous differential equation that has the following solution:

$$I = \frac{V_0}{(C_1 - C_2)L}\left(e^{C_1 t} - e^{C_2 t}\right)$$

$$C_1 = -\frac{R}{2L} + \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} \qquad C_2 = -\frac{R}{2L} - \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}$$

.....2.86

The RLC circuit has a unique set of characteristics that are categorized as follows:

When $R^2 > 4L/C$, a condition within the circuit known as *overdamping.*

Taking R^2 to be much greater than $4L/C$, the approximate solution for the current is

$$I \approx \frac{V_0}{R}\left(e^{-t/RC} - e^{-Rt/L}\right)$$

.....2.87

However, when $R^2 = 4L/C$, *the* condition obtained is *critical damping,*

which has the solution

$$I = \frac{V_0 t}{L} e^{-Rt/2L}$$

.....2.88

And finally, if $R2 < 4L/C$, *it* is called *underdamping,* which has the solution

$$I = \frac{V_0}{\omega L} e^{-Rt/2L} \sin(\omega t)$$

.....2.89

Figure 2.32 shows the overdamping, critical damping, and underdamping conditions.The underdamped *RLC* circuit is called a *resonant circuit* because of its oscillatory behavior. The angular frequency ω of oscillation can be related to the cycling frequency *f* by $\omega = 2\pi f$.. In the underdamped case, electrical energy is switching between being stored in the electrical fields within the capacitor and being stored in the magnetic fields about the inductor .However, because of the resistor, energy is gradually lost to heating, so there is an exponential decay. A simple *LC* circuit, without the resistor, is also a resonant circuit.

CHAPTER 3
DIGITAL ELECTRONICS

CONTENTS

3.1 Introduction

3.2 Number systems

3.2.1 Number system representation

3.2.2 Number system conversion

3.2.3 Number formats

3.3 Binary system

3.3.1 Binary arithmetic

3.3.2 Complements

3.3.3 Signed binary numbers

3.4 Logic gates

3.5 Boolean logic

3.6 Minimization of digital circuits

3.6.1 Using Boolean laws and theorems

3.6.2 Karnaugh map method

3.7 Combinational logic circuits

3.7.1 Adder

3.7.2 Subtractor

3.7.3 Decoder

3.7.4 Encoder

3.7.5 Multiplexer

3.7.6 Demultiplexer

3.8 Sequential logic circuits

3.8.1 Flip-flops

3.8.2 Counters

3.8.3 Shift registers

3.9 Memories

3.10 Clock timing parameters

3.11 Logic families

3.11.1 Characteristics

3.11.2 TTL,ECL,CMOS Logic

3.12 Summary

Feedback questions

Data sheets(Shift register,Buffer)

References

DIGITAL ELECTRONICS

3.1 Introduction:

The study of electronics is divided into analog and digital circuits. Digital electronics is concerned with the design of digital electronic circuits. Digital circuits are employed in the design and construction of systems such as digital computers,data communication,data processing,control systems and measurement.

Digital circuits are widely available from simple logic gates in small-scale integrated (SSI) circuits to very complex digital functions in very large-scale integrated (VLSI) circuits. In almost all digital circuits, the semiconductor devices diodes and transistors act as switches and operate in two modes, on or off, carrying current or not carrying current. This is why digital circuits is often called as switching circuits. To effectively use semiconductor devices, it is necessary to have a basic understanding of digital electronics. This chapter focuses on topics ranging from number system to logic families with the goal of providing fundamental concepts that is the prerequisite for IC testing.

3.2 Number systems:

3.2.1 Number system representation:

A number is a mathematical quantity, usually correlated in electronics to a physical quantity such as voltage, current, or resistance. In general, A number with decimal point is represented by a series of coefficients as follows.

$$X_nX_{n-1}X_{n-2}\ldots\ldots X_2X_1X_0 \bullet X_{-1}X_{-2}X_{-3}\ldots\ldots X_{-m+1}X_{-m}$$

Commonly, there are various number systems in use such as decimal number system,binary number system,octal,hexadecimal and the like. The number system is defined by means of the *radix* or *base*. Any positive integer can be chosen as the radix. If the base is r,then r digits(0,1,2,3,....,r-1) are used .The digits are called *Symbols*.

In general, a number expressed in a base r system has coefficients multiplied by powers of r

$$x_n*r^n + x_{n-1}*r^{n-1} + ... + x_2*r^2 + x_1*r^1 + x_0 + x_{-1}*r^{-1} + x_{-2}*r^{-2} + ... + x_{-m+1}*r^{-m} \quad \text{----1.1}$$

where the coefficients x_j range in value from 0 to r-1, r is the base of the number system.

j(=n ,n-1,....2,1,0,-1,-2...-m+1,-m)is the power of the radix and it is equal to the position of the coefficient(place value) .

The decimal number system is said to be of base,or radix,10 because it uses 10 digits(0,1,2,.....,9) and the coefficients are multiplied by powers of 10.From the equation 1.1 the x_j coefficients are any of the 10 digits and the subscript value j gives the place value and,hence,the power of 10 by which the coefficient must be multiplied. This can be expressed as

$10^n x_n+10^{n-1}x_{n-1}+...+10^2x^2+10^1x^1+10^0x_0+10^{-1}x_{-1}+10^{-2}x_{-2}+10^{-m}x_{-m}$ ----------eqn 1.2

The binary system is a another number system. Here the radix r is 2 and the coefficients of the binary number system have only two possible values: 0 or 1. Each coefficient x_j is multiplied by 2^j.

Therefore the binary number system can be expressed as

$2^n x_n+2^{n-1}x_{n-1}+...+2^2x^2+2^1x^1+2^0x_0+2^{-1}x_{-1}+2^{-2}x_{-2}+2^{-m}x_{-m}$ ----------eqn 1.3

Example: 111011,11011.111

The dot (•) in a binary number is called a *binary point* unlike a decimal point in decimal number system .

Similarly for octal number system, the radix is 8 and the symbols are 0,1,2...,7 .and for hexadecimal number systems 16 is the radix and the symbols are 0 to 9 ,A,B,C,D,E and F. The alphabets are used for the digits 10,11,12,13,14 and 15 respectively. The octal and hexadecimal number systems can be expressed as in equation 1.4and 1.5.

$8^n x_n+8^{n-1}x_{n-1}+...+8^2x^2+8^1x^1+8^0x_0+8^{-1}x_{-1}+8^{-2}x_{-2}+8^{-m}x_{-m}$ ----------eqn 1.4

$16^n x_n+16^{n-1}x_{n-1}+...+16^2x^2+16^1x^1+16^0x_0+16^{-1}x_{-1}+16^{-2}x_{-2}+16^{-m}x_{-m}$ ----------eqn 1.5

To distinguish between numbers of different bases, the coefficients are enclosed by parentheses and write a subscript equal to the base used.

eg. $(2413)_8$, $(2459.27)_{10}$,

3.2.2 Number system conversion:

Any radix to decimal conversion:

The conversion of a number from the base r number system to decimal is performed by expanding the number using the general format shown in equation 1.1 and adding all the terms to get the decimal form .

Example: convert $(11001101)_2$ to decimal form:

bits = 1 1 0 0 1 1 0 1

weight = (1*2^7)+(1*2^6)+(0*2^5)+(0*2^4)+(1*2^3)+(1*2^2)+(0*2^1)+(1*2^0)=$(205)_{10}$

Convert $(101.011)_2$ to decimal form:

bits = 1 0 1 . 0 1 1

weight = (1*2^2)+(0*2^1)+(1*2^0)+(0*2^-1)+(1*2^-2)+(1*2^-3)=$(5.375)_{10}$

Example: Convert the octal number $(245.37)_8$ in to decimal.

octal digits = 2 4 5 . 3 7

weight = (2*8^2) (4*8^1) (5*8^0). (3*8^-1) (7*8^-2) = $(165.484375)_{10}$

Example: Convert the hexadecimal number $(30F.A9)_{16}$ to decimal:

hexadecimal

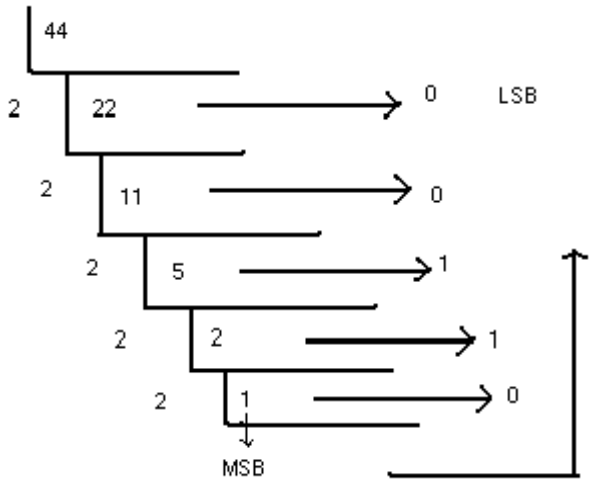
digits = 3 0 F . A 9

weight = (3*16^2) (0*16^1) (F*16^0) . (A*16^-1) (9*16^-2) = $(783.66015625)_{10}$

Decimal to any radix conversion:

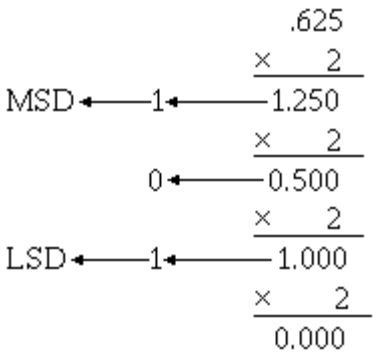
The decimal integer can be converted into an equivalent base r number by continually dividing the decimal number by the base (r) and arranging the remainders from last to first. The conversion of a decimal fraction to radix r is accomplished by continually multiplying the fractional part alone by the radix r and arranging the integer parts from first to last until multiplication can not be continued or until desired accuracy is reached.

Example: convert the decimal number $(44)_{10}$ into binary form



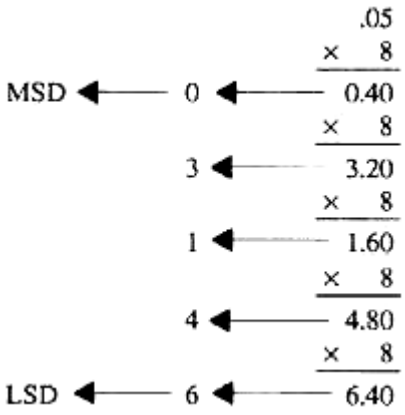
The binary bits are assembled from the remainders of the successive division steps, beginning with the MSB and proceeding to the LSB. In this case, we arrive at a binary notation of $(101100)_2$.

Example: Convert the decimal fraction number 0.625 to binary form



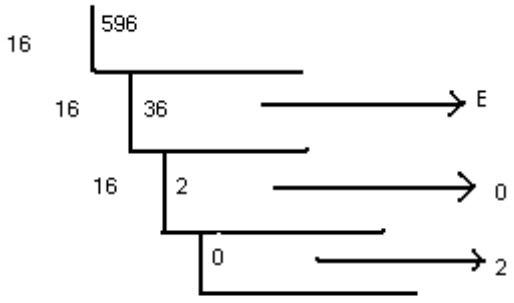
$.625_{10}$ equals $.101_2$

Example: convert the decimal number 0.05 to octal form



Write the solution from MSD to LSD $=0.03146_8$

Example: Convert the number 590 Decimal to Hexadecimal



$(590)_{10} = (20E)_{16}$

Binary to octal / hexadecimal conversion:

Binary equivalent of each octal digit corresponds to three binary digits and each hexadecimal digit corresponds to four binary digits. To convert binary to octal , group the bits in three's, from the binary point left, and from the binary point right, adding (implied) zeros as necessary to make complete 3-bit groups. Each octal digit was translated from the 3-bit binary groups. Similarly for binary to hexadecimal conversion, group the bits in four's.

Example:Convert $(10110111.1)_2$ to octal:

010 110 110 .101

Convert each group of bits ---- ---- ---- . ----

to its octal equivalent: 2 6 6 . 5

$(10110110.101)_2 = (266.5)_8$

Example: Convert $(10110111.1)_2$ to hexadecimal:

1011 0110 .1000

Convert each group of bits ------ ------ . ------

to its hexadecimal equivalent: B 6 . 8

$(10110110.1)_2 = (B6.8)_{16}$

Likewise, the conversion from either octal or hexadecimal to binary is done by taking each octal or hexadecimal digit and converting it to its equivalent binary (3 or 4 bit) group, then putting all the binary bit groups together.

3.2.3 Number formats:

Fixed and floating point numbers:

A fixed-point number has fixed number of digits after and before the radix point . In positional notation, it is represented as

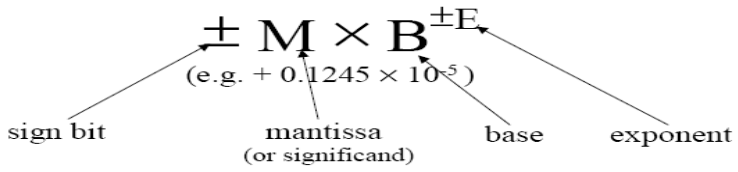
$B = b_{n-1}b_{n-2}\ldots b_1 b_0 \bullet b_{-1}b_{-2}\ldots b_{-k}$ with a corresponding value of

$$V(B) = \sum_{i=-k}^{n-1} b_i \times 2^i$$

The position of the radix point is assumed to be fixed. Fixed-point numbers have a range that is limited by the significant digits used to represent the number. Logic circuits that deal with fixed-point numbers are essentially the same as those used for integers

A floating point number has its radix point floating i.e., it can be placed anywhere relative to the significant digits of the number. The advantage of floating-point representation over fixed-point and integer representation is that it can support a much wider range of values with high precision. The numbers are represented by a *mantissa* comprising the significant digits and an *exponent* of the radix *r.*

The format is *Mantissa* x *R Exponent*



The numbers are usually *normalized* such that the radix point is placed to the right of the first non-zero digit. For example, 5.234×10^{43} or 3.75×10^{-35}.

IEEE 754 - Floating point representation standard:

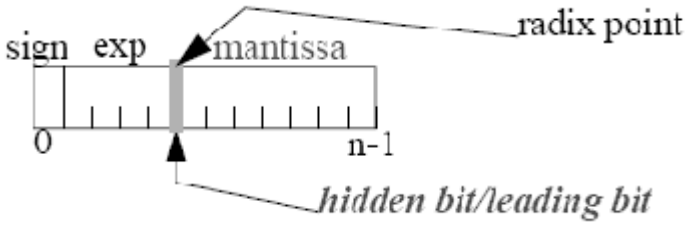


Figure 3.1 IEEE 754 standard of floating point representation

	Single	Double
Mantissa bits(incl hidden)	24	53
sign bits	1	1
exponent bits	8	11
Bias/Excess	127	1023
Max Exponent	127	1023
Min Exponent	-126	-1022

Table 3.1 IEEE Floating point standard

3.3 Binary system:

3.3.1 Binary arithmetic:

addition:

The basic rules for the addition of binary bits are as follows:

0 + 0 = 0

1 + 0 = 1

0 + 1 = 1

1 + 1 = 10

Just as with decimal addition, when the sum in one column is a two-bit (two-digit) number, the least significant figure is written as part of the total sum and the most significant figure is "carried" to the next left column.

Consider the following examples:

```
                    11    1 <------Carry bits----->11
  1001100          1001101            1000110
+ 0010011        + 0011001          + 0010110
  -------          -------            -------
  1011111          1100110            1011100
```

Subtraction:

The basic rules for the subtraction of binary bits are as follows

0 - 0 = 0

0 - 1 = 1, and borrow 1 from the next more significant bit

1 - 0 = 1

1 - 1 = 0

Example: $(10110110)_2 - (10100100)_2$

```
    10110110
    10100100
    --------
    00010010
    --------
```

Multiplication:

The basic rules for the multiplication of binary bits are as follows

0 x 0 = 0

0 x 1 = 0

1 x 0 = 0

1 x 1 = 1, and no carry or borrow bits

Multiplication is achieved by adding a list of shifted multiplicands according to the digits of the multiplier.

```
      1 0 1 1 multiplicand (4 bits)
    X 1 1 0 1 multiplier (4 bits)
  ------------------
      1 0 1 1
    0 0 0 0
  1 0 1 1
1 0 1 1
--------------------
1 0 0 0 1 1 1 1
--------------------
```

Division:

To perform binary division follow the following steps:

1. Align the divisor (Y) with the most significant end of the dividend. Let the portion of the dividend from its MSB to its bit aligned with the LSB of the divisor be denoted X.

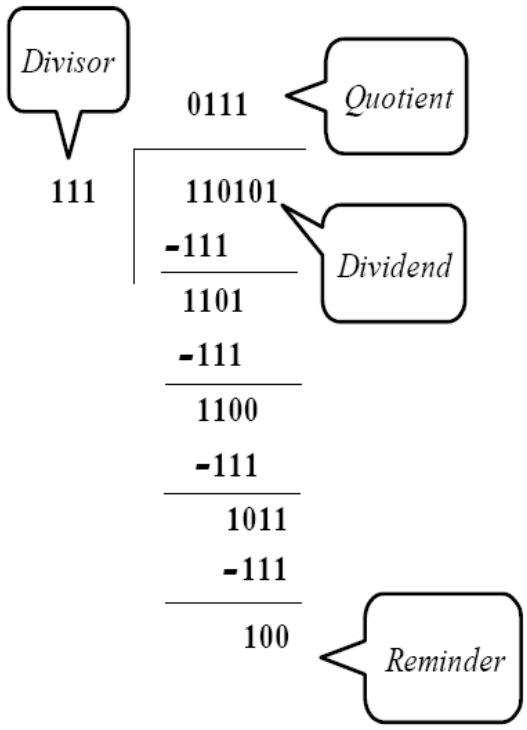
2. Compare X and Y.

a) If X >= Y, the quotient bit is 1 and perform the subtraction X-Y.

b) If X < Y, the quotient bit is 0 and do not perform any subtractions.

3. Shift Y one bit to the right and go to step 2.

Example:



X		Y	Q	Action
110	<	111	$Q_1 = 0$	Do Not Subtract
1101	>	111	$Q_2 = 1$	Subtract
1100	>	111	$Q_3 = 1$	Subtract
1011	>	111	$Q_4 = 1$	Subtract

3.3.2 Complements:

Diminished radix complement and radix complement:

The diminished radix complement of a number is obtained by subtracting each digit of the number from the highest symbol of the radix.

On the other hand the radix complement of a number is obtained by:

1.subtracting each digit of the number from the highest symbol of the radix.

2.Adding '1' to the result of step 1.

Example: In case of Binary number systems

The 1's complement (diminished radix complement) of a binary number is obtained by:

1.Replacing all the 1's with 0's and the 0's with 1's.

The 2's complement(radix complement) of a binary number is obtained by:

1.Replacing all the 1's with 0's and the 0's with 1's.(one's complement)

2. Adding 1 to this number.

2's complement addition rules:

Add magnitudes, ignore end around carry. The sign of the result is the sign found by the result addition.

2's Complement subtraction :

To subtract a smaller number from a larger one, the 2's complement method is applied as follows.

1. Determine the 2's complement of the smaller number.
2. Add the 2's complement to the larger number.
3. Discard the carry (there is always a carry in this case).

To subtract a larger number from a smaller one, the 2's complement method is applied as follows.

1. Determine the 2's complement of the larger number
2. Add the 2's complement to the smaller number.
3. There is no carry. The result is in 2's complement form and is negative.
4. To get an answer in true form ,take the 2's complement of the result and change its sign.

Example:

The decimal subtraction 29 - 7 = 22 is the same as adding (29) + (-7) = 22

```
  00000111   (decimal 7)
  11111000   (one's complement)
+ 00000001   (add 1)
  11111001   (two's complement)
```

11111001 now represents -7.

1. Add

29	00011101
+ -7	11111001
22	(1)00010110

Note that the final carry 1 is ignored.

3.3.3 Signed binary numbers:

Unsigned arithmetic does not involve a sign-bit, and therefore can express larger absolute numbers, because the MSB is merely an extra digit rather than a sign indicator. However to represent negative numbers we need to introduce a sign bit.

For example the subtraction of 7 - 5 is essentially the same as the addition 7 + (-5). The minus 5 can be represented by placing a sign bit directly to the left of the most significant digit.

$(0101)_2 = (5)_{10}$ (positive)

$(1101)_2 = (5)_{10}$ (negative)

Extra bit, representing sign (0=positive, 1=negative)

The signed binary systems are not useful for digital computing. Hence signed binary complement systems has been developed.

3.4 Logic gates:

Logic gates are electronic circuits that operate on one or more input signals to produce an output signal. Electrical signals such as voltages or currents exist throughout a digital system in either of two recognizable values. Voltage-operated circuits respond to two separate voltage levels that represent a binary variable equal to logic 1 or logic 0. The basic gates are AND,OR,NOT

AND gate:

The AND operation is mathematically defined as the product of two Boolean values. A truth table provides a direct mapping between the possible inputs and outputs. A basic AND operation has two inputs with four possible combinations of output .It is listed in the truth table 3.2. From the truth table it is clear that **"the AND gate output will be a low if any one of the inputs is low and the output is high when all the inputs are high".** The logic symbol for an AND gate is shown in the figure 3.2. The example for an AND gate is IC DM 74LS08. It is a quad two input ,dual-in-line package IC and its pin detail is shown

in the figure 3.3. The logic expression for the AND gate output is expressed as

Output = AB

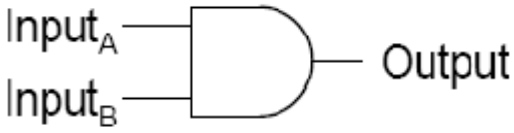


Figure 3.2 AND gate symbol

A	B	A AND B
0	0	0
0	1	0
1	0	0
1	1	1

Table 3.1 Truth table for AND gate

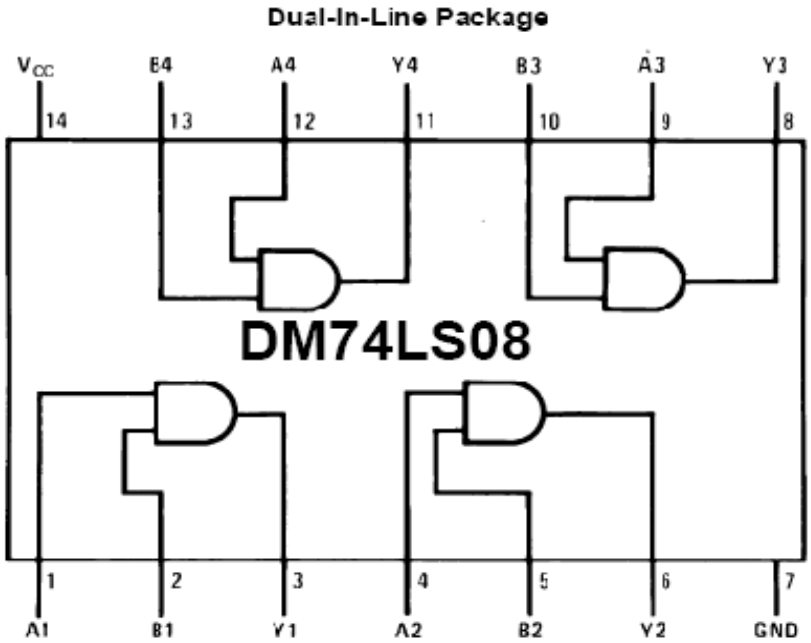


Figure 3.3 Pin configuration of the AND gate DM74LS08

OR gate:

The OR operation is mathematically defined as the logical sum of two Boolean values and its operation in Boolean algebra is listed in table 3.3. From the truth table it is clear that "**the OR gate output will be a high if any one of the inputs is high and the output is low when all the inputs are low".** The logic expression for the OR gate output is expressed as

output = A+B

The logic symbol for an OR gate is shown in the figure 3.4. The example for an OR gate is IC DM 74LS32. It is a quad two input ,dual-in-line package IC and its pin detail is shown in the figure 3.5.

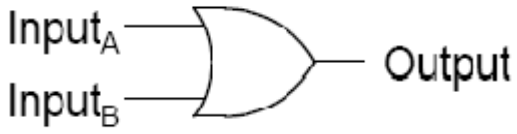


Figure 3.4 OR gate symbol

A	B	A OR B
0	0	0
0	1	1
1	0	1
1	1	1

Table 3.3 Truth table for OR gate

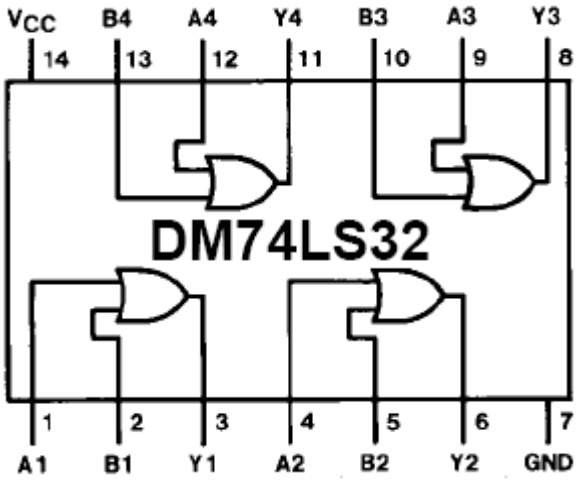


Figure 3.5 Pin configuration of the OR gate DM74LS32

NOT gate:

The NOT gate returns the complement of the input: i.e.1 becomes 0, and 0 becomes 1 and its operation is described in the truth table shown in table 3.4. NOT is a unary operator , meaning that it requires only one operand. But AND and OR are referred to as binary operators, because they require two operands. The logic symbol for an NOT gate is shown in the figure 3.6. The example for an NOT gate is IC DM 74LS04. It is a Hex inverter, dual-in-line package IC and its pin detail is shown in the figure 3.7.

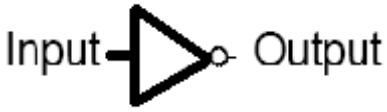


Figure 3.6 NOT gate symbol

Input	Output
0	1
1	0

Table 3.4 Truth table for NOT gate

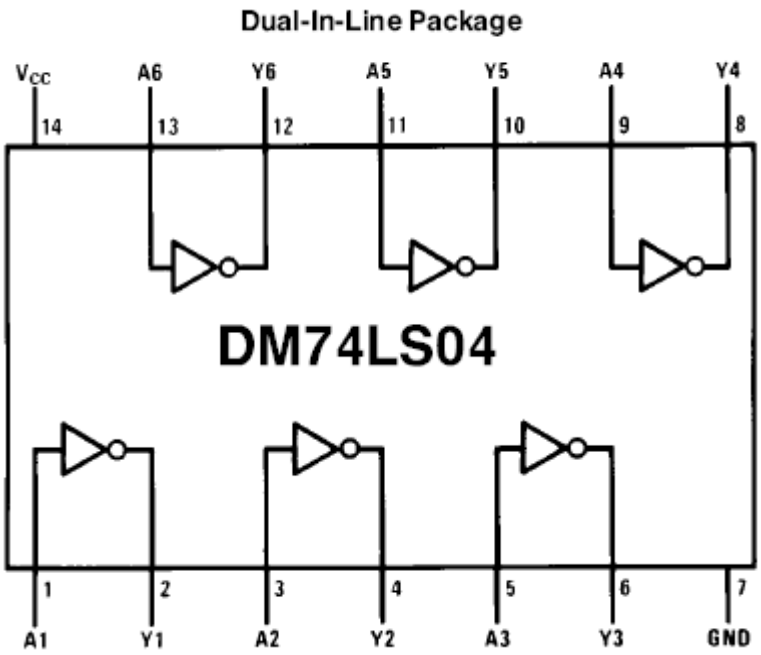


Figure 3.7 Pin configuration of the NOT gate DM74LS04

NAND gate:

The operation of a NAND gate is described in the truth table and it is shown in the table 3.5. From the truth table it is clear that **"the NAND gate output will be high when any of the inputs are low and will be low when all the inputs are high."** The logic expression for the OR gate output is expressed as

output = (AB)'

The logic symbol and equivalent circuit for a NAND gate are shown in the figures 3.8 figure 3.9 respectively. The example for a NAND gate is IC DM 74LS00. It is a quad two input ,dual-in-line package IC and its pin detail is shown in the figure 3.10.

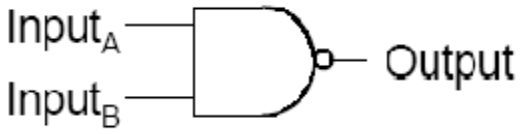


Figure 3.8 NAND gate symbol

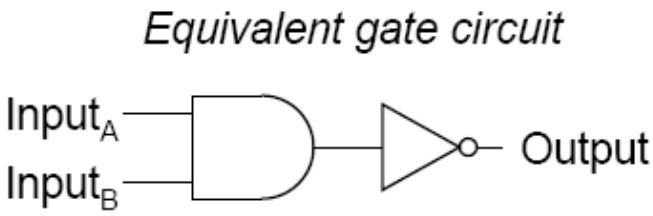


Figure 3.9 NAND gate equivalent circuit

A	B	Output
0	0	1
0	1	1
1	0	1
1	1	0

Table 3.5 Truth table for NAND gate

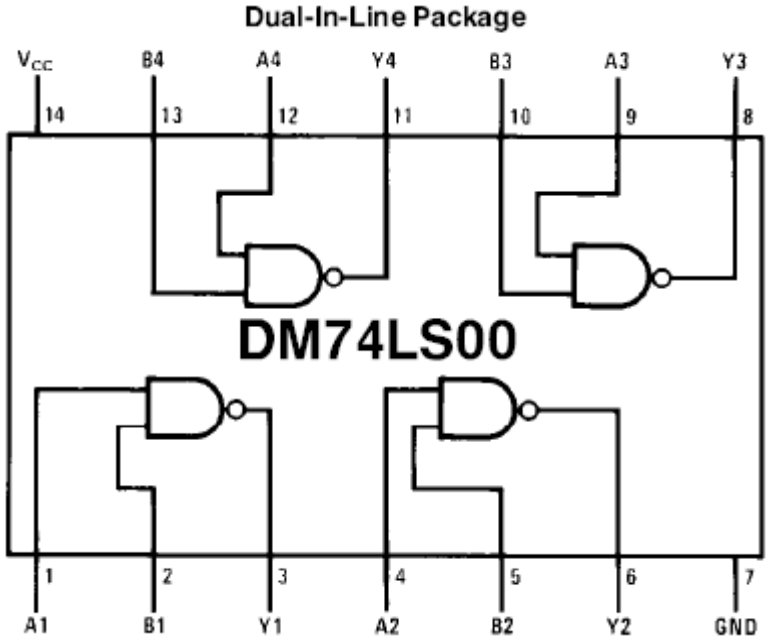


Figure 3.10 Pin configuration of the NAND gate DM74LS00

NOR Gate:

The operation of a NOR gate is described in the truth table shown in table 3.6. From the truth table it is clear that **"the NOR gate output will be high when all the inputs are low and if any one of the input is high output is low."** The logical expression for the NOR gate output is expressed as

output = (A+B)'

The logic symbol and equivalent circuit for a NOR gate are shown in the figures 3.11 figure 3.12 respectively. The example for a NOR gate is IC DM 74LS02. It is a quad two input ,dual-in-line package IC and its pin detail is shown in the figure 3.13.

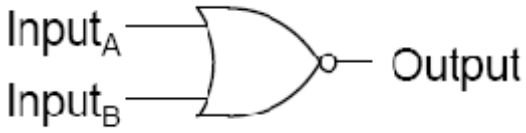


Figure 3.11 NOR gate symbol

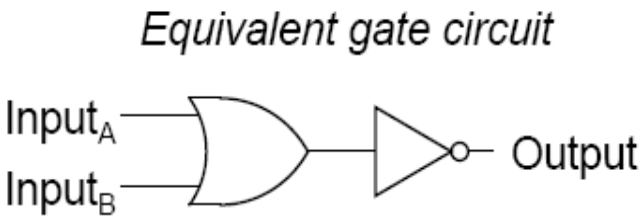


Figure 3.12 NOR gate equivalent circuit

A	B	Output
0	0	1
0	1	0
1	0	0
1	1	0

Table 3.6 Truth table for NOR gate

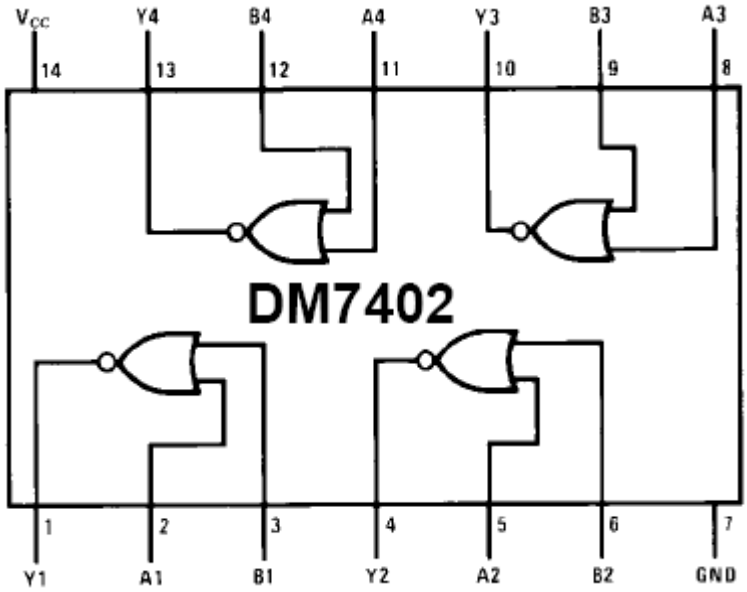


Figure 3.13 Pin configuration of the NOR gate DM74LS02

Exclusive-OR Gate:

The operation of Exclusive-OR (EX-OR) gate is described in the truth table shown in table 3.7. From the truth table **Ex-OR outputs a high logic level if odd number of inputs are at logic high levels. Conversely, it outputs a low logic level if even number of inputs are at the logic high levels.** The logical expression for the EX-OR gate output is expressed as

Output = A'B + AB'

The logic symbol for an EX-OR gate are shown in the figures 3.14 . The example for an EX-OR gate is IC DM 74LS86. It is a quad two input ,dual-in-line package IC and its pin detail is shown in the figure 3.15

Figure 3.14 EX-OR gate symbol

A	B	Output
0	0	0
0	1	1
1	0	1
1	1	0

Table 3.7 Truth table for EX-OR gate

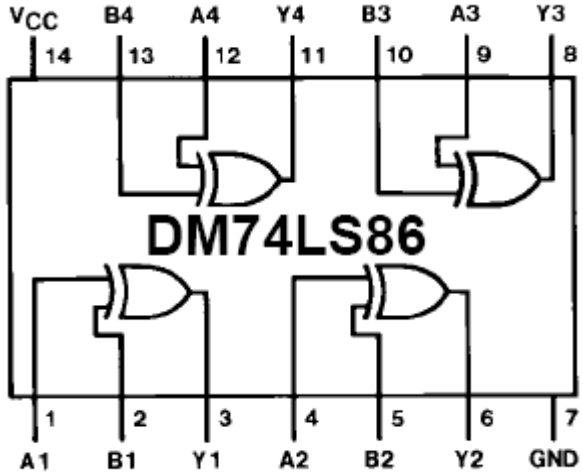


Figure 3.15 Pin configuration of the EX-OR gate DM74LS86

XNOR Gate:

XNOR gate is equivalent to an Exclusive-OR gate with an inverted output. The truth table is described in the table shown in table 3.8 . The schematic symbol for XNOR gate and its equivalent circuit are shown in the figure 3.16, figure 3.17 respectively. The logical expression of this gate is defined as

output = A'B' + AB

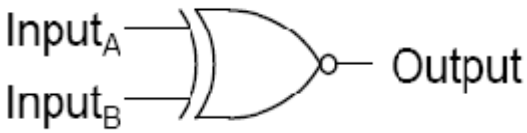


Figure 3.16 XNOR gate symbol

A	B	Output
0	0	1
0	1	0
1	0	0
1	1	1

Table 3.8 Truth table for XNOR operation

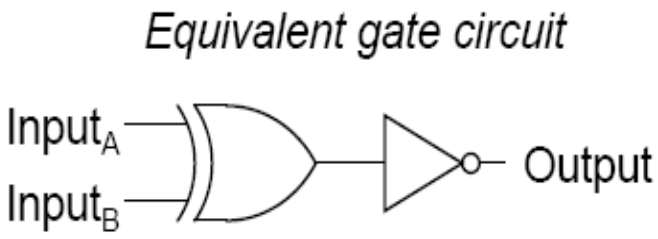
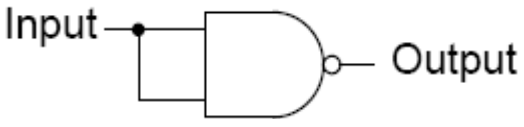


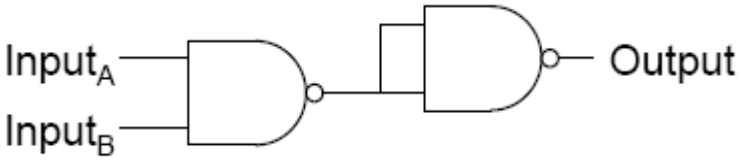
Figure 3.17 XNOR gate equivalent circuit

Universal gates:

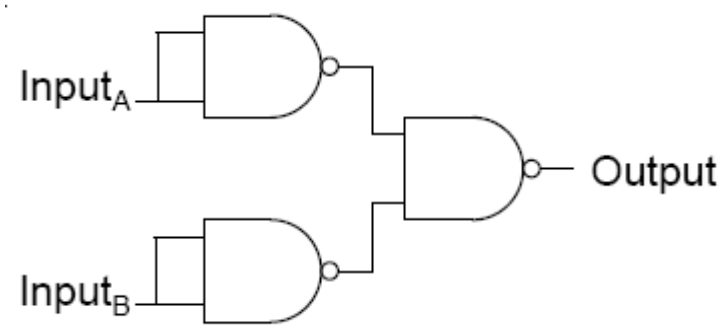
NAND and NOR gates are known as universal gates because either one of these gates can be used to construct any other logic gate or function. The gates of NOT,AND,OR and NOR function are implemented by using only NAND gate as shown in the figure 3.18a to 3.18d. Similarly NOT,AND,OR,NAND function implementation is done by using only NOR gate as shown in the figure 3.19a to 3.19.d.



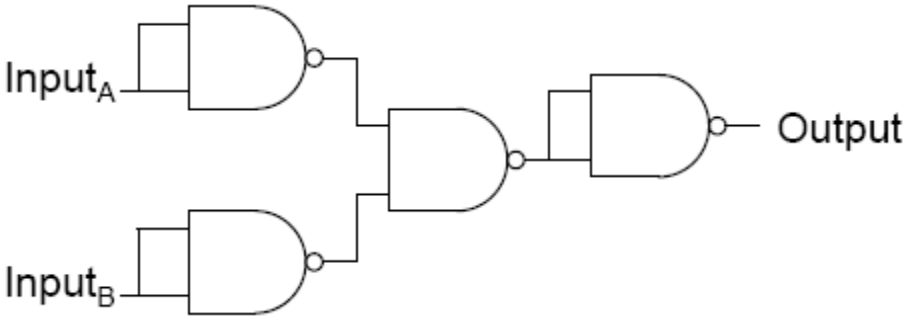
a)



(b)

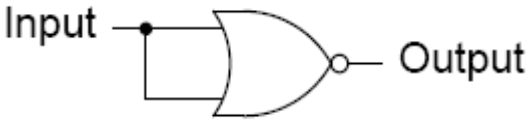


(c)

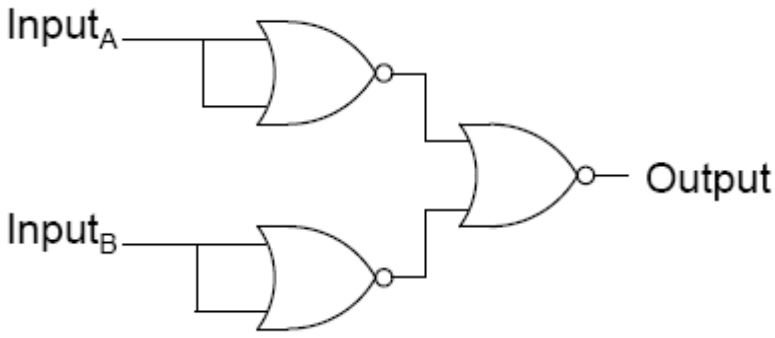


(d)

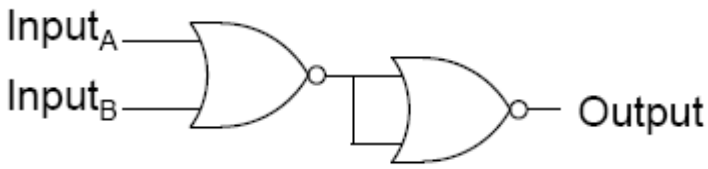
Figure 3.18 Implementation of a)NOT b)AND c)OR d)NOR using NAND gate



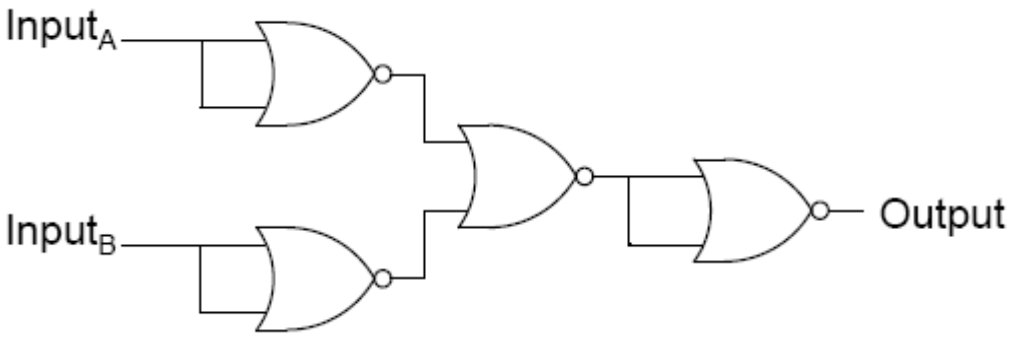
(a)



(b)



(c)



d)

Figure 3.19 Implementation of a)NOT b)AND c)OR d)NAND using NOR gate

3.5 Boolean logic:

Boolean algebra in digital systems deals with binary variables and logical operations. A boolean function consists of binary variables,0's and 1's and logical operators or symbols. Boolean logic is a branch of mathematics that was discovered in the nineteenth century by an English mathematician named George Boole. The basic theory of this boolean logic is that the logical relationships can be modeled by algebraic equations. The basic laws and theorems of boolean algebra are explained as follows.

Boolean laws and theorems:

Boolean laws and theorems are helpful to simplify the boolean expression and also to construct a complex digital circuit with less number of logic gates.

Basic Boolean algebraic identities

Additive

A + 0 = A

A + 1 = 1

A + A = A

A + A' = 1

Multiplicative

0A = 0

1A = A

AA = A

AA' = 0

Boolean algebra also has commutative, associative, and distributive properties as listed below:

Commutative: A * B = B * A and A + B = B + A

Associative: (A * B) * C = A * (B * C) and (A + B) + C = A + (B + C)

Distributive: A * (B + C) = A * B + A * C

Two variable theorem:

A + AB = A

A + A'B = A + B

A + A'B = A + B, can be illustrated using the truth table shown in Table 3.9

A	B	$\overline{A}B$	$A + \overline{A}B$	A + B
0	0	0	0	0
0	1	1	1	1
1	0	0	1	1
1	1	0	1	1

Table 3.9 Truth table for the Boolean expression A + A'B = A + B

Demorgan's theorem:

Theorem 1:

This theorem states that the complement of a sum is equal to the product of the individual complements. Expressed in equation form

(A+B)'=A'.B'

Theorem 2:

This theorem states that the complement of a product is equal to the sum of the individual complements. Expressed in equation form

(A.B)'=A'+B'

3.6 Minimization of digital circuits:

Simplification of Boolean functions is mainly used to reduce the gate count of a digital circuit. Less number of gates means less power consumption, less cost, fast performance,etc. There are several ways to represent a logic design and some of them are explained below.

Sum of product(SOP) form:

Let Z=AB+BC+CA . In the above expression, each term is a product of two variables. This expression is the sum of three product terms. But it is not a standard SOP form. A expression is said to be a S*tandard SOP* form when all product terms have all the variables. example. Z=ABC+ABC'+A'B'C'+A'BC'. Each term of a standard SOP expression is termed as a *Minterm*.

Product of Sum form:

Let Z=(A+B)(B+C)(C+A) . In the above expression, each term is a sum of two variables. This expression is the product of three sum terms. But it is not a standard POS form. A expression is said to be a S*tandard POS* form when all sum terms have all the variables. example. Z=(A+B+C)(A+B+C')(A'+B'+C')(A'+B +C').Each term of a standard POS expression is termed as a *Maxterm*.

3.6.1 Using boolean laws and theorems:

A digital circuit can be simplified using boolean laws and theorems. These laws and theorems can be used wherever necessary and it is explained with the following examples.

Example 1: Simplify: C + (BC)':

Expression	Rule(s) Used
C + (BC)'	Original Expression
C + (B' + C')	DeMorgan's Law.
(C + C') + B	Commutative, Associative Laws.
1 + B	Compliment Law.
1	Identity Law.

Example 2: Simplify: (AB)'(A' + B)(B' + B):

Expression	Rule(s) Used
(AB)'(A' + B)(B'+B)	Original Expression
(AB)'(A' + B)	Compliment law, Identity law.

(A' + B')(A + B)	DeMorgan's Law
A' + B'B	Distributive law. This step uses the fact that or distributes over and. It can look a bit strange since addition does not distribute over multiplication.
A'	Compliment, Identity.

3.6.2 Karnaugh map method:

Karnaugh map(K-map) method is a powerful technique to simplify Boolean expression more quickly and easily compared to Boolean algebra. This method is a systematic method.. Steps to be followed to simplify a boolean expression are as follows.

1. K-map format

   The K-map consists of 2^n cells if the total number of variables in a given boolean expression is n.

2. Plotting

   Once a Boolean expression is in the SOP form, convert it to standard SOP form. Now plot the K-map by placing a 1 in each cell corresponding to a minterm in the standard -SOP.

3. Grouping

   a. The 1s in adjacent cells must be combined in groups of 1,2,4,8,16 and so on. (2^n where n=0,1,..,k)

   b. Each group should consist of the largest number of adjacent cells possible.

   c. Every '1' on the map must be included in at least one group.

   d. There can be overlapping of groups if they include common 1's.

4. Simplifying the expression

   Each group of 1's creates a product term. A pair of adjacent 1's eliminates one variable. A quad(4 adjacent 1's)eliminates two variables. An octet(group of eight 1's) eliminates three variables and so on. The final simplified expression is formed by summing the product terms of all the groups.

The simplification using K-map method is explained with examples under the following.

Example:1

Obtaining of simplified Boolean expression using K-map method for the Truth table 3.10 as given below.

A	B	Output
0	0	0
0	1	1
1	0	1
1	1	1

Table 3.10

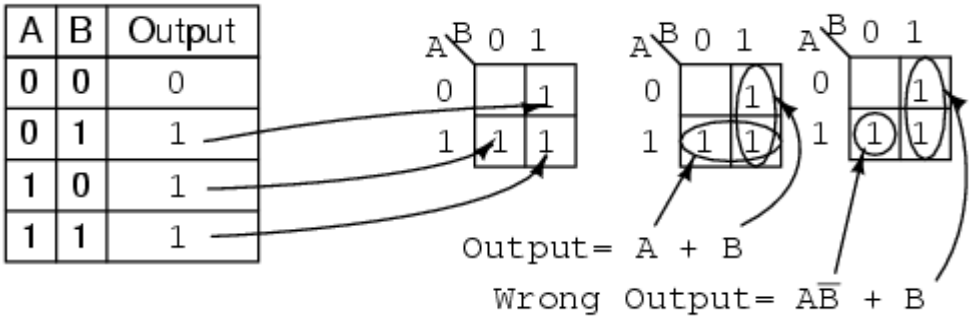


Figure 3.20 Example of mapping a truth table on a karnaugh map

Example:2

Simplification of the given Boolean expression using Karnaugh map

Output=A'B + AB' + AB

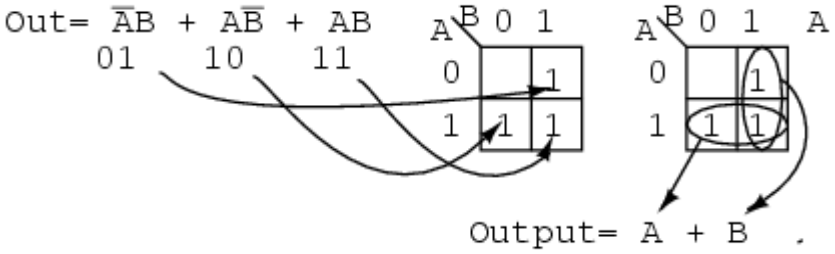


Figure 3.21Example of plotting a boolean expression on a karnaugh map

Output=A+B

Example:3

Simplify the logic diagram below shown in the figure 3.22 using K-map

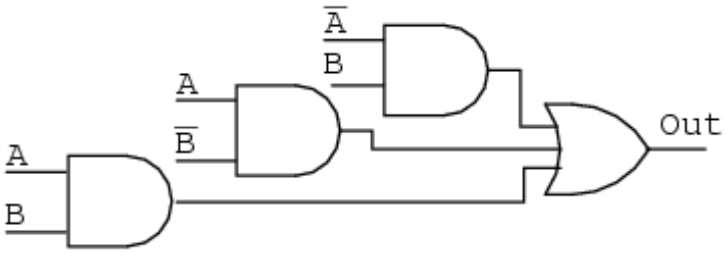


Figure 3.22

Solution:

The solution shown in the figure 3.23

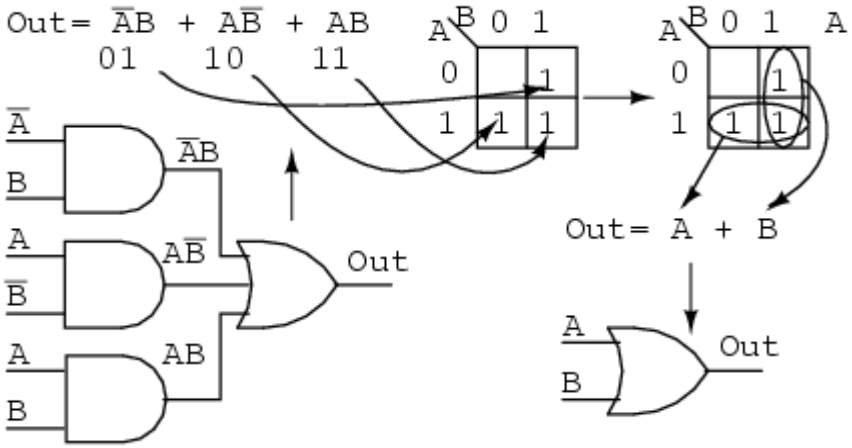


Figure 3.23

Example :4

$$Out= \overline{A}\,\overline{B}CD+\overline{A}BCD+ABCD+A\overline{B}CD+AB\overline{C}\,\overline{D}+AB\overline{C}D+ABC\overline{D}$$

The solution of above expression shown in the figure 3.24

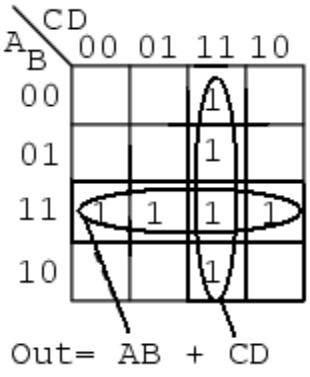


Figure 3.24

3.7 Combinational logic circuits:

A combinational circuit consists of logic gates whose outputs at any time are determined from the present combination of inputs. A combinational logic circuit performs an operation that can be specified logically by a set of Boolean functions. Some of the combinational logic circuits such as Adder, Subtractor,Decoder,Encoder ,Multiplexer and demultiplexer.

3.7.1 Adder:

Half adder:

Half adder is a circuit that can add only two binary digits. The outputs of this circuit are sum and carry. Truth table of half adder is shown in the table 3.11.The logical diagram of half-adder circuit is shown in the figure 3.25

A	B	Σ	C_{out}
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

Table 3.11 Truth table for Half adder

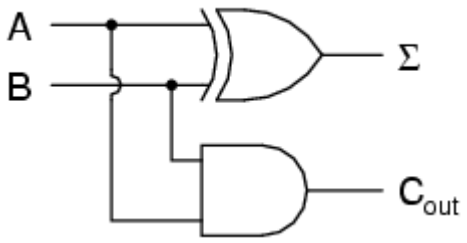


Figure 3.25 Half adder logic diagram

Sum = A'B + AB' , C_{out} = AB

Full adder:

A full adder is a logical circuit that performs an addition operation on two binary values taking into account the value of previous carry. The full adder produces a sum(S) and carry(Co) value. Two of the input variables, denoted by A and B,represent the two significant bits to be added. The third input,C_I,represents the carry from the previous lower significant position. The truth table of the full adder is listed in table 3.12

$$\Sigma = (A \oplus B) \oplus C_I$$

$$C_O = AB + (A \oplus B)\, C_I$$

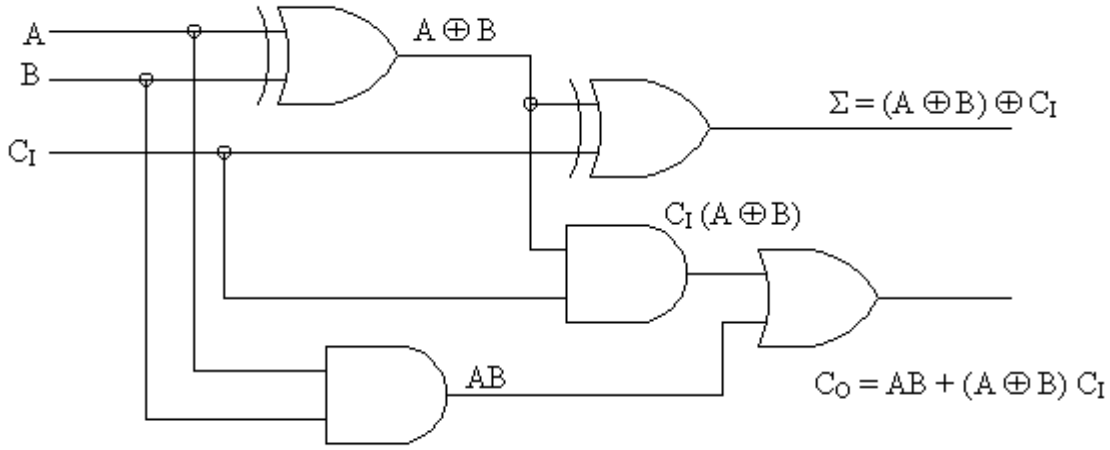


Figure 3.26 Full adder logic diagram

Input			Output	
A	B	C_I	C_O	S
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

Table 3.12 Truth table for Full adder

3.7.2 Subtractor:

Half-Subtractor:

The half-subtractor is a combinational circuit which performs subtraction of two bits. It has two inputs, A (minuend) and B (subtrahend) and two outputs D (difference) and B (borrow). The truth table of the half-subtractor is listed in table 3.13

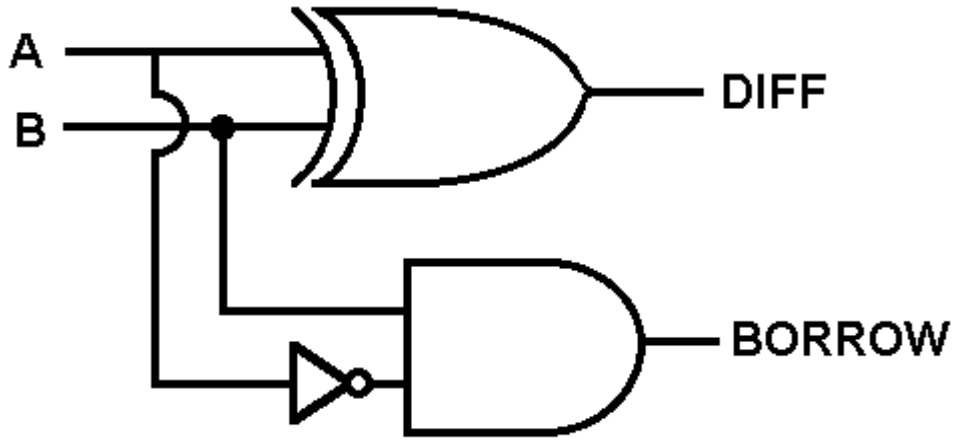


Figure 3.27 Half subtractor logic diagram

A	B	DIFF	BORROW
0	0	0	0
1	0	1	0
0	1	1	1
1	1	0	0

Table 3.13 Truth table for Half subtractor

Full-Subtractor:

The full-subtractor is a combinational circuit which performs subtraction of two bits taking into account the previous borrow. The truth table of the full-subtractor is listed in table 3.14. From the truth table, It has three inputs, A and B and BOR_{IN} and two outputs D (difference) and BOR_{OUT} (borrow).

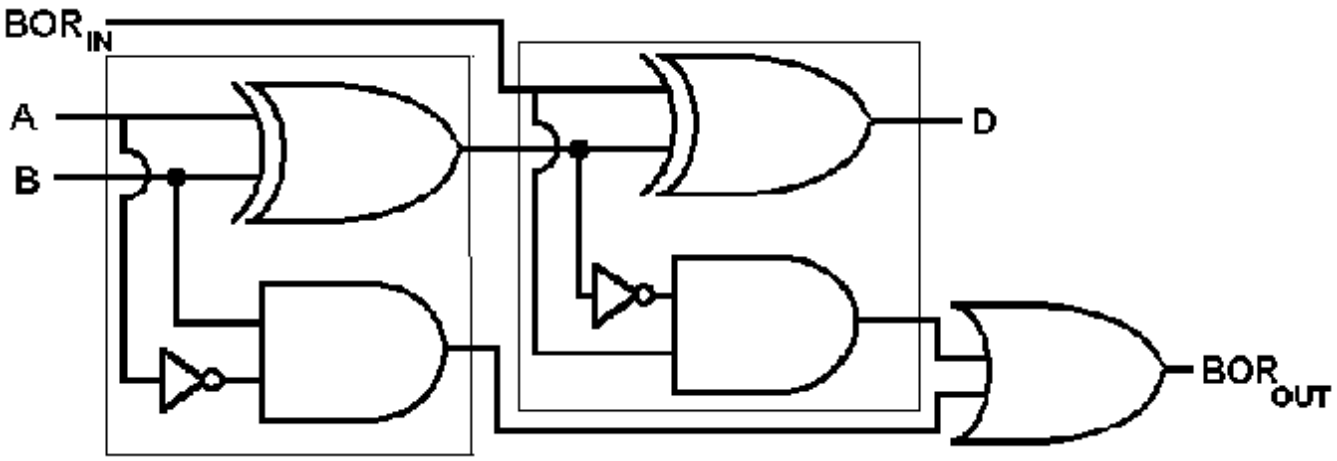


Figure 3.28 Full subtractor logic diagram

A	B	BOR_{IN}	D	BOR_{OUT}
0	0	0	0	0
0	0	1	1	1
0	1	0	1	0
0	1	1	0	0
1	0	0	1	1
1	0	1	0	1
1	1	0	0	0
1	1	1	1	1

Table 3.14 Truth table for Full subtractor

3.7.3 Decoder:

A decoder is a combinational circuit that converts binary information from n input lines to a maximum of 2^n unique output lines. e.g. BCD decoders. The truth table of 2-to-4 line decoder is listed in table 3.15

A_1	A_0	D_3	D_2	D_1	D_0
0	0	0	0	0	1
0	1	0	0	1	0
1	0	0	1	0	0
1	1	1	0	0	0

Table 3.15 Truth table for Decoder

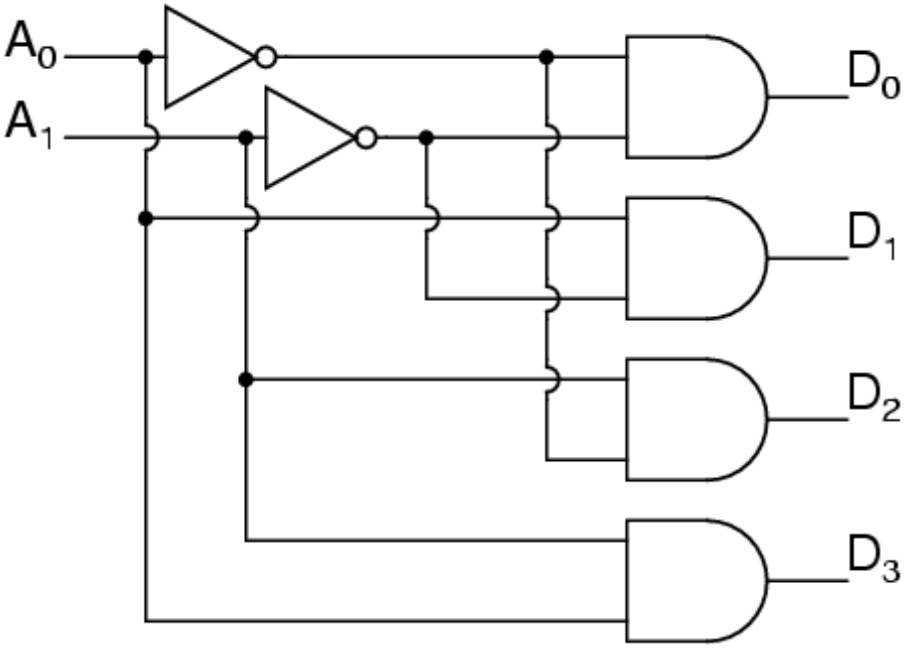


Figure 3.29 Logic diagram for 2 to 4 Decoder

The logical diagram for 2-to-4 line decoder shown in the figure 3.29. Here A0 and A1 can select any of the four outputs. Decoders are commonly used to address memories and to convert one code to another.

3.7.4 Encoder:

An encoder is a circuit that performs the inverse operation of a decoder. An encoder has 2^n or fewer input lines and n output lines. The binary code is generated as the output of a encoder corresponding to the input value. The encoder has the limitation that only one input can be active at any given time. If two inputs are simultaneously active, the output produces an undefined combination.

The Decimal-to-BCD Encoder:

The logic symbol for the 10-line-to-4-line Encoder is presented in Figure 3.30 and the associated conversion table listed in Table 3.16 with A_3 representing the most significant bit. Logic diagram of Encoder is shown in figure 3.31

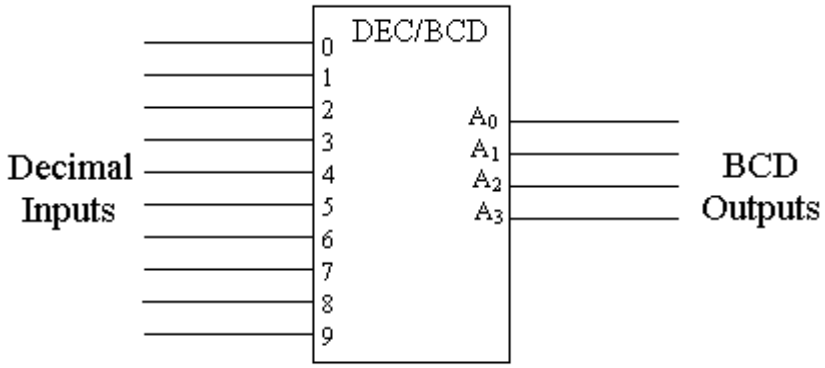


Figure 3.30 Logic symbol Decimal-to-BCD Encoder

Decimal digit	BCD code			
	A3	A2	A1	A0
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1

Table 3.16 Truth table for Decimal-to-BCD encoder

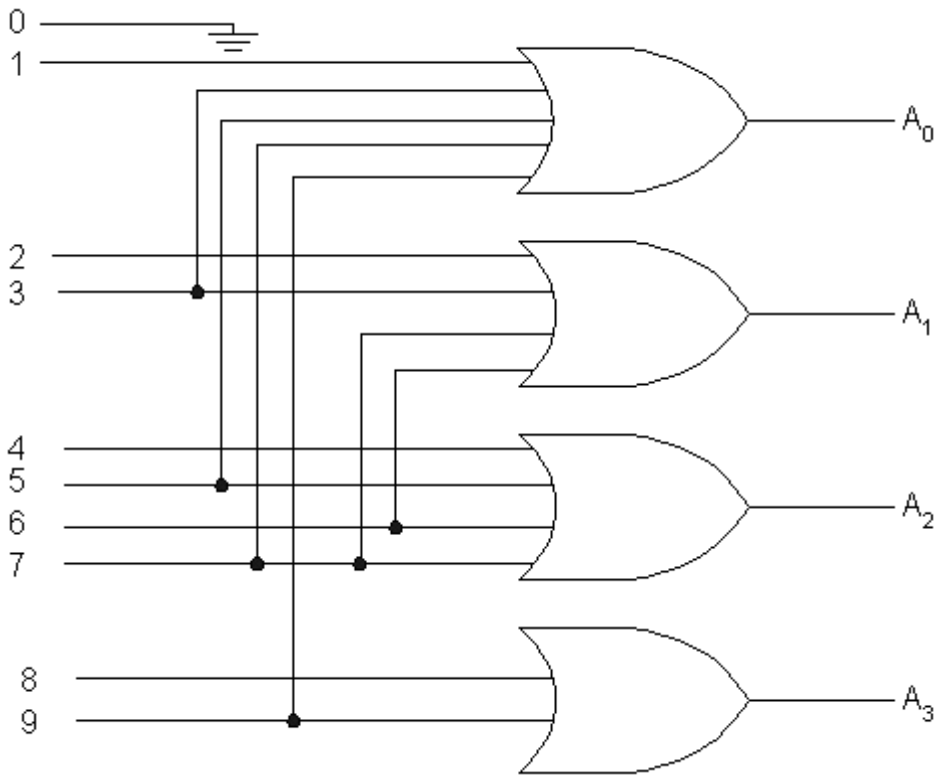


Figure 3.31 Logic diagram for Decimal-to-BCD encoder

3.7.5 Multiplexer:

Multiplexer(MUX) is a combinational logic circuit that selects one of many digital input signals and outputs that into a single line. The selection of a particular input line is controlled by a set of selection lines. There are 2^n input lines and n selection lines and 1 output line. A multiplexer makes it possible for several signals to share one expensive device or resource, for example one communication line or channel, instead of having one channel per input signal. The multiplexer also called as data selector.

The schematic diagram for a 4:1 multiplexer is shown in the figure 3.32. and its truth table is listed in table 3.17 . From the figure 3.32, D0-D3 are data inputs, S0,S1 are selection lines and Y is the output line.

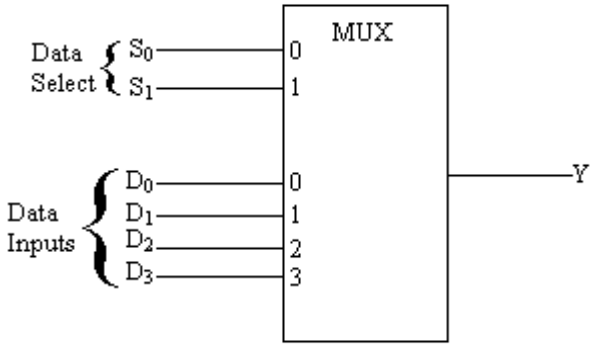


Figure 3.32 4:1 Logic symbol for 4:1 Multiplexer

Data Select Inputs		Input Selected
S1	S2	
0	0	D0
0	1	D1
1	0	D2
1	1	D3

Table 3.17 Truth table for 4:1 Multiplexer

The logical diagram of 4 : 1 multiplexer shown in the figure 3.33. One example of a 4:1 multiplexer is DM74LS153 .It is a Dual 4-Line to 1-Line , dual in line package IC and its pin details is shown in the figure 3.34

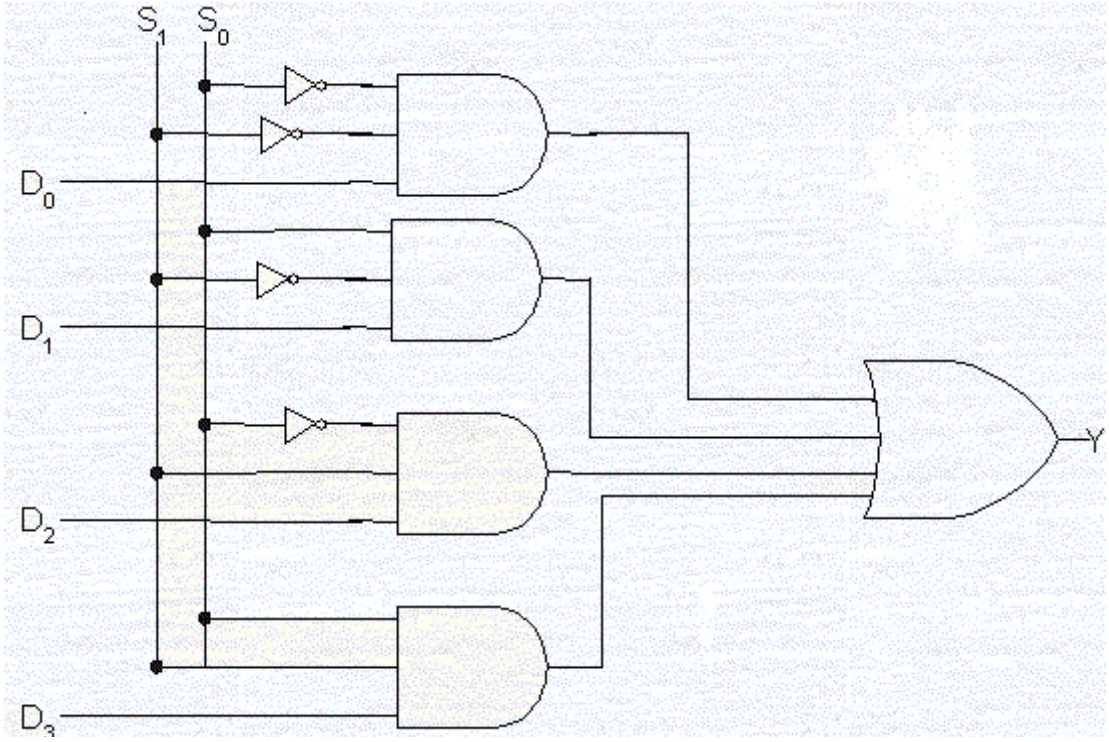


Figure 3.33 4:1 logical diagram for 4:1 Multiplexer

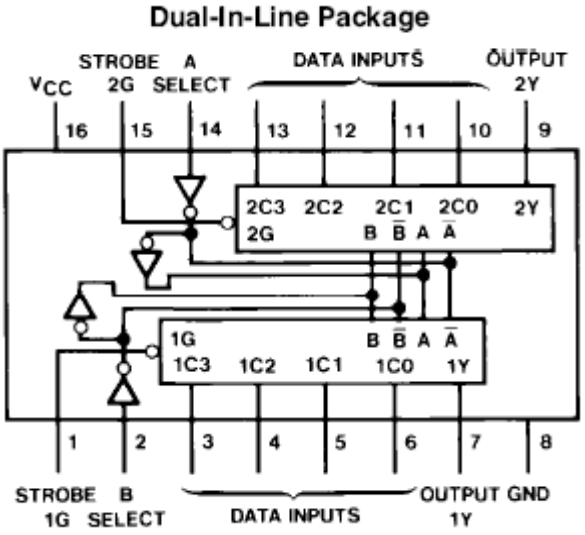


Figure 3.34 Pin configuration of DM74LS153 Dual 4-Line to 1-Line Data Selectors/Multiplexer

3.7.6 Demultiplexer:

A demultiplexer (or DEMUX) is a combinational logic circuit takes a single input signal and transfers this signal to one of 2^n possible output lines. The selection of a specific output line is controlled

by the digital code applied to n selection lines. The schematic diagram of 1:4 demultiplexer is shown in the figure 3.35. D is the input and S0,S1 are selection lines and Y0,Y1,Y2 and Y3 are output lines. The truth table for a 1:4 demultiplexer is listed in table 3.18.

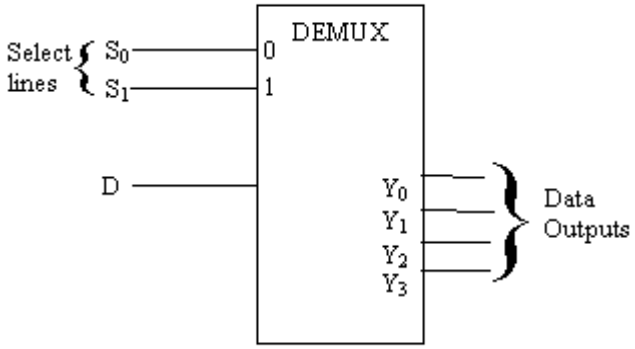


Figure 3.35 Logic symbol for 4:1 Demultiplexer

Data	Address		Outputs			
	S1	S2	Y0	Y1	Y2	Y3
D	0	0	D	0	0	0
D	0	1	0	D	0	0
D	1	0	0	0	D	0
D	1	1	0	0	0	D

Table 3.18 Truth table for 4:1 Demultiplexer

The logical diagram of 1:4 demultiplexer is shown in the figure 3.36. One example of a 1:4 De-multiplexer IC is DM74LS156.It is a Dual 2-Line to 4-Line Decoders/Demultiplexers, dual-in-line package IC and its pin details is shown in the figure 3.37.

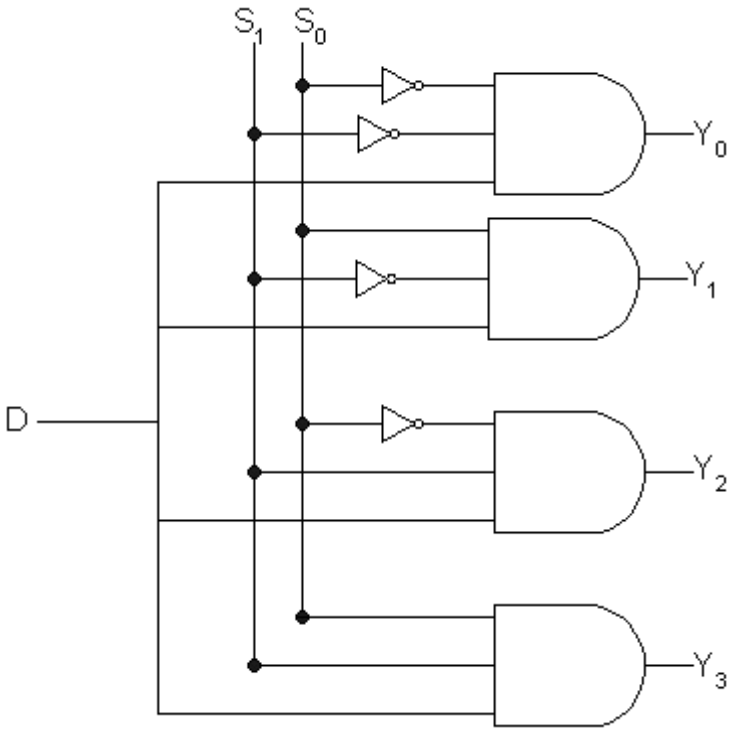


Figure 3.36 Logic diagram for 1:4 Demultiplexer

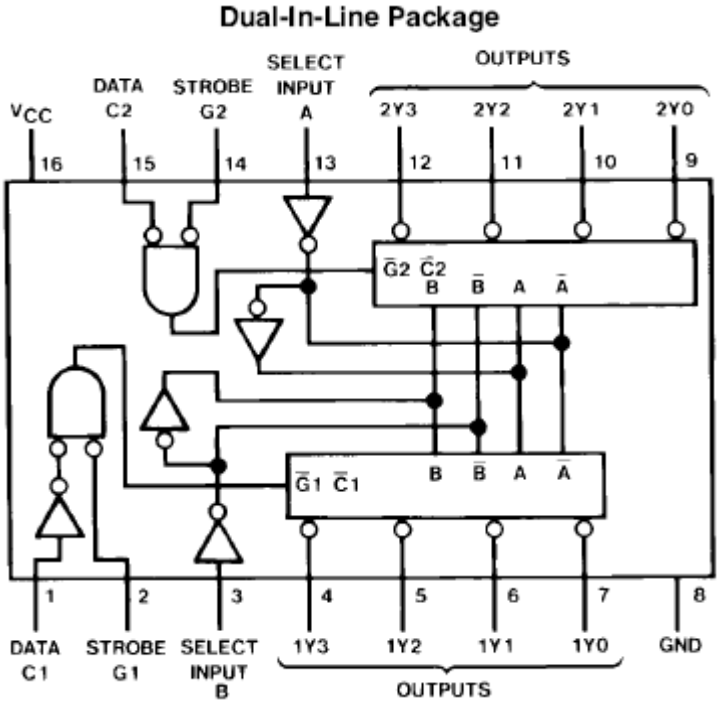


Figure 3.37 Pin configuration of DM74LS156 Dual 2-Line to 4-Line Decoder/Demultiplexer

3.8 Sequential Logic Circuits:

Sequential logic circuits employ storage elements in addition to combinational logic circuit. Their outputs are a function of the inputs and the state of the storage elements(previous output). The state of storage elements, in turn, is a function of previous inputs. The circuit behavior must be specified by a time sequence of inputs and internal states.

Depending on the timing of the signal,sequential circuits can be classified in to synchronous and Asynchronous sequential circuits. A synchronous sequential circuit is a system whose behavior can be defined from the knowledge of its signals at specified instants of time. The behavior of an asynchronous sequential circuit depends upon the input signals at any instant of time and the order in which the inputs change. The storage elements commonly used in asynchronous sequential circuits are time-delay devices

3.8.1 Flip-Flops:

The storage elements used in clocked sequential circuits are called Flip-Flops. A flip-flop is a binary storage device capable of storing one bit of information. The state of the flip-flop changes during pulse transition or pulse duration of a clock pulse. When a clock pulse is not active the feedback loop is broken because the flip-flop outputs cannot change even if their inputs change. The flip-flops discussed below are level triggered i.e. flip-flops respond during the entire pulse duration. The level triggered flip-flops are called *latches.*

SR flip-flop:

SR flip-flop circuit can be constructed from either using two NOR gates or two NAND gates. NOR gate flip-flops are shown in figure 3.38 Each flip-flop has two outputs(Q and Q') and two inputs(set S and reset R). This type of flip-flop is referred to as SR flip-flop or SR latch. The flip-flop in figure 3.38 has two useful states. When Q=1 and Q'=0, it is in the set state (or 1-state). When Q=0 and Q'=1, it is in the clear state (or 0-state). The outputs Q and Q' are complements of each other and are referred to as the normal and complement outputs respectively. The binary state of the flip-flop is taken to be the value of the normal output. When a 1 is applied to both the set and reset inputs of the flip-flop in figure 3.38, both Q and Q' outputs go to 0. This condition violates the fact that both outputs are complements of each other. In normal operation this condition must be avoided by making sure that 1's are not applied to both inputs simultaneously.

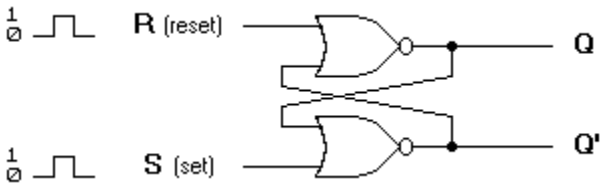


Figure 3.38 SR flip-flop using NOR gate

SR Flip-Flop			
S	R	Qn+1	Comment
0	0	Qn	No change
0	1	0	Reset
1	0	1	Set
1	1	--------	Indeterminate

Table 3.19 Truth table for SR flip-flop

Clocked SR Flip-Flop:

The clocked SR flip-flop shown in Figure 3.39 consists of a basic SR flip-flop and two AND gates. The outputs of the two AND gates remain at 0 as long as the clock pulse (or CP) is 0, regardless of the S and R input values. When the clock pulse goes to 1, information from the S and R inputs passes through to the basic flip-flop. With both S=1 and R=1, the occurrence of a clock pulse causes both outputs to momentarily go to 0. When the pulse is removed, the state of the flip-flop is indeterminate, ie., either state may result, depending on whether the set or reset input of the flip-flop remains at logic 1 for a longer period at the end of the pulse. Hence the state S=R=1 should be avoided. The example for a SR flip-flop is IC DM74LS279. It is a quad ,dual-in-line package IC and its pin detail is shown in the figure 3.40

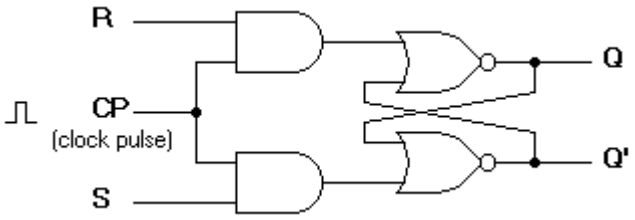


Figure 3.39 Clocked SR flip-flop

	SR Flip-Flop			
CP	S	R	Qn+1	Comment
⎍	0	0	Qn	No change
⎍	0	1	0	Reset
⎍	1	0	1	Set
⎍	1	1	--------	Indeterminate

Table 3.20 Truth table for Clocked SR flip-flop

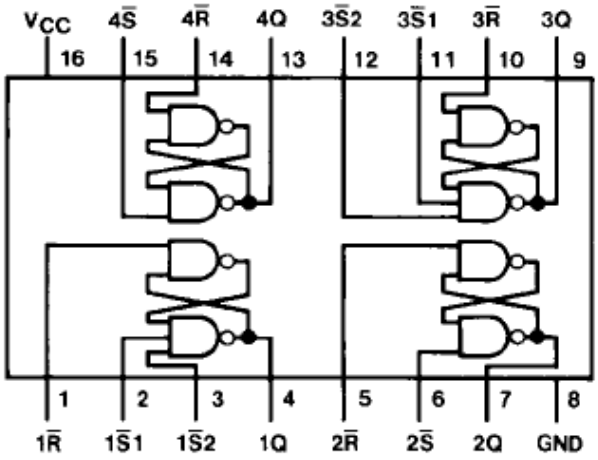


Figure 3.40 Pin configuration of DM74LS279 Quad S-R Latch

D-Flip-flop:

The D flip-flop shown in Figure 3.41 is a modification of the clocked SR flip-flop. The D input goes directly into the S input and the complement of the D input goes to the R input. The D input is sampled during the occurrence of a clock pulse. If it is 1, the flip-flop is switched to the set state . If it is 0, the flip-flop switches to the clear state. The logic diagram for D-flipflop is shown in figure 3.41 and its truth table is shown in figure 3.21. The example for a D flip-flop is IC DM74LS75. It is a quad ,dual-in-line package IC and its pin detail is shown in the figure 3.42.

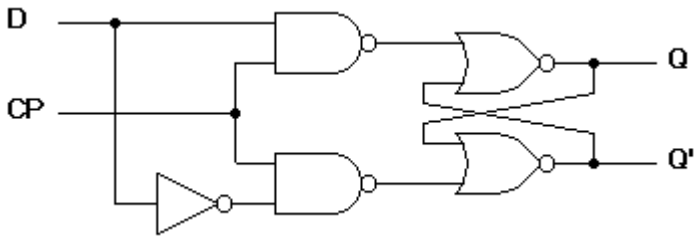


Figure 3.41 D-flip-flop

D Flip-Flop			
CP	D	Qn+1	Comment
⎍	0	0	Reset
⎍	1	1	Set

Table 3.21 Truth table for D-flip-flop

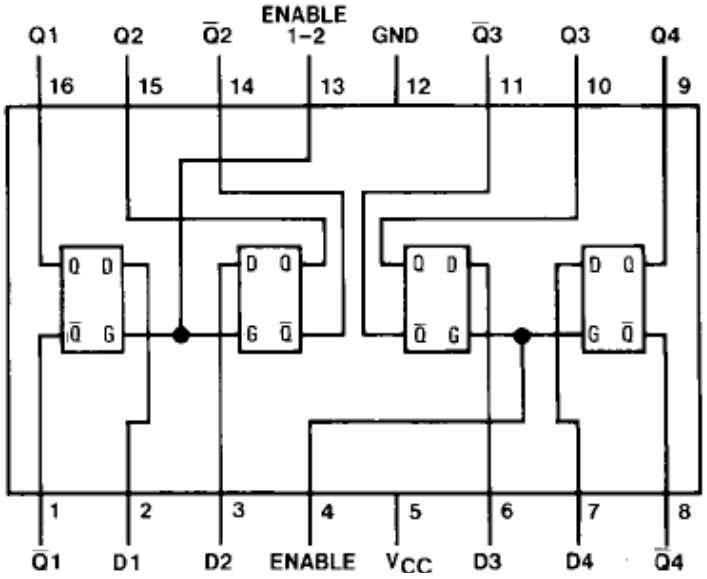


Figure 3.42 Pin configuration of DM74LS75 Quad D-Latch

JK Flip-Flop:

A JK flip-flop is a refinement of the SR flip-flop in which the indeterminate state of the SR type is defined. Inputs J and K behave like inputs S and R to set and clear the flip-flop. When logic 1 inputs are applied to both J and K simultaneously, the flip-flop switches to its complement state, ie., if Q=1, it switches to Q=0 and vice versa.

In clocked JK flip-flop shown in Figure 3.43.,To reset the flip-flop the output Q is AND-ed with inputs K and CP(Clock Pulse), so that the flip-flop is cleared during a clock pulse only if Q was previously 1. Similarly, output Q' is ANDed with J and CP inputs so that the flip-flop is set with a clock pulse only if Q' was previously 1. Note that because of the feedback connection in the JK flip-flop, a CP signal which remains a 1 (while J=K=1) after the outputs have been complemented once will cause repeated and continuous transitions of the outputs called *race around* condition. To avoid this, the clock pulses must have a time duration less than the propagation delay through the flip-flop. The restriction on the pulse width can be eliminated with a master-slave or edge-triggered construction. The same reasoning also applies to the T flip-flop presented next. The example for a JK flip-flop is IC SN74LS78A. It is a quad ,dual-in-line package IC and its pin detail is shown in the figure 3.44.

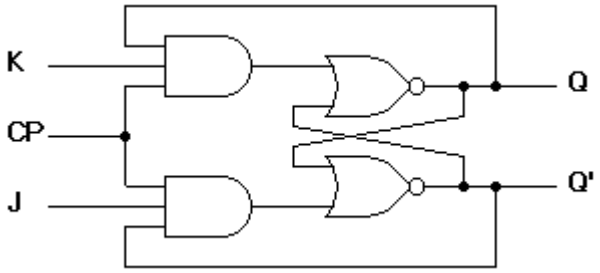


Figure 3.43 JK flip-flop

	JK Flip-Flop			
CP	J	K	Qn+1	Comment
⎍	0	0	Qn	No change
⎍	0	1	0	Reset
⎍	1	0	1	Set
⎍	1	1	Qn'	Complement

Table 3.22 Truth table for JK flip-flop

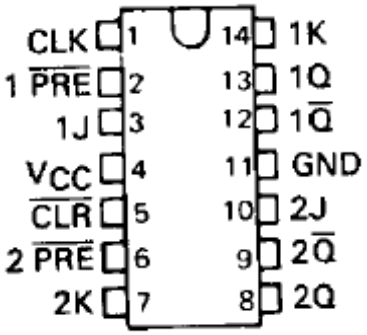


Figure 3.44 Pin configuration of SN74LS78A Dual JK Flip-flops

T Flip-Flop:

The T flip-flop is a single input version of the JK flip-flop. As shown in figure 3.45, the T flip-flop is obtained from the JK type if both inputs are tied together. The output of the T flip-flop toggles the input with each clock pulse.

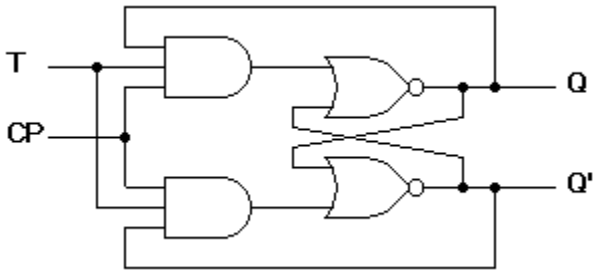


Figure 3.45 T flip-flop

T Flip-Flop			
CP	T	Qn+1	Comment
⎍	0	Qn	No change
⎍	1	Qn'	Complement

Table 3.23 Truth table for T Flip-flop

Triggering of Flip-flops:

The state of a flip-flop is changed by a momentary change in the input clock signal. This momentary change is called a *trigger*. The basic flip-flop circuit shown in figure 3.39 require an input trigger defined by a change in signal level. This level must be returned to its initial level before a second trigger is applied. As shown in figure 3.46 , the flip-flops can also be triggered when the clock pulse goes through signal transitions: either from 0 to 1(positive edge) or return from 1 to 0(negative edge). This is called *edge triggering*.

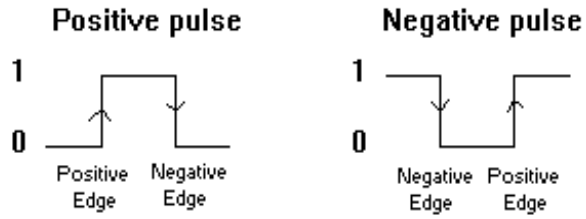


Figure 3.46 Definition of clock pulse transition

Master-Slave Flip-Flop:

The race around condition in level triggered JK flip-flop is overcome in Master slave flip-flop design. The Master-Slave Flip-Flop is basically two J-K flip-flops connected together in a series configuration with the outputs from Q and Q' from the "Slave" flip-flop being fed back to the inputs of the "Master" as shown in figure 3.47

The input signals J and K are connected to the "Master" flip-flop which "locks" the input while the clock (Clk) input is high at logic level "1". The outputs from the "Master" flip-flop are only "seen" by the "Slave" Flip-flop when the clock input goes "LOW" to logic level "0". Therefore on the "High-to-Low" transition of the clock pulse, the locked outputs of the "Master" Flip-flop are fed through to the J-K inputs of the "Slave" flip-flop making this type of flip-flop edge triggered. Thus, the circuit accepts input data when the clock signal is "HIGH", and passes the data to the output on the falling-edge of the clock signal. In other words, the Master-Slave J-K Flip-flop is a "Synchronous" device as it only passes data with the timing of the clock signal. The example for a Master-Slave negative edge triggered JK flip-flop is IC DM74LS73A. It is a quad ,dual-in-line package IC and its pin detail is shown in the figure 3.48.

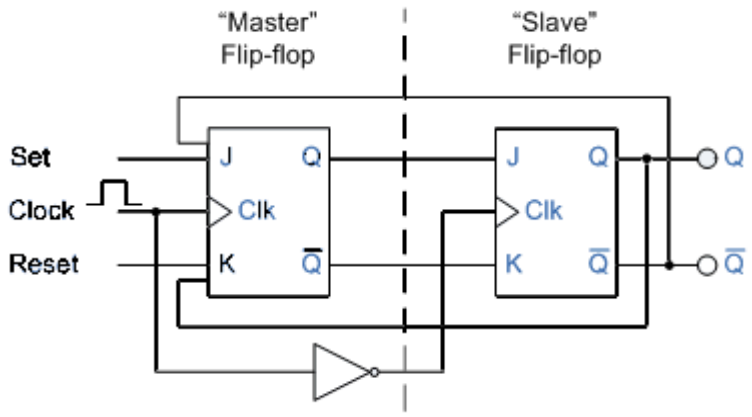


Figure 3.47 Logic diagram of a master-slave flip-flop

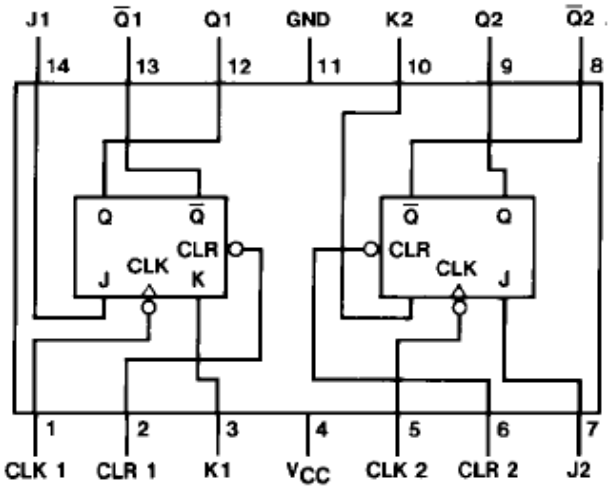


Figure 3.48 Pin details of DM74LS73A Dual Negative-Edge-Triggered Master-Slave J-K Flip-Flops

3.8.2 Counters:

A counter is a sequential logic circuit that provides an output corresponding to the number of pulses applied to its input. Counters can be classified in to two categories *asynchronous and synchronous* based on how they are clocked. In asynchronous counter also called ripple counter, the first flip-flop is clocked by the external clock pulse and then each successive flip-flop is clocked by the output of the previous flip-flop. In synchronous counter all the flip-flops are clocked simultaneously by an external clock. Digital counters are found in modern electronic equipment, that are digitally controlled or having digital numeric displays. In a digital circuit, counters are used to do functions such as timing, sequencing and counting.

Example:A digital voltmeter, as a test instrument consists simply of a string of decade counters that count the pulses of an input signal for a known period of time, and displays that count on a seven-segment

display. A10-MHz frequency may be divided to produce a pulse train of a much lower frequency, say 1 Hz. This is required in a precision digital clock. Counters prove useful as timing circuits in this case

How are counters made?.

Counters are generally made up of flip-flops and logic gates. The value of a state is expressed as a multi digit binary number, whose `1's and `0's are usually derived from the outputs of internal flip-flops that make up the counter. The number of states a counter may have is limited by the number of flip-flops used in that counter. For counting 2^n clock cycles, n number of flip-flops are required. The main types of flip-flops used are J-K or T flip-flops.

Asynchronous two bit binary counter:

An asynchronous two bit binary counter can be constructed from J-K flip-flops by connecting the output of first flip-flop to the clock input of the next. The J and K inputs of each flip-flop are set to 1 to produce a toggle at each cycle of the clock input. For each two toggles of the first flip-flop, a toggle is produced in the second flip-flop. This produces a binary number equal to the number of cycles of the input clock signal. The maximum number of clock cycles can be counted by this counter is 4.

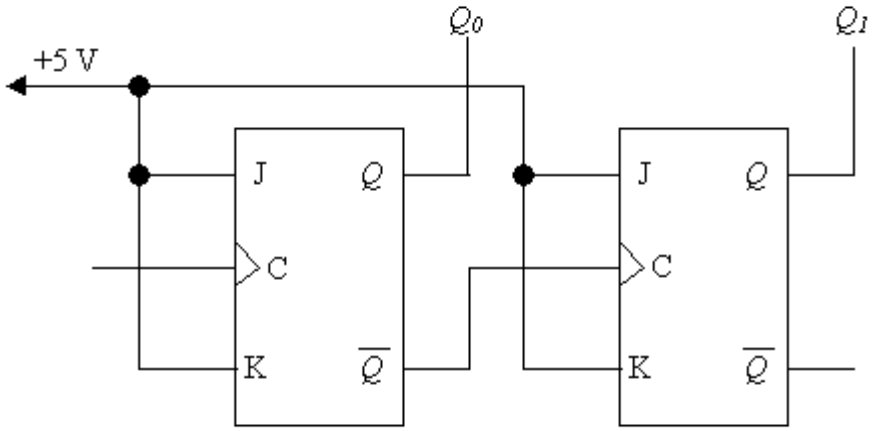


Figure 3.49 Asynchronous two bit binary counter

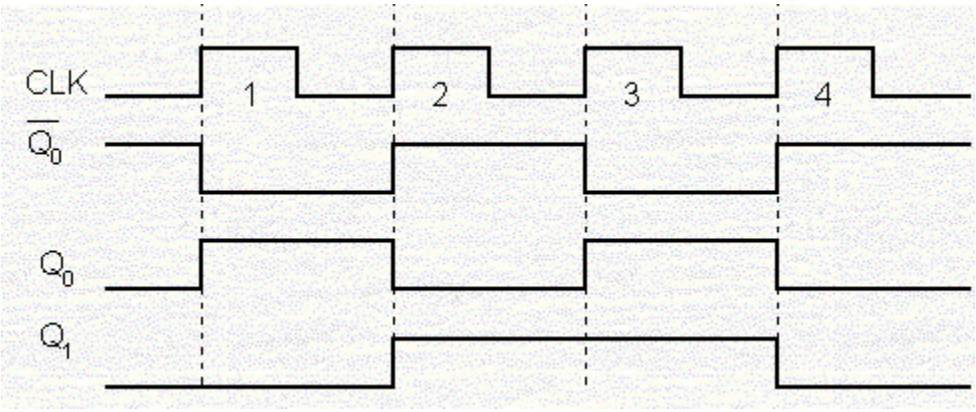


Figure 3.50Timing diagram of 2 bit asynchronous counter

synchronous two bit binary counter:

Synchronous digital counters have a common clock which results in all the flip-flops being triggered simultaneously. Consequently, there are no cumulative delays that result because the clock signal must ripple through the stages as in the asynchronous counters. Synchronous counters can be designed to count up and down in numerical order. In addition, they may be used to produce count sequences of non-consecutive numbers. The count sequence produced by synchronous counters is not dependent on the trigger characteristics of the flip-flops that comprise the count stages. The count sequence is achieved by applying the required logic function into the flip-flops.

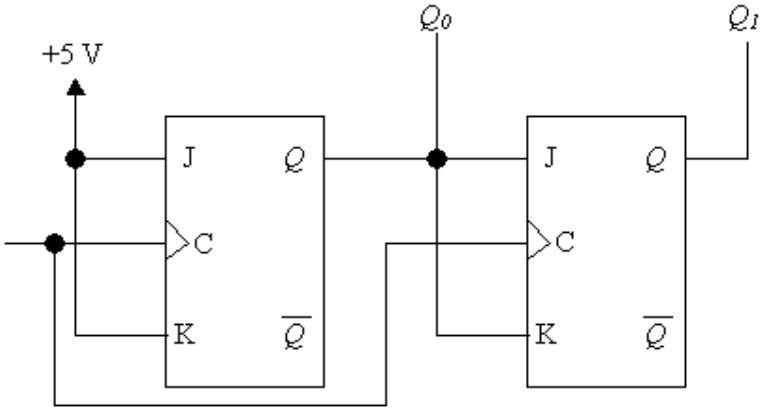


Figure 3.51 Two bit synchronous counter

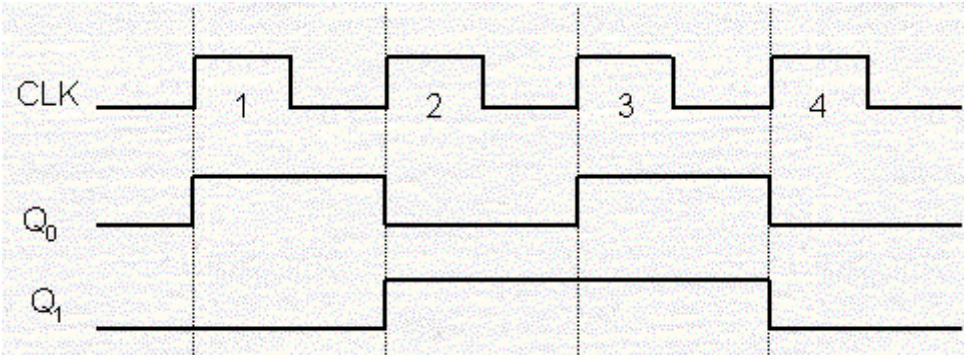


Figure 3.52 Timing diagram of 2 bit synchronous counter

3.8.3 Shift registers:

Shift registers are mainly used to store data and to convert data from a serial to parallel or parallel to serial format. A common clock (Clk) drives all the D flip-flops in the shift register making it as a synchronous sequential circuit. The number of D flip-flops used to make up a shift register decides the number of bits to be stored. Shift registers are generally provided with a preset or clear input so that they can be SET or RESET when required. Generally, Shift registers operate in one of four different ways:

- Serial In -Parallel Out (SIPO)
- Serial In - Serial Out (SISO)
- Parallel In - Parallel Out (PIPO)
- Parallel In - Serial Out (PISO)

Serial In - Parallel Out:

A 4-bit Serial In -Parallel Out (SIPO) Shift Register is shown in the figure 3.53. Here

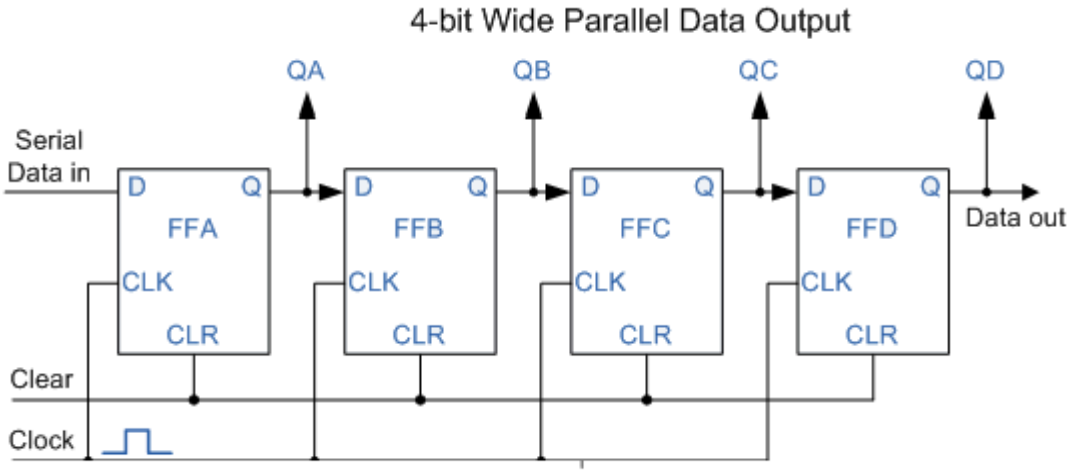


Figure 3.53 Serial In - Parallel Out shift register

Assume all the flip-flop outputs are at logic "0". The process of storing “1010” is explained as follows. The lsb “0” is forced to the DATA input pin of FFA. On applying the first clock pulse the output of FFA (QA) will be set to logic "0". Next DATA input pin of FFA is forced a logic "1". The next clock pulse will change the output of FFA to logic "1" and the output of FFB(QB) to logic "0". The logic "0" has been shifted one place to the right. Like wise upon the arrival of the fourth clock pulse all the four bits are serially shifted in as shown in the table3.24. Now the parallel outputs QA - QD directly denote the parallel data “1010”.

Clock Pulse No	QA	QB	QC	QD
0	0	0	0	0
1	0	0	0	0
2	1	0	0	0
3	0	1	0	0
4	1	0	1	0

Table 3.24 shift register operation

Serial In - Serial Out:

A Serial In - Serial Out (SISO**)** **S**hift register accepts input data serially one bit at a time. It gives the stored information on its output also in serial form. A 4-bit SISO Shift Register is shown in the figure 3.54

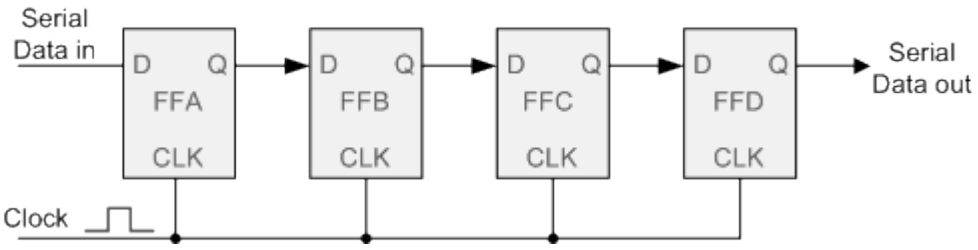


Figure 3.54 Serial In - Serial Out shift register

This type of register acts as a temporary storage device or as a time delay device, with the amount of time delay being controlled by the number of stages in the register and the input clock frequency.

Parallel In - Serial Out :

In this Parallel In -Serial Out (PISO) Shift Register the DATA is applied in parallel form(simultaneously) to the parallel input pins D0 to D3 of the register and the output is read out sequentially from the register one bit at a time from the last flip-flop output.

4-bit Parallel-in to Serial-out (PISO) Shift Register shown in the figure 3.55

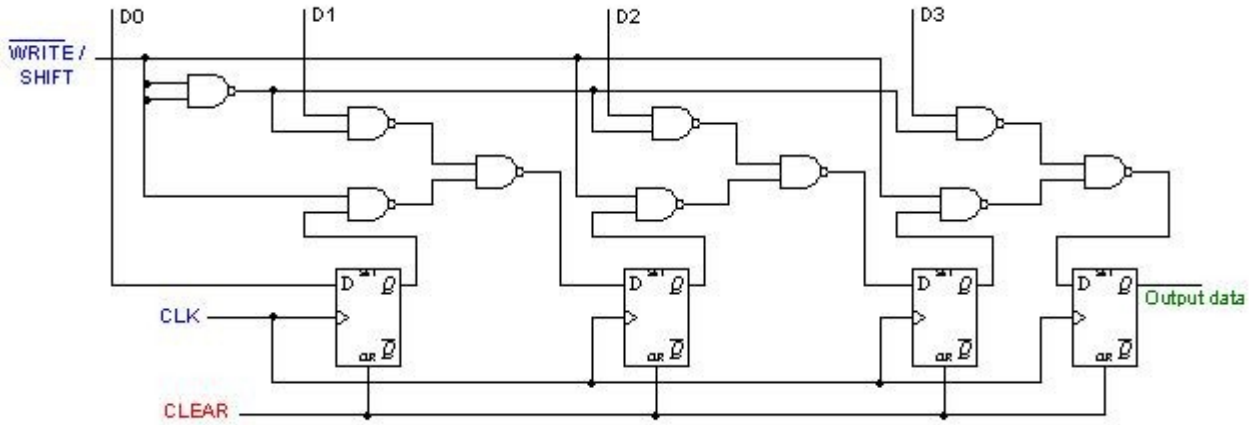


Figure 3.55 Parallel In - Serial Out shift register

D0, D1, D2 and D3 are the parallel inputs, where D0 is the most significant bit and D3 is the least significant bit. To write data in, the mode control line is taken to LOW and the data is clocked in. The data can be shifted when the mode control line is HIGH as SHIFT is active high. The register performs right shift operation on the application of a clock pulse. The example for a PISO shift register is IC SN54/74LS165. It is a dual-in-line package IC and its pin detail is shown in the figure 3.56

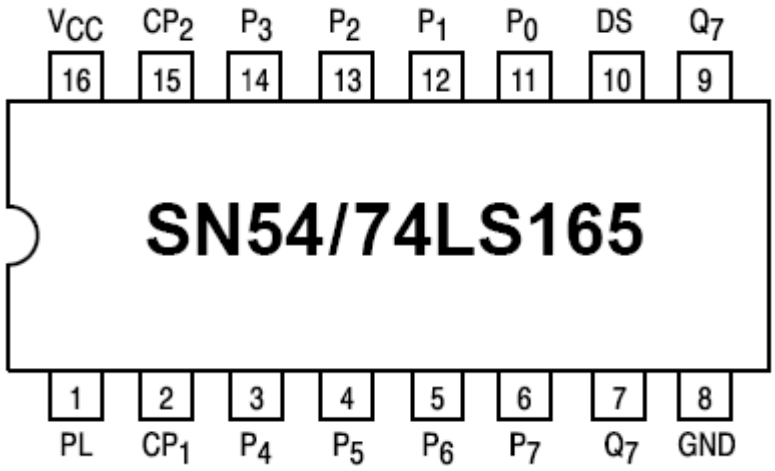


Figure 3.56 Pin details of SN54/74LS165 8-Bit Parallel-To-Serial Out shift register

Parallel In – Parallel Out:

In this register ,binary data is shifted in to the register in parallel mode and the stored data is shifted out of the register in parallel mode. A 4-bit Parallel In- Parallel Out (PIPO) Shift register is shown in the figure 3.57

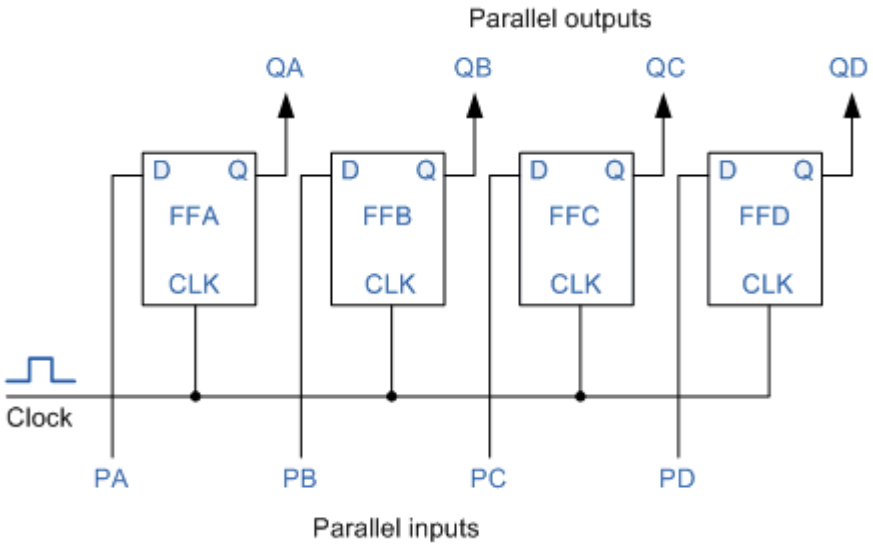


Figure 3.57 Parallel In – Parallel Out shift register

3.9 Memories:

In digital systems, Memories are used to store, retain, and subsequently retrieve information for processing. Memories are made up of storage locations in which data can be stored. Each location is identified by an address. Each storage location can accommodate one or more bits. The total number of bits that a memory can store is its capacity. Memories can be broadly classified in to semiconductor memories and magnetic memories based on the material used to make them. Here the semiconductor memories are discussed below.

Semiconductor memories are made up of storage elements such as flip-flops or capacitors. A storage element is called a cell. The two basic types of semiconductor memories are the Read Only Memory(ROM)and the Random Access Memory(RAM). In a ROM, the data are permanently stored that can only be read randomly but it can not be rewritten. There is no write operation because specified data to be stored are manufactured in to the device or programmed in to device by the user and can not be altered. The ROM is a *Nonvolatile memory* because it retains its contents indefinitely, even when there is no power . Example: The memory used to hold the boot code for a computer is ROM

ROMs can be divided in to three types based on how they are programmed: Mask ROM, PROM and EPROM.

In Mask ROM the contents are permanently stored(programmed) in the memory during the manufacturing process at factory according to the desired specification and no changes are possible afterwards. In the PROM the data are electrically stored by the user with the use of specialized equipment called PROM programmer. Once programmed the contents can not be changed. Mask ROM and PROM are manufactured with (using) both TTL and MOS technologies. EPROM is electrically programmable by the user,but the stored data can be erased either by means of ultraviolet(UV) radiation(UV EPROM) or by electrical pulses(EEPROM).EPROM is manufactured only with MOS technologies. One type of EEPROM is the Flash memory. Example USB drive.

Random Access Memory(RAM) has both read and write capability. In RAM, any storage location can be accessed in any order to read or write data. RAMs are *Volatile memories* that lose their contents when power is turned off. Example :The memory used to store dynamic variables including the stack and other programs in a computer is RAM.

RAM can also be separated into two subcategories: devices whose contents are nonvolatile for as long as power is applied. These devices are referred to as static RAM and devices whose contents require periodic refreshing to avoid the loss of data even while power is present. These devices are referred to as dynamic RAM.

3.10 Clock Timing parameters:

Commonly a digital signal is a series of pulses. The clock period(T) of a digital signal is defined as the time taken to complete a full cycle which consists of on time and off time. The reciprocal of period is equal to the frequency. The characteristics of the pulse is described below and is illustrated in the figure 3.58

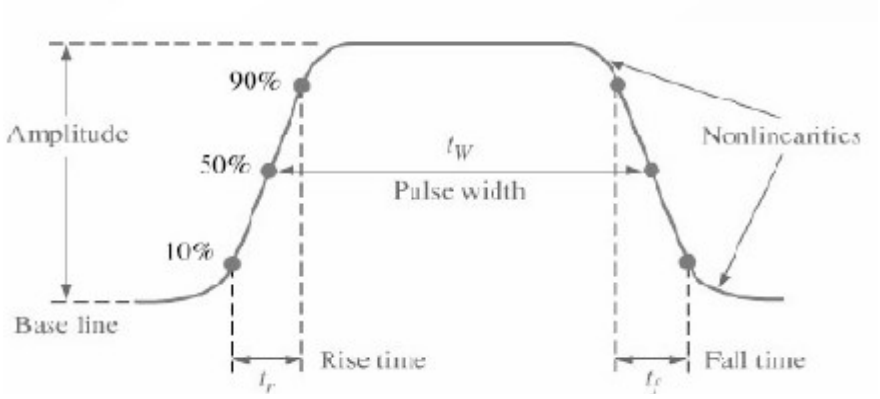


Figure 3.58 characteristics of the pulse

Rise time:

The rise time is the time required for a pulse to rise from 10% to 90% of its final value.

Fall time:

The fall time is the time required for a pulse to fall from 90% to 10% of its final value.

Pulse width:

It is the time duration between 50% levels of the rising edge and falling edge transitions.

Setup time :

It is the minimum amount of time the applied input signal must be valid and stable prior to a clock transition. Any violation in this minimum required time causes incorrect output data to be captured and is known as setup violation. It is illustrated in the figure 3.59

Hold time:

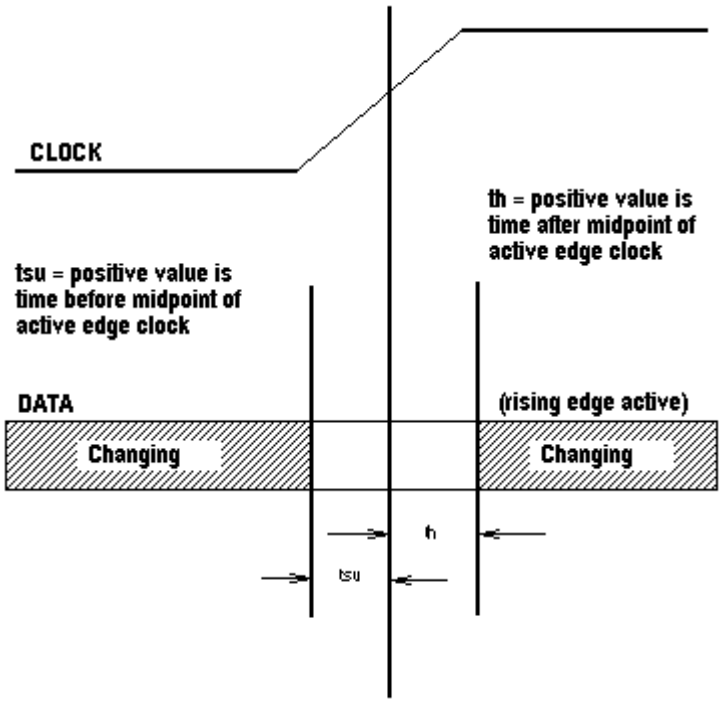


Figure 3.59 setup time and hold time

Hold time is defined as the minimum amount of time the input data must be stable after the clock

transition. Any violation in this required time causes incorrect input data to be latched and is known as hold violation. This is illustrated in the figure 3.60

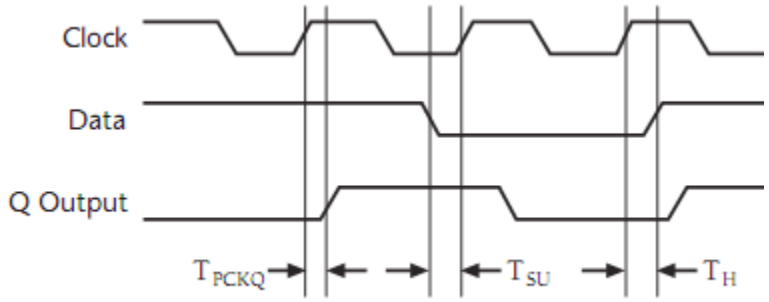


Figure 3.60 setup time and hold time

Jitter:

Jitter is the random fluctuations present in a signal with respect to its ideal characteristics. there are several types of jitter viz., delay jitter,amplitude jitter and frequency jitter. Practically consecutive pulses of a clock signal does not arrive at exact intervals, the delay in arrival is known as delay jitter., which is illustrated in the figure.3.61. The delay jitter is also called as phase jitter. Jitter may be induced and coupled onto a clock signal from several different sources like internal circuitry of the clock source, various noise sources affecting clock generation and distribution.,Traces and cables, changes in circuit impedance .It is not uniform over all frequencies. In digital systems, jitter can lead to a violation of timing margins as illustrated in the figure 3.62 causing circuits to behave improperly. For example the effect of a jitter on set up time of a data is illustrated in the figure 3.63. Here the presence of the jitter in

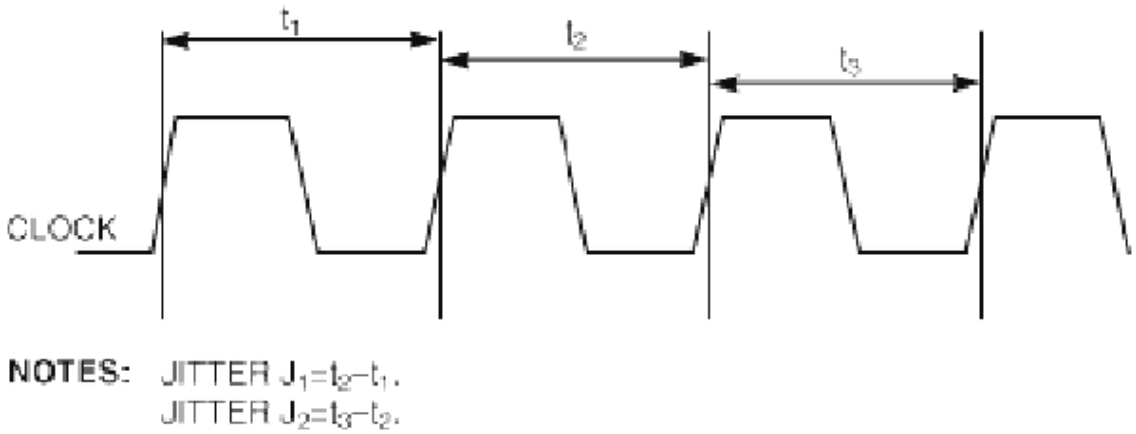


Figure 3.61 Delay Jitter

the clock signal sampled the data wrongly. Therefore the accurate measurement of jitter is necessary for ensuring the reliability of a system.

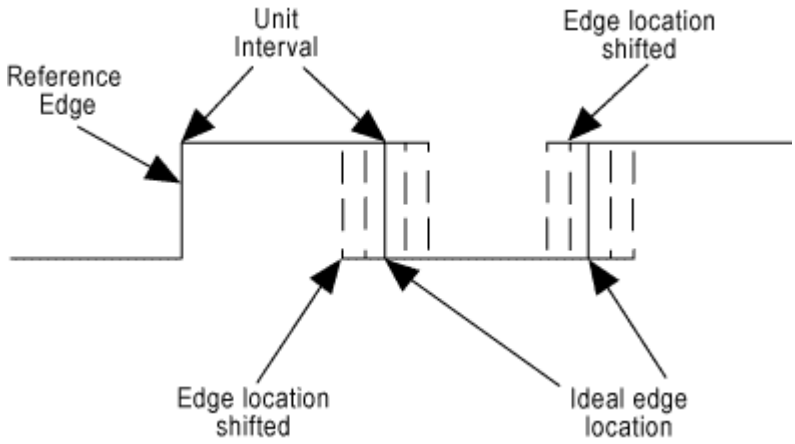


Figure 3.62 jitter violation of timing margins

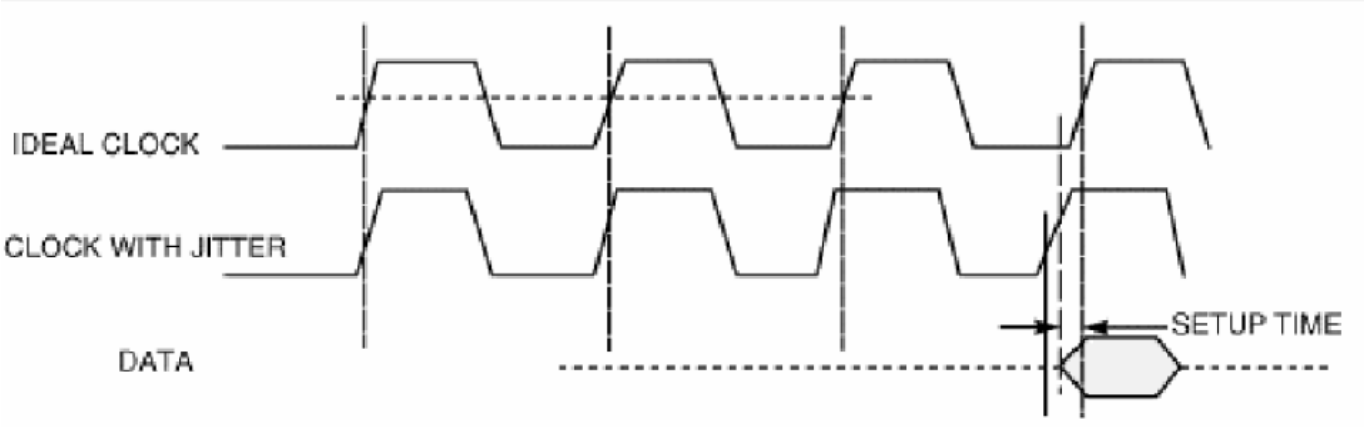


Figure 3.63 Effect of delay jitter on set up time

Clock skew:

When multiple circuits are synchronized with a single clock, clock must arrive at all memory elements in time to load data .If there is a critical path that extends routing of the clock line or a gate clock with a logic gate installed on the clock line then there is a clock delay generated and therefore the setup time and hold time cannot be satisfied. Consequently, the clock cannot sample the data in the correct timing. This phenomenon is called clock skew.

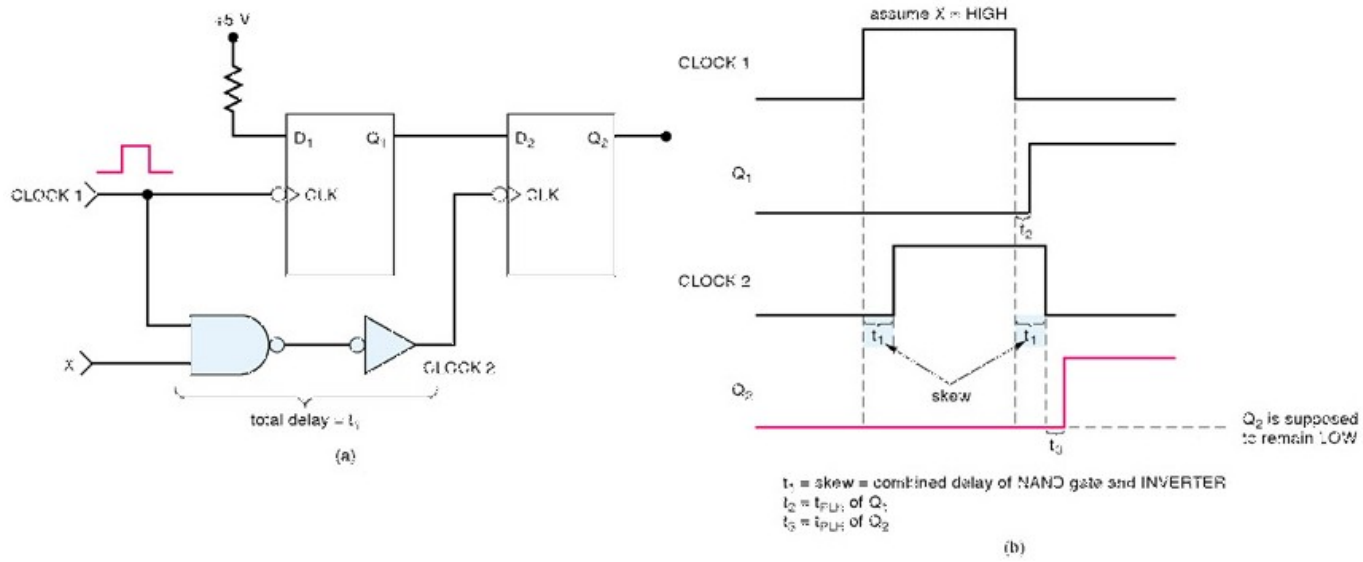


Figure 3.64 Illustration of a Clock skew

Clock skew is the difference in time that a single clock signal takes to reach two different sequential circuit memory elements(flip-flops). From the figure 3.64., clock skew occurs when two flip-flops that are supposed to be clocked simultaneously are clocked at slightly different times due to a delay in the arrival of the clock signal at the second flip-flop. For example, if the clock takes 1 ns to get to first flip-flop, but 1.25 ns to get to second flip-flop, then the skew is .25 ns.

Causes:

Clock skew caused by extended path, adding multiple logic circuits along that line ,mismatch in buffer load sizes,temperature fluctuation,, material imperfections. A source of clock skew arises when there are too many clock loads to be driven by a single clock source. Multiple clock drivers are necessary in these situations **with small variations in electrical characteristics between each driver. Clock skew usually reduces the frequency of operation of a synchronous circuit**

3.11 Logic families:

A logic family is referred as a group of compatible IC's with the same logic levels and supply voltages which perform various logic functions, and are fabricated as per a specific circuit configuration.

Logic families are mainly classified as:

1. Bipolar logic families

2. Unipolar logic families

A bipolar IC mainly uses bipolar devices like diodes and transistors, in addition to passive elements like resistors and capacitors. There are two types of bipolar IC's :Saturated logic example. RTL,DTL,TTL., Unsaturated logic example:ECL.

A unipolar IC uses unipolar devices like MOSFETS, in addition to passive elements. Example- CMOS family.

3.11.1 Characteristics of digital IC's:

The performance of different digital IC's can be analyzed by comparing their characteristics.

The major characteristics are:

1. Propagation delay
2. Noise margin
3. Power dissipation
4. Fan-in and fan-out

Propagation delay:

The propagation delay of a gate is the average transition-delay time for the signal to propagate from input to output. Propagation delay is measured in nanoseconds (ns). The signals that travel from the inputs of a digital circuit to its outputs pass through a series of gates. The sum of the propagation delays through the gates is the total delay of the circuit. When speed of operation is important, each gate must have a short propagation delay and the digital circuit must have a minimum number of gates between inputs and outputs.

Noise margin:

Spurious electrical signals also called noise can induce undesirable voltages on the connecting wires between logic circuits. All gates are designed to tolerate a certain amount of noise on their input

and output ports. *Noise margin* is the maximum noise voltage added to an input signal of a digital circuit that does not cause an undesirable change in the circuit output. The ability of circuits to operate reliably in a noise environment is important in many applications. Noise margin is expressed in volts and represents the maximum noise signal that can be tolerated by a gate. It derives from I/P-O/P voltage characteristic, measured under different operating conditions. It is normally supplied from manufacturer in the gate data sheet.

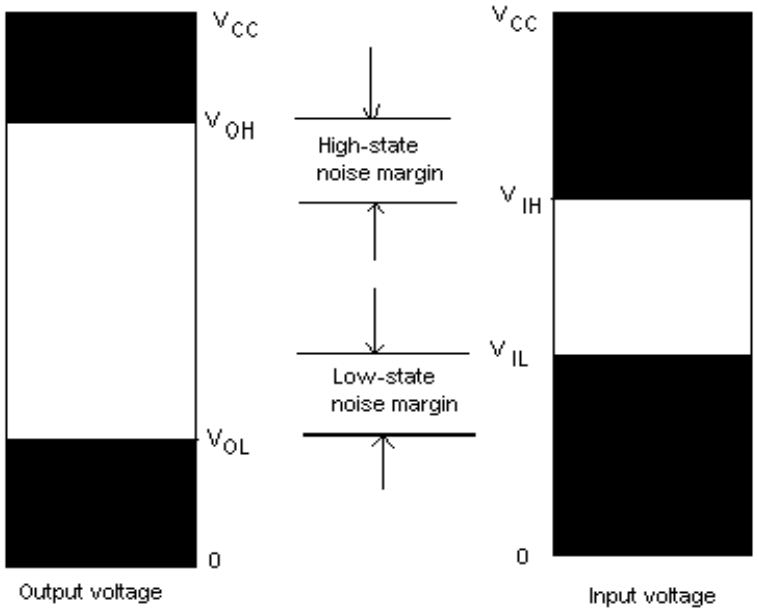


Figure 3.65 Noise margin

From the figure 3.65 Logic levels are the voltage levels for logic high and logic low.

VO_{Hmin} : The minimum output voltage in HIGH state (logic '1'). VO_{Hmin} is 2.4 V for TTL and 4.9 V for CMOS.

- VO_{Lmax} : The maximum output voltage in LOW state (logic '0'). VO_{Lmax} is 0.4 V for TTL and 0.1 V for CMOS.
- VI_{Hmin} : The minimum input voltage guaranteed to be recognized as logic 1. VI_{Hmin} is 2 V for TTL and 3.5 V for CMOS.
- VI_{Lmax} : The maximum input voltage guaranteed to be recognized as logic 0. VI_{Lmax} is 0.8 V for TTL and 1.5 V for CMOS.

- **Low noise margin (LNM):** The largest noise amplitude that is guaranteed not to change the output voltage level when superimposed on the input voltage of the logic gate (when this voltage is in the LOW interval). LNM=VI_{Lmax}-VO_{Lmax}.
- **High noise margin (HNM):** The largest noise amplitude that is guaranteed not to change the output voltage level if superimposed on the input voltage of the logic gate (when this voltage is in the HIGH interval). HNM=VO_{Hmin}-VI_{Hmin}

Power dissipation:

Every electronic circuit requires a certain amount of power to operate. The power dissipation represents the amount of power needed by the gate and is expressed in milliwatts (mW). Since each gate can be in a High, Transition or Low state, there are three different currents drawn from power supply. They are

1.ICCH: Current drawn during HIGH state.

2.ICCT: Current drawn during HIGH to LOW, LOW to HIGH transition.

3.ICCL: Current drawn during LOW state.

For TTL, ICCT the transition current is negligible, in comparison to ICCH and ICCL. If we assume that ICCH and ICCL are equal then,

Average Power Dissipation = Vcc * (ICCH + ICCL)/2

For CMOS, ICCH and ICCL current is negligible, in comparison to ICCT. So the Average power dissipation is calculated as below.

Average Power Dissipation = Vcc * ICCT.

For TTL family, power dissipation does not depend on frequency of operation, and for CMOS the power dissipation depends on the operating frequency.

Moreover, power dissipation can be classified into Static power dissipation and Dynamic power dissipation.

1.Static Power Dissipation (Ps):

It is defined as the power consumed when the output or input are not changing or rather when clock is turned off. Normally static power dissipation is caused by leakage current. As the transistor size is reduced below 90nm, leakage current could be as high as 40% of total power dissipation.

2.Dynamic Power Dissipation(Pd):

It is defined as the total *p*ower consumed by a gate during its operation.

Thus total power dissipation = static power dissipation + dynamic power dissipation.

Fan-in and Fan-out:

Fan-in is the maximum number of inputs that can be connected to a gate without affecting the logic levels. The *Fan-out* of a gate specifies the maximum number of standard loads that can be connected to the output of the gate without degrading its normal operation. It is advantage when a gate has high fan-out, since it results in reduction of additional driver gates,and there is saving of hardware.

DL (diode logic):

All the logic is implemented using diodes and resistors. The OR gate is explained in the figure 3.66

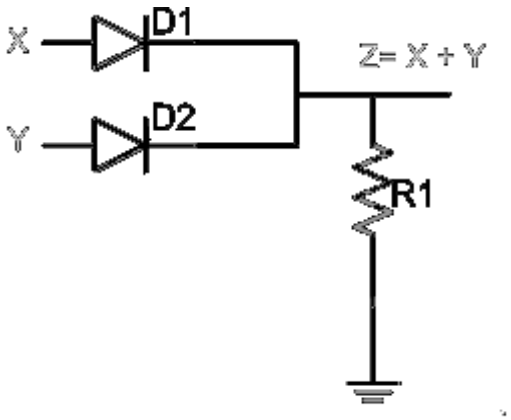


Figure 3.66 DL logic OR gate

When both inputs are low ,the output Z is low, as diodes D1,D2 do not conduct. When either X or Y, or both are driven high, the corresponding diode gets forward biased and thus conducts. The current flow results in a voltage drop across the resister which causes the HIGH output.

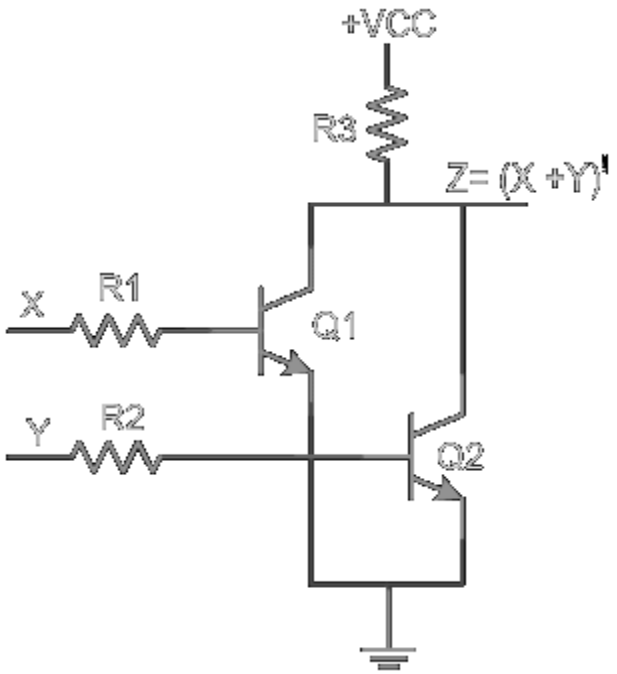
- Diode Logic suffers from voltage degradation from one stage to the next.

- Diode Logic only permits OR and AND functions.
- Diode Logic is used extensively but not in integrated circuits.

Resistor Transistor Logic(RTL):

In RTL (resistor transistor logic), all the logic are implemented using resistors and transistors. Below is the example of a NOR gate.

A basic circuit of an RTL NOR gate consists of two transistors Q1 and Q2, connected as shown in the figure 3.67. When either input X or Y is driven HIGH, the corresponding transistor goes to saturation and output Z is pulled to LOW.



3.67 RTL logic NOR gate

Diode Transistor Logic(DTL):

In DTL (Diode transistor logic), all the logic functions are implemented using diodes and transistors. A NAND gate is illustrated in the figure below. Each input is associated with one diode. The diodes and the 4.7K resistor form an AND gate. If either of the inputs X, Y or Z is low, the corresponding diode conducts. This makes the diode D4 and D5 not to conduct. Hence the transistor does not conduct, resulting in a HIGH output. If all the inputs X, Y, Z are driven high, the diodes in series conduct, driving the transistor into saturation. Thus output out is Low.

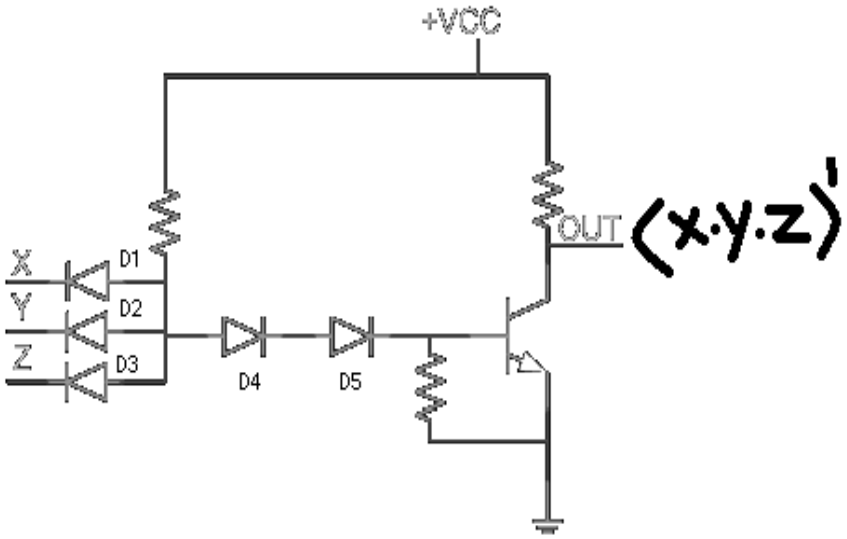


Figure 3.68 DTL logic NAND gate

3.11.12 TTL,ECL,CMOS Logic:

TTL (Transistor Transistor Logic):

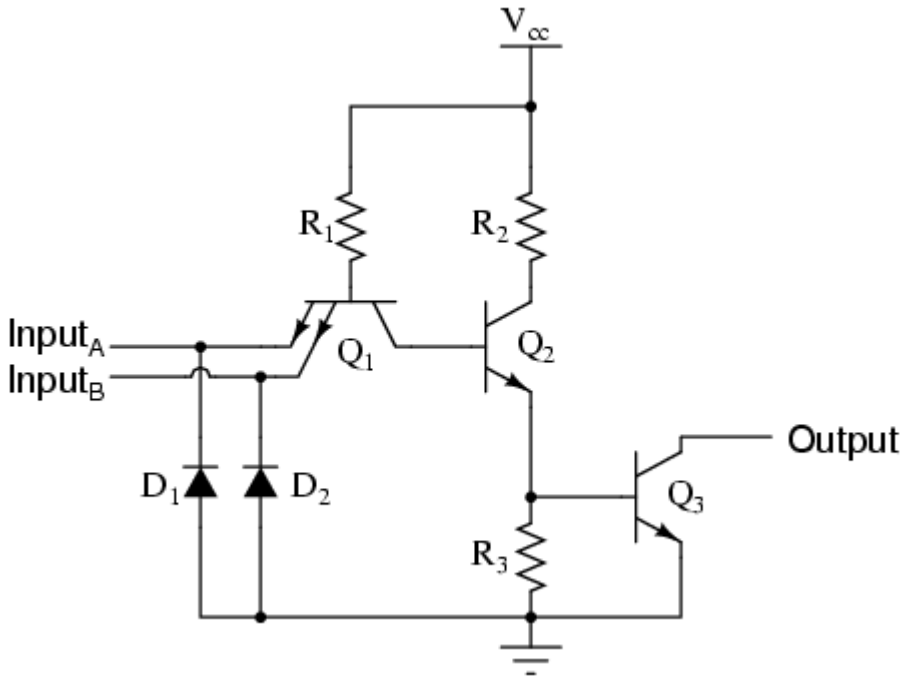


Figure 3.69 TTL NAND gate

Transistor Transistor Logic(TTL) is a saturated bipolar logic family which uses only transistors as shown in figure 3.69 .TTL logic is superior to DTL and RTL. The circuit diagram of a 2-input TTL NAND gate is shown in figure 3.69. The multi-emitter transistor Q1 as seen in the figure is a unique characteristic of TTL logic.

If the input A is grounded Q1 turns on whose collector potential is insufficient to turn on Q2. hence Q3 is in cutoff state resulting in an "high" (1) state. In the case of the open-collector output configuration, this "high" state simply "floats". If the input A is floating or connected to V_{cc} then Q1 turns off whose collector potential turns on Q2. Hence Q3 conducts resulting in the "low" or 0 state. Thus a low input resulted in a high output and vice versa. Similarly a low at input B produces a high output and vice versa. Thus each of the inputs will have the same effect on the output. In all if either of the inputs are grounded, transistor Q_2 will be forced into cutoff, turning Q_3 off which produces floating output (output goes "high").

TTL AND gate with open- collector output:

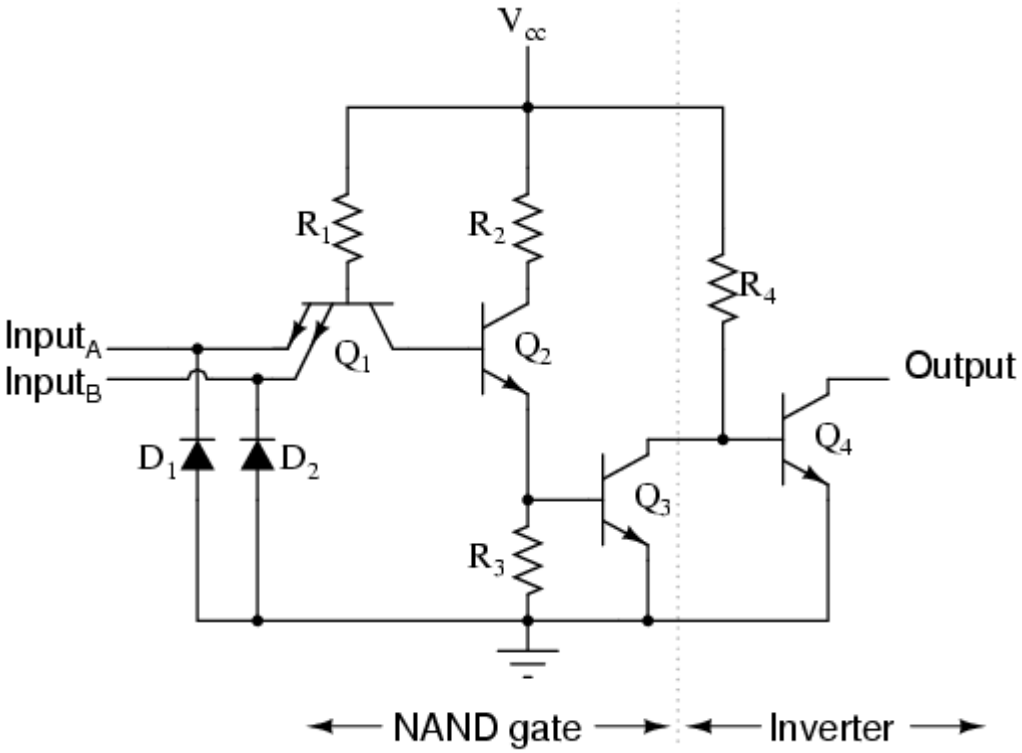


Figure 3.70 TTL AND gate open-collector circuit

The figure 3.70 shows the TTL AND gate circuit in which the TTL NAND gate is connected with the transistor Q4 acts as an inverter. The operation of TTL AND gate is similar to the circuit shown in figure 3.69 and only the difference is the inverted output of the TTL NAND gate operation due to the transistor Q4.

TTL NOR gate with open collector output:

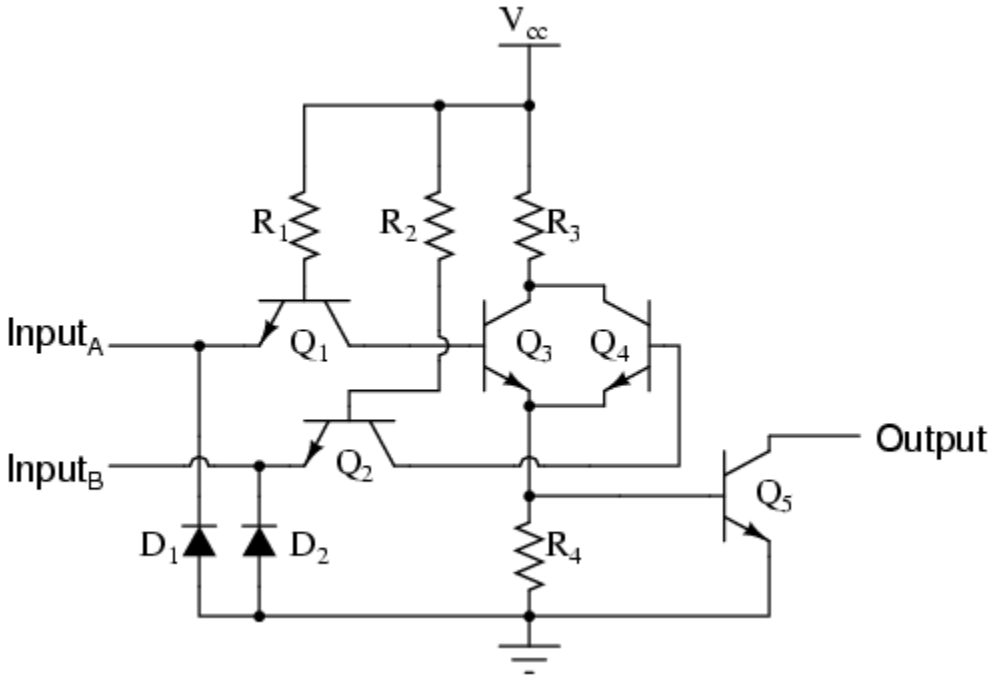


Figure 3.71 TTL NOR gate with open collector output

If input A is left floating (or connected to V_{cc}), current will go through the base of transistor Q_3, saturating it. If input A is grounded, that current is diverted away from Q_3's base through the left steering diode of "Q_1," thus forcing Q_3 into cutoff. Similarly input B drives Q_4 to either saturation or cutoff resulting in low or high output respectively. In all if either or both of the inputs are high, the output will be low . If both the inputs are low the output will be high.

TTL OR gate with open-collector output:

The figure 3.72 shows the TTL OR gate circuit in which the TTL NOR gate is connected with the transistor Q6 acts as an inverter. The operation of TTL OR gate is similar to the circuit shown in figure 3.71 and only the difference is the inverted output of the TTL NOR gate operation due to the transistor Q6.

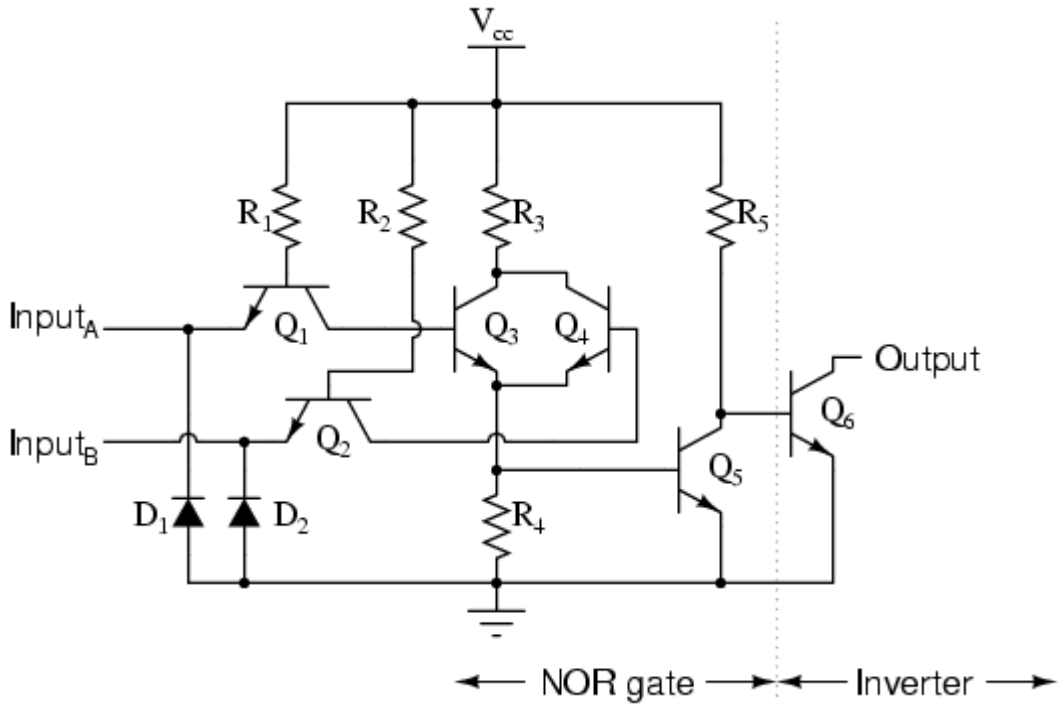


Figure 3.72 TTL OR gate

ECL (Emitter Coupled Logic):

Emitter-coupled logic (ECL) derives its name from the differential-amplifier configuration in which one side of the differential-amp consists of multiple-input bipolar transistors with their emitters tied together. An input bias on the opposite side of the differential-amp causes the amplifier to operate continuously in the active mode. Consequently, ECL consumes a relatively substantial amount of power in both states (one or zero) but also results in the fastest switching speeds of all logic families. An inherent benefit of ECL is the narrow switching level swing between devices (approximately 800 mV) which helps to reduce noise generation.

A typical ECL OR/NOR gate circuit shown in figure 3.73 which consists of a differential amplifier input circuit , a bias circuit ,emitter-follower outputs

Emitter follower outputs provide the OR logic function and its NOR complement . Because of the low output impedance of the emitter follower and the high input impedance of the differential amplifier input,high fanout operation is possible.In this type of circuit the saturation is not possible this results in higher power consumption and limited voltage swing. The lack of saturation in the ECL circuit permits high frequency switching

For best operation the Vcc pin is normally connected to ground and -5.2 volt from the power supply is connected to V_{EE} . Notice that in figure 3.73 (a). the output varies from LOW level of -1.75 volt to a HIGH level of -0.9 volt with respect to ground.

Beginning with LOWs are off because the base emitter junctions are reverse biased and that transistor Q3 is conducting but not saturated. The bias circuit holds the base of Q3 at -1.29 V and therefore its emitter is approximately 0.8 V below the base at -2.09 V. The voltage differential from base to emitter of the input transistors Q1 and Q2 is -2.09 V - (-1.75 V) = -0.34 V. This is less than the forward bias voltage of these transistors and they are therefore of this condition is shown in figure 3.73 (b).

When anyone or all of the inputs are raised to HIGH level (-0.9 V),that transistor (or transistors) will conduct. When this happen the voltage at the emitters of Q1 ,Q2 and Q3 increases from -2.09 V to -1.7 V (one base emitter drops below the -0.9 V base). Since the base of Q3 is held at a constant -1.29 V by the bias circuit, Q3 turns off. The resulting collector voltages are coupled through the emitter followers to the output terminals. Because of differential action of Q1 and Q2 with Q3 , Q3 is off when any one of Q1 and Q2 or both conduct, thus providing simultaneously complementary outputs. This is shown in figure 3.73(b). When all of the inputs are returned to the LOW state ,Q1 and Q2 are again cutoff and Q3 conducts.

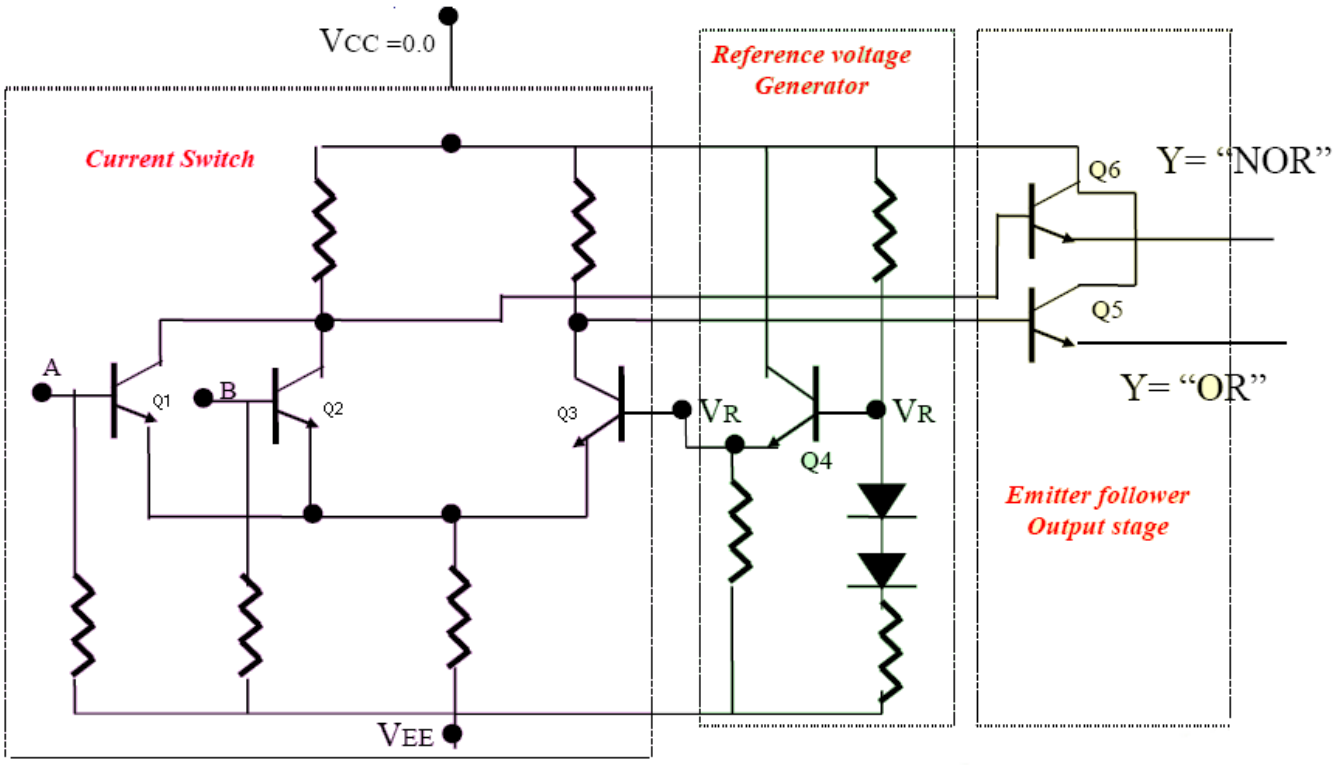


Figure 3.73 ECL OR/NOR gate circuit

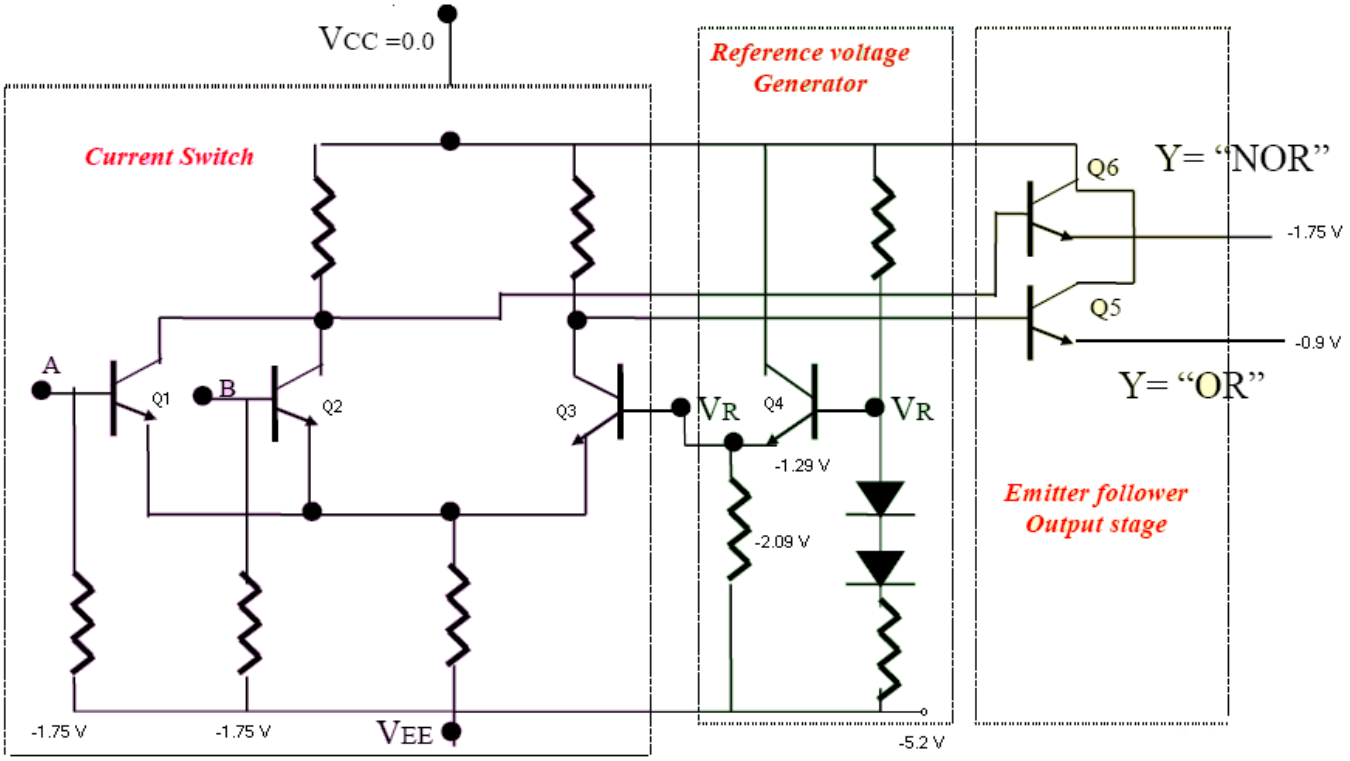


Figure 3.73 (a) ECL OR/NOR gate circuit operation

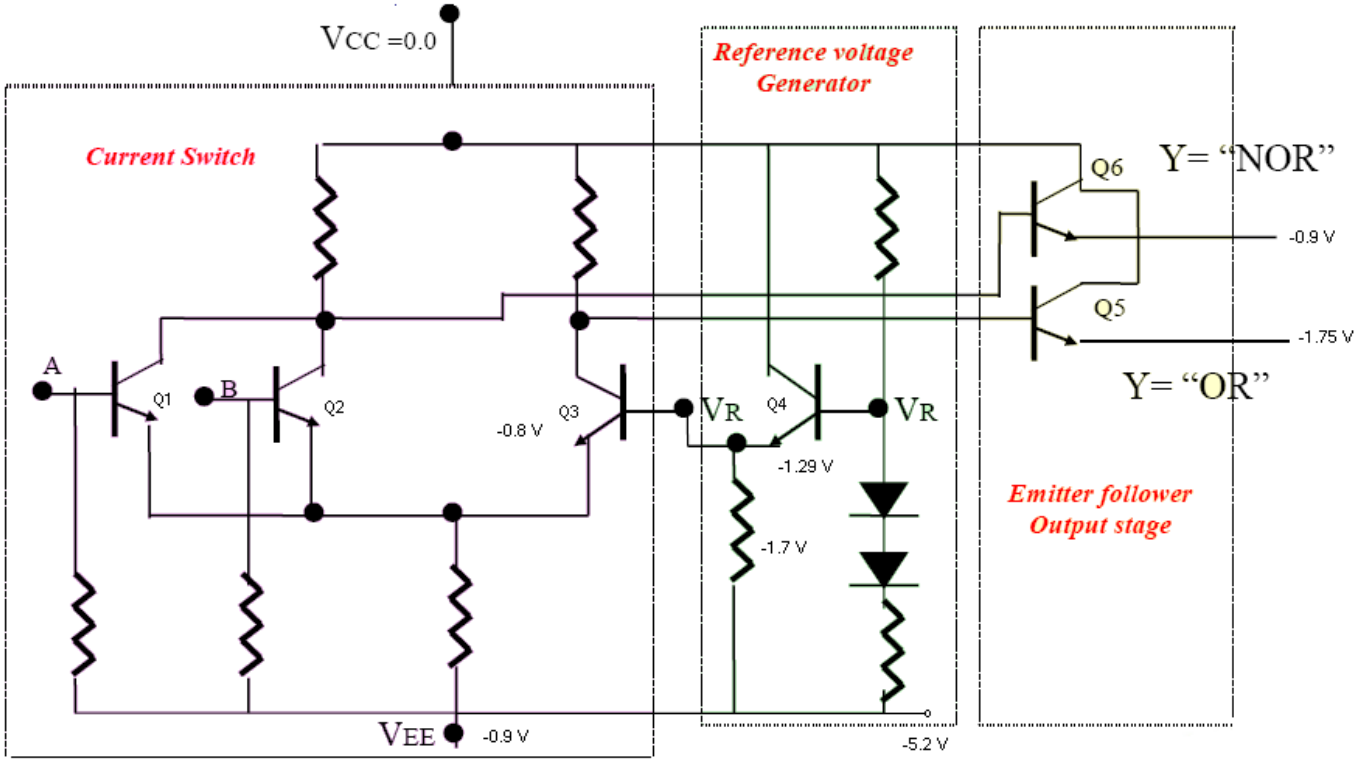


Figure 3.73 (b) ECL OR/NOR gate circuit operation

CMOS (Complementary Metal–Oxide–Semiconductor) Logic :

CMOS uses complementary and symmetrical pairs of p-type and n-type metal oxide semiconductor field effect transistors (MOSFETs) for logic functions. Two important characteristics of CMOS devices are high noise immunity and low static power consumption. Significant power is only drawn when the transistors in the CMOS device are switching(transition) between on and off states. Consequently, CMOS devices do not produce as much waste heat as other forms of logic do.

CMOS inverter:

CMOS or Complementary Metal Oxide Semiconductor logic is built using both NMOS and PMOS. Below is the basic CMOS inverter circuit, which follows these rules:

- NMOS conducts when its input is HIGH.
- PMOS conducts when its input is LOW.

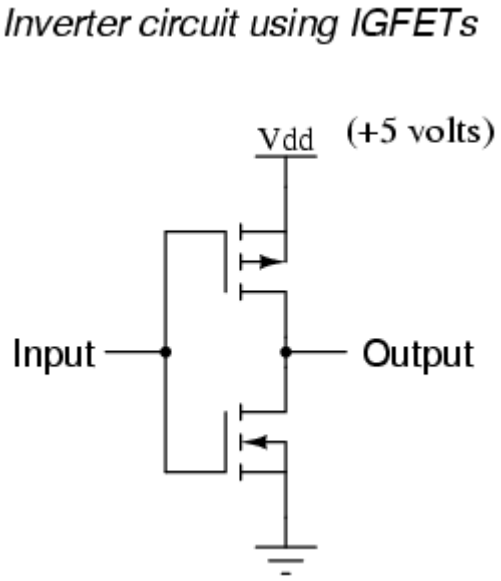


Figure 3.74 CMOS inverter

When input is HIGH, NMOS conducts, and thus output is LOW; when input is LOW PMOS conducts and thus output is HIGH. Commonly Standard CMOS logic circuit is divided into pull down (NMOS network) and pull up networks(PMOS network). The procedure to design any standard CMOS logic circuit is follows.

1.To design Pull down network

a. For AND operation,NMOS transistors are to be connected in series.

b. For OR operation ,NMOS transistors are to be connected in parallel.

2.To design a pull up network

a. For AND operation,PMOS transistors are to be connected in parallel.

b. For OR operation ,PMOS transistors are to be connected in series.

3.The actual output of the circuit is the inverted output.

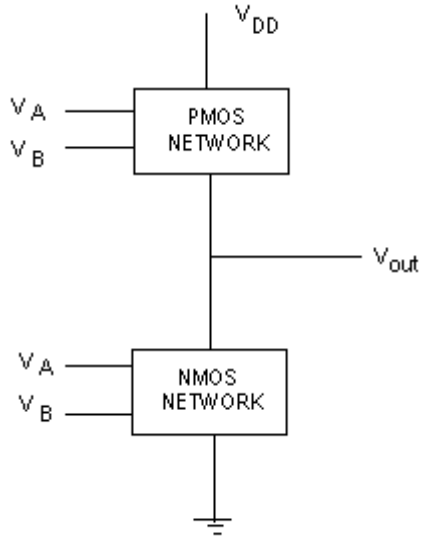


Figure 3.75 Basic PMOS and NMOS network

The implementation of NAND and NOR gates using the CMOS logic is described below.

CMOS NAND gate:

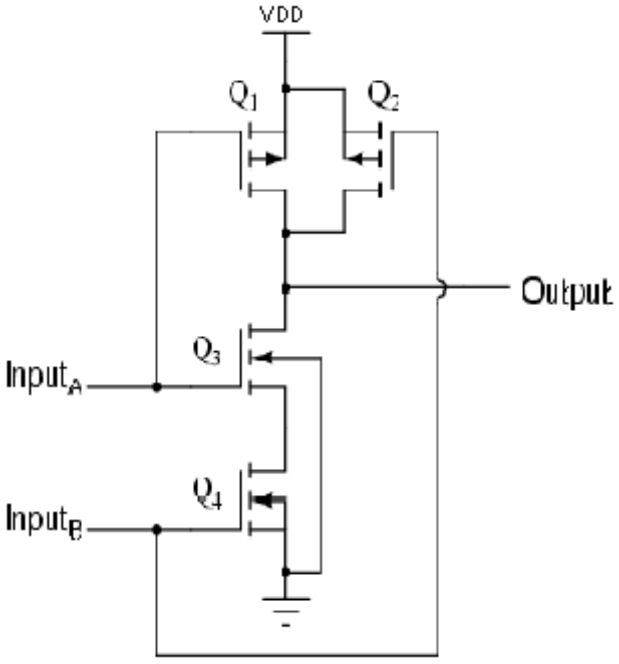


Figure 3.76 CMOS NAND gate

When A=B=0, PMOS transistors Q1& Q2 are turning on and Q3&Q4 are turned OFF so output isHIGH.
If A=0 & B =1 the Q1 is ON, Q2 is OFF ,Q3 is OFF and Q4 is ON therefore output is HIGH.

If A=1 & B =0 the Q1 is OFF, Q2 is ON ,Q3 is ON and Q4 is OFF therefore output is HIGH.

If A=1 & B =1 the Q1 is OFF, Q2 is OFF ,Q3 is ON and Q4 is ON therefore output is LOW.

AND gate operation can be implemented by connecting a CMOS inverter at the output of the NAND gate CMOS logic. The schematic diagram of CMOS AND gate is shown in the figure 3.77.

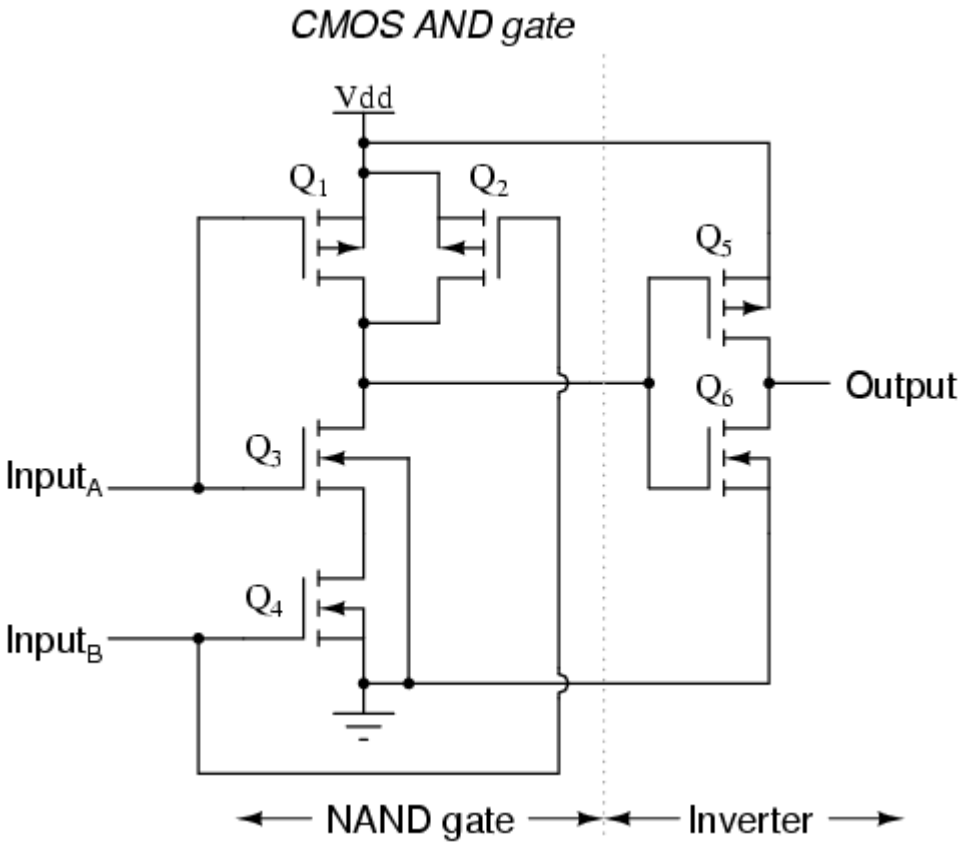


Figure 3.77 CMOS AND gate

CMOS NOR gate:

A CMOS NOR gate circuit uses four MOSFETs just like the NAND gate, except that its transistors are differently arranged. Instead of two paralleled sourcing (upper) transistors connected to V_{dd} and two

series-connected sinking (lower) transistors connected to ground, the NOR gate uses two series-connected sourcing transistors and two parallel-connected sinking transistors. This is illustrated in the figure 3.78

As with the NAND gate, transistors Q_1 and Q_3 work as a complementary pair, as do transistors Q_2 and Q_4. Each pair is controlled by a single input signal.

If either input A or input B are "high" (1), at least one of the lower transistors (Q_3 or Q_4) will be saturated, thus making the output "low" (0). Only in the event of both inputs being "low" (0) will both lower transistors be in cutoff mode and both upper transistors be saturated, the conditions necessary for the output to go "high" (1). This behavior, of course, defines the NOR logic function.

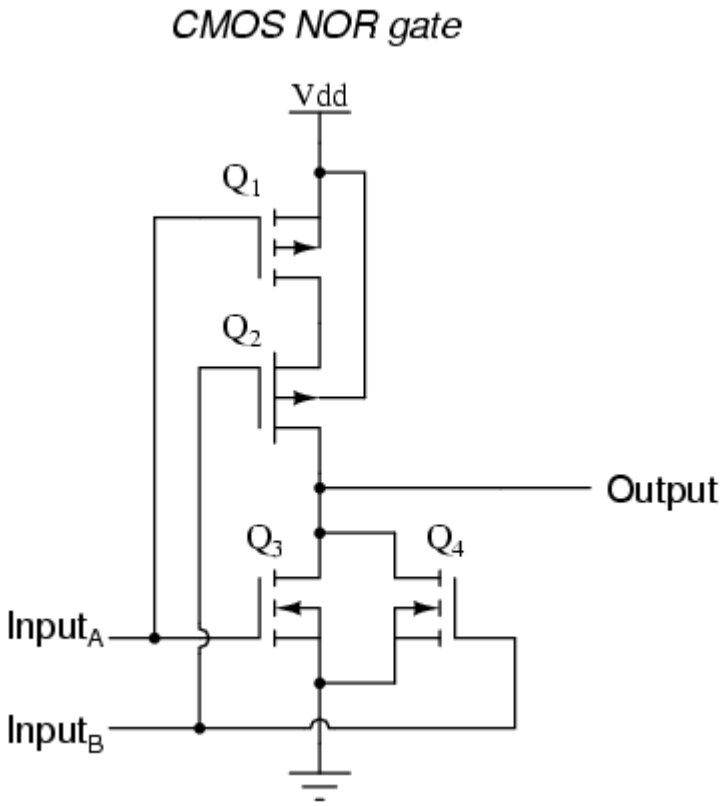


Figure 3.78 CMOS NOR gate

3.11.3 Comparison of logic families:

The table below compares some typical characteristics of several popular logic families available in the market today. The following sections provide brief explanations of the various parameters.

Typical Commercial Parameter (0°C to +70°C)	**Logic Families**												
	TTL				**CMOS**					**ECL**			
	LS	**ALS**	**ABT**	**FAST**	**MG**	**HC**	**FACT**	**LVC**	**LCX**	**10H**	**100K**	**ECL in PS(3)**	**E-Lite**
Speed													
Gate Prop Delay (ns)	9	7	2.7	3	65	8	5	3.3	3.5	1	0.75	0.33	0.22
Flip-Flop Toggle Rate (MHz)	33	45	200	125	4	45	160	200	200	330	400	1,000	2800
Output Edge Rate (ns)	6	3	3	2	50	4	2	3.7	3.6	1	0.7	0.5	0.25
Power Consumption Per Gate (mW)													
Quiescent	5	1.2	0.005	12.5	0.0006	0.003	0.0001	0.003	1E-04	25	50	25	73
Operating (1 MHz)	5	1.2	1.0	12.5	0.04	0.6	0.6	0.8	0.3	25	50	25	73
Supply Voltage (V)	+4.5 to +5.5	+4.5 to +5.5	+4.5 to +5.5	+4.5 to +5.5	+3 to +18	+2 to +6	+1.2 to +3.6	+2 to +3.6	+2 to +6	-4.5 to -5.5	-4.2 to -4.8	-4.2 to -5.5	-4.2 to -5.5
Output Drive (mA)	8	8	32/64	20	1	4	24	24	24	50 ohm load	50 ohm load	50 ohm load	50 ohm load
5V Tolerant													
Inputs	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	Yes	N/A	N/A	N/A	N/A
Outputs	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No	Yes	N/A	N/A	N/A	N/A
DC Noise Margin (1)													
High Input %	22	22	22	22	30	30	30	30	30	27	41	28/41	33
Low Input %	10	10	10	10	30	30	30	30	30	31	30	31/31	33
Packaging(4)													
DIP	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	No
SO	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes
LCC	No	Yes	No	Yes	No	No	Yes	No	No	Yes	No	Yes	No
SSOP	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No
TSSOP	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	No	No	No

Table 3.25 comparison of logic families

Chapter 4

Linear Integrated Circuits

Contents

4.1 Introduction to Integrated Circuits

4.2 Theory and operation of op-amp

4.2.1 Block Diagram of an op-amp

4.2.2 Open Loop op-amp configurations

4.2.3 Closed loop op-amp configurations

4.3 Parameters of an op-amp

4.3.1 DC Parameters

4.3.2 AC Parameters

4.4 Applications Of Operational Amplifier

4.4.1 Comparator

4.4.2 Voltage follower

4.4.3 Summer

4.4.4 Subtractor

4.4.5 Differentiator

4.4.6 Integrator

4.4.7 Active Filters

4.5 Analog to Digital Converter

4.5.1 Types of Analog to Digital converter

4.6 Digital to Analog Converter

4.6.1 Types of Digital to Analog Converter

4.7 Summary

Feedback Questions

References

4.1 Introduction to Integrated Circuits:

The integrated circuit (IC) is a miniature form of complex electronic circuit whose components are fabricated on a single chip[1]. The ICs are fabricated using Monolithic or Hybrid technology. Most of the ICs are produced by the monolithic process. In this process all the active components like transistors and all the passive components like resistors,capacitors are fabricated on a single piece of semiconductor material called substrate. The main advantage of integrated circuits are small size,low cost,less weight,low supply voltages,low power consumption and highly reliability. In addition to these advantages,the ICs are used in system design because of their versatility,flexibility,dependability and scalability.

Integrated Circuits could be classified based on their operation as Analog Ics, Linear ICs and Digital ICs. In Linear ICs ,the output varies in proportion to the input. There is a linear relationship between input and output (eg.op-amp IC 741) .Digital ICs operate at only a few defined levels ,rather than over a continuos range of signal amplitudes(eg. logic gates).Analog Ics produce continuous output for continuous input. Also ICs can be classified according to the number of components integrated on a single chip as follows:

Small-Scale integration(SSI)	Upto 100 components per chip
Medium-Scale integration(MSI)	From 100 to 3000 components per chip
Large-Scale integration(LSI)	From 3000 to 10,000 components per chip
Very Large-Scale integration(VLSI)	From 10000 to 1000000 components per chip
Ultra Large Scale Integration (ULSI)	More than one million components per chip

The components density of a chip has increased together with some increase in the chip area.(The chip areas range from 1600mil^2 (1mm^2) for the SSI chip to 160,000mil^2 (1cm^2)for the LSI chip.) Generally the size of the chips is measured by *mil. 1mil = 0.001 inch = 25.4 μm =.0254 mm.*

The basic processes involved in fabrication of ICs are

1. Substrate (Silicon wafer) preparation
2. Epitaxial growth
3. Oxidation

1 Chip stands for closely and highly integrated package.

4. Photolithography
5. Diffusion
6. Ion implantation
7. Isolation technique
8. Metallization
9. Assembly processing and packaging

IC Packages

The package functions as both a carrier and as an enclosure. As a carrier it allows the device's functionality to be fully accessed and electrically interconnected to other devices. As an enclosure it provides physical and chemical protection to the circuits and facilitates assembling.

The basic IC package types(figure 4.1) are insertion type (through hole) and Surface mounting type which are described in table 4.1 & 4.2. Common package materials are ceramic and plastic. In through hole package type pins(or leads) are inserted into a hole in the circuit boards and soldered in place from the opposite side of the board. SMT packages have leads that are soldered directly to corresponding exposed metal lands on the surface of the circuit board. The advantages of SMT are

- Elimination of holes
- Components can be mounted on both sides of the PCB
- Smaller dimensions
- Improved package parasitic components
- Increased circuit-board wiring density

Table 4.1 Classifications of Through hole packages

Style	Package		Characteristics	Material
Dual In Line	DIP	DIP	Leads are on the both sides of the package	plastic
		SDIP	DIP with a shrunk lead pitch	Plastic
		HDIP	DIP with heat sink	Plastic
Single In Line	SIP	SIP	Leads are on the single side of the package in the straight line	Plastic
		SSIP	SIP with a shrunk lead pitch	Plastic
	ZIP	ZIP	Leads protude from two opposite sides of the package in a straggered configuration.	Plastic
		SZIP	ZIP with shrunk lead pitch	Plastic
		HZIP	ZIP with heat sink	Plastic

Table 4.2 Classifications of Surface Mount packages

Style	Package		Characteristics	Material
Dual In Line	SOP	SOP	Leads are on two opposite sides of the package in a gull wing configuration.	Plastic
		SSOP	SOP with a lead pitch of 1.0 mm or less	Plastic
		HSOP	Thermally-enhanced SOP with either a radiation fin on its leads or an internal heat spreader under the die	Plastic
Single In Line	SIP	HSIP	SIP with a heat sink	Plastic

Style	Package		Characteristics	Material
Quad Type	QFP	QFP	Leads are on four sides of the package in a gull wing configuration.	Plastic/Ceramic/ Glass Shield
		LQFP	QFP with a seated height of 1.7 mm	Plastic
		HQFP	Thermally-enhanced QFP with either a radiation fin on its leads or an internal heat spreader under the die	Plastic
		TQFP	QFP with a seated height of 1.2 mm or less	Plastic
	QFH	QFH		
	QFN	QFN	Leads are on four sides and the bottom of the package.	Plastic
		HQFN		plastic

Style	Package		Characteristic	Material
Area Array	BGA	UFBGA	0.5mm<UFBGA≤0.65mm BGA with a minimum ball pitch of 0.8 mm or less, a minimum seated height of over 0.5 mm and a maximum seated height of 0.65 mm or less	Plastic
		XFBGA	XFBGA≤0.50mm BGA with a minimum ball pitch of 0.8 mm or less and a maximum seated height of 0.5 mm or less	Plastic
	LGA	MLGA		Plastic
		XLGA	XFLGA≤0.50mm LGA with a minimum ball pitch of 0.8 mm or less and a maximum seated height of 0.5 mm or less.	Plastic
		ULGA	0.5mm<UFLGA≤0.65mm LGA with a minimum ball pitch of 0.8 mm or less, a minimum seated height of over 0.5 mm and a maximum seated height of 0.65 mm or less	Plastic

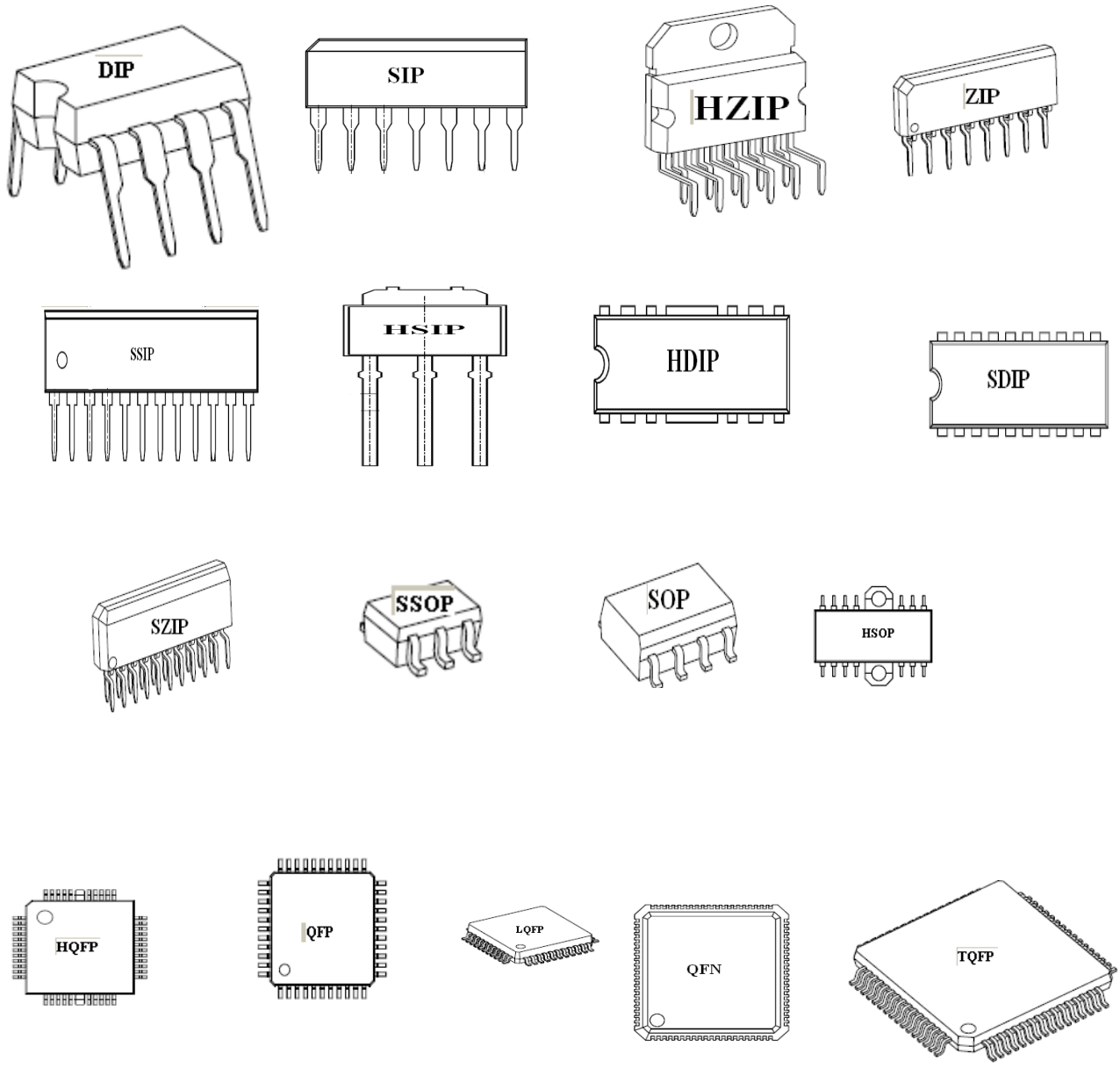


Figure 4.1 IC Packages

Operational amplifier:

An operational amplifier popularly known as o*p-amp* is a direct-coupled high gain amplifier consisting of one or more differential amplifiers followed by a level translator and an output stage. The output stage is generally a push-pull or complementary-symmetry pair. An op-amp is available as a single integrated circuit package .

The first op-amp was manufactured by Fairchild (μA741).This op-amp is also manufactured by various other manufacturers like National Semiconductor (LM741), Motorala (MC1741),Texas Instruments (SN32741) and etc. The schematic diagram of an op-amp is shown in figure:4.2 . The pin details of an IC 741 is shown in the figure 4.3.

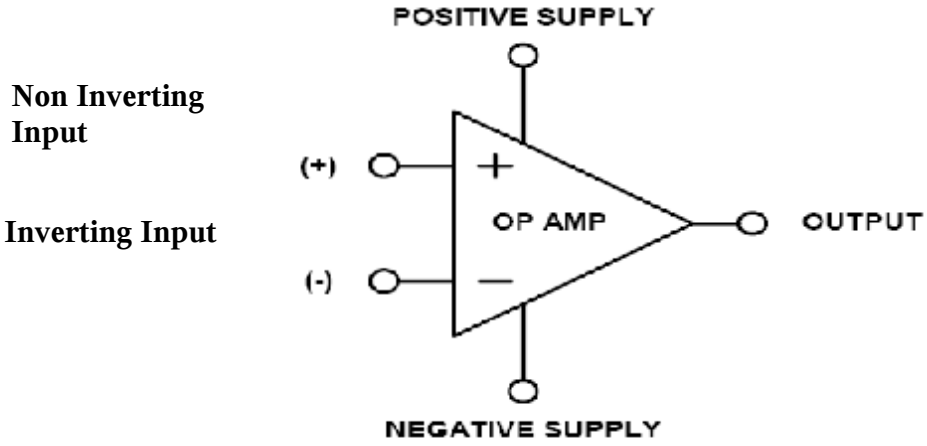


Figure 4.2 Schematic diagram of an Op-amp

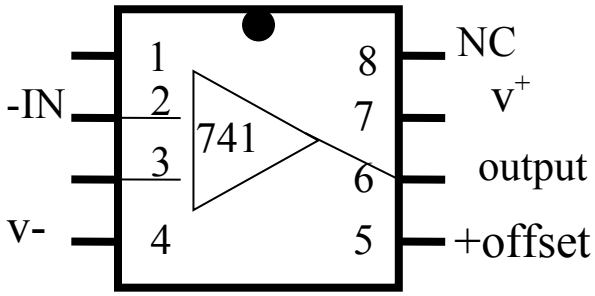


Figure 4.3 Pin details of an IC 741

Input and Output Terminals:

An op-amp has two inputs called inverting and non-inverting terminals. The input at inverting terminal results in the opposite polarity(antiphase or 180° phase difference) output. While the input at non-inverting terminal results in the same polarity output.(0° phase difference).

Power Supply connections

The op-amp is usually powered by a dual polarity balanced power supply in the range of +/- 5 volts to +/- 30 volts. The dual polarity is called balanced,if the voltage of the positive supply and negative supply are same in magnitude. If the two voltage magnitudes are not same in the dual supply it is called as unbalanced dual supply. This is shown in figure 4.4and 4.5

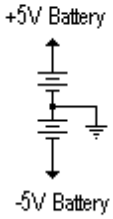


Figure 4.4 Balanced Dual Supply

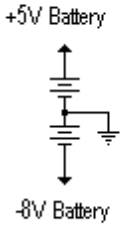


Figure 4.5 Unbalanced Dual Supply

Op-amp Packages

Complex integrated circuits involving many op-amps and other ICs can now be fabricated on a single large chip or by interconnecting many large chips and placing them in a single package. The three basic types of Linear IC packages are

1. The flat pack
2. The metal can or transistor pack
3. The dual-in-line package

Flat Pack

In the flat pack, the chip is enclosed in a rectangular ceramic case with terminal leads extending through the sides and ends. The flat pack is available with 8,10,14,16 leads or pins. These pins accommodate the power supplies, inputs, outputs and special connections required to complete the circuit.

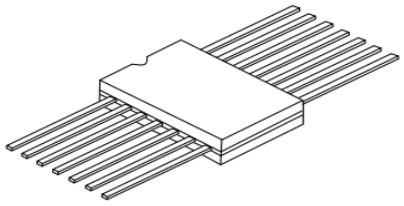


Figure 4.6 Flat pack

Metal Can or Transistor Pack

In the metal can or transistor pack, the chip is enclosed in a metal or plastic case. The transistor pack is available with 3,5,8,10 or 12 leads or pins. The metal can package is best suited for power amplifiers because metal is good heat conductor and has better dissipation capability than the flat pack or dual-in-line package. The metal can package permits the use of external heat sinks.

Figure 4.7 Metal Can

Dual- in- Line Package

In this type the chip is mounted inside a plastic or ceramic case.The DIP is the most widely used package type because it can be mounted easily.The DIP is available with 8,12,14,16 and 20 pins. In general, as the density of components integrated on the single chip increases, the number of pins also goes up.

Figure 4.8 Dual-in-line Package

Properties of an ideal op-amp:

Infinite Open loop Voltage Gain($A_{OL} = \infty$)

Open-Loop Gain A_{OL} is the gain of the op-amp without positive or negative feedback. In the ideal op-amp A_{OL} is infinite. Typical values range from 20,000 to 200,000 in practical devices

Infinite Input impedance($Z_i = \infty$)

Input Impedance is the ratio of input difference Voltage(Vid) to input current(Ii)

$$Z_i = \frac{V_{id}}{I_i} \qquad \text{----------(4.1)}$$

When Z_i is infinite, the input current I_i=0. Real op-amps have the input impedance in the range of megaohms.

Zero Output impedance($Z_{out} = 0$)

The ideal op-amp acts as a perfect internal voltage source with no internal resistance. But practical op-amps have internal resistance in series with the load thus reducing the output voltage available to the load. The output-impedance in the range of 50 to 200Ω.

Zero offset voltage(Vo=0)

The output offset voltage is the output voltage of an amplifier when both inputs are grounded. This is illustrated in the figure 4.9. The ideal op-amp has zero output offset, but real op-amps have very small amount of output offset voltage in the range of few mV. (+/-15mV)

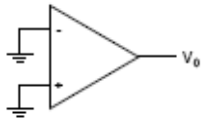


Figure 4.9 Zero offset voltage Circuit

Infinite Bandwidth(BW= ∞)

The range of frequency over which the amplifier performance is satisfactory is called as its bandwidth. It defines the operating frequency of the op-amp. The bandwidth of an ideal op-amp is infinite. This ensures that the gain of the op-amp will be constant over the frequency range from d.c. (zero frequency) to infinite frequency. So op-amp can be used to amplify dc as well as ac signals.

Infinite CMRR(CMRR= ∞)

The ability of a differential amplifier to reject a common mode signal is expressed by a ratio called CMRR. In other words CMRR measures the relative sensitivity of an op-amp to a

difference signal as compared to a common mode signal. It is defined as the ratio of the differential voltage gain Ad to common mode voltage gain Ac,

$$\therefore \quad \mathbf{CMRR} = \left|\mathbf{(Ad/Ac)}\right| \qquad \text{----------(4.2)}$$

When the same voltage is applied to both the inputs,the differential amplifier is said to be operated in a common mode configuration. Example:Many disturbance signals, noise signals appear as a common input signal to both the input terminals of the differential amplifier. Such a common signal should be rejected by the differential amplifier. Ideally the common mode voltage gain is zero,hence the ideal value of CMRR is infinite. For a practical differential amplifier Ad is large and Ac is small hence the value of CMRR is very large.

CMRR is also expressed in dB as,

$$\mathbf{CMRR\ in\ dB = 20log\ \left|(Ad/Ac)\right| dB} \qquad \text{----------(4.3)}$$

Virtual Ground

In the ideal op-amp, a voltage applied to one input also appears at the other input. Consider the input signal is applied to the inverting terminal through resistor and the non-inverting terminal is grounded. The difference input voltage is ideally zero;that is, the voltage at the inverting terminal is approximately equal to that at the non-inverting terminal. In other words, the inverting terminal voltage is approximately at ground potential. Therefore the inverting terminal is said to be at virtual ground.

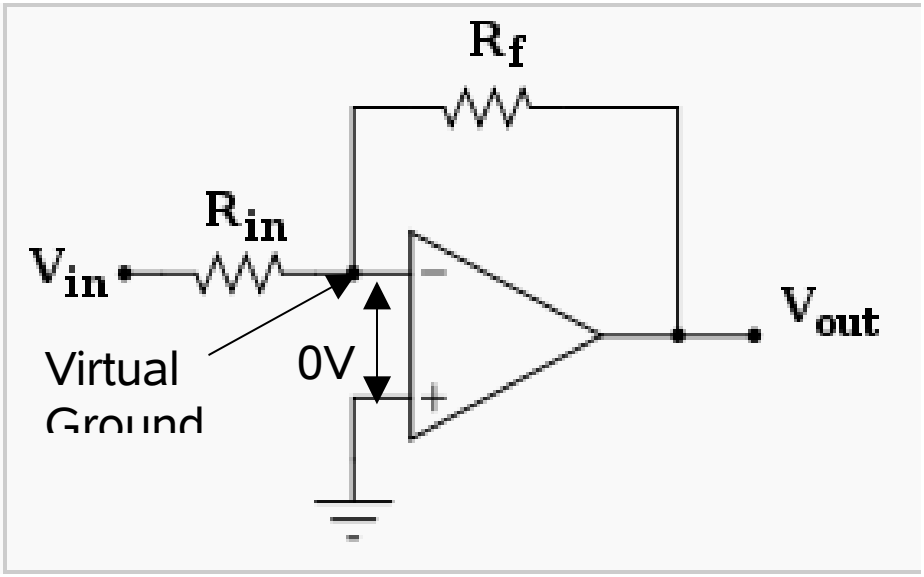


Figure 4.10 Op-amp Inverting Amplifier

Equivalent circuit of an op-amp

The equivalent circuit is useful in analyzing the basic operating principles of the op-amps**.** The figure 4.11 shows the equivalent circuit of an op-amp. The Operational Amplifier, or Op-amp is one of the basic building blocks of Analog Electronic Circuits. It is a linear device that has all the properties required for ideal dc. amplification and is used extensively in signal conditioning, filtering or to perform mathematical operations such as addition, subtraction, integration and differentiation. An Operational Amplifier is a 3-terminal device that consists basically of two high impedance inputs, one an Inverting input marked with a negative sign ("-") and the other a Non-inverting input marked with a positive plus sign ("+"). An ideal Operational Amplifier also has a Low Output impedance that is referenced to a common ground terminal. They can also operate from either a single or a dual (±) power supply.

An Operational Amplifier on its own has a very high open-loop d.c. gain and by applying some form of negative feedback we can produce an amplifier circuit that has a very precise gain characteristic. An operational amplifier only responds to the difference between the voltages on its two input terminals, known commonly as the Differential Input Voltage and not to their common potential. Thus, the gain of the Op-amp is known as the Open Loop Differential Gain, and is given the symbol (A_{OL}).

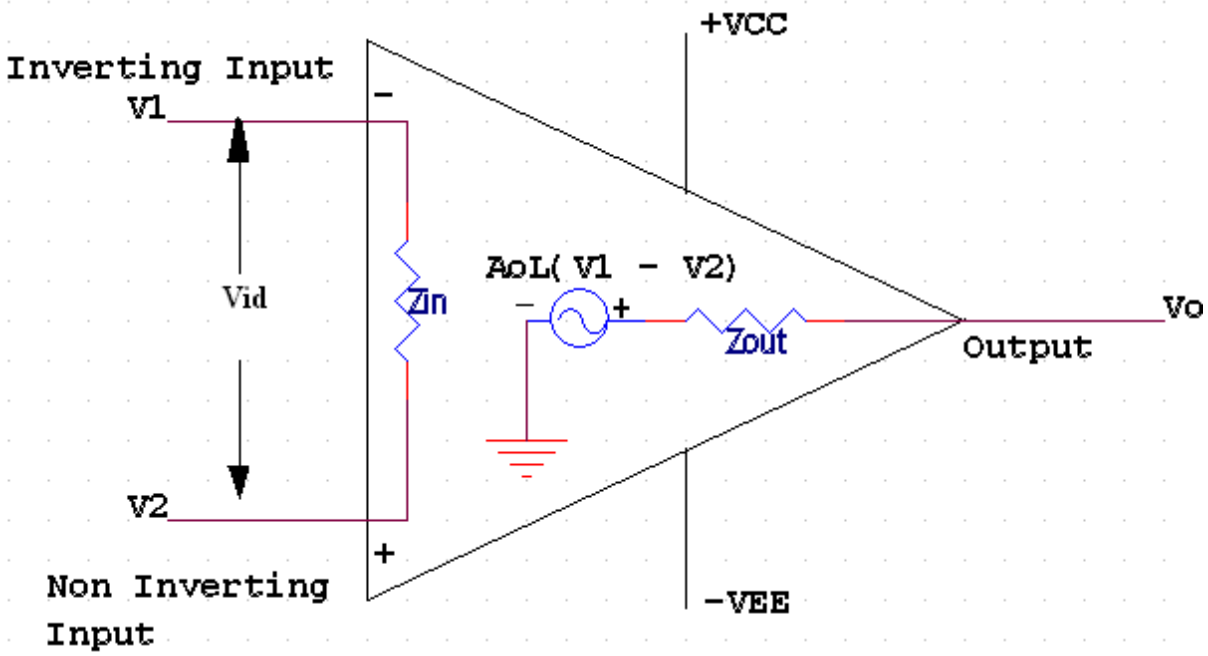


Figure 4.11 Equivalent circuit of an op-amp

The main function of an op-amp is to amplify the input signal and the more open loop gain it has the better.For an ideal amplifier the gain will be infinite.The input impedance(Zin) is assumed to be infinite to prevent current flowing from the source supply into the amplifiers input circuitry. The output impedance(Zout) of an ideal operational amplifier is assumed to be zero so that it can supply as much current as necessary to the load.

The open circuit output voltage(Vo), is proportional to the difference of the voltages at the non-inverting terminal(Vi1) and at the inverting terminal(Vi2) . This can be expressed by the equation The output voltage is

$$Vo=A\,Vid = A(Vi1 - Vi2) \qquad \text{----------(4.4)}$$

where A = Amplification factor or Gain of the amplifier

Vid = difference input voltage

Vi1 = voltage at the non inverting input terminal with respect to ground.

Vi2 = voltage at the inverting input terminal with respect to ground.

This equation is useful in studying characteristics of the op-amp and in analyzing different circuit configurations using op-amp that employ feedback .

Ideal Voltage Transfer Curve

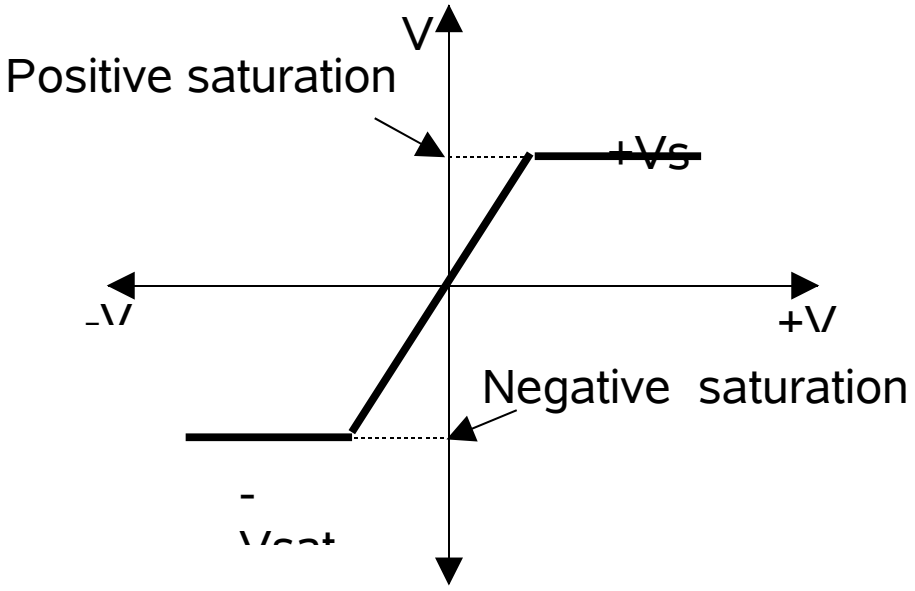


Figure 4.12 Ideal Voltage Transfer Curve

An ideal voltage transfer curve shown in figure 4.12 relates the output voltage (Vo) and the input difference voltage keeping the gain A as constant. The output voltage(Vo) is directly proportional to the input difference voltage (Vid) only until it reaches the saturation voltages and that thereafter output voltage remains constant. Note, the output voltage cannot exceed the positive and negative saturation voltages. These saturation voltages are specified by an output voltage swing rating of the op-amp for given values of supply voltages.

4.2 Theory and Operation of Op-amp:

4.2.1 Block Diagram of an Op-amp:

An op-amp usually consists of four cascaded blocks. The block diagram of IC op-amp is shown in the figure 4.13.

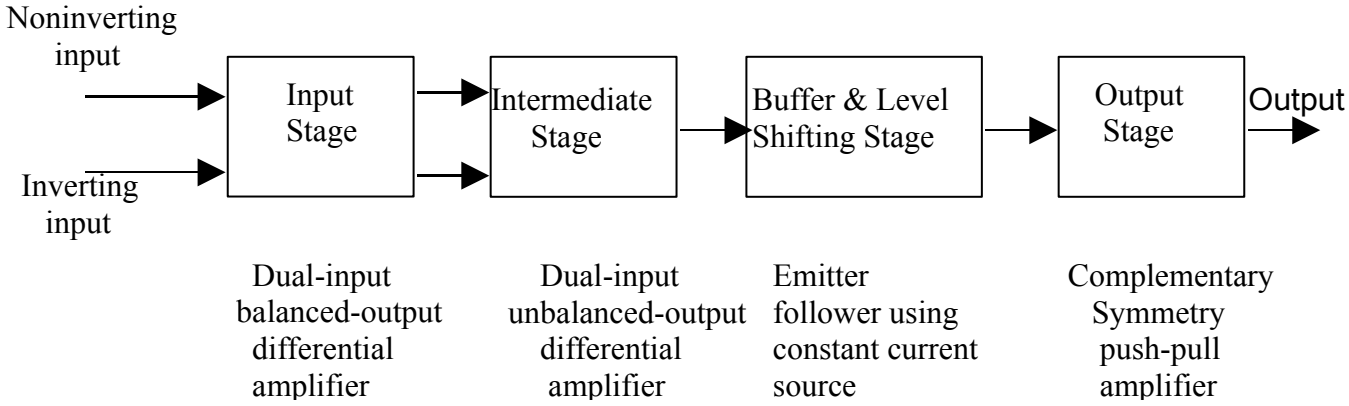


Figure 4.13 Internal Block Schematic of an Op-amp

Input Stage:

The input stage requires high input impedance to avoid loading on the sources. It requires two input terminals. It also requires low output impedance. All Such requirements are achieved by using the dual input,balanced output differential amplifier as the input stage. The function of a differential amplifier is to amplify the difference between the two input signals. The differential amplifier has high input impedance. This stage provides most of the voltage gain of the amplifier.

Intermediate Stage:

The output of the input stage drives the next stage which is an intermediate stage. This is another differential amplifier with dual input,unbalanced i.e. Single ended output. The overall gain requirement of the op-amp is very high. The input stage alone cannot provide such a high gain. The main function of the intermediate stage is to provide an additional voltage gain required. Pratically, the intermediate stage is not a single amplifier but the chain of cascade amplifiers called as multistage amplifiers.

Level Shifting Stage:

All the stages are directly coupled to each other. As the op-amp amplifies dc signals also,the coupling capacitors are not used to cascade the stages. Hence the dc quiescent voltage level of previous stage gets applied as the input to the next stage. Hence stage by stage dc. Level increases well above ground potential. Such a high dc voltage level may drive the transistor into saturation. This further may cause distortion in the output due to clipping. This may limit the maximum ac. Output voltage swing without any distortion. Hence before the output stage,it is necessary to bring such a high dc. Voltage level to zero volts with respect to ground.

The level shifter stage brings the dc. Level down to ground potential,when no signal is applied at the input terminals. Then the signal is given to the last stage which is the output stage.

The buffer is usually an emitter follower whose input impedance is very high. This prevents loading of the high gain stage.

Output Stage:

The basic requirements of an output stage are low output impedance,large a.c.output voltage swing and high current sourcing and sinking capability.

The push-pull complementary amplifier meets all these requirements and hence used as an output stage. This stage increases the output voltage swing and keeps the voltage swing symmetrical with respect to ground. The stage raises the current supplying capability of the op-amp.

The schematic circuit diagram for an op-amp as shown in the figure 4.14 is explained as follows. From the circuit an op-amp consists of four stages.1) an input stage; 2) an intermediate stage; 3) a level-shifting stage; and 4) an output stage.

The biasing current provides the device for driving all of the transistors in the active mode. The biasing branch consists of Q12, R5, and Q11.The current generated is approximately

($V_{CC} - V_{EE} - 2Vbe$) / R5.This current is delivered to input stage through Q11.The current at the base of Q11 is mirrored to base of Q10. This current is then delivered to base of Q3 and Q4.

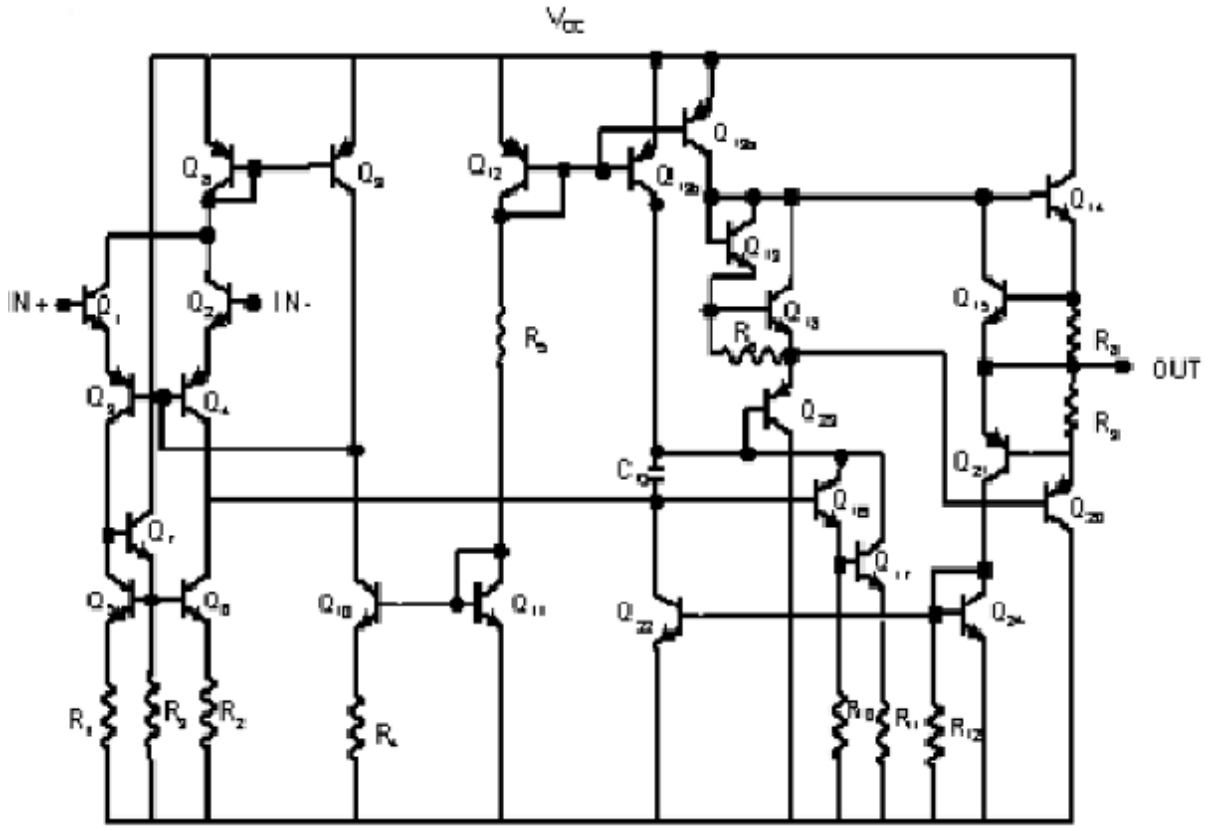


Figure 4.14 Schematic circuit diagram for an op-amp IC741

Differential Input Stage:

The differential input stage consists of Q1 and Q2 in series with Q3 and Q4.This serial combination provides high gain per stage needed to achieve the adequate open-loop gain in a two stage amplifier Any differential signal across either one of these inputs will flow current through the input stage and then to Q8 and Q9.Q1 and Q2 are connected as emitter followers. This provides the high input impedance to the input stage.Q3 and Q4 provide DC level shifting. They also protect Q1 and Q2 from emitter-base junction breakdown. For NPN transistors, the emitter-base junction breakdown happens at about 7 volts of reverse bias. For PNP transistors, this breakdown happens at about 50 volts.

Intermediate Stage :

The transistors Q5,Q6 and Q7 along with the resistors R1, R2, and R3 form the active load for Q3 and Q4.This is a high impedance load and this load converts the differential signal of the input transistors into a single ended signal at Q6 collector. The single ended signal is then delivered to the second stage of the amplifier, which is the gain stage.

Voltage Gain Stage:

Signal from collector at Q6 is fed into Q16 base. Q16 is connected as an emitter follower,which gives this stage very high input impedance. Since it minimizes the loading of the input stage of the amplifier, it prevents gain loss at this stage.Q17 is connected as a common emitter amplifier. The load of Q17 is the output resistance of Q13b and reaches the PNP transistor and parallels with the input resistance of Q23, which is another PNP transistor. Output of the second stage of the amplifier is taken as collectors of Q17.

Level Shifting Stage:

A transistor current source used as a load resistance is called a active load. This achieves very high gains without the actual need for high resistors and helps to save chip area. A large value resistor would require a large supply voltage source. Whereas an active load, do not need a high supply voltage. An example of this active load is the load on collector of Q17. The network consisting of transistors Q18,Q19 and R10 is a fixed voltage level shifter, shifting the voltage output of Q17 by a fixed amount on its way to the output complementary stage formed by Q14. The level shifter network is designed to bias the output stage in the linear region. Transistor Q22 performs two functions. It serves as a buffer between Q17 and Q20 and also provides a negative feedback to Q16. A compensation capacitor is used in this part of the circuit. It's a Miller compensation capacitor, which provides frequency compensation for the amplifier. It's connected in the feedback path and it creates a dominant pole at about 5 Hz. This is going to shift the other poles of the amplifier further out. It also gives the op-amp a fairly uniform 20dB per decade drop.

Output Stage:

Q14 the source transistor and Q20 the sink transistor form the output complementary stage of the amplifier. When the output of the amplifier is pulled positive, Q14 conducts more and pulls the output more towards the positive supply. The opposite happens when the output is pulled negative, it means Q20 is going to conduct more and the output is actually pulled more towards the negative supply. This amplifier has short circuit protection and not all amplifiers have short circuit protection. Q15 provides short circuit protection for the source transistor Q14 and Q21 provides short circuit protection for the sink transistor Q20. Also in this output stage Q18 and Q19 bias the output transistors in the linear region and these two transistors are fed by Q13a.

Q23 is connected as an emitter follower. This minimizes the loading effect on the output stage of the amplifier. The output of the amplifier has very low output impedance.

4.2.2 Open Loop op-amp configurations

The term open loop signifies that there is no connection, either direct or via another network, between the output and input terminals. That is , the output signal is not fedback in any form to the input terminal. and the loop that would have been formed otherwise with feedback is open. When connected in open loop configuration,the op-amp simply functions as a high gain amplifier.

There are three open loop configurations:

1.Differential Amplifier

2.Inverting Amplifier

3.Noninverting Amplifier

Differential Amplifier:

The differential amplifier amplifies the difference between two input voltage signals. Hence it is called difference amplifier.

Consider an differential amplifier shown in the figure 4.15

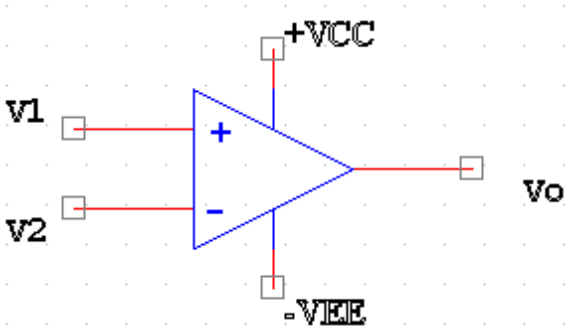


Figure 4.15 Differential amplifier

V1 and V2 are the two input signals while Vo is the single ended output. Each signal is measured with respect to the ground. In a differential amplifier,the output voltage Vo is proportional to the difference between the two input signals. Hence we can write,

$$Vo \ \alpha \ (V1-V2) \qquad \text{-------(4.5)}$$

From the equation (4.5) we can write,

$$V0=A_{OL}(V1 - V2) \qquad \text{-------(4.6)}$$

The difference between the two inputs (V1-V2) is generally called difference voltage(Vid).

$$V0 = A_{OL}\, Vid \qquad \text{-------(4.7)}$$

Hence the open loop differential gain can be expressed as,

$$A_{OL} = V0/Vd \qquad \text{------(4.8)}$$

Generally the open loop differential gain is expressed in its decibel (db) Value as,

$$A_{OL} = 20\log_{10}(A_{OL}) \text{ in dB} \qquad \text{------(4.9)}$$

Inverting Amplifier:

In the inverting amplifier input is applied to the inverting input terminal. The noninverting input terminal is grounded as shown in the figure 4.16. Since V1=0 and V2=Vin,then the output voltage becomes,

$$Vo = -A_{OL}V_2 = -A_{OL}V_{in} \qquad \text{------(4.10)}$$

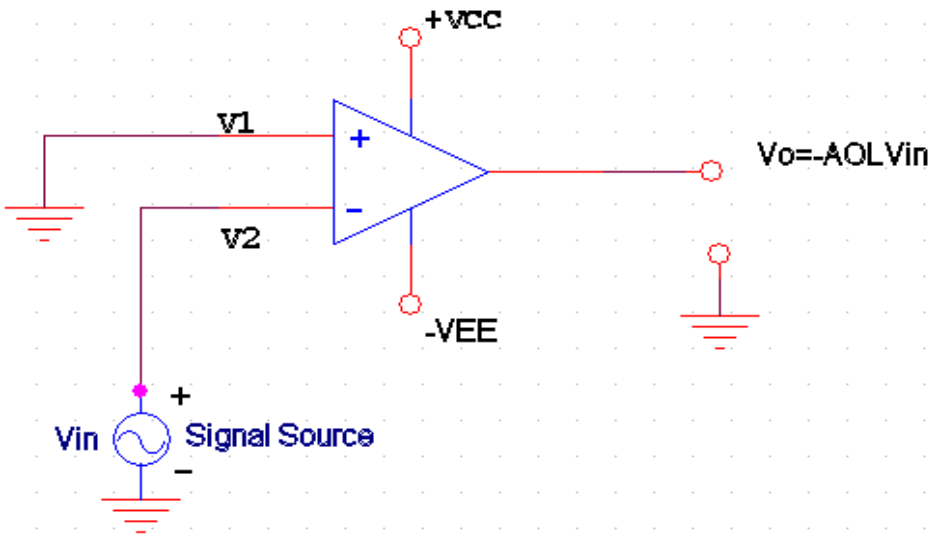


Figure 4.16 Inverting Amplifier

The negative sign indicates that the output voltage is out of phase(180°) with respect to input Thus in the inverting amplifier the input signal is amplified by gain AoL and is also inverted at the output.

Noninverting Amplifier

In this configuration the input is applied to the noninverting input terminal,and the inverting terminal is connected to ground as shown in the figure 4.17.

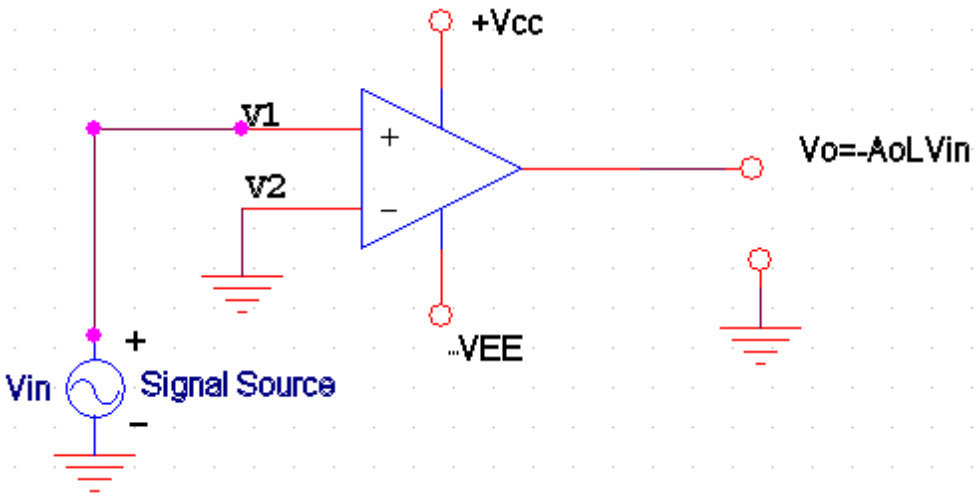


Figure 4.17 Noninverting Amplifier

Since V1=Vin and V2=0V. Then the output becomes,

$$V0=A_{OL}Vin \qquad \text{------(4.11)}$$

From the equation 4.11, the output voltage is larger than the input voltage by gain AoL and is in phase with the input signal. In all three open-loop configurations any input signal(differential or single) that is only slightly greater than zero drives the output to saturation level. This results from the very high gain(A_{OL}) of the op-amp. Thus,when operated open-loop,the output of the op-amp is either negative or positive saturation or switches between positive and negative saturation levels. For this reason, open-loop op-amp configurations are not used in linear applications.

4.2.3 Closed loop op-amp configurations

As seen before, the op-amps cannot operate linearly in open loop configuration. But the utility of an op-amp can be considerably increased by operating it in closed loop configuration.

The closed loop configuration is performed with the help of feedback.Typically output of the op-amp is controlled either by negative feedback,which largely determines the output voltage gain,or by positive feedback,which is mainly used in oscillators.

In the linear applications, op-amp is always used with negative feedback. The main reason for using op-amp with negative feedback, is that the relationship between the input and output is dependent on

external components such as resistors and not on the internal circuitry of amplifiers. Hence, it is easy to predict the output and overcome non-linearity and distortion in the system. The negative feedback is possible by adding a resistor as shown in figure 4.18 called feedback resistor.

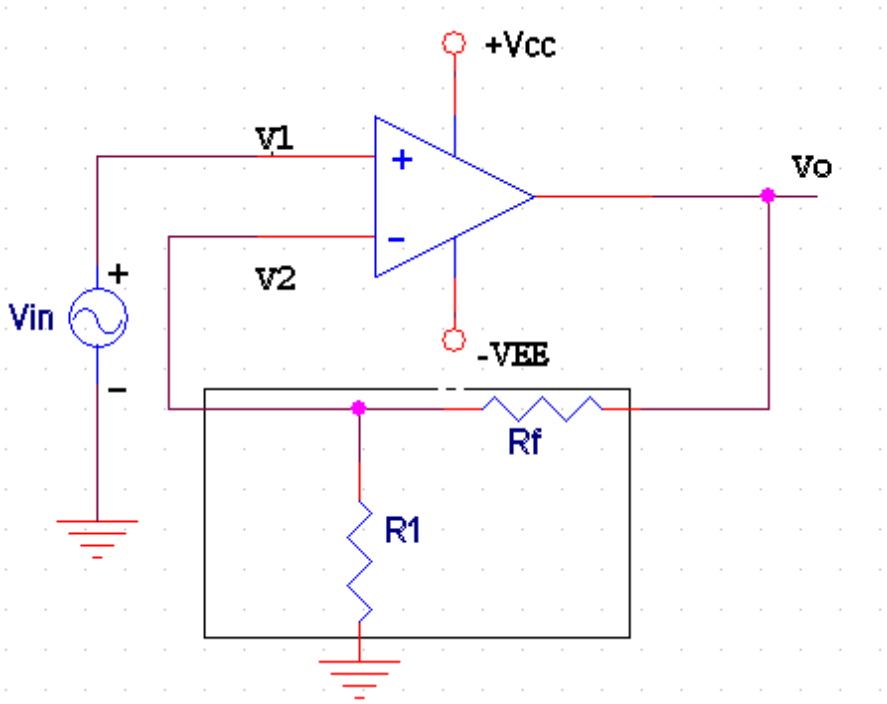


Figure 4.18 Op-amp with Negative Feedback

The gain resulting with feedback is called closed loop gain of the op-amp. The closed loop gain is much less than the open loop gain due to the feedback resistance.

Inverting Amplifier:

In inverting closed loop configuration, a resistor is used to feedback part of the output signal into the inverting input of the amplifier. In this configuration the noninverting input is grounded. The circuit characteristics are determined by values of feedback resistor and input resistor. Since no current can flowing into the amplifier, all the current from the feedback path go through the input resistor. So, basically, the input current and the feedback current are equal. Since the noninverting input is grounded, the inverting input at virtual ground. The feedback current is equal to -Vout /RF. The negative sign indicates that the output signal is completely out of phase with the input signal. The input current is Vin / Rin. Since these two currents are equal, we find the gain relationship of the operation amplifier to be Vout / Vin = -RF / Rin. This is referred to as the

operational amplifier's closed loop gain.

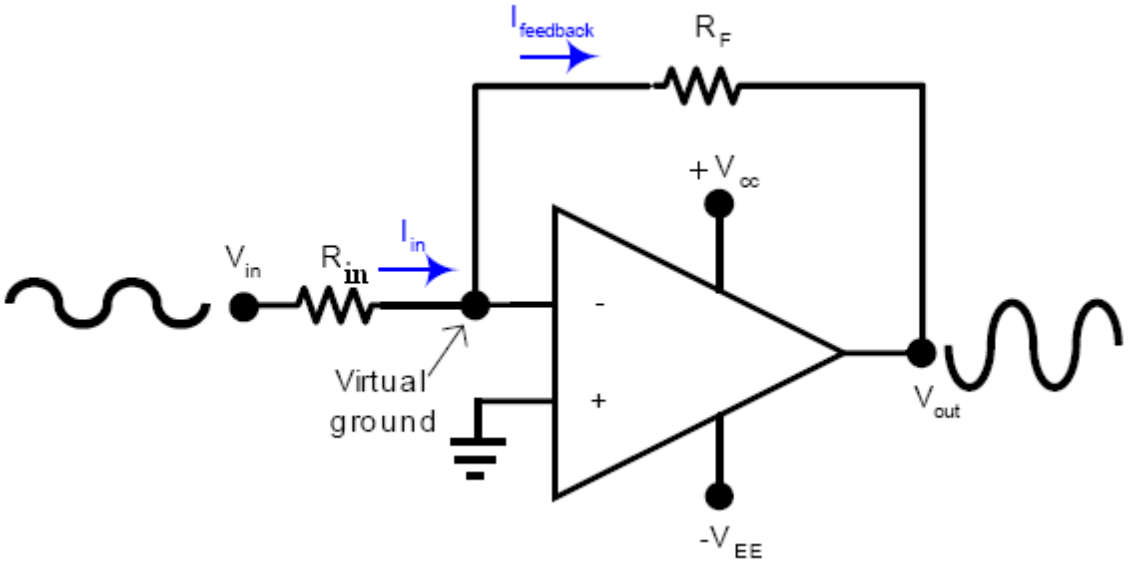


Figure 4.19 Inverting Amplifier

Noninverting Amplifier:

In the non-inverting closed loop configuration, a feedback resistor is used to feedback part of the output signal back into the inverting input. The input signal is applied to the non-inverting input of the amplifier.Since the feedback path is around the inverting pin, negative feedback is applied in this configuration also.

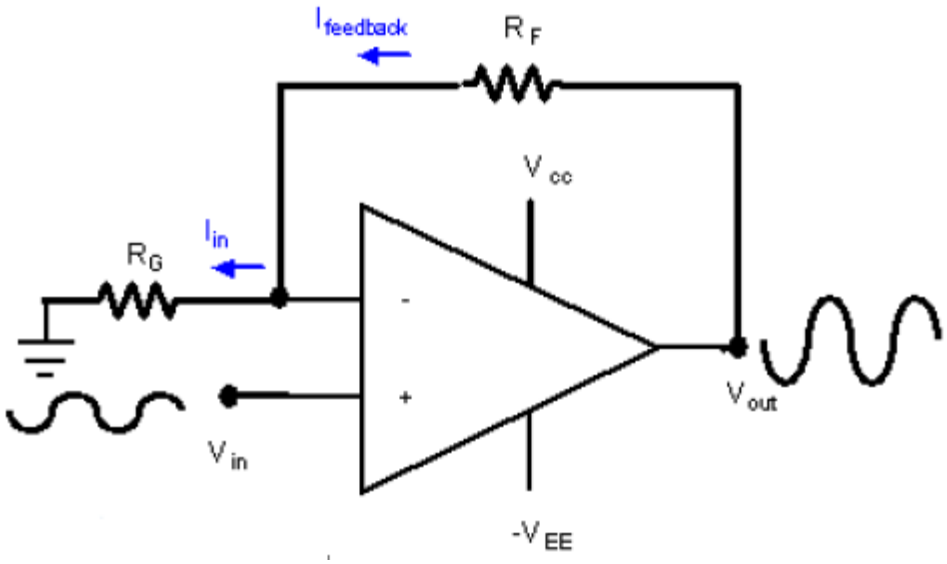


Figure 4.20 NonInverting Amplifier

It is seen that the feedback current is equal to the input current. Also the voltage on the inverting input is equal to that on the non-inverting input. From the circuit as shown in the figure4.20, since R_f and R_{in} form a voltage divider, the feedback voltage is given by

feed back voltage= $V_{out}R_{in}/(R_f + R_{in})$. ------(4.12)

Since the two input voltages to the amplifier are the same, the input vin is equal to $V_{out}R_{in}/(R_f + R_{in})$.

Therefore the gain of this circuit is given by

$V_{out}/V_{in} = (R_f + R_{in})/R_{in}$,

$= 1+(R_f/R_{in})$ ------(4.13)

4.3 Parameters of an op-amp

4.3.1 DC Parameters

Input Bias Current

In an ideal op-amp, it is assumed that no current is drawn from the input terminals. However, practically input terminals do conduct a small value of dc current even though both the input terminals are grounded . A positive input bias current(I_B^+) refers to small current that's seen on the non-inverting input of the amplifier and similarly, negative input bias current(I_B^-) is the small amount of current that's seen on the inverting terminal of the amplifier. Input bias current refers to the average of these two values. Simply put, input bias current is the average of the two input currents of the amplifier.

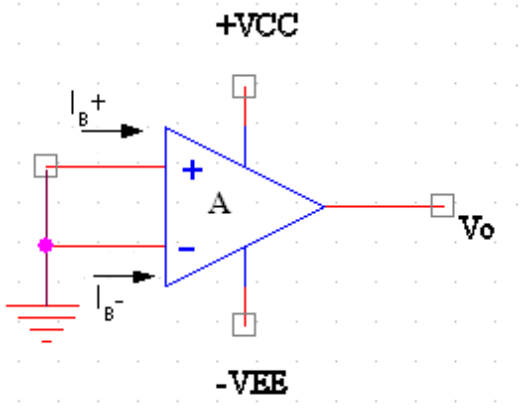


Figure 4.21 Input Bias Current

So, $I_B = (I_B^+ + I_B^-)/2$ ------(4.14)

The input bias current (I_B) has a maximum value of 500nA.

Input Offset Current

The algebraic difference between the bias currents entering into the inverting input terminal and the non inverting input terminal is referred to as input offset current. The input offset current is used to determine(as an indicator of) the degree of mismatching between these two bias currents. It is 200nA maximum for IC 741.

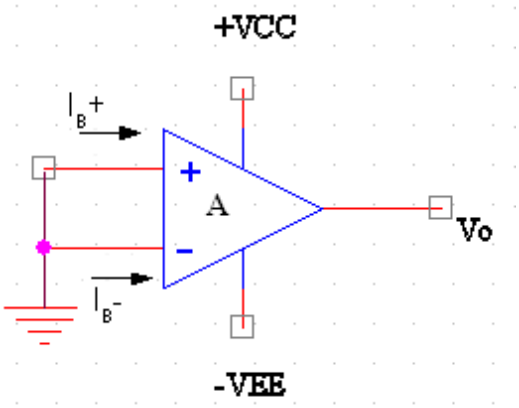


Figure 4.22 Input Offset Current

$$Iio = |I_B^+ - I_B^-| \qquad \text{------(4.15)}$$

Input Offset Voltage

It is the differential input voltage that exist between two input terminals of an op-amp without any external inputs applied. It is the voltage that must be applied between the input terminals of an op-amp to nullify the output since this output is not zero with zero input voltages. This is illustrated in the figure 4.23. The input offset voltage can range from microvolts to millivolts and can be of either polarity. The input offset voltage for 741IC is 150uV maximum.

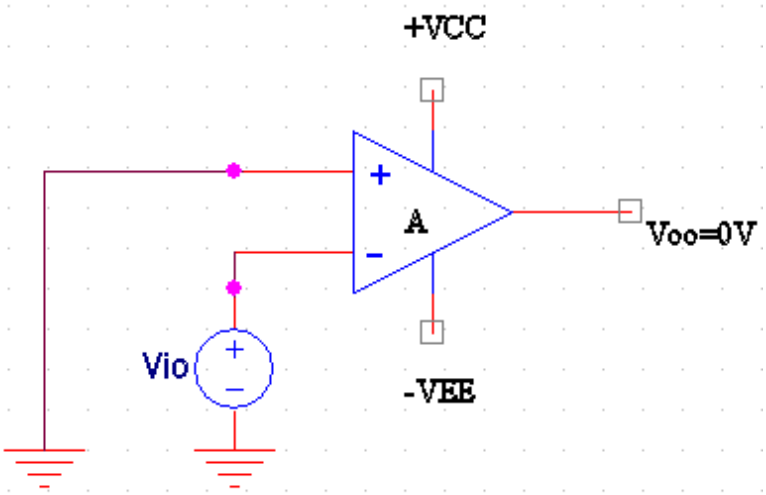


Figure 4.23 Input Offset Voltage

Output Offset Voltage

In practice op-amp has some dc output error voltage called output offset voltage even though both inverting and non inverting input terminals are grounded. This output offset voltage is caused by mismatching between two input terminals because of the two transistors in the input differential amplifier stage may not have the same characteristics. It is either positive or negative dc voltage depending on whether the potential difference between two terminals is positive or negative. Typically for IC 741 output offset voltage range is +/-15mv. Many op-amps provide offset compensation pins to nullify the offset voltage. So it is necessary to refer the manufacturer specifications while using the offset null connections. The manufacturers recommend that a 10Kohm potentiometer (can be used to nullify the output offset voltage) to be placed across offset null pins 1 and 5 and the wiper of the potentiometer be connected to the negative supply pin 4.It is illustrated in the figure 4.25. If the given op-amp does not have these offset null pins,external balancing techniques can be used.

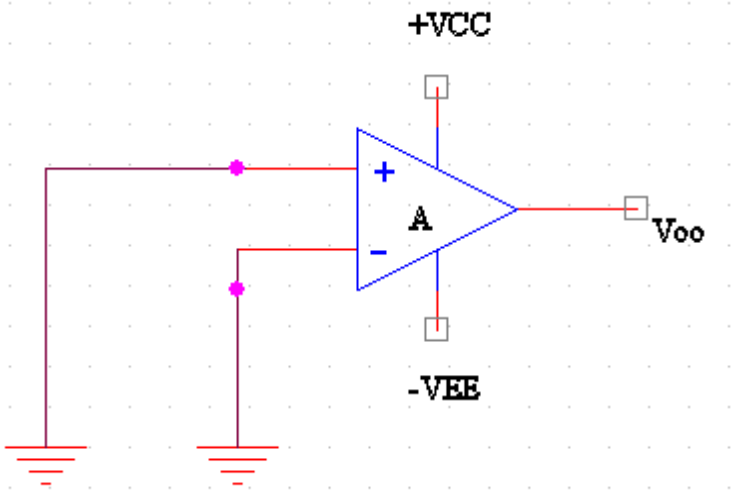


Figure 4.24 Output Offset Voltage

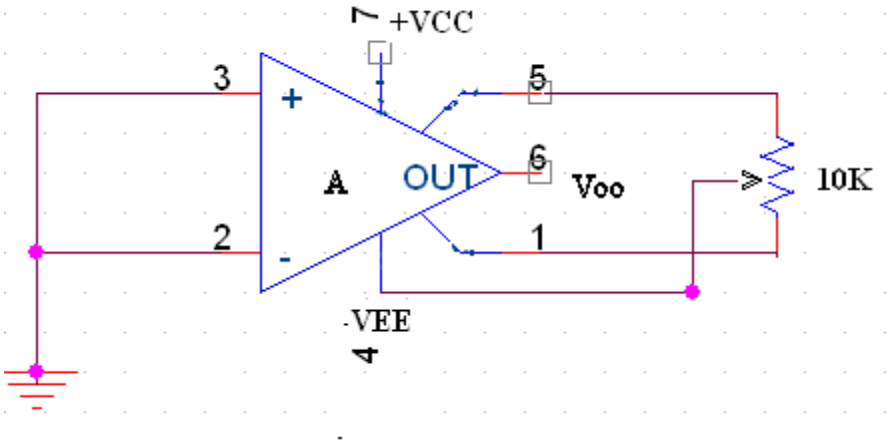


Figure 4.25 Output Offset Voltage Compensation technique

Power Supply Rejection Ratio PSRR

The power supply rejection ratio (PSRR) is defined as the ratio of the change in input offset voltage due to the change in supply voltage producing it, keeping other power supply voltage constant. It is also called Power Supply Sensitivity(PSS) or Supply Voltage Rejection Ratio(SVRR). The basic amplifier is shown in figure 4.26

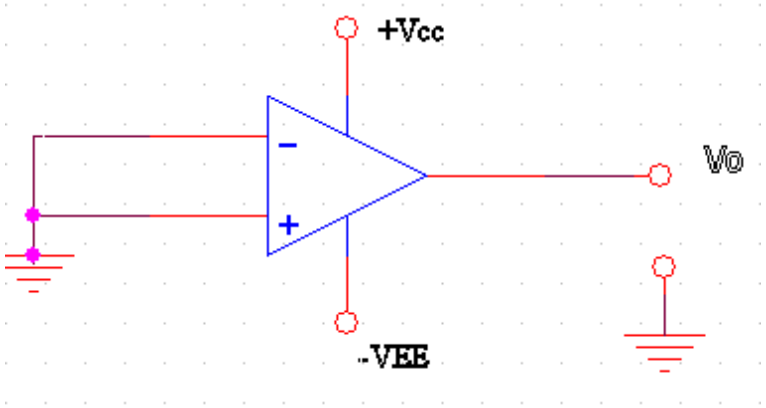


Figure 4.26 Basic Amplifier

Now if VEE is constant and due to certain change in VCC, there is change in input offset voltage then PSRR is defined as PSRR = ΔVios/ΔVCC ---------(4.16)

For a fixed VCC, if there is a change in VEE then

PSRR = ΔVios/ΔVEE ----------(4.17)

As input offset voltage is very small, PSRR is specified in μV/V or in decibels. The typical value of PSRR for IC741 op-amp is 30μV/V. The value of the PSRR is as low as possible for a good op-amp.

4.3.2 AC Parameters

Slew Rate(SR)

The slew rate (SR) is defined as the maximum rate of change of the output with respect to time of an op amp circuit. It is specified in V/us .The slew rate in general describes the degradation effect on the high frequency response of the active amplifier. The slew rate is caused due to limited charging rate of the compensating capacitor present in the internal circuit, the op-amp output voltage is not responding immediately to a fast change in input since the capacitor requires a finite amount of time to charge and discharge.

The maximum rate at which the capacitor charges is the Slew rate and the voltage across the capacitor tis the output voltage.

$$V=1/c\int Idt \qquad \text{----------(4.18)}$$

Therefore SR = dV/dtmax = Imax/C ----------(4.19)

4.4 Applications Of Operational Amplifier:

4.4.1 Comparator:

A comparator is a circuit which compares a signal voltage applied at one input of an op-amp with a known reference voltage at the other input,and produces either a high or a low output voltage,depending on which input is higher. As comparator output has two voltage levels,either high or low,it is not linearly proportional to input voltage. An open loop configuration or sometimes in a positive feedback configuration of op-amp can be used as a comparator.

Non Inverting Comparator:

In comparator,when the noninverting voltage is larger than the inverting voltage,the output is high(positive saturation voltage Vsat+). While if the noninverting input is less than the inverting input,the comparator produces a low output voltage(negative saturation voltage Vsat-).The simplest type of comparator is the basic noninverting comparator using op-amp,as shown in the figure 4.27.Since the inverting input terminal is grounded the reference voltage is zero for this comparator. As long as Vin is positive (above ground potential), the output voltage is +Vsat;but when Vin become negative(below ground potential) ,the output is -Vsat.

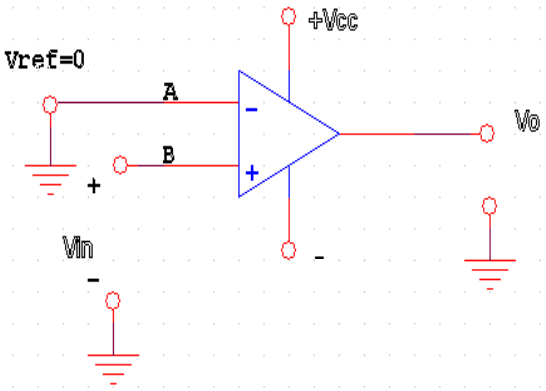


Figure 4.27 Noninverting Comparator

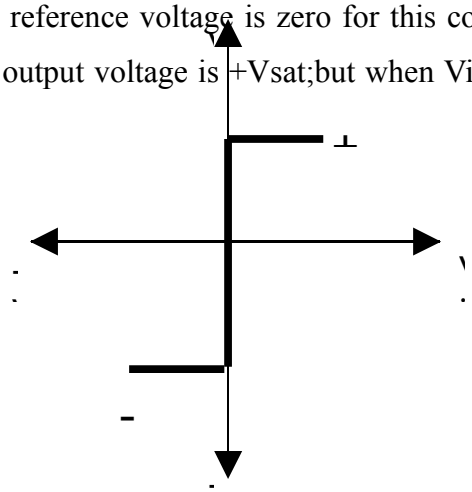


Figure 4.28 Transfer Characteristics

The op-amp voltage gain A is very large. So when inverting input is grounded,very small input voltage Vin in the range of microvolt is enough to saturate the op-amp. The ±Vsat the saturation voltage levels of op-amp are mentioned in the data sheet. Hence knowing Vsat and the voltage gain A,we can determine the minimum input voltage level required to saturate op-amp as,

$$Vin(min) = Vsat/A \; . \qquad \text{------(4.20)}$$

The input output voltage transfer characteristic of an ideal noninverting comparator as shown in the figure4.28 . As gain A is very large, for very very small positive or negative Vin, the output saturates. Hence at Vin=0,the transfer characteristics is almost a straight line. as shown in figure 4.28

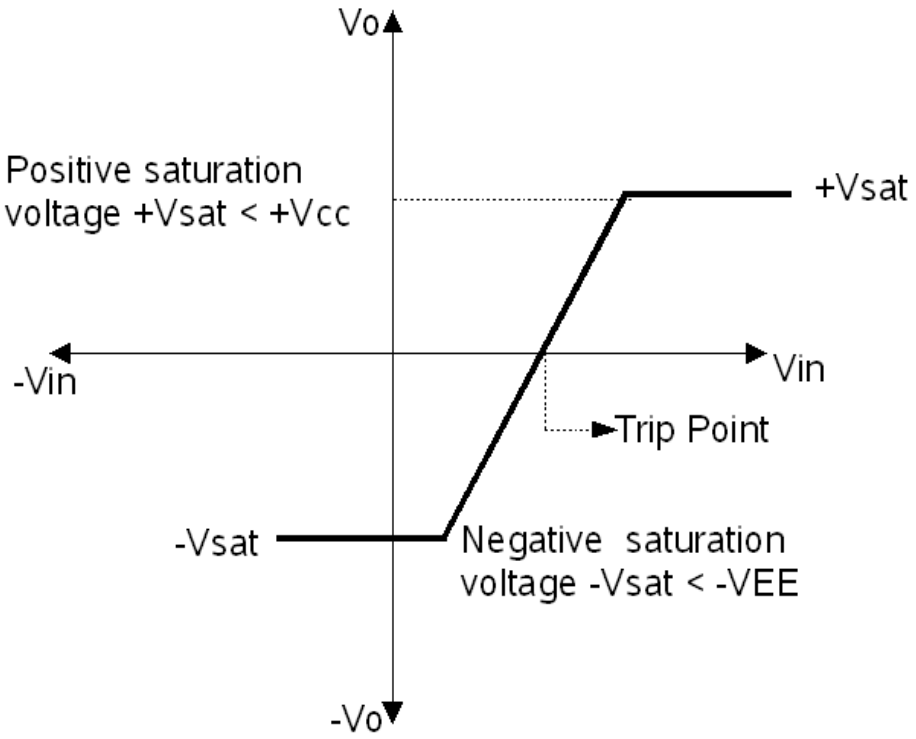


Figure 4.29 Transfer Characteristics Of Practical NonInverting Comparator.

The figure 4.29 shows the input output voltage transfer characteristics of a practical non inverting comparator. The input voltage at which the output changes its states from low to high or high to low is the termed as trip point . In the ideal comparator this trip point is zero as at Vin=0,the output changes its states. So when Vin is greater than trip point,the output is high while if Vin is less than the trip point the output is low. As this change over occurs at Vin=0,the basic comparator can be used to detect occurrence of zero in the input voltage. Hence this circuit is called *Zero crossing detector.*

Inverting Comparator:

The Figure 4.30 shows inverting comparator in which the reference voltage Vref is applied to the noninverting (+) input and signal voltage (Vin) is applied to the inverting (-) input of the op-amp.

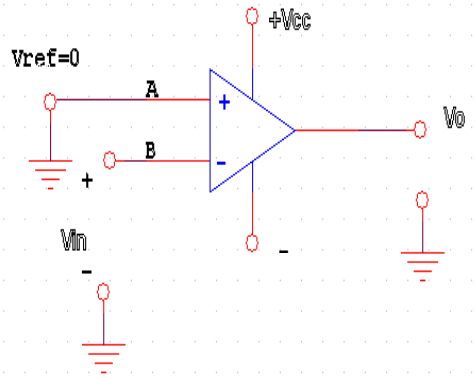


Figure 4.30 Basic Inverting Comparator.

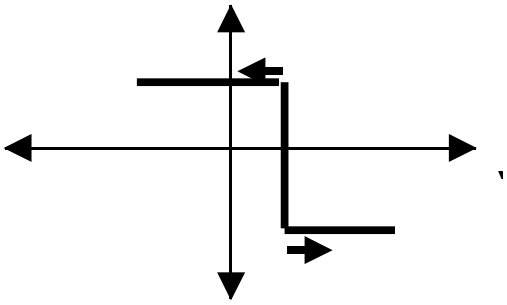


Figure 4.31 Transfer characteristics

When Vin is less than Vref,the output voltage Vo is at +Vsat(≅ +Vcc)because the voltage at the inverting input(-) is less than that at the noninverting (+) input. On the other hand,when Vin is greater than Vref, the noninverting (+)input becomes negative with respect to the inverting (-) input, and Vo goes to -Vsat (≅ $-V_{EE}$)

Transfer characteristics for inverting comparator with +Vref is shown in the **figure 4.31**

Applications of comparator

Important applications of the comparators are Zero crossing detector, Window detector,Time marker generator,Phase meter. Some of the applications are explained here

Zero crossing detector

The basic inverting comparator can be used as the zero crossing detector provided that Vref=0 as shown in figure 4.32. The output voltage Vo is driven into negative satuartion when the input signal vin passes through zero in the positive direction and when the vin passes through zero in the negative direction, the output voltage switches and saturate positively.

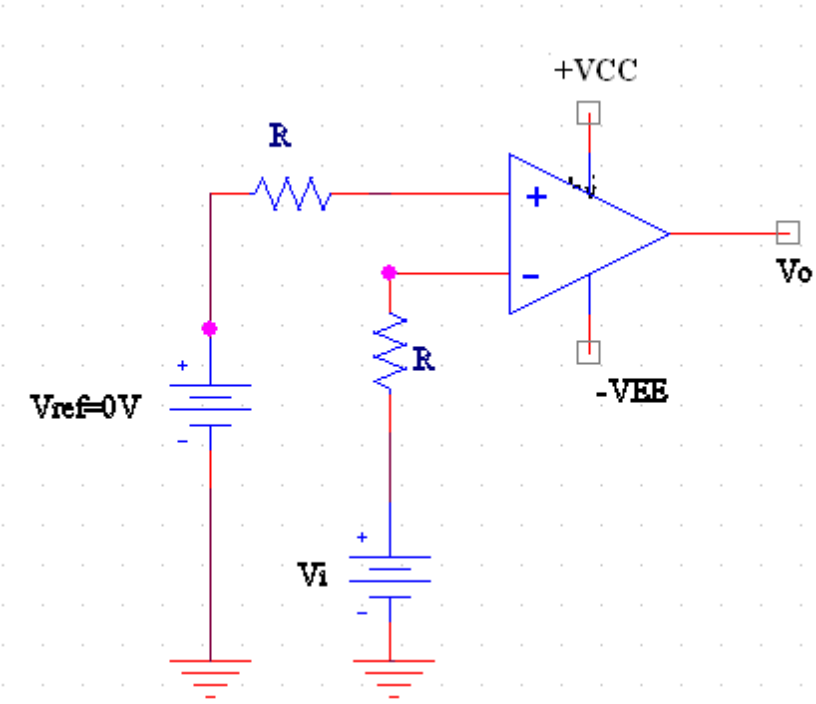


Figure 4.32 Zero Crossing Detector

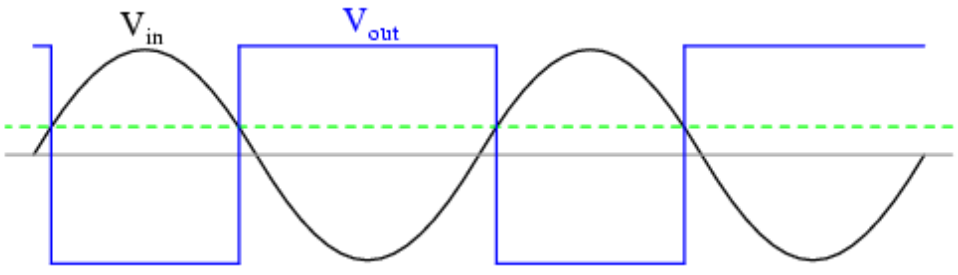


Figure 4.33 Input and Output Waveform

This circuit is also called a sine to square wave generator.

4.4.2 Voltage follower

A circuit in which the output voltage equals the input voltage is called voltage follower. The voltage follower circuit using op-amp is shown in figure 4.34

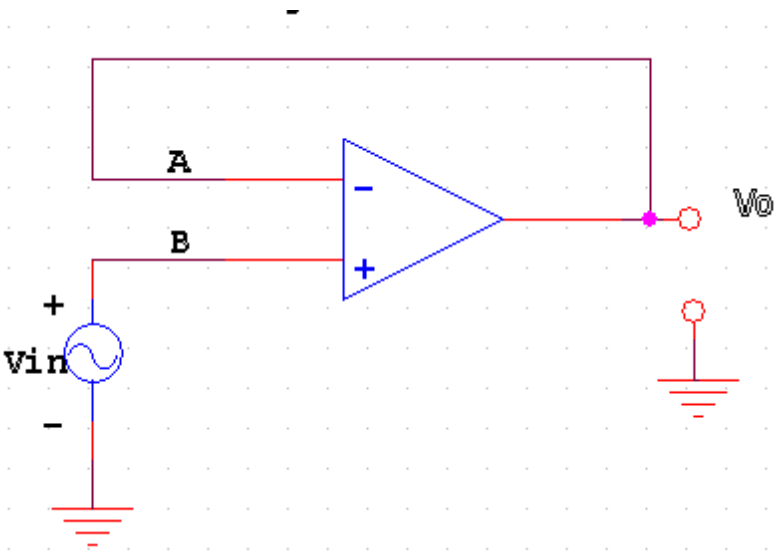


Figure 4.34 Voltage Follower

The node B is at potential Vin. Now Node A is also at the same potential as B i.e Vin.

$$\therefore \quad VA = VB = Vin \qquad \text{------(4.21)}$$

Now Node A is directly connected to the output, Hence we can write,

$$V0 = VA \qquad \text{------(4.22)}$$

Equating the equation (4.21) and (4.22),

$$V0 = Vin \qquad \text{------(4.23)}$$

For this circuit, the voltage gain is unity. Thus the output voltage V0 is equal to the input voltage Vin. If Vin increases(or decreases), Vo also increases(or decreases). Voltage follower is also called source follower,unity gain amplifier,buffer amplifier or isolation amplifier.

The main advantages of such voltage follower circuit are,

1. Very large input resistance, of the order of MΩ, Low output impedance, almost zero. Hence it can be used as a buffer to connect high impedance source to a low impedance load. Therefore the signal source is not loaded. Example: This property is most suitable in the instrumentation

amplifier, which is used to amplify the signal output from a transducer.

2. It has large bandwidth.
3. The output follows the input exactly without a phase shift since it works in noninverting mode.

If the output of a transducer is directly applied to the input of an amplifier which does not have high input impedance, then there is every possibility that the transducer may get loaded. Loading alters the characteristics of the transducer. To prevent this, the voltage follower circuit can be used in between the transducer and the amplifier. Such an use of the voltage follower is called voltage follower or an as isolator or a buffer.

4.4.3 Summer

As the input impedance of an op-amp is extremely large,more than one input signal can be applied to the op-amp. Such circuit gives the addition of the applied signals at the output. Hence it is called Summer or Adder circuit. Depending upon the sign of the output,the summer circuits are classified as inverting Summer and noninverting Summer**.**

Inverting Summer

The inverting summer configuration as shown in the figure 4.35,has three input voltages v1,v2 and v3 , three input resistors R1,R2,R3 and one feedback resistor Rf.. The output of the circuit is the sum of all input signals and the output signal that is proportional to the algebraic sum of the three input voltages V_1, V_2 and V_3. The output voltage Vout is obtained by using Kirchhoffs current law at node X.

$IF = I1 + I2 + I3 = -(V1/Rin + V2/Rin + V3/Rin)$ ------(4.24)

$Vout = -(RfV1/Rin + RfV2/Rin + RfV3/Rin)$ ------(4.25)

$Vout = -Rf(V1/Rin + V2/Rin + V3/Rin)$ ------(4.26)

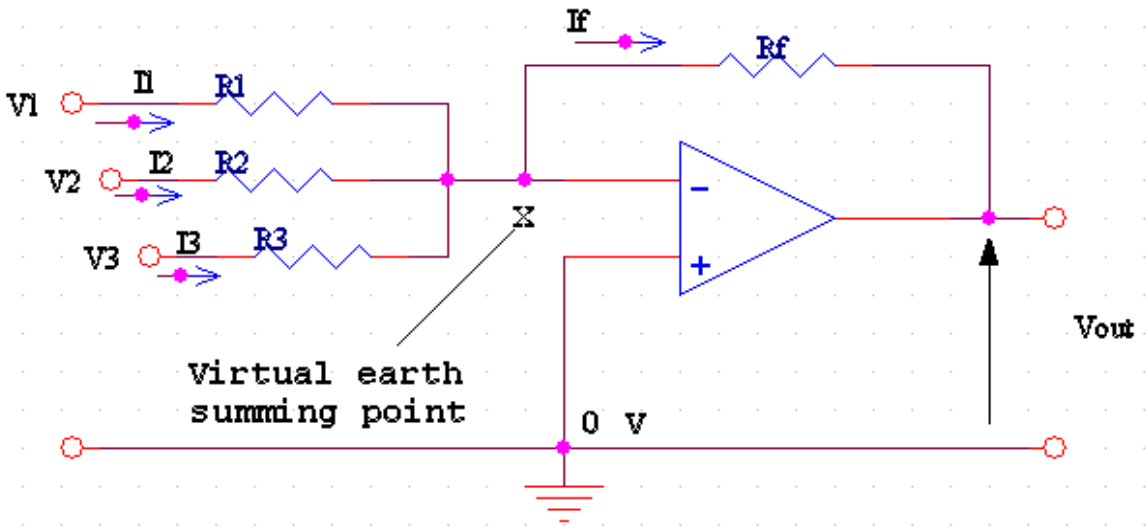


Figure 4.35 Inverter Summer

when all the resistances are of equal value(R1=R2=R3=Rf), Then the output voltage can be expressed as

$$V_{out}=-(V1+V2+V3)\ . \qquad \text{------(4.27)}$$

This type of summing amplifier is called as unity gain inverting adder. If the input resistors are of different values a scaling summing amplifier is produced which gives a weighted sum of the input signals.

Noninverting Summer

A summer that gives noninverted sum of the input signals is called noninverting summing amplifier. The circuit is shown in the figure 4.36

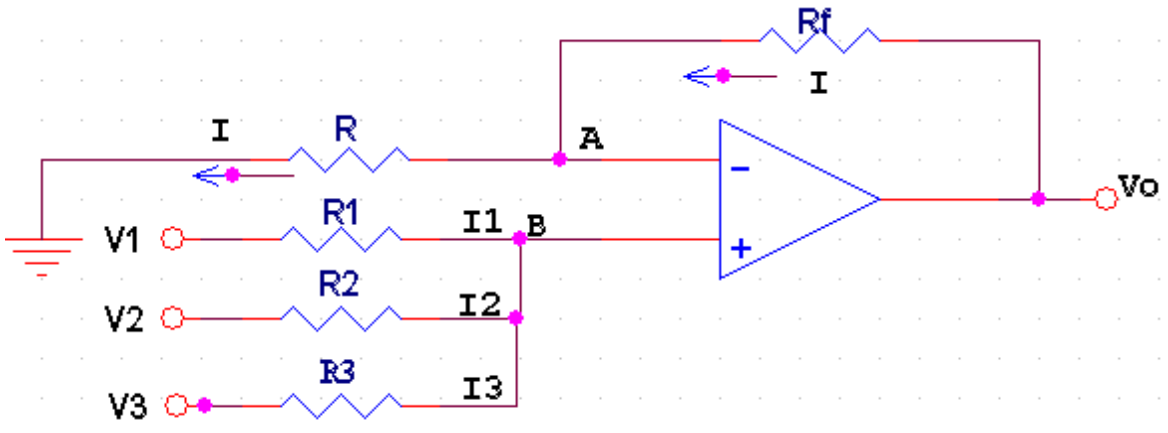


Figure 4.36 Noninverting Summer

Using the superposition theorem, the voltage Vb at the noninverting terminal is

(V1-Vb)/R1 * (R2 R3)/(R2+R3) +(V2-Vb)/R2 * (R1 R3)/(R1 + R3) + (V3-Vb)/R3 * (R1 R2)/(R1+R2) =0

If a noninverting summing amplifier has the input resistances R1=R2=R3=R ,

$$Vb= (V1+V2+V3)/3 \qquad \text{--------(4.28)}$$

The output voltage Vo is

$$Vo=(1+Rf/R)Vb \qquad \text{-------(4.29)}$$

$$Vo=(1+Rf/R1)(V1+V2+V3)/3 \qquad \text{---------(4.30)}$$

Then the output voltage $Vo=(1+Rf/R)(V1+V2+V3)/3$ --------(4.31)

Here the output voltage is equal to the average of all input voltages times the gain of the circuit(1+Rf/R).Hence this amplifier is called averaging amplifier.

If the gain (1+Rf/R1) is equal to the number of inputs, the output voltage becomes equal to the sum of all input voltages. That is, if (1+Rf/R1)=3 ,then the output voltage

$$Vo= Va+Vb+Vc \qquad \text{----------(4.32)}$$

Hence the circuit is called the noninverting summing amplifier.

4.4.4 Subtractor

A subtractor is basically a differential amplifier circuit in which one input is given to noninveritng terminal and another input is given to inverting terminal as shown in the figure 4.37 .The resultant output voltage will be proportional to the difference between the two input signals V1 and V2.

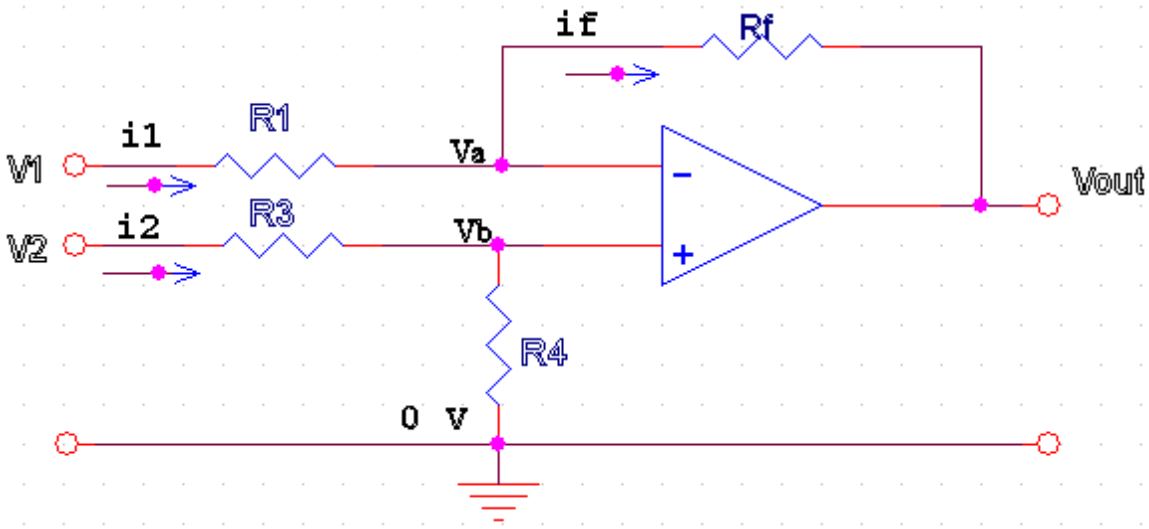


Figure 4.37 Basic Subtractor Amplifier

By applying superposition theorem,

Make V1=0(short circuit) , the circuit acts as a noninverting amplifier .Now

the voltage at node Vb is equal to Vb= V2 R4/(R3+R4).Therefore

Vo1=(1+Rf/R1)Vb ------(4.33)

=(1+Rf/R1) V2 R4/(R3+R4) ------(4.34)

Make V2=0(short circuit) , the circuit acts as inverting amplifier. So

Vo2= -Rf/R1(V1) ------(4.35)

Thus the output voltage vout due to both the inputs V1 and V2 can be expressed as

Vout=V01+V02=(1+Rf/R1) V2 R4/(R3+R4)-Rf/R1(V1) ------(4.36)

If all the resistors are all equal then the circuit will become a Unity Gain differential amplifier or subtractor and the gain of the amplifier will be 1 or Unity.

The output voltage is

$$Vout = V2 - V1 \qquad \text{---------(4.37)}$$

hence the circuit is called a subtractor.

4.4.5 Differentiator

The circuit which produces the differentiation of the input voltage at its output is called differentiator as shown in figure 4.38.

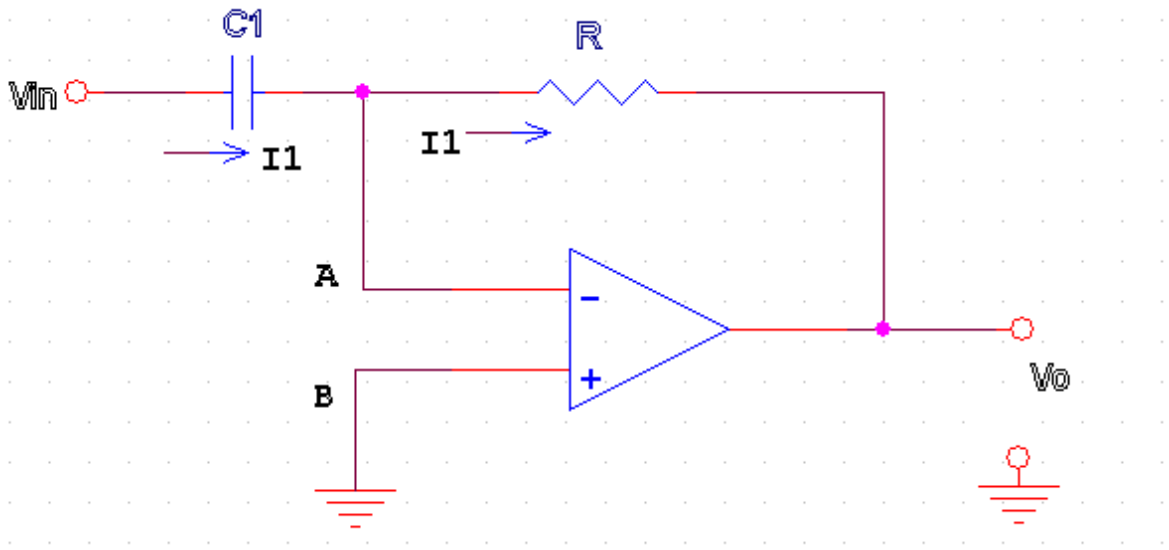


Figure 4.38 Differentiator Amplifier circuit

This circuit preforms the mathematical operation of differentiation, that is it produces a voltage output which is proportional to the rate-of-change input voltage and the current flowing through the capacitor C1. In other words the faster the change in the input voltage signal, the greater the input current, the faster change in the output voltage which seems to a spike in shape.

Since the node voltage of the operational amplifier at its inverting input terminal is zero(virtual ground potential i.e V_N =0.) , the current, i flowing through the capacitor is given as:

$$ic = C1\ d/dt(vin - vN) = C1\ dvin/dt \qquad \text{------(4.38)}$$

The current through the feedback resistor is v0/Rf and there is no current into the op-amp. Therefore, the nodal equation at Node N is,

$$C1\ dvin/dt + v0/Rf = 0 \qquad \text{---------(4.39)}$$

from which we have

$$v0 = -Rf\,C1\,dvin/dt \qquad \text{---------(4.40)}$$

Thus the output voltage v0 is a constant (-Rf C1) times the derivative of the input voltage vin . The minus sign indicates a 180° phase shift of the output vo with respect to the input signal.

Differentiator Waveforms

If we apply a continuously changing signal such as a Squarewave, Triangular or Sinewave signal to the input of a differentiator amplifier circuit the resultant output signal will be changed and whose final shape is dependant upon the RC time constant of the C1 and Rf combination. The waveforms are shown in figure 4.39

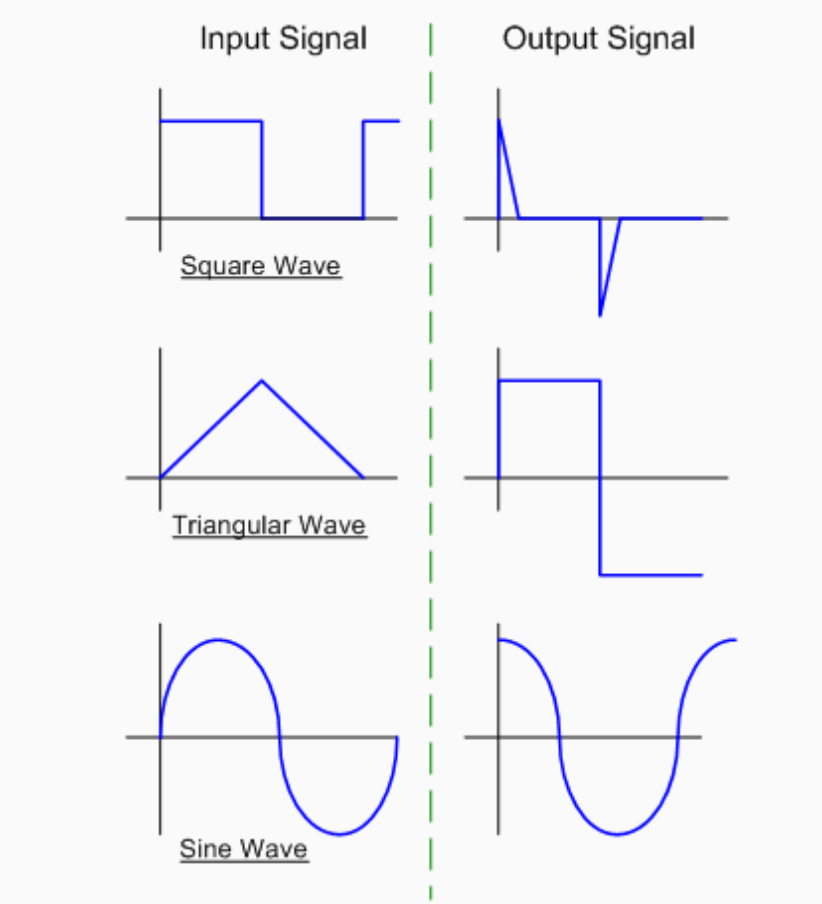


Figure 4.39 Differentiator Waveform

The differentiator amplifier circuit in its basic form has two main disadvantages. One is that it suffers from instability at high frequencies and the other is that the capacitive input makes it very susceptible to random noise signals and any noise or harmonics present in the circuit will be amplified more than the input signal itself.

At low frequencies the reactance of the capacitor is high resulting in a low gain (Rf/Xc) and low output voltage from the op-amp. At higher frequencies the reactance of the capacitor is much lower resulting in a higher gain and higher output voltage from the op-amp. However, at high frequencies a differentiator circuit becomes unstable and will start to oscillate. To avoid stability problem, the high frequency gain of the circuit needs to be reduced by adding an additional small value capacitor across the feedback resistor Rf and a series resistor, Rin to the input as shown in the figure4.40

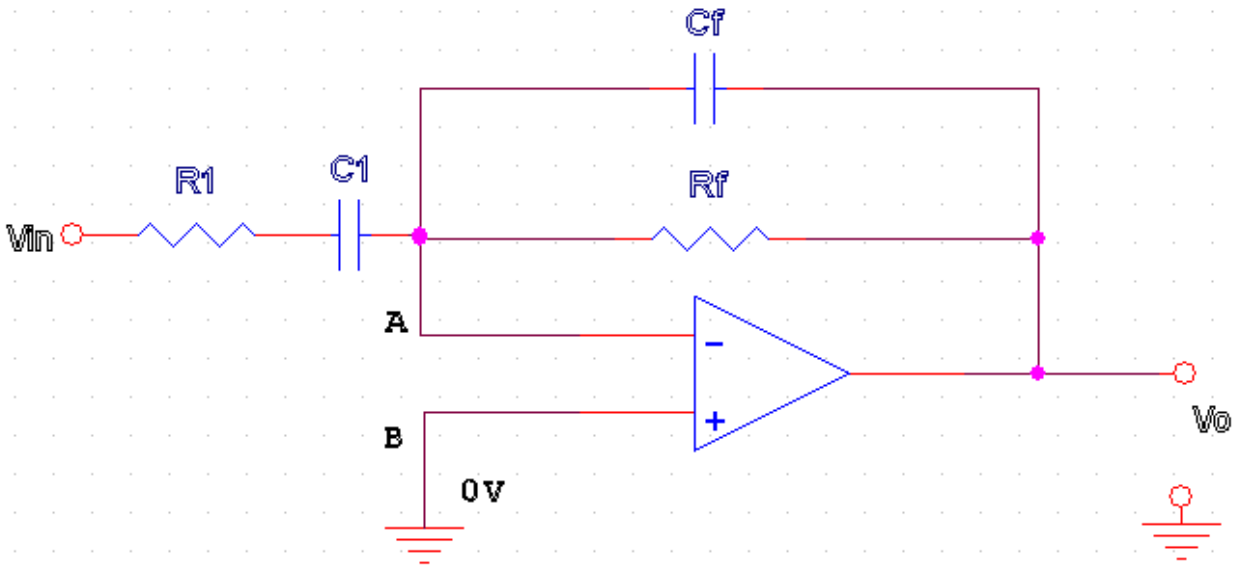


Figure 4.40 Improved Differentiator Circuit

4.4.6 Integrator

In an integrator circuit, the output voltage is the integration of the input voltage. The integrator circuit can be obtained without using active devices like op-amp, transistor etc. In such a case an integrator is called passive integrator. While an integrator using an active devices like op-amp is called active integrator. Here the position of the capacitor and resistor have been reversed and now the Resistor , R in is connected to the input terminal of the inverting amplifier while the capacitor, C forms the negative feedback element across the operational amplifier. We now have a resistor and

capacitor forming an RC Network across the operational amplifier as shown in figure 4.41

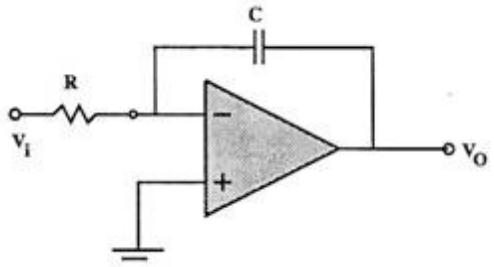


Figure 4.41 Integrator Amplifier Circuit

When an input signal, *vi*(*t*), is applied to the input terminal the device will generate at the output terminal the integral respect to time of the input waveform multiplied by a constant. In equation form

$V0 = -1/RC \int Vi(t)\, dt$ ------(4.41)

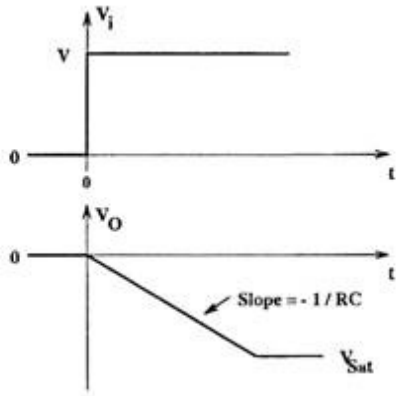


Figure 4.42 Integrator Input and Output Waveform

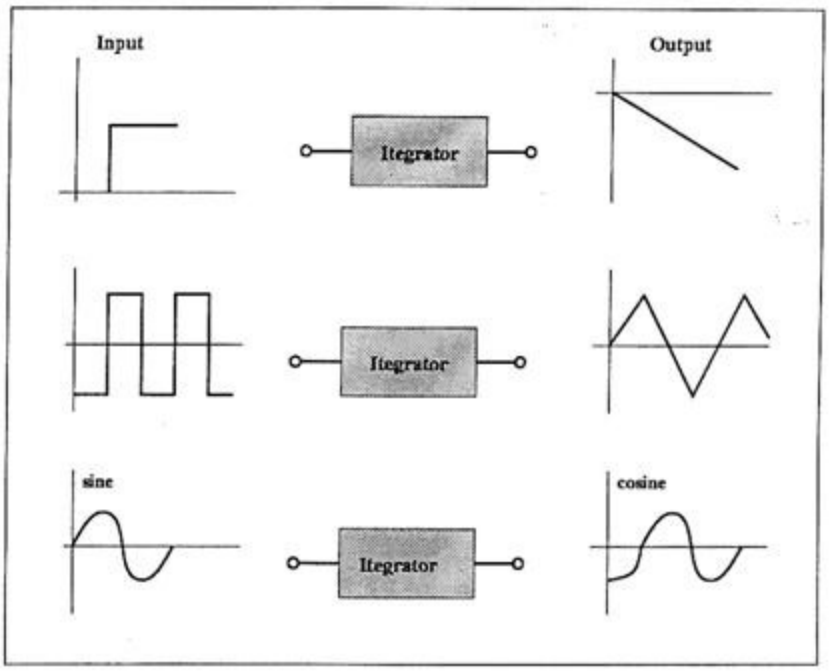


Figure 4.43 Basic Integrator Responses

The rate at which the output voltage increases is determined by the values of the resistor and the capacitor, "RC time constant". By changing this RC time constant, output voltage saturation can be changed .Figure 4.43 shows the output produced when several input functions are applied at the input terminal of an integrator. As you can see a constant voltage applied to the input of an integrator generates a voltage with a constant negative slope (a ramp), a square wave produces a triangular wave, and a sine functions generates a negative cosine function.

4.4.7 Active Filters

A filter is a frequency selective circuit that passes a specified band of frequencies and blocks signals of frequency outside this band. Filters may be classified in many ways.

1.Passive or Active filters(depending upon the components)

2.Analog or Digital filters(depending upon the signal)

3.Audio or Radio frequency filters(depending upon the operating frequency)

Passive filters are designed using passive components like resistors,capacitors and inductors. The passive components used in the filter determine its operating frequency. For example RC filters are used for audio or low frequency operation whereas LC or crystal filters are employed at RF or high frequencies. Crystal filters provides more stable operation at high frequencies because of the high Q

value for crystals. The main disadvantage of passive filters is that the amplitude of the output signal is less than that of the input signal, ie, the gain is never greater than 1. With filter circuits containing multiple stages, this loss in amplitude called attenuation can become quiet severe.

On the other hand active filters employ transistors or op-amp in addition to passive components. They have many advantages over passive filters such as easy design,good performance characteristics, very good accuracy with a steep roll-off,low noise , no loading effect and low cost. Active filters provide filtering and amplification to the input signal.

Operational amplifiers can be used to shape or alter the frequency response of the circuit by producing a more selective output response by making the output bandwidth of the filter more narrower or even wider.

First order Butter worth Low pass filter

The principle of operation and frequency response is exactly the same as that of a passive low pass filter.The only difference being it uses an op-amp for amplification and gain control. The simplest form of a low pass active filter is to connect an inverting or non-inverting amplifier to the basic RC low pass filter as shown in figure 4.45

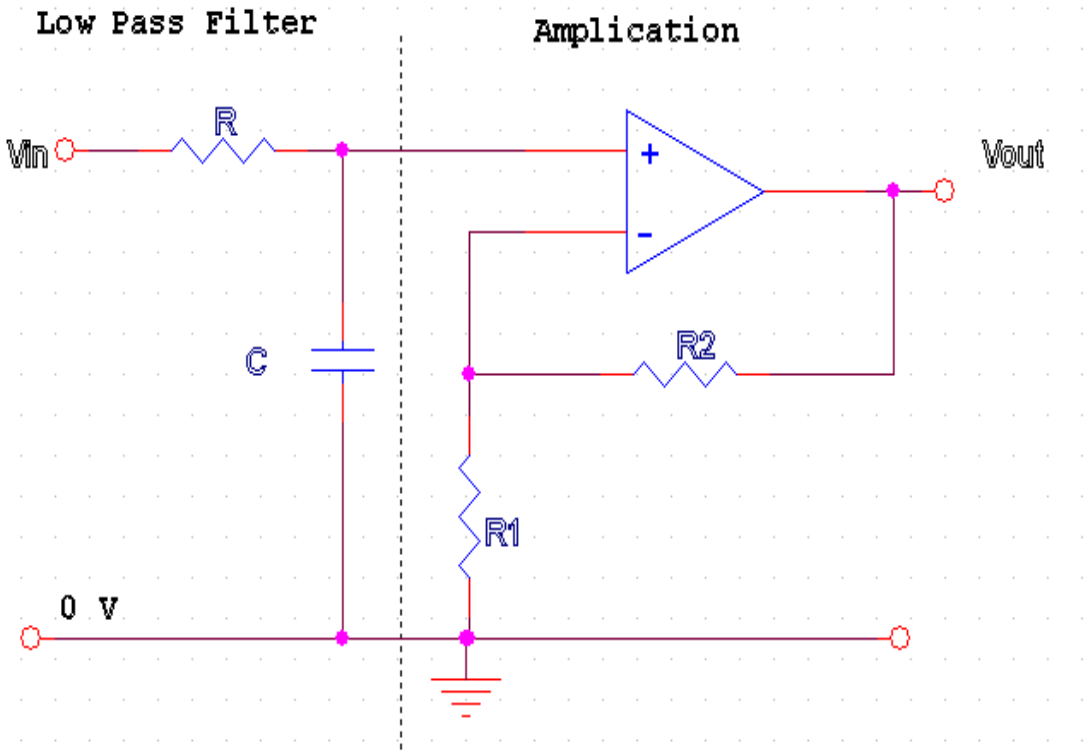


Figure 4.45 First order Butter worth Low Pass Filter

This 1st order low pass Butter worth filter, consists simply of a passive RC filter connected to the input of a non-inverting operational amplifier. The frequency response of the circuit will be the same as that of the passive RC filter, except that the amplitude of the output signal is increased by the voltage gain of the non-inverting amplifier and is given as: 1 + R2/R1.

The voltage gain of the filter is generally expressed in decibels and is a function of the feedback resistor (R2) divided by its corresponding input resistor (R1) value and is given as:

$$\text{Voltage Gain} = 20\log(V_{out}/V_{in}) = A_F/\sqrt{(1 + (f/fc)^2)} \qquad \text{----------(4.43)}$$

Where:

A_F = the Gain of the filter, (1 + R2/R1)

f= the Frequency of the Input Signal in Hertz, (Hz)

fc = the Cut-off Frequency in Hertz, (Hz)

phase shift $\varphi = -\tan^{-1}(f/fc)$

The frequency response of first order Low Pass Filter is shown in figure 4.46

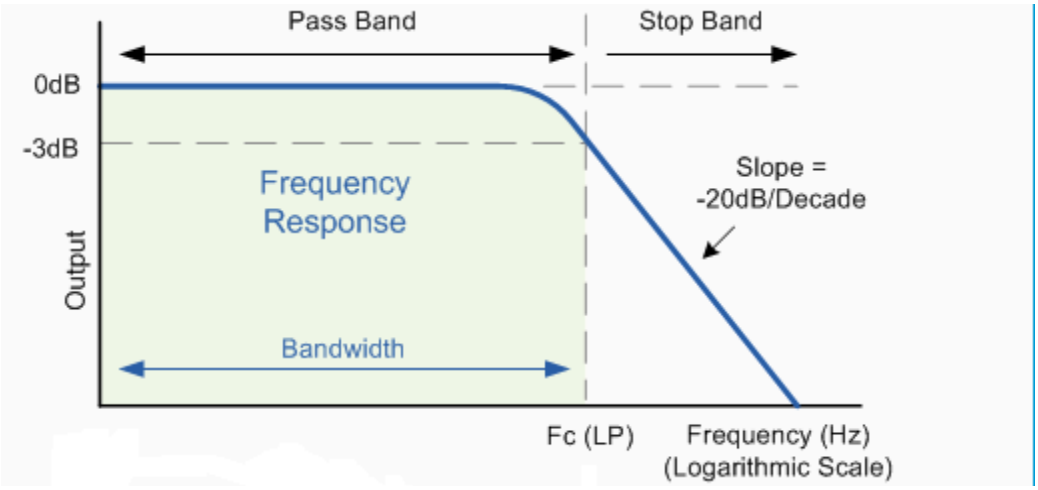


Figure 4.46 Frequency Response of First order High Pass Filter

The low pass filter has a constant gain A_F from 0 HZ to the cutoff frequency fc. At fc the gain is $0.707A_F$ and after fc it decreases at a constant rate with an increase in frequency. When the frequency is increased tenfold (one decade), the voltage gain is divided by 10. In other words, the gain decreases 20db (=20 log10) each time the frequency is increased by 10. Hence the rate at which the gain rolls off after fc is 20db/decade or 6db/octave, where octave signifies a twofold increase in frequency. The

frequency f=fc is called the cutoff frequency because the gain of the filter at this frequency is down by 3db(=20log .707) from 0Hz.

First order Butter worth High Pass Filter

The basic operation of an active high pass filter (HPF) is exactly the same as that for its equivalent RC passive filter circuit, except that this type of circuit has an op-amp for amplification and gain control. Like the active low pass filter circuit, the simplest form of an active high pass filter is to connect a standard inverting or non-inverting operational amplifier to the basic RC high pass passive filter circuit.

The maximum pass band frequency response of an active High Pass Filter is limited to the characteristics of the op-amp being used unlike passive high pass filters which have an "infinite" frequency response.

A first order high pass active filter as its name implies, attenuates low frequencies and passes high frequency signals. It consists simply of a passive filter section followed by a non-inverting operational amplifier. The frequency response of the circuit is the same as that of the passive filter, except that the amplitude of the signal is increased by the gain of the amplifier and for a non-inverting amplifier the value of the passband voltage gain is given as 1 + R2/R1.

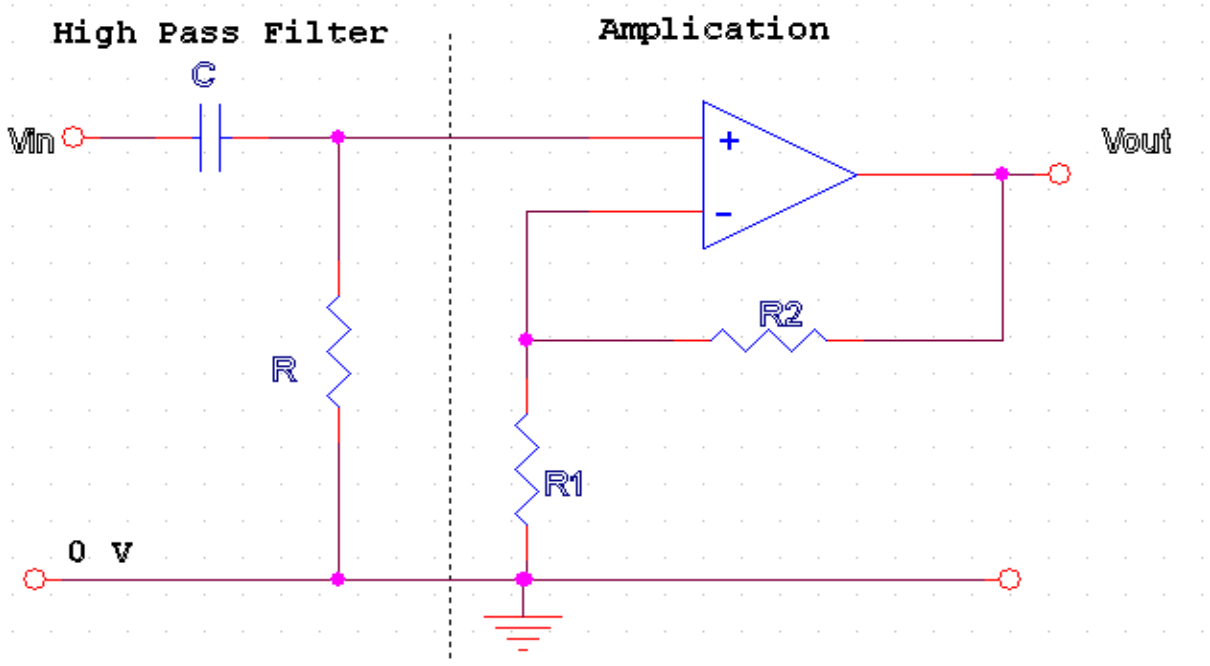


Figure 4.47 First Order High Pass Filter

The voltage gain of the HPF is given as:

$$\text{Voltage Gain} = 20\log(\text{Vout/Vin}) = A_F(f/fc)/\sqrt{(1 + (f/fc)^2} \qquad \text{----------(4.45)}$$

Where:

A_F - the Passband Gain of the filter, (1 + R2/R1)

f - the Frequency of the Input Signal in Hertz, (Hz)

fc - the Cut-off Frequency in Hertz, (Hz)

For a 1st order filter the frequency response curve of the filter increases by 20dB/Decade or 6dB/octave upto the determined cut-off frequency point which is always at -3dB below the maximum gain value. The cut-off or corner frequency (fc) can be found by using the equation

$$fc = 1/2\pi RC \quad \text{Hz} \qquad \text{———(4.46)}$$

The phase angle of the output signal leads the input signal as same as the passive RC filter and is given as:

$$\text{phase shift } \varphi = \tan^{-1}(1/2\pi RC) \qquad \text{----------(4.47)}$$

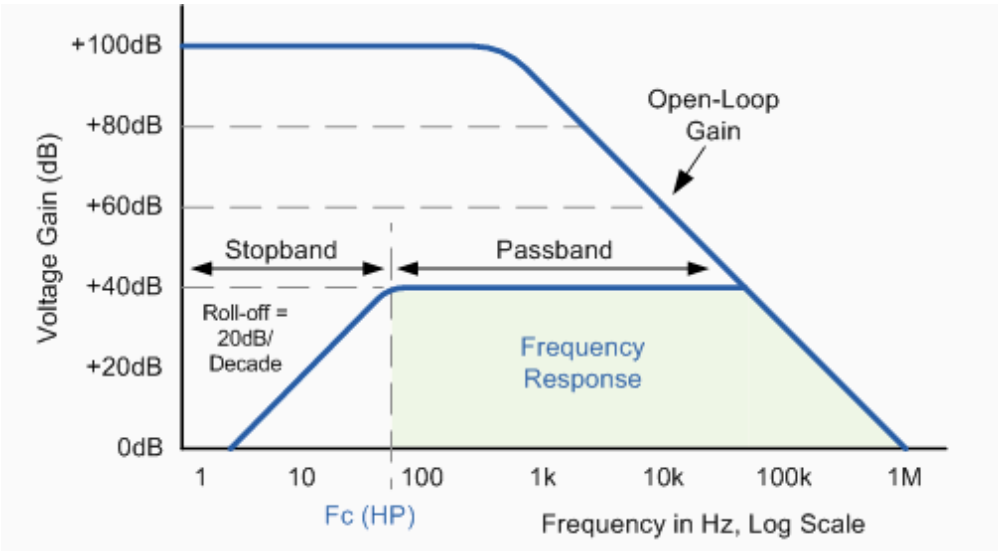


Figure 4.48 Frequency Response of First order High Pass Filter

Applications of high pass active filters are in audio amplifiers and speaker systems to direct the high frequency signals to the smaller tweeter speakers or to reduce any low frequency noise or “rumble" type distortion. When used like this in audio applications the high pass active filter is sometimes called a "Treble Boost" filter.

4.5 Analog to Digital Converter

An analog-to-digital converter (abbreviated **ADC**, **A/D** or **A to D**) is an electronic integrated circuit, which converts continuous signals to discrete digital numbers. The reverse operation is performed by a digital-to-analog converter (**DAC**).

Notice that the ADC input can be a voltage or current, depending upon the particular ADC. Also, some ADCs have a reference voltage and others have a reference current. The ADC inputs can be either voltage or current irrespective of reference input.

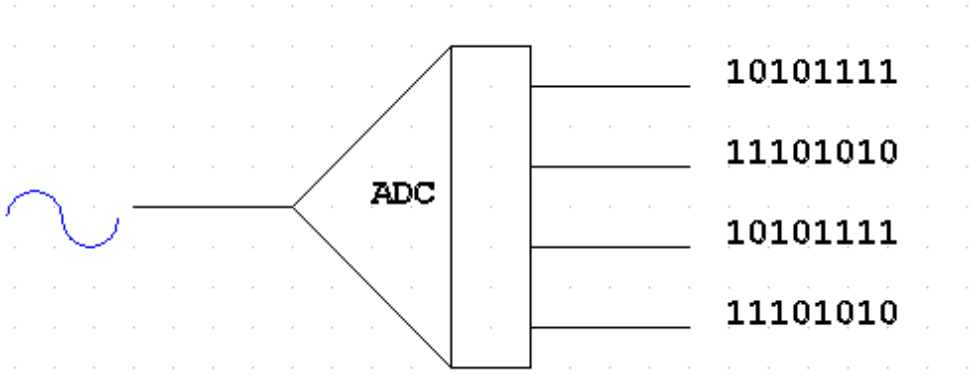


Figure 4.49 Analog to Digital Converters

4.5.1Types of Analog-to-Digital Converters

There are several types of A to D converters viz Successive approximation, Flash ,Tracking, Dual slope integrating, Charge balancing ,Sigma-delta converter .

Successive Approximation Converter

A successive approximation ADC provides a fast conversion of a momentary value of the input signal. It works by first comparing the input with a voltage which is half the input range.If the input is above this level it compares it with three-quarters of the range otherwise it compares with the one-fourth of the range ,and so on. Twelve such steps gives 12-bit resolution. While these comparisons are taking place the signal is frozen in a *sample and hold circuit*. After A-D conversion the resulting bytes are placed into either a pipeline or buffer store.

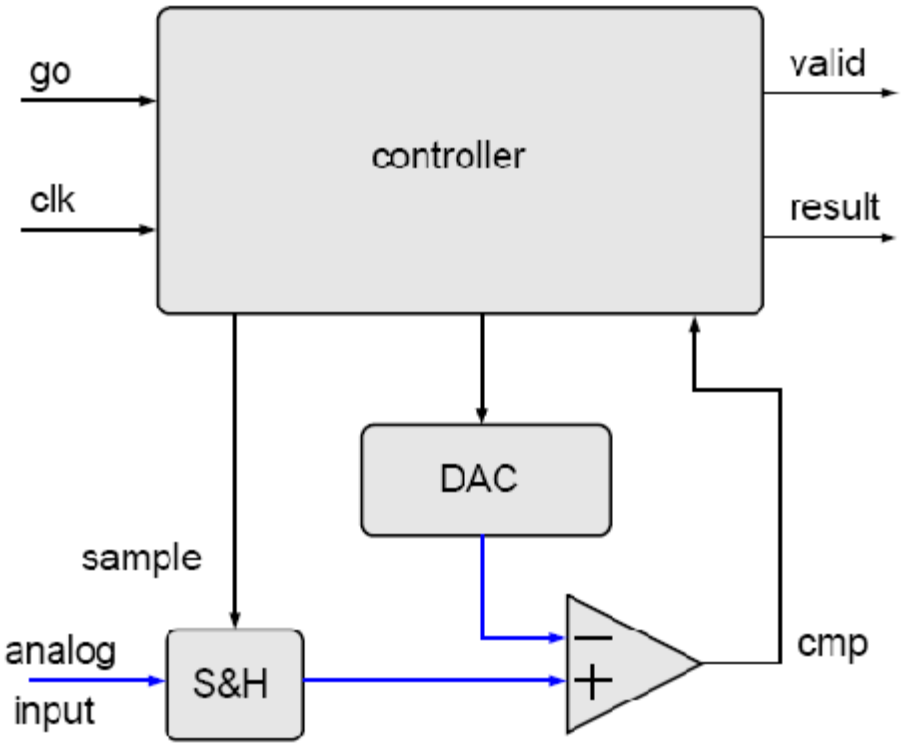


Figure 4.50 Analog to Digital Converters

A successive approximation ADC works by using a digital to analog converter (DAC) and a comparator to perform a binary search to find the input voltage. A sample and hold circuit (S&H) is used to sample the analog input voltage and hold (i.e. keep a non-changing copy) the sampled value while the binary search is performed. The binary search starts with the most significant bit (MSB) and works towards the least significant bit (LSB). For a 8-bit output resolution, 8 comparisons are needed in the binary search, taking a least 8 clock cycles. The sample and hold circuit samples the analog input on a rising edge of the control signal given by the controller. The comparator output (cmp) is a logic 1

if the sampled analog voltage is greater than the output of the DAC, 0 otherwise. In figure xxx “go” is the control signal. At any point if “go” is 0 the circuit is reset. When the “go” becomes 1 the process of sampling and conversion takes place. When a conversion is finished “valid” is set to 1 and the converted output is available at the “result” pin.

A pipeline store enables the A-D converter to do another conversion while the previous data is transferred to the computer. Buffered A-D converters place the data into a queue held in buffer memory. The computer can read the converted value immediately, or can allow values to accumulate in the buffer and read them when it is convenient. This frees the computer from having to deal with the samples in real time, allowing them to be processed in convenient batches without losing any data.

Flash Converter

A flash converter is the fastest of all A to D converters . Like the successive approximation converter it works by comparing the input signal to a reference voltage but a flash converter has as many comparators as there are steps in the comparison. For a 3-bit flash converter 2^3 -1, comparators are required. The analog voltage which need to be converted is applied simultaneously to a bank of voltage comparators which have scaled reference voltages (V/8,2V/8,3V/8...7V/8).The output of the comparator is high when the input voltage is less than the threshold voltage(reference voltage).Similarly the output of the comparator is low when the input voltage is higher than the threshold voltage. The comparator outputs connect to the inputs of a priority encoder circuit, which then produces a binary output as shown in figure 4.51. The priority encoder generates a binary number based on the highest-order active input, ignoring all other active inputs. Flash converter is also called the *parallel* A/D converter since the input voltage is compared with all the reference levels simultaneously.

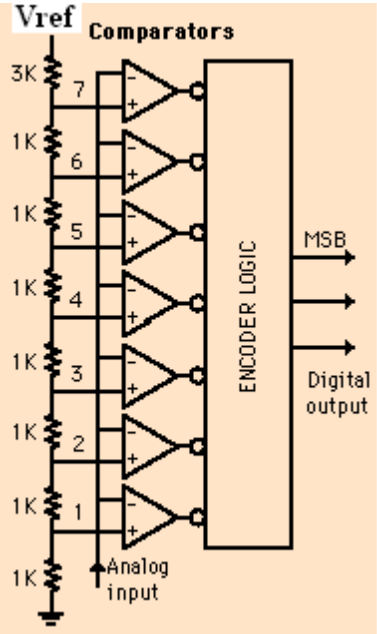


Figure 4.51 Flash A/D Converter

Encoder									
Input							Output		
B7	B6	B5	B4	B3	B2	B1	D2	D1	D0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	1
0	0	0	0	0	1	1	0	1	0
0	0	0	0	1	1	1	0	1	1
0	0	0	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	0	1
0	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	1	1

Table 4.1 Truth table of Flash Converter

The truth table for the output of the converter is shown in the table 4.1.

Tracking ADC

The tracking ADC as shown in figure 4.52 has an up-down counter that feeds a digital to analog converter (DAC). The input signal and the DAC output both go to a comparator. The comparator controls the counter. The circuit uses negative feedback from the comparator to adjust the counter until the DAC's output is close enough to the input signal. If the DAC voltage is less than the voltage at the analog input, the comparator produces a positive output which in turn puts the counter to

count in up direction thus the DAC output voltage is kept increasing until the voltage which make the comparator to give a low signal. Due to the low signal at the output of comparator the counter counts down only by one count . This cause the control to count up and the count increases by 1LSB. This process goes on being repeated and the digital output changes back and forth by +- 1 LSB around the correct value. This type of converters are often very good choices to read real-world signals.

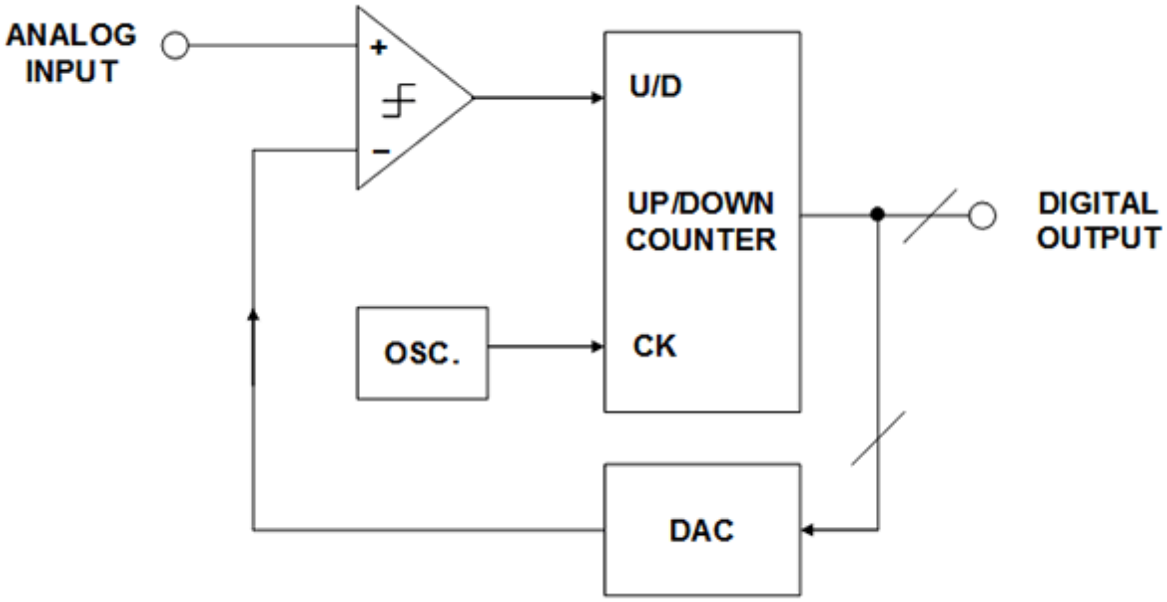


Figure 4.52 Tracking ADC

Pipeline ADC Architecture

The pipelined analog-to-digital converter (ADC) has become the most popular ADC architecture for sampling rates from a few megasamples per second (Msps) up to 100Msps. Resolutions range from eight bits at the faster sample rates and up to 16 bits at the lower rates. The architecture for pipelined analog-to-digital converter (ADC) is shown in the figure 4.53 and its operation is described below.

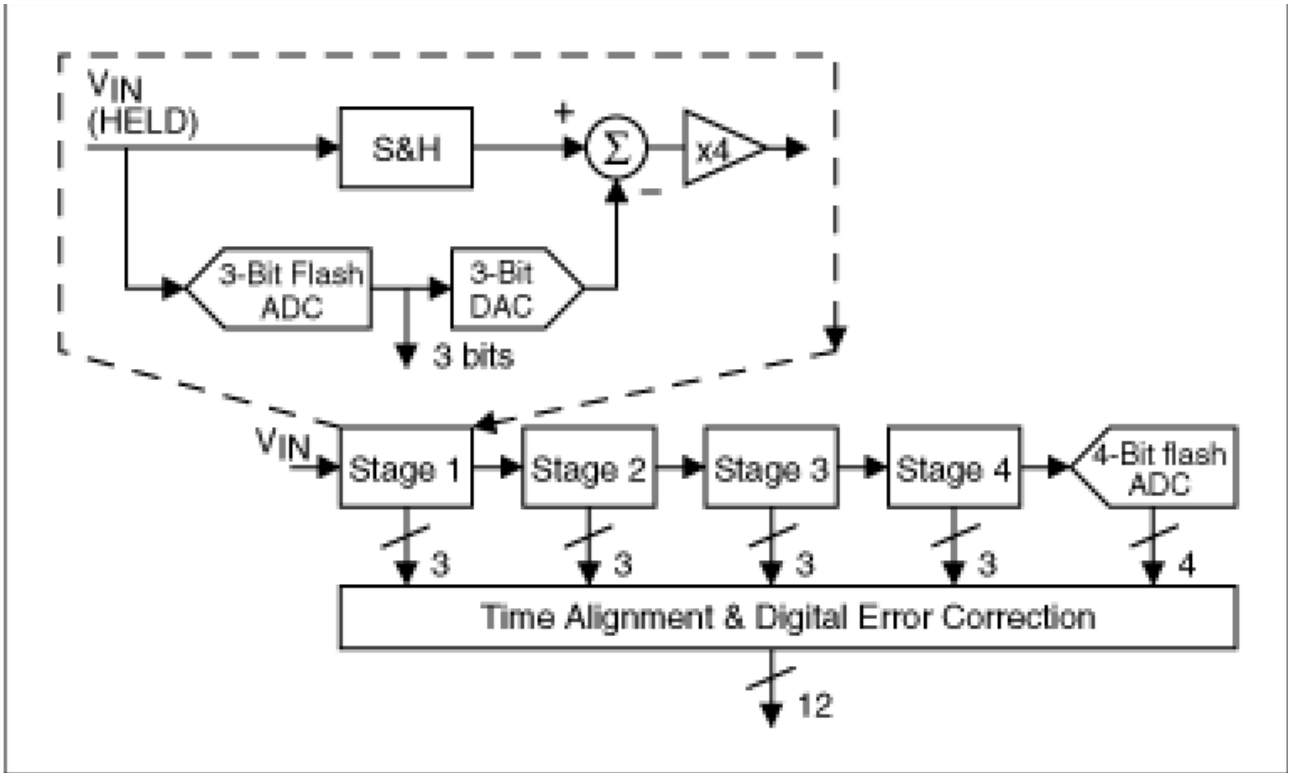


Figure 4.53 Pipelined ADC with four 3-bit stages (each stage resolves two bits).

The analog input VIN is first sampled and held steady by a sample-and-hold (S&H), while the flash ADC in stage 1 quantizes it to three bits. The 3-bit output is then fed to a 3-bit DAC , and the analog output is subtracted from the input. This "residue" is then gained up by a factor of four and fed to the next stage (Stage 2). This gained-up residue continues through the pipeline, providing three bits per stage until it reaches the 4-bit flash ADC, which resolves the last 4LSB bits. Because the bits from each stage are determined at different points in time, all the bits corresponding to the same sample are time-aligned with shift registers before being fed to the digital-error-correction logic. When a stage finishes processing a sample, determining the bits, and passing the residue to the next stage, it can then start processing the next sample received from the sample-and-hold embedded within each stage. This pipeline action is the reason for the high throughput.

Because each sample must propagate through the entire pipeline before all its associated bits are available for combining in the digital-error-correction logic, data latency is associated with pipelined ADCs.

Most modern pipelined ADCs employ a technique called "digital error correction" to greatly reduce the accuracy requirement of the flash ADCs. The 3-bit residue at the summation-node output has a dynamic range one-eighth that of the original Stage 1 input (VIN), yet the subsequent gain is only 4.

Therefore, the input to Stage 2 occupies only half the range of the 3-bit ADC in Stage 2 that is, when there is no error in the first 3-bit conversion in Stage 1.If one of the comparators in the first 3-bit flash ADC has a significant offset when an analog input close to the trip point of this comparator is applied, then an incorrect 3-bit code and thus an incorrect 3-bit DAC output would result, thus producing a different residue. As long as this gained-up residue does not over range the subsequent 3-bit ADC, it can be proven that the LSB code generated by the remaining pipeline will give the correct ADC output code. The implication is that none of the flash ADCs in Figure 4.53 has to be as accurate as the entire ADC. In fact, the 3-bit flash ADCs in Stages 1 through 4 require only about four bits of accuracy. The digital error correction will not correct for errors made in the final 4-bit flash conversion. Any error made at that conversion is suppressed by the large (44) cumulative gain preceding the 4-bit flash. Thus the final stage only needs to be more than 4-bits accurate. Although each stage generates three raw bits in the Figure 1 example, because the inter stage gain is only 4, each stage (Stages 1 to 4) effectively resolves only two bits. The extra bit is simply to reduce the size of the residue by one half, allowing extra range in the next 3-bit ADC for digital error correction, as mentioned above. This process is called "1-bit overlap" between adjacent stages. The effective number of bits of the entire ADC is therefore 2 + 2 + 2 + 2 + 4 = 12 bits.

4.5.2 Specifications of Analog-to-Digital Converters

Usually several specifications for A-D converters are quoted by hardware manufacturers. They are namely resolution, linearity, offset errors, sample and hold acquisition time, throughput, integration time and re-calibration discussed below.

Resolution

The resolution of the A-D converter is the number of steps the input range is divided into. The resolution is usually expressed as bits (n) and the number of steps is 2^n-1. A converter with 12-bit resolution, for instance, divides the range into 2^{12}, or 4096, steps. In this case a 0-10 V range will be resolved to 0.25 mV, and a 0-100 mV range will be resolved to 0.0025 mV. Although the resolution will be increased when the input range is narrowed. To increase the resolution , the input range is decreased below the noise level of the system results in unstable readings.

Linearity

Ideally an A-D converter with n-bit resolution will convert the input range into 2^n-1 equal steps (4095 steps in the case of a 12-bit converter). In practice the steps are not exactly equal, which leads to non-linearity in a plot of A-D output against input voltage. The linearity of a converter is an important measure of its accuracy and tells us how close the converter output is to its ideal transfer characteristics.

Sample and Hold Acquisition Time

A sample and hold circuit freezes the analogue input voltage at the moment the sample is required. This voltage is held constant whilst the A-D converter digitizes it. The acquisition time is the time between releasing the hold state and the output of the sample circuit settling to the new input voltage value.

Throughput

The throughput is the maximum rate at which the A-D converter can output data values. In general it will be the inverse of the sum of conversion time and acquisition time of the A-D converter. For example a converter that takes 3 microseconds to acquire and 7 microseconds to convert will be able to generate about 100000 samples per second. Throughput can be increased by using a pipelined A-D converter, in which second conversion can start while the first is still in progress. Throughput may be slowed down, by the factors which prevent data transfer at the full rate.

Re-Calibration

Some A-D converters are able to re-calibrate themselves periodically by measuring a reference voltage, and compensating for offset and gain drifts. This is useful for long term monitoring since drifts do not accumulate.

4.5.3 DC & AC Parameters

DC Parameters

1.Offset error

2.Gain error

3.Full Scale Range

4.LSB Size

5.DNL

6.INL

Offset error

The offset error is defined as the nonzero level of the output voltage when all inputs are zero. It adds a constant value to all output values. It is due to the presence of offset voltage and leakage currents in the converters.

Gain Error

Gain error is defined as the difference between the ideal full scale voltage and the actual full scale voltage with the offset error removed. This error can be represented in terms of voltage or percentage or LSB. Gain error is due to the errors in the feedback resistor.

$$V_{G\text{-}error} = (V_{FS\ actual} - V_{ZS\ actual}) - V_{FSR\_ideal} \quad \text{------(4.48)}$$

$$\text{Gain error } G_{error\ \%} = ((V_{FS\ actual} - V_{ZS\ actual}) / V_{FSR\_ideal} - 1) * 100 \quad \text{------(4.49)}$$

$$G_{error\_LSB} = ((V_{fs\ actual} - V_{ZS\ actual}) - V_{FSR\_ideal}) / V_{LSB} \quad \text{------(4.50)}$$

Full Scale (FS) Range

Full Scale Range of the ADC is the input range of voltages over which the ADC will digitize the input. In other words full scale range voltage is the difference between the Vfst and Vzst plus 2VLSB.

Vfst – Full scale transistion voltage.

Vzst – Zero scale transistion voltage.

LSB Size

For an ADC the LSB size equals the full-scale voltage range divided by 2^N, where N is the number of bits. For a 12-bit ADC with a unipolar full-scale voltage of 2.5V, 1LSB = $(2.5V/2^{12})$ = 610μV

DIFFERENTIAL NON-LINEARITY (DNL)

In an ideal converter, the code-to-code transition points are exactly 1 LSB apart. However successive inputs produce deviation in the step size. DNL is a measure of the maximum deviation from the ideal step size (1 LSB). *The difference between the ideal step size and the worst case actual input output code transitions is called Differential Non-Linearity*. DNL is commonly measured at the rated clock frequency with a ramp input. In an 8-bit ADC, for example, these changes are separated from each other by 1 LSB which is 1/256 of full scale voltage.

INTEGRAL NON-LINEARITY (INL)

It is a measure of the deviation of each individual code from a line drawn from zero scale or negative full scale (1⁄2 LSB below the first code transition)through positive full scale (1⁄2 LSB above the last code transition). The deviation of any given code from this straight line is measured from the center of that code value. INL is commonly measured at rated clock frequency with a ramp input. The end point test method is used.

AC Parameters

1.Signal to noise Ratio
2.Signal to Noise and Distortion Ratio(SINAD)
3.Total Harmonic Distortion
4.Inter-Modulated Distortion

SIGNAL TO NOISE RATIO (SNR)

It is ratio of the output signal amplitude to the output noise level, not including harmonics or dc and expressed in dB. SNR usually degrades as frequency increases because the accuracy of the comparator(s) within the ADC degrades at higher input slew rates. This loss of accuracy shows up as noise at the ADC output. In an A/D converter, noise comes from several sources:

(1) quantization noise, (2) noise generated by the converter itself, (3) application circuit noise and (4) jitter.

TOTAL HARMONIC DISTORTION (THD)

THD gives an indication of a circuit's linearity in terms of its effect on the harmonic content of a signal. THD is the ratio of the sum of the powers of all harmonics to the power of the fundamental component.

THD = 20 log((Vf2 +Vf3 +.....)/Vf1) ------(4.51)

where Vf1 is the fundamental amplitude, Vf2 is the second harmonic amplitude, etc.

As a practical matter, there is no such thing as a completely linear input to output transfer function. This non-linearity leads to output distortion. As the input signal swing increases in amplitude, the output becomes more and more distorted. The result is an increase in distortion as the input amplitude increases. THD performance degrades with increasing frequency because the effects of jitter get worse and because the input circuitry becomes slew limited.

THD can be expressed as a percentage or in dB.

SIGNAL TO NOISE PLUS DISTORTION (S/(N+D) or SINAD)

It is defined as the ratio between the power of sum of all harmonics and sum of all noise present in the spectrum but excluding dc to the power of the fundamental signal. RMS value of the output signal to the RMS value of all of the other spectral components below half the clock frequency, including harmonics but excluding dc.

SINAD = 20 log((Vf2 +Vf3 +...)+(Vfnoise))/Vf1) ------(4.52)

INTERMODULATION DISTORTION (IMD)

It is the creation of additional spectral components as a result of two sinusoidal frequencies being applied to the ADC input at the same time. It is defined as the ratio of the power in the intermodulation products to the power in one of the original input frequencies. IMD is usually expressed in dB.

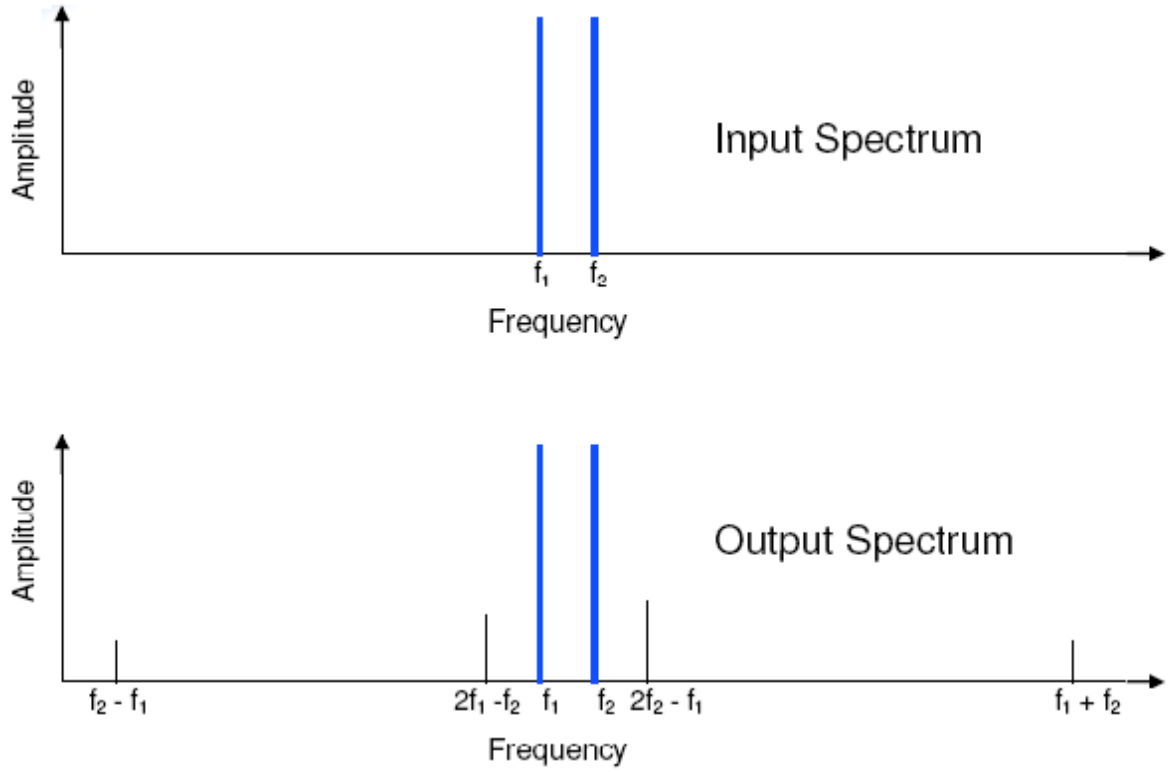


Figure 4.54 Intermodulation Distortion(IMD)

Any complex signal contains components at several frequencies simultaneously. Non-linearity in the converter's transfer function will not only cause distortion of a pure tone; it will also cause two or more signal frequencies to interact with each other to produce intermodulation products called Intermodulation Distortion (IMD). IMD can be expressed as the ratio of the power in the intermodulation products to the power in one of the original input frequencies.

Ideally, input frequencies f1 and f2 should produce output frequencies of only f1 and f2. Device non-linearity, however, cause the production of new frequencies at the sum and difference frequencies of the input signals and their harmonics. i.e. (f1 + f2), (f2 - f1), (f1 + 2f2), (2f1 + f2), (2f1 - f2), (2f2 - f1), etc. That is, the frequencies (intermodulation products) produced are Σ(mf1 + nf2)where "m" and "n" can take on any integer values. The amplitudes of the various intermodulation products will depend upon the nature of the non-linearity involved.

In RF applications the third-order difference products (2f1 - f2 and 2f2 - f1) are important because they are closest to the input frequencies, where other terms that are farther from the input frequencies can be digitally filtered out. For this reason the terms other than 3rd order terms are often ignored where IMD is specified for RF applications.

4.6 Digital to Analog Converter

A Digital-to-Analog Converter, (also written as DAC , D/A Converter and simply, D/A) is an electronic device or circuit used to convert a digital word into an analog voltage or current.

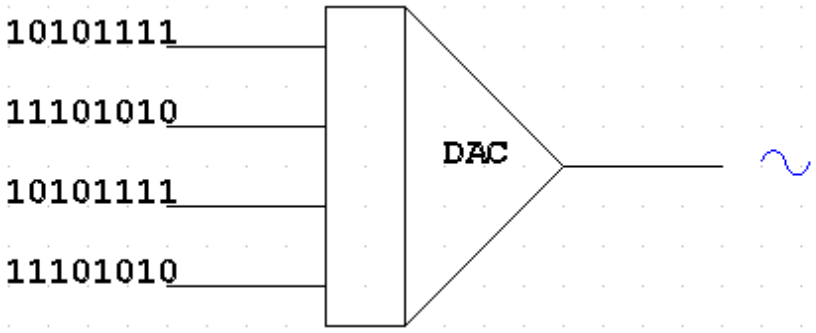


Figure 4.55 Digital to Analog Converter

1.The DAC output can be a **voltage** or a **current**, depending upon the particular DAC. Also, some DACs have a reference voltage and others have a reference current. Whether a DAC has a voltage or current output has no necessary relationship to whether it has a voltage or current reference input.

2.A DAC has an **analog reference** voltage or current which is multiplied by the digital input word. So, basically, the DAC is a **multiplier**.

A 3-bit D/A converter is shown in the figure 4.55.There are 8 possible input codes, each producing a different output voltage or current. The difference between each output voltage is, in the ideal case, the size of the LSB, which is G x VREF / 2^n, where "G" is the DAC gain factor and VREF is the DAC reference voltage.

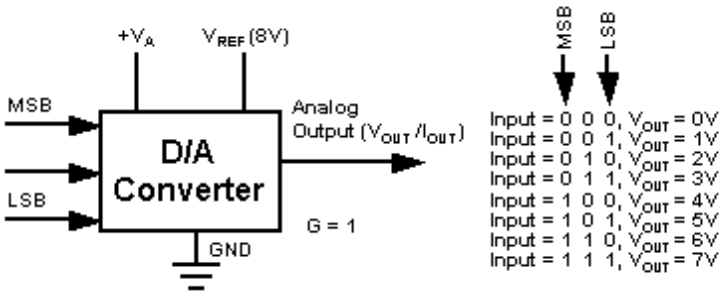


Figure 4.56 Digital to Analog Converter

If the output response has no errors, when the input code increases by one count, the output voltage will increase by one LSB. If the reference voltage is 8V and gain is 1 then the LSB represents 1 Volt, which is the smallest increment that this converter can resolve.

If the reference voltage reduces to 0.8V, the LSB size would then become 100mV, allowing the output to take on a smaller range of voltages (0 to 0.8V) with greater accuracy.

The resolution of a DAC is the number of input bits it has(3 bits, in this example) .If number of bits are more then the resolution becomes high. Resolution may also be defined as the size of the smallest output voltage or current step, or one LSB. Generally higher resolution means a smaller LSB size. When the reference voltage is reduced the noise present at the output becomes a larger part of the LSB .

4.6.1 Types of Digital-to- Analog Converters

DAC with Binary weighted Resistor

The binary weighted DAC using op-amp contains resistors for every bit of the DAC connected to a summing point as shown in figure 4.57. The op-amp is connected in inverting mode. Since this DAC has 4 binary inputs, there are 16 (ie. 2^4) possible analog outputs. When switch bo is closed , the voltage across the resistance R is 10V because V2 = V1 = 0V.Therefore the current through R is 10V/10K ohm =1mA. Hence the current through the feedback Resistor Rf =1mA which in turn produces the output voltage of -1V(-1K ohm * 1mA). Now the op-amp is working as a current to voltage converter. When the switch b1 is closed and b0 is opened, the resistance R/2 is now connected to the positive supply of +10V causing twice the current (2mA) flowing through the feedback resistance Rf which in turn produce the output voltage of -2V. Similarly when both the switches b0 and b1 are closed, the current through Rf will be 3mA which will be converted to an output voltage of -3V. When all the switches are closed, obviously the output will be maximum. The output voltage equation is expressed as

$$V0= - Rf(b0/R + b1/R/2 + b2/R/4 + b3/R/8) \quad \text{------(4.53)}$$

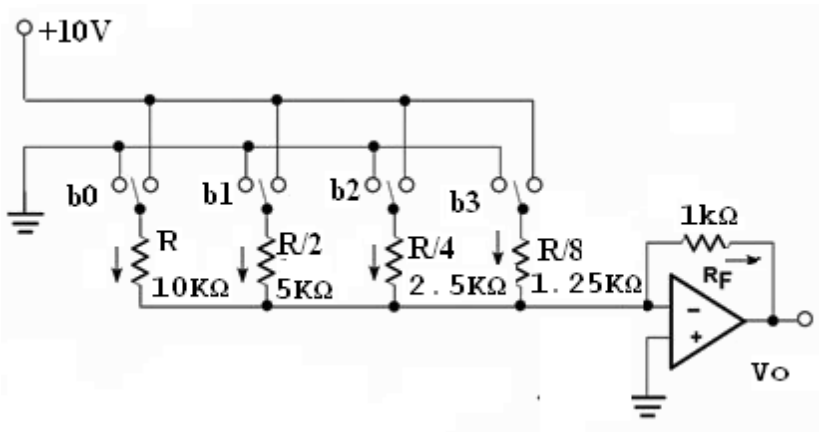


Figure 4.57 Binary weighted resistor DAC

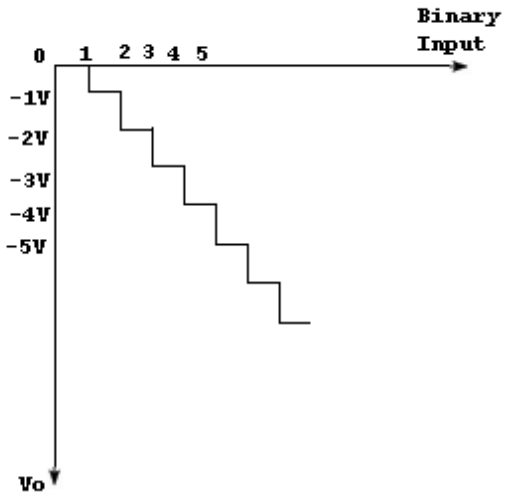


Figure 4.58 Output Waveform of Binary weighted resistor DAC

Figure 4.58 shows analog outputs versus possible combination of inputs. The output is negative going staircase waveform with 15steps of -1V each. The steps size may not all be the same size because of the variations in logic high voltage levels due to DNL. The step size may also depend on the value of feedback resistance, provided that the maximum output voltage does not exceed the saturation levels of an op-amp. The problem with this type of DAC is that it requires binary weighted resistance, which may not be readily available, especially the number of inputs is more than four. An alternative is to use R-2R resistors for the DAC since it requires only two sets of precision resistance values.

Ladder Type DAC

The DAC with R and 2R resistance as shown in Figure 4.59. The binary inputs are fed through switches b0 through b3, the analog output is proportional to the binary inputs. Assume that the

MSB switch b3 is connected to positive supply +10V and other switches are connected to ground. Applying thevenin's theorem the equivalent resistance Rth is

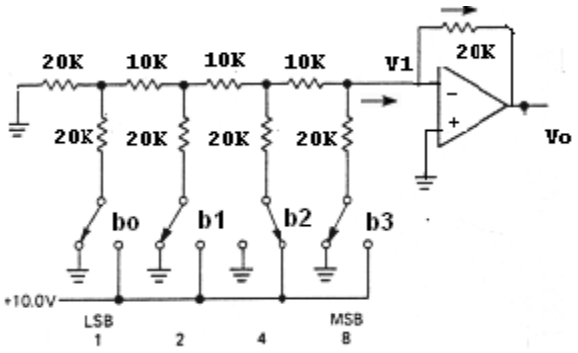


Figure 4.59 DAC with R-2R Resistors

Rth = [{[2R || 2R + R)|| 2R] + R} ||2R] + R

= 2R

From the figure 4.59 the inverting input is at virtual ground(V2=0). The current through 2R resistance connected to +5V is 5V/20K ohm = .25mA.

The same current flows through Rf and in turn produces the output voltage

V0=-(20k) * (0.25mA) = -5V

Using the same analysis, the output voltage corresponding to all possible combinations of binary inputs can be calculated.

Therefore the output equation can be generally expressed as

Vo= -Rf(b3/2R + b2/4R + b1/8R + b0/16R) ------(4.54)

where each of the inputs b0 through b3 may be either high (10V)or low(0V).

4.6.2 DAC Transfer & Output Characteristics

DAC Transfer Characteristic

The basic DAC transfer function is a straight line with output values between a lower (negative) and upper (positive) reference.

1) All DACs have a Positive Reference and Negative Reference.Negative reference is usually at ground potential (0.0V).

2) The analog output range for a DAC with a voltage reference is

$$G*V_{RN} \leq Output \leq G*V_{RP} \quad \text{------(4.55)}$$

Where VRP is the positive or top reference voltage, VRN is the negative or bottom reference voltage and "G" is the DAC gain factor. Also, the total reference is VREF = VRP - VRN.

Assuming the gain is 1, then the analog output range equation reduces to

$$V_{RN} \leq Output \leq V_{RP} \quad \text{------(4.56)}$$

3) Further assuming that VRN is 0V, as in the example of the figure above, these two relationships reduce to

$$0V \leq Output \leq G*V_{RP} \quad \text{------(4.57)}$$

and

$$0V \leq Output \leq G*V_{RP} \quad \text{------(4.58)}$$

So, with a gain of G = 1, a VRN of 0V and an input code of 0000 0000, the output of an ideal 8-bit DAC is 0V. With an input code of 1111 1111, the output voltage = G x (2^8 - 1) x VREF / 2^8 = G x 255 x VREF / 256, which is the full-scale output value. With an input code of 1000 0000, the output voltage = G x 2^7 x VREF / 2^8 = G x 128 x VREF / 256 = G x VREF / 2, which is the half-scale output value.

4.6.3 DC & AC Parameters

DAC are tested for DC and AC parameters. Some times DC parameters is also known as Intrinsic Parameter.

DC Parameters

1. Offset.
2. Gain error
3. Full Scale Range
4. LSB Size
5. DNL
6. INL

AC Parameters

1. Signal to Noise Ratio
2. Signal to Noise and Distortion ratio
3. Third harmonic distortion
4. Total Harmonic Distortion
5. Inter-modulation distortion

Offset Error

When the DAC is supplied with zero code the expected out out is 0.0mv, but it produced non zero value. Thus the offset error can be defined as the difference between the actual and the expected DAC output value while a zero code is supplied to the DAC digital input.

$$V_{offset} = V_{(ZS\ actual)} - V_{(ZS\ ideal)} \qquad \text{------(4.59)}$$

Full scale range

The full scale range of the DAC is defined as the difference between the maximum output voltage and the minimum output voltage that can be produced at the out put of the DAC.

$$V_{FSR} = V_{(FS\ actual)} - V_{(ZS\ actual)} \qquad \text{------(4.60)}$$

Gain Error

Gain Error is measured with all digital input at 1 so the output of the DAC is at full scale reading. Gain error is defined as the difference between the Full scale voltage and the actual full scale voltage (with the offset error removed). The error can be represented in terms of voltage or percentage or LSB

$$V_{G\_error} = (V_{(FS\ actual)} - V_{(ZS\ actual)}) - V_{(FSR\_ideal)} \qquad \text{------(4.61)}$$

$$G_{error\%} = \left(\left(\frac{V_{(FS\,actual)} - V_{(ZS\,actual)}}{V_{(FSR\_Ideal)}} \right) - 1 \right) \times 100 \quad \text{------(4.62)}$$

$$G_{error\_LSB} = \left(\frac{(V_{(FS\,actual)} - V_{(ZS\,actual)}) - V_{(FSR\_ideal)}}{V_{LSB}} \right) \quad \text{------(4.63)}$$

LSB size

The LSB size is computed by the ratio between the full scale range and the total no of code transitions. It means the expected amount of voltage change at the output of the DAC due to the change in the Input digital code by 1 bit.

$$V_{LSB} = \frac{FSR}{2^n - 1} \text{----(4.64)}$$

Differential non-linearity

The DNL represents the error in each step size and it is expressed in fraction of LSB.

$$DNL(i) = \frac{V_{(i+1)actual} - V_{(i)actual}}{V_{LSB}} - 1 \quad \text{----------(4.65)}$$

DNL is calculated by taking the difference between the successive samples and dividing it by LSB , then subtract the result with 1

Integral non-linearity

The amount of deviation of the measured transfer function of an digital-to-analog converter from the ideal transfer function (defined as a straight line drawn from zero to full scale). INL can be calculated by subtracting actual voltage of the DAC for a given code and the Ideal DAC voltage for that code

$$INL(i) = \frac{V_{(i)actual} - V_{(i)\,Ideal}}{V_{LSB}} \quad \text{-----------(4.66)}$$

CHAPTER 5

BASICS OF SIGNALS AND COMMUNICATION THEORY

Contents:

5.1 Introduction

5.2 Types of signals

5.3 Classification of signals

5.3.1 Continuous time and discrete time signals

5.3.2 Deterministic and random signal

5.3.3 Energy and power signal

5.3.4 Periodic and non-periodic signals

5.3.5 Odd and even signals

5.4 Frequency Analysis of signals

5.4.1 Analysis of Continuous Time(CT) signals

5.4.2 Sampling

5.4.3 Analysis of Discrete Time(DT) signals

5.4.4 Analysis of random signals

5.5 Logarithmic unit of Measurement - dB, dBm and dBc

5.6 Introduction to communication theory

5.6.1 Analog modulation techniques

5.6.2 Digital modulation techniques

5. 7 Summary

Feedback questions

Appendices

References

5.1 Introduction:

The analyses of signals play a vital role in the fields of communications, image processing, robotics and the like. A Signal can be defined as a function that conveys information,generally about the state or behavior of a physical system. It is a representation of variation of a physical quantity. Signals are represented mathematically as functions of one or more independent variables. The independent variable may be time, spatial coordinates and the like. If a signal amplitude or value depends on a single independent variable then it is said to be *one dimensional*.

Example: speech signal x(t) ,whose amplitude varies with time as shown in figure 5.1 .

When the signal value depends on more than one independent variable then it is said to be *multi dimensional*.

Example: image signal f(x,y) is a two dimensional whose brightness is a function of spatial coordinates. .

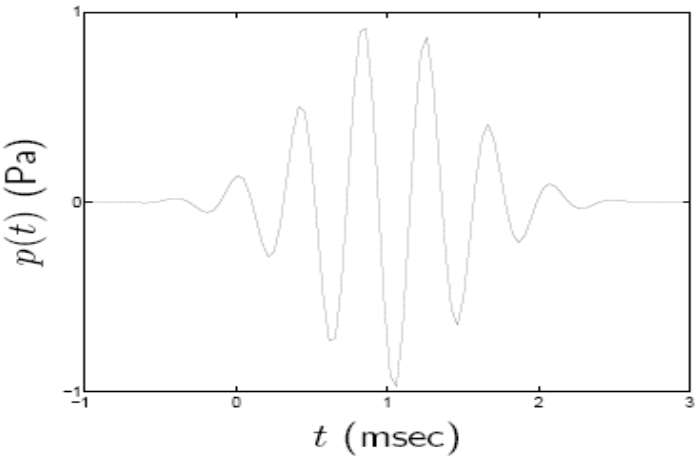


Figure 5.1 graphical representation of a signal

5.2 Types of Signals:

There are many types of signals such as sinusoidal,exponential,ramp,staircase,step,impulse and the like. Some basic types are given below,

Unit step function: The continuous-time version of unit step function is defined in the equation (5.1)

u(t)=0 for t<0

=1 for t>=0. (5.1)

It is illustrated in figure 5.2

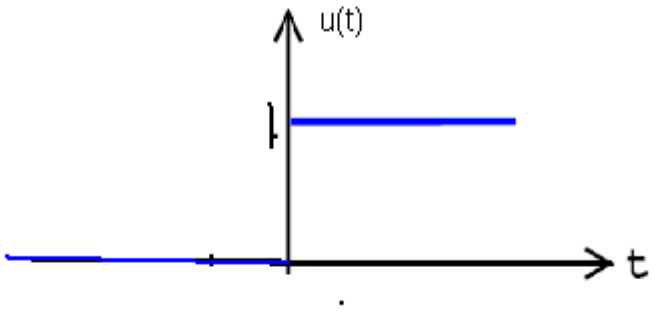


Figure5.2Unit step function

Step function: The continuous-time version of step function is defined in the equation (5.2)

x(t) =0 for t<0

=A for t>=0 (5.2)

It is illustrated in figure 5.3

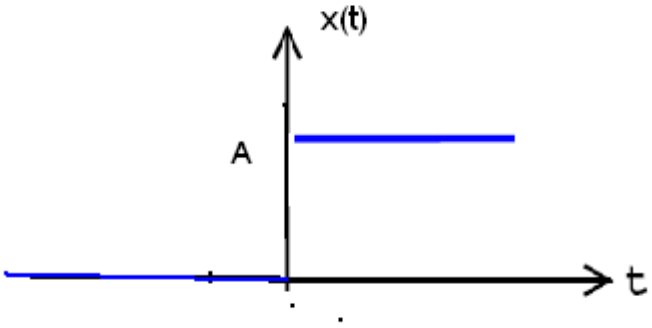


Figure 5.3 Step function

Unit ramp function: The continuous-time version of unit ramp function is defined in the equation (5.3)

r(t)=0 for t<0

=1 for t>=0 **(**5.3)

It is illustrated in figure 5.4

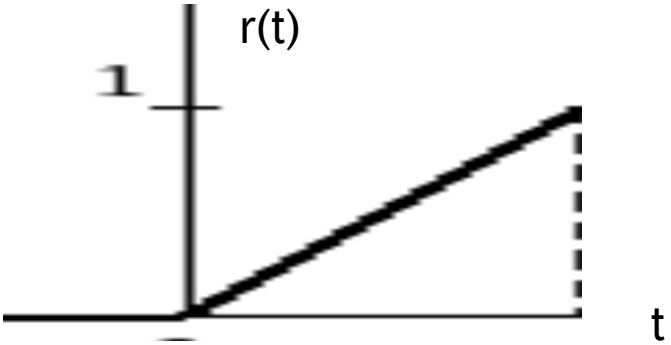


Figure 5.4 Unit ramp function

Ramp function: The continuous-time version of ramp function is defined in equation (5.4)

x(t)=0 for t<0

=kt for t>=0 where k is the slope (5.4)

It is illustrated in figure 5.5

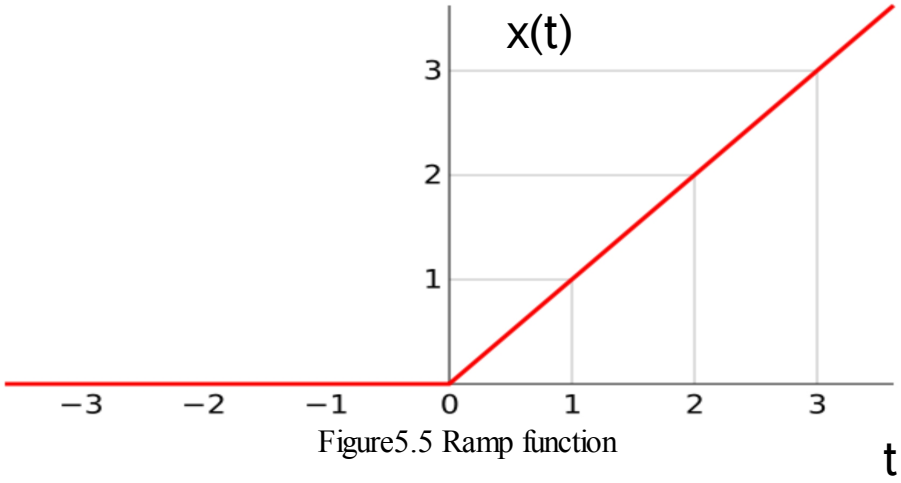


Figure5.5 Ramp function

Unit impulse function: The unit impulse also called Dirac delta function is a function representing an infinitely sharp peak bounding unit area i.e., total area under unit impulse is unity $\int \delta(t)dt=1$. A function δ(t) that has the value zero everywhere except at t = 0 where its value is infinitely large in such a way that its total integral is called unit impulse function Figure 5.6 shows the graphical view of the impulse function with unit strength.

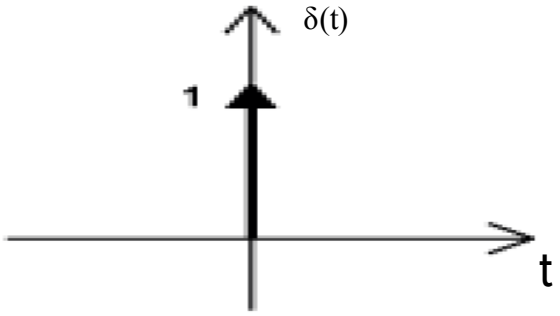


Figure 5.6 unit impulse function

Impulse function: The continuous-time version of impulse signal is defined in the equation (5.5)

,

$$x(t)=A\delta(t) \text{ at } t=0 \text{ where A is the strength value}$$

$$=0 \quad \text{at } t\neq 0 \ . \qquad (5.5)$$

which is illustrated in figure 5.7 for an impulse signal with strength A. The area under the pulse defines the strength of the impulse.

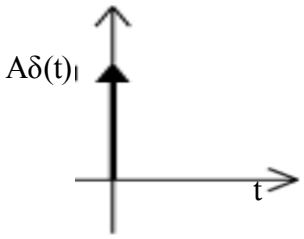


Figure5.7 Impulse function

Impulse function is an hypothetical waveform which can not be generated physically as the amplitude

is infinite at t=0 and zero elsewhere .However,the impulse function helps to provide an approximation to a physical signal of high amplitude with short duration. It can be used as a test signal to determine the characteristic of a given system.

Relationship between unit ramp r(t),unit step u(t) and unit impulse function δ(t) : The relationship between unit step and unit ramp function are expressed as follows,

$$d/dt(r(t))=u(t) \quad (5.6)$$

$$\int u(t)dt=r(t) \quad (5.7)$$

The unit step and unit impulse functions can be related as

$$d/dt(u(t))=\delta(t) \quad (5.8)$$

$$\int \delta(t)dt=u(t) \quad (5.9)$$

then the relationship between unit impulse and unit ramp functions is expressed by

$$\iint \delta(t)dt=r(t) \quad (5.10)$$

Finally the overall relationship between all the above three functions are summarized in equation (5.11) and (5.12)

δ(t)------∫-->u(t)----∫----->r(t) (5.11)

r(t)-----d/dt--->u(t)---d/dt-->δ(t) (5.12)

5.3 Classification of signals:

5.3.1 continuous and discrete time signals:

Signals can be categorized as continuous and discrete based on how they are defined as a function of time (t). A signal is said to be *continuous-time signal* whose amplitude varies as function of time .It has the value continuously for all time t.. It is denoted as x(t). Figure 5.8 shows a continuous -time signal in which the value x(t) defined for all time.

Example: ECG signals,speech signals,Television signals etc.

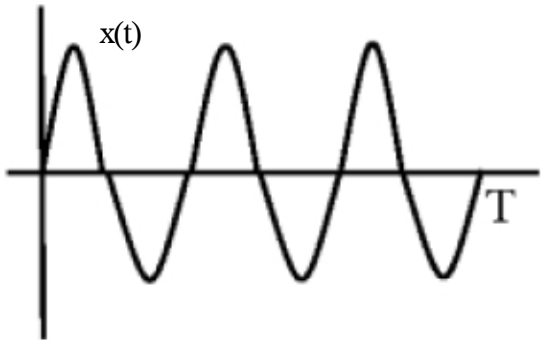


Figure:5.8 continuous signal

Discrete-time signals that have continuous amplitude but only exist at discrete times as shown in figure 5.9. The values or amplitudes are defined only on a discrete instants of time. These signals are commonly denoted as x[n] and represented as sequence of numbers.

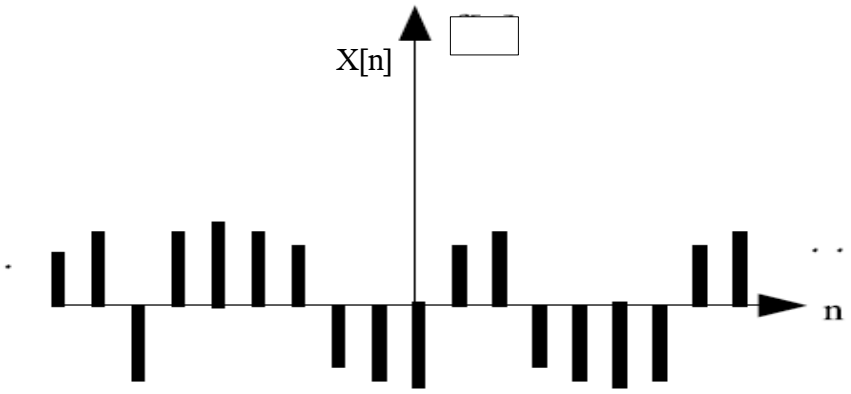


Figure 5.9 Discrete signal

Digital signal is a signal whose amplitude and time are discrete. It has only two amplitude levels as shown in figure:5.10. The values which are specified as one of two possibilities such as 1 or 0, HIGH or LOW, TRUE or FALSE and so on.

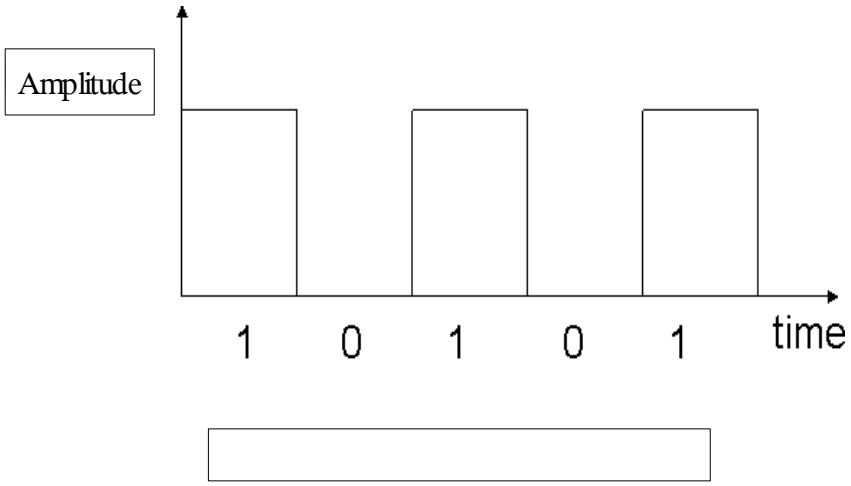


Figure 5.10 Digital signal

Based on the properties, signals can also be classified as follows

1.Deterministic and random signal

2.Energy and power signal

3.Periodic and aperiodic signals

4.Odd and even signals

5.3.2.Deterministic and random signal :

A *deterministic signal* is a signal in which each value of the signal is fixed and can be determined by a mathematical expression, rule, or table. Because of this the future values of the signal can be calculated from past values with complete confidence. There is no uncertainty with respect to its value at any time as shown in figure 5.11.

Example: Sinusoidal signal.

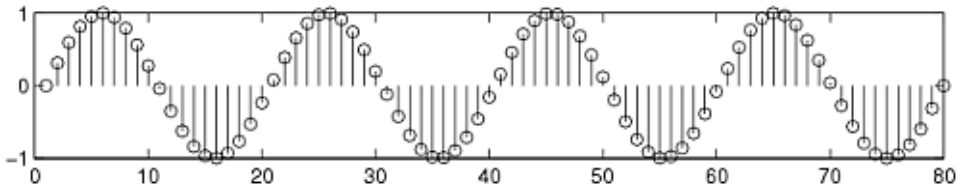


Figure 5.11 Deterministic signal

Random signal has a lot of uncertainty about its behavior. The future values of a random signal cannot be accurately predicted and can usually only be guessed based on the averages of sets of

signals. Random signal is shown in figure 5.12

Example:noise.

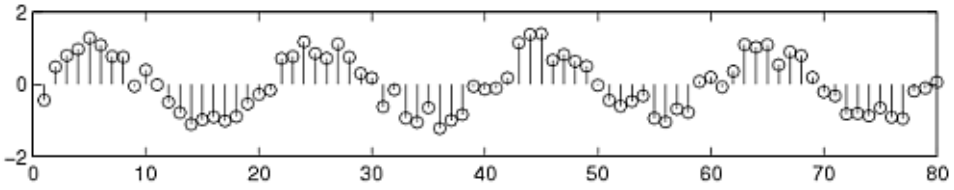


Figure 5.12 Random signal

5.3.3 **Energy and power signal:**

A signal x(t) is said to be an *energy signal* if and only if the total energy E is finite and non-zero(i.e., $0<E<\infty$). Example: Deterministic and Non-periodic signals .

Power of energy signal is zero over infinite time. Total energy is expressed as in equation (5.13)

$$E_f = \int_{-\infty}^{\infty} (|f(t)|)^2 dt$$

(5.13)

A signal x(t) is said to be *power signal* if and only if the average power P is finite and non-zero.(i.e.,$0<P<\infty$).Power is a time average of energy (ie.energy per unit time). Example :Periodic signal and random signals. Energy of power signal is infinite over infinite time. Average power is expressed as in the equation: (5.14)

$$P_f = \lim_{T\to\infty} \frac{1}{T} \int_{-\left(\frac{T}{2}\right)}^{\frac{T}{2}} (|f(t)|)^2 dt$$

(5.14)

5.3.4 Periodic and non-periodic signals:

Periodic Signals are signals that repeat themselves after a certain amount of time. More formally, a function x(t) is periodic if it satisfies the condition in equation (5.15)

$$\mathbf{x(t + T) = x(t)} \qquad (5.15)$$

for any instant time T for all t. The classic examples of a periodic function are shown in figure 5.13

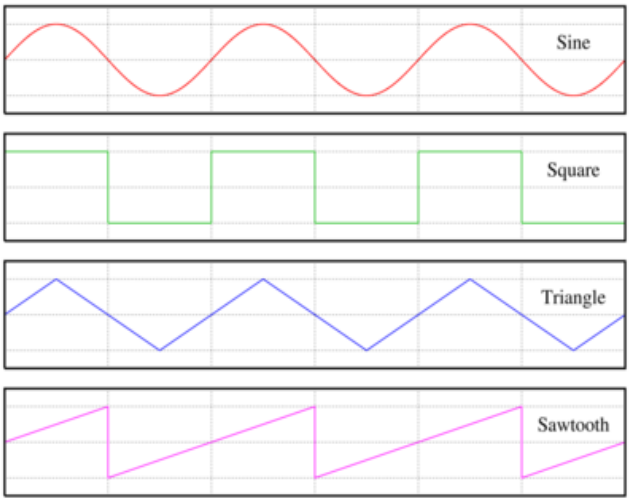


figure:5.13 periodic signal

A signal is said to be *non-periodic* if there is no value of T that satisfies the condition in equation (5.15). The signal that does not repeat itself as shown in figure 5.14 is a non-periodic signal.

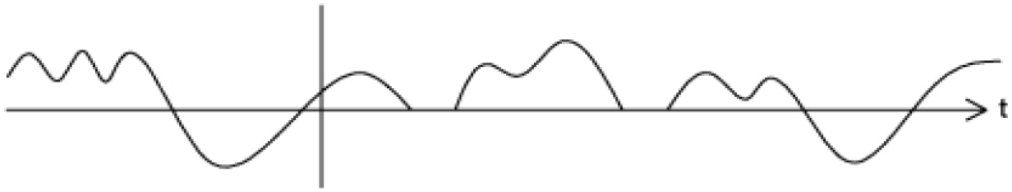


Figure 5.14 Non-periodic signal

5.3.5 Odd and Even signals:

A signal x(t) is said to be an *even signal* if it satisfies the condition in equation (5.16)

$$\mathbf{x(t) = x(-t)}. \qquad (5.16)$$

Even signals are symmetric around the vertical axis. The signal x(t) is said to be an *odd signal* if it satisfies the condition in equation (5.17)

$$\mathbf{x(t) = -x(-t)} \qquad (5.17)$$

Odd signals are antisymmetric around the vertical axis .The examples are shown in figure:5.15a and 5.15b respectively.

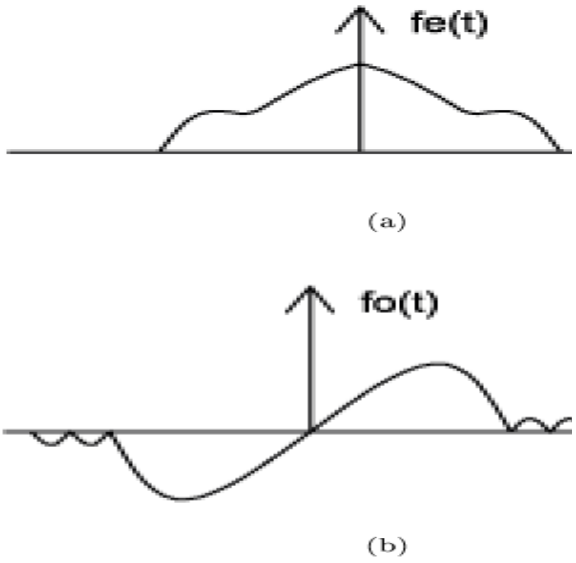


Figure 5.15 Even and Odd signal

5.4 Frequency analysis of signals

The time and frequency domains are alternative ways of representing signals as shown in figure:5.16 In order to come up with appropriate signal processing techniques, which enable to extract the desired signal from a distorted and noisy version of the transmitted signal, there must be ways to represent the signal for analysis. Frequency or Spectrum analysis of a signal involves the resolution of the signal into its frequency components. Signal can be analyzed with mathematical tools like Fourier series, Fourier Transform, Laplace Transform, Z-Transform etc., by decomposing the signals in terms of sinusoidal components or complex exponential. The process of obtaining the spectrum of a given signal using mathematical tools is known as *Frequency or spectrum analysis* .

Time domain Signals physically exist in the time domain and are usually expressed as a function of the time parameter. The signal has time as the independent variable. The time domain representation displays changes in signal amplitude with time. The time domain plot can be visualized with the help of an oscilloscope.

Frequency domain representation also called *spectrum* shows the relationship between amplitude and frequency . This can be displayed with the help of *spectrum analyzer.* Spectrum gives an information about where the different frequency components including *harmonics* lie within each given frequency band .It provides information on the phase shift that must be applied to each sinusoidal in order to recombine the frequency components to recover the original time signal.

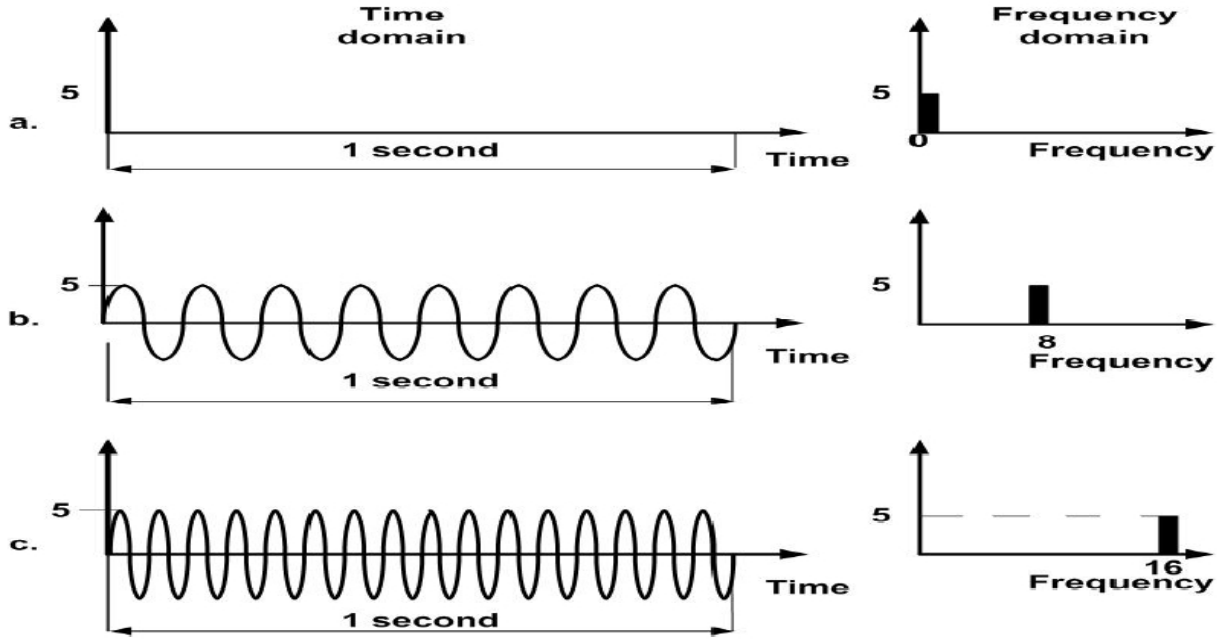


Figure 5.16 Time and frequency domain representation

5.4.1 Analysis of CT signal:

The analysis of CT signal can be performed using mathematical tools such as Fourier series,Fourier Transform,and Laplace transform. Fourier series is used to analyze the CT periodic signals and Fourier Transform is used to analyze the CT non-periodic & periodic signals..

Fourier series which are used to provide a representation of periodic signals. This has some application to circuit analysis for periodic signals in terms of infinite number of sinusoidal components(harmonics). Fourier series representation of CT periodic signal x(t) is given as,

$$x(t) = \sum_{k=-\infty}^{\infty} X[k]\, e^{jk\omega 0t} \tag{5.18}$$

$$X[k]=1/T \int_{(T)} x(t)\, e^{-jk\omega 0t}\, dt \tag{5.19}$$

$$x(t)\text{--------}FS{:}\omega 0\text{---------}>X[k] \tag{5.20}$$

Fourier transform :is an operation that transforms one function of a real variable into a function of a complex variable. The new function is the frequency domain representation of the original function in time domain. Fourier Transform maps a time series (example audio samples) into the series of frequencies (their amplitudes and phases) that composed the time series. To determine the original function in time domain x(t), Inverse Fourier Transform is taken for the function X(jω). The two functions are inverses of each other.

$$x(t)=1/2\prod \int_{-\infty}^{\infty} X(j\omega)\, e^{jk\omega t}\, d\omega \qquad (5.21)$$

$$X(j\omega)=\int_{-\infty}^{\infty} x(t)\, e^{-jk\omega t}\, dt \qquad (5.22)$$

Note:e^xi=cosx+isinx.In general FT is the complex valued function of ω.The frequency response X(jω) can be expressed either in rectangular form or in polar form as follows,

X(jω)=XR(jω)+jXI(jω) and

X(jω)=|X(jω)| angle X(jω)

Dirichlet's conditions:

The Fourier series representation for a periodic signals can be convergent if the signal satisfies some conditions so called Dirichlet's conditions. They are used as the test conditions on signals before representing the Fourier series .The conditions are stated as follows,

1.The function x(t) should be single valued within the interval T0.

2.The function x(t) should have finite number of discontinuities in the interval T0.

3.The function x(t) should have finite number of maxima and minima over any period *T0* .

4.The function x(t) should be absolutely integrable over any period as in equation (5.23).

$$\int_{To} |X(t)|\, dt < \infty \qquad (5.23)$$

5.4.2 Sampling:

Sampling is the process of translating continuous-time signals into discrete time signals of uniform time interval. A sample refers to a value at a point in time or space. Figure 5.17 shows the

sampled signal y(t) obtained as a product of original continuous time signal x(t) and an impulse train δ(t) .The representation is commonly termed as impulse(ideal) sampling .However practical sampling differs from ideal sampling in that the sampled wave consists of pulses having finite amplitude and duration rather than impulses.

Equation (5.24) shows the sampled signal y(t) contains the original signal x(t) and Impulse signal δ(t)

$$y(t)=\int_{n=-\infty}^{\infty} x(\tau)\ \delta(t-\tau)\, d\tau \qquad (5.24)$$

where τ is a sampling interval.

Therefore the effect of sampling is determined by relating the FT of y(t) as in equation (5.25)

$$Y(j\omega)=1/\tau \sum_{k=-\infty}^{\infty} X(j(\omega-k\,\omega s)) \qquad (5.25)$$

Quadrature/trignometric fourier series:

$$s(t) = \sum_{n=0}^{\infty} b_n \cos(n\omega_0 t) + \sum_{n=1}^{\infty} c_n \sin(n\omega_0 t)$$

(5.26)

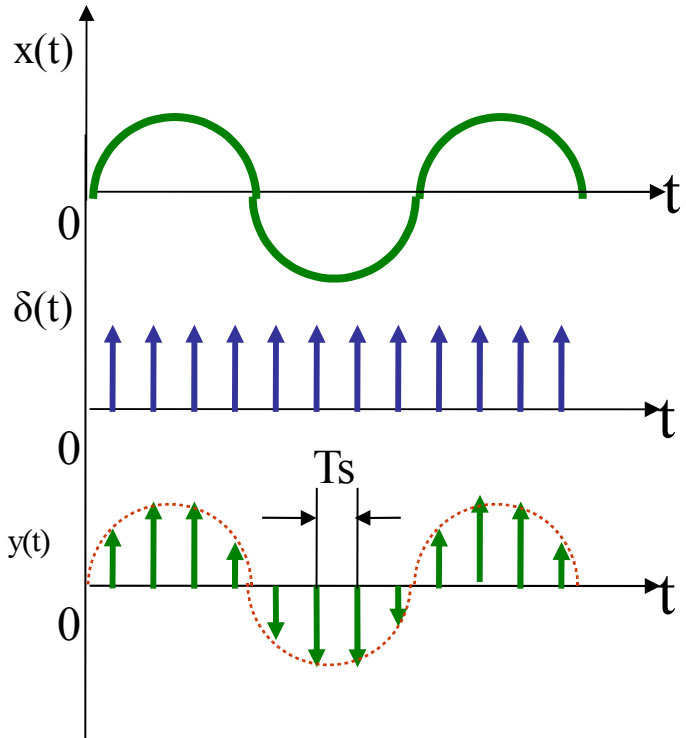
where,

$$b_0 = (1/T)\int_{-T/2}^{T/2} s(t)\,dt$$

$$b_n = (2/T)\int_{-T/2}^{T/2} s(t)\cos(n\omega_0 t)\,dt, \qquad n = 1, 2, \ldots,$$

$$c_n = (2/T)\int_{-T/2}^{T/2} s(t)\sin(n\omega_0 t)\,dt, \qquad n = 1, 2, \ldots,$$

$\omega_0 = 2\pi/T$



x(t)

Figure 5.17 Sampling process

Sampling Theorem states that when x(t) is a band limited signal which has a highest frequency component of fm,then the sampling frequency fs must be greater than or equal to 2 fm as in equation (5.27) in order to reconstruct the original signal .

$$fs >= 2fm \quad (5.27)$$

The minimum sampling frequency is known as the Nyquist sampling rate or Nyquist rate.

Aliasing: When the signal is sampled at a rate less than Nyquist rate it is called as under sampling. Under sampling causes frequency components that are higher than half of the sampling frequency to overlap with the lower frequency components as shown in figure 5.18**.** As a result, the higher frequency components also roll into the reconstructed signal which leads to distortion of the signal. This signal distortion is called aliasing.

Due to aliasing there is an overlap of shifted, periodic copies of original signal's Fourier transform components. Also, aliasing makes it impossible to correctly determine the strength at a particular frequency in the overlapped region.

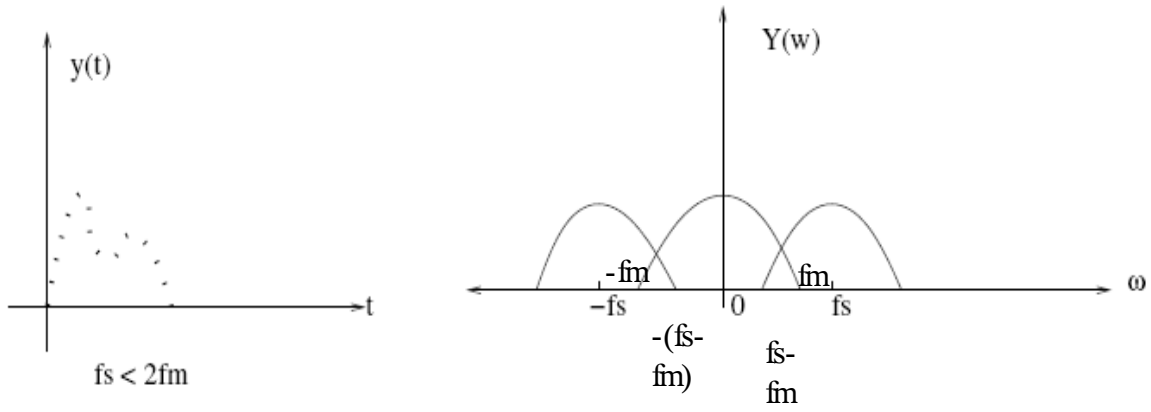


Figure 5.18 Aliasing

Signal reconstruction:

A continuous-time signal of frequency fm is sampled at a frequency of fs to produce a sampled

signal. Figure 5.19 shows that reconstruction of original signal from sampled signal If fs > 2fm there is no loss of information due to sampling. The goal of reconstruction is to apply some operation on y(jω) to convert it back to X(jω).

$$X(j\omega) = Y(j\omega)H(j\omega) \tag{5.28}$$

where H(jω) is the transfer function of reconstruction filter.It is defined in equation (5.29)

$$H(j\omega) = \begin{cases} \tau, & |\omega c| <= \omega s/2 \\ 0, & |\omega c| > \omega s/2 \end{cases} \tag{5.29}$$

where ωc is the cutoff frequency of reconstruction filter.

In practical reconstruction, low-pass filter with cutoff frequency ωc is used to obtain the reconstructed signal which is in the range of ωm <ωc <ωs - ωm .

.

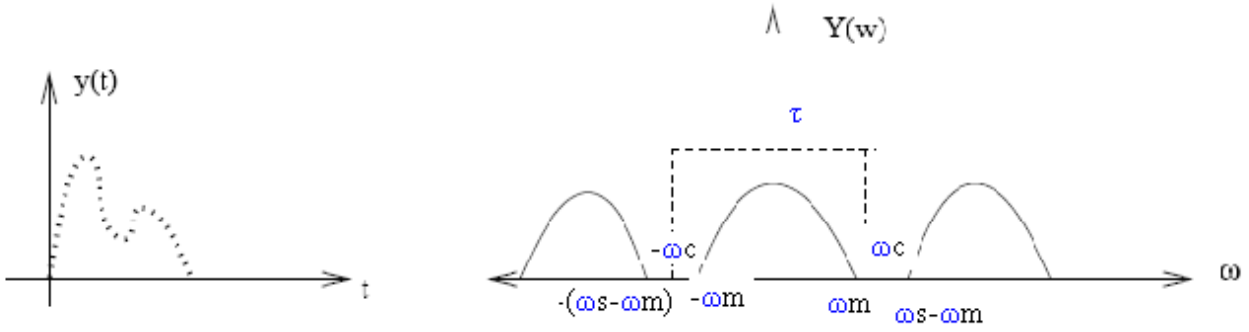


Figure 5.19 Reconstruction of original signal from sampled signal

If a sinusoidal signal is sampled at exactly twice its frequency then some information may be lost. To understand this consider an arbitrary sinusoidal signal, represented as the sum of a sine and cosine components operating at the same frequency, as shown in equation (5.30)

$$C\sin\left(2\pi f_o t + \phi\right) = A\cos\left(2\pi f_o t\right) + B\sin\left(2\pi f_o t\right)$$

(5.30)

Figure 5.20 illustrates the samples derived from a sinusoidal signal sampled at twice its frequency. As it is evident all the samples from the sine wave are zero, whereas those from the cosine signal have sampled values. Therefore sampling at exactly twice the Nyquist rate is not advisable.

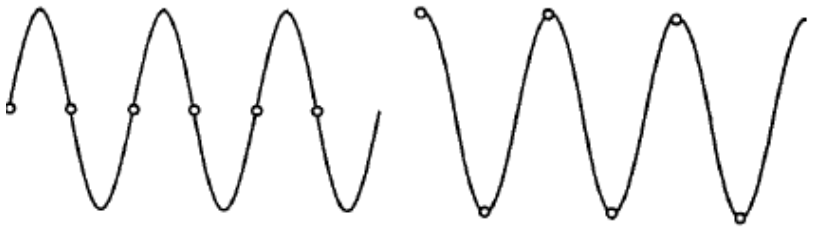


Figure 5.20 Sine and cosine waves sampled at twice the signal frequency

Similarly sample sequence obtained by under sampling(which results in aliasing) can represent different CT signals during reconstruction as shown in figure 5.21.

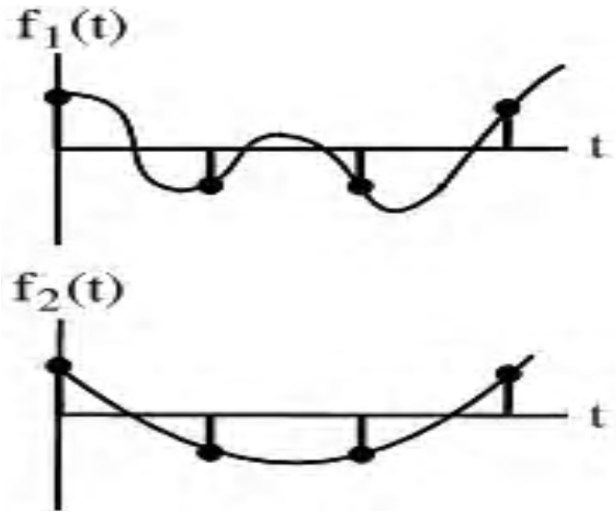


Figure 5.21 different signals of same sample points

Therefore if the processed signal is to be converted back to CT signal then proper reconstruction is necessary. By dynamically selecting the sampling rate near exact reconstruction is possible. Decimation and Interpolation are methods for changing the effective sampling rate of DT signal. Decimation reduces the effective sampling rate and Interpolations increases the effective sampling rate.

Sub-Sampling:

Sampling is also performed frequently on discrete time signals to change the effective data rate. Sub-sampling is the process of reducing the data-rate**.** The impact of sub-sampling can be examined by

comparing the DTFT of the sampled signal to the DTFT of the original signal.

The sub-sampling is used to relate DTFT of x[n] to the DTFT of y [n] where y[n] is a subsampled version of x[n] .Let us represent the x[n] as the sampled version of x(t) and y[n] as the sampled version of y(t).

The x[n] is obtained as integral mutiples of τ as shown in equation (5.31),

$$x[n]=x(n\tau) \tag{5.31}$$

Therefore $$y[n]=x[qn] \tag{5.32}$$

$$=x(nq\tau)$$

where τ is sampling interval and q is the integer for sample.

Decimation:

Decimation is a technique for reducing the number of samples in a discrete-time signal. Decimation is commonly called as downsampling. Decimation is a two-step process:

1.Low-pass anti-aliasing filtering

2.Down sampling

Consider x1[n] and x2[n] are the sampled signals of the two identical CT signals with different intervals and the signal x2[n] to be converted to x1[n]. One way to do this ,to convert the DT sequence of x2[n] back to CT signal and again resample with the rate same as that of x1[n].But this method leads to distortion in reconstruction operation.

Another method,subsampling operates directly on DT signals to change the sample rate. Subsampling the x2[n] directly results in the aliasing of noise into frequency less than W. This aliasing problem is prevented by applying a low pass filter to x2[n] prior to subsampling .

Figure 5.22 depicts that decimation system. The input signal X[n] corresponds to oversampled signal. The low pass filter output xp[n] has (q-1) zero values between nonzero values.

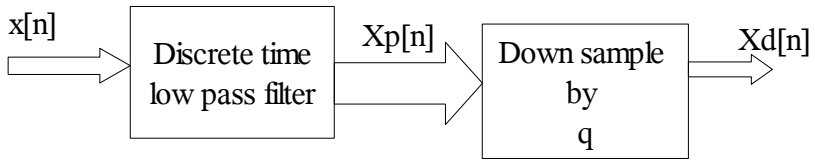


Figure 5.22 Decimation system

Figure 5.23 shows that ***decimation in time domain.***

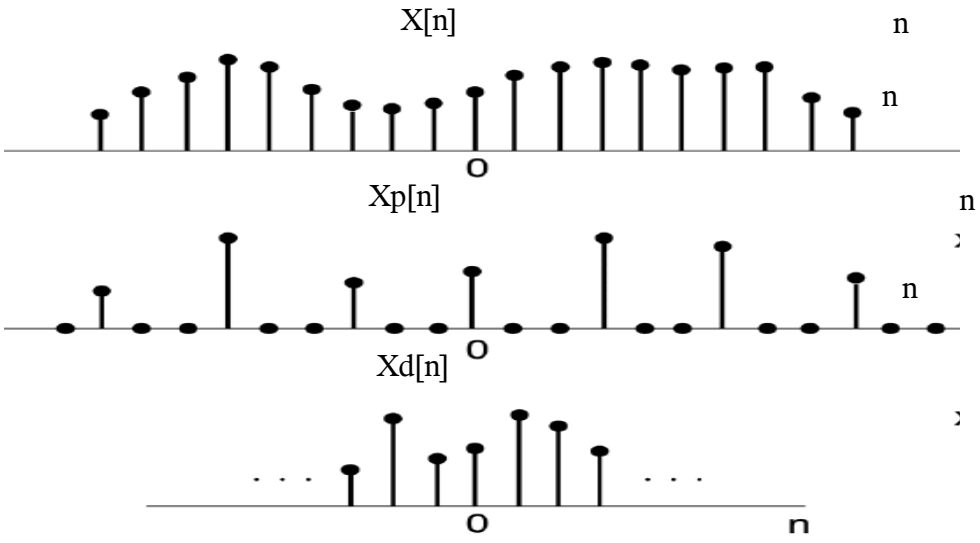


Figure 5.23 Decimation in time domain

Figure 5.24 shows that ***decimation in frequency domain.***

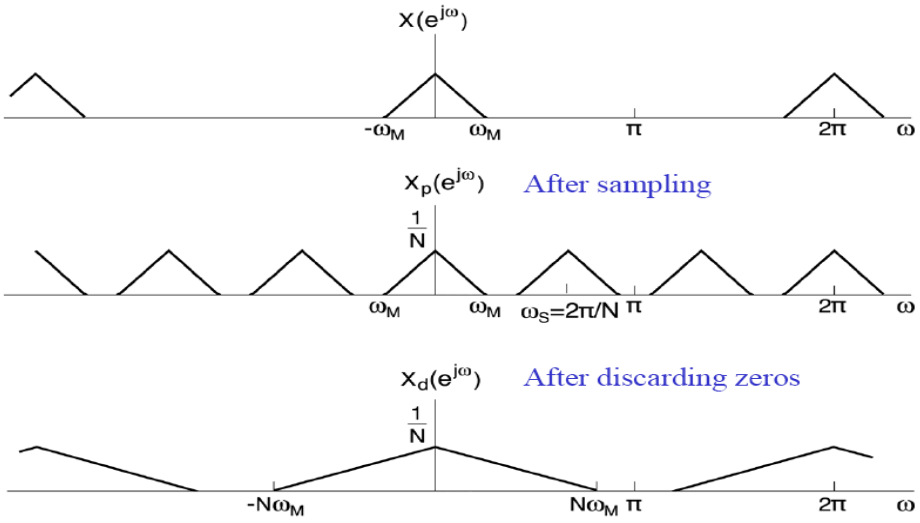


Figure 5.24 Decimation in frequency domain

SuperSampling::

Supersampling is the process of sampling at a frequency higher than the target sampling frequency . The key idea of SuperSampling is the same as oversampling. However, in supersampling,

no analog filtering or quantization is involved. Supersampling when combined with filtered decimation is a simple and effective method for anti-aliasing when prefiltering (filtering the continuous signal before any sampling) is difficult.

Oversampling:

The process of increasing the effective sampling rate of a DT signal is called oversampling. It is done to relax the anti-aliasing filter and allows a wide transition band in the anti-aliasing filter.

An anti-aliasing filter is used to prevent aliasing by limiting the signal band width prior to sampling. The signal of interest having ,maximum frequency fm, then CT signal will have energy at higher frequency due to presence of noise and other non-linear characteristic. The anti-aliasing filter is chosen to prevent the noise from aliasing back down into band of interest.

Interpolation: Interpolation is the process of upsampling and filtering a signal to increase its effective sampling rate by inserting redundant samples of zero amplitude.

Interpolation is a two-step process:

1.Upsampling

2.Low pass filtering.

The frequency domain of x1[n] is converted into x2[n] by increasing the sampling rate by a integer factor ie., $\tau 2=\tau 1/q$. The interpolation by a factor q can be accomplished by inserting q-1 zeros in between each sample of x1[n] and then low pass filtering.This interpolation procedure has a time domain interpretation analogous to that of CT reconstruction process.

Figure 5.25 shows an interpolation system where the decimated signal are converted into up sampled signal by inserting N-1 zeros in xb[n]

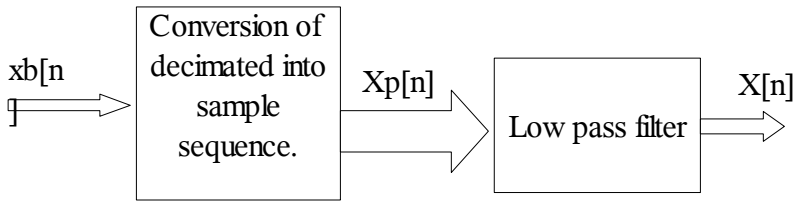


Figure 5.25 Interpolation system

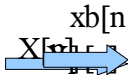


Figure 5.26 shows the ***interpolation in time domain*** function

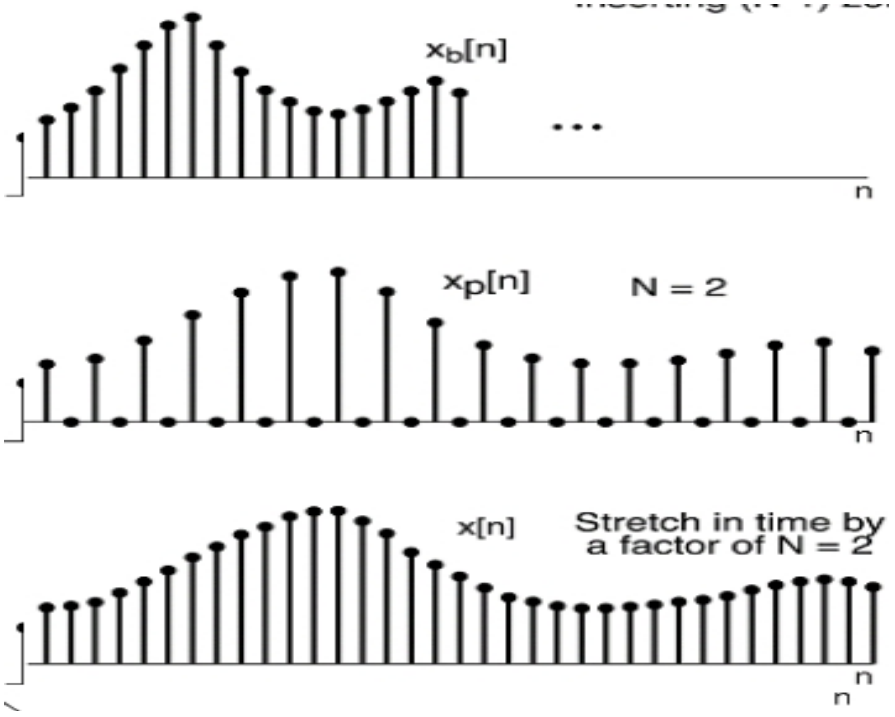


Figure 5.26 Interpolation in time domain

Figure 5.27 shows that the ***interpolation in frequency domain***

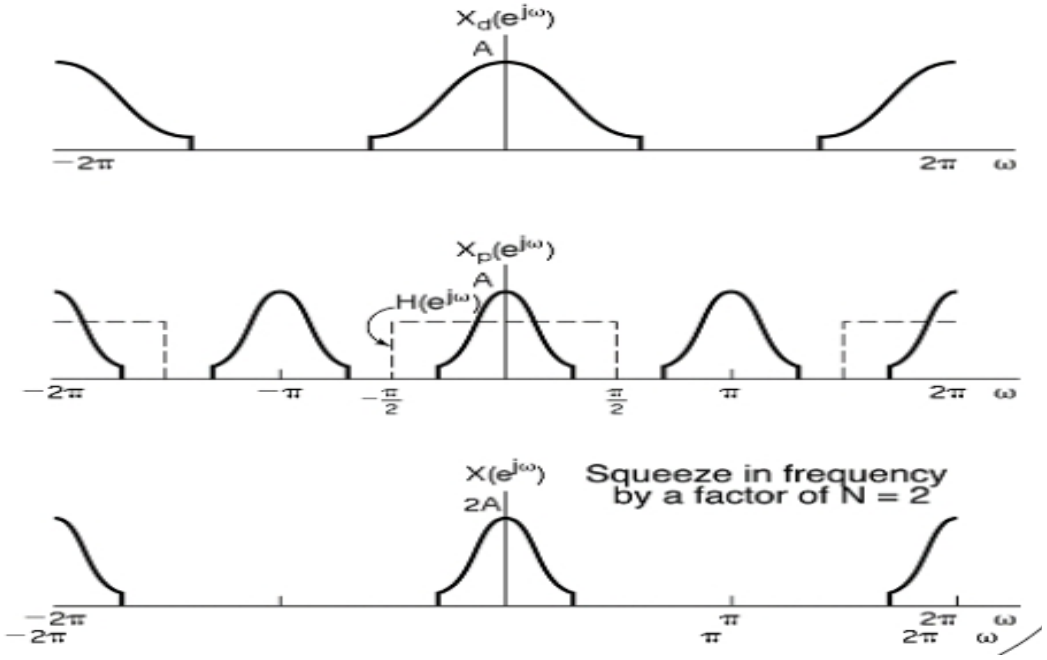


Figure 5.27 Interpolation in frequency domain:

5.4.3 Analysis of DT signals:

The discrete time signals are analyzed with the help of Discrete Time Fourier Transform(DTFT),Discrete Time Fourier Series(DTFS),Discrete Fourier Transform(DFT) and Z-Transform. Normally Discrete Time Fourier Series(DTFS) is used to analyze periodic signals whereas Discrete Time Fourier Transform(DTFT) is useful to analyze non periodic signals. The DFT and Z-Transform are used for analyzing both the signals.

DFT is a Fourier representation of a finite-length time domain sequence. It depicts samples that are equally spaced in frequency of the Fourier transform of the signal. The sequence $x[n]$ of N samples $x_0, \ldots, x_{N-1}$ is transformed into frequency domain X_k by DFT and it is expressed as I equation (5.33)

$$X[k]= \sum_{n=0}^{N-1} x[n]\ e^{-j2\pi kn/N} \qquad \text{where } k=0,1,2....N-1 \qquad (5.33)$$

The Inverse Discrete Fourier Transform (IDFT) is given in equation (5.34)

$$x[n]= 1/N \sum_{k=0}^{N-1} X[k]\ e^{-j2\pi kn/N} \qquad \text{where } n=0,1,2....N-1 \qquad (5.34)$$

Fast Fourier Transform (FFT) An FFT computes the DFT of a sequence in a much faster way and produces exactly the same result as evaluating the DFT definition directly. The main difference is that the number of computations in FFT is less. Also, in the presence of round-off error, many FFT algorithms are more accurate than evaluating the DFT directly.

FFT functionally decomposes the set of data to be transformed into a series of smaller data sets. The DFT of the first N/2 points and DFT of the second N/2 points are combined algorithmically to produce a single N-point DFT. Each of these N/2-point DFTs is calculated using smaller DFTs in the same way. Therefore a radix-2 FFT begins by calculating N/2 2-point DFTs. These are combined to form N/4 4-point DFTs. The next stage produces N/8 8-point DFTs, and so on, until a

single N-point DFT is produced. For example, an FFT of size 32 is broken into 2 FFT's of size 16, which are broken into 4 FFT's of size 8, which are broken into 8 FFT's of size 4, which are broken into 16 FFT's of size 2. Figure 5.28 shows that Three stages in the computation of an $N = 8$-point DFT.

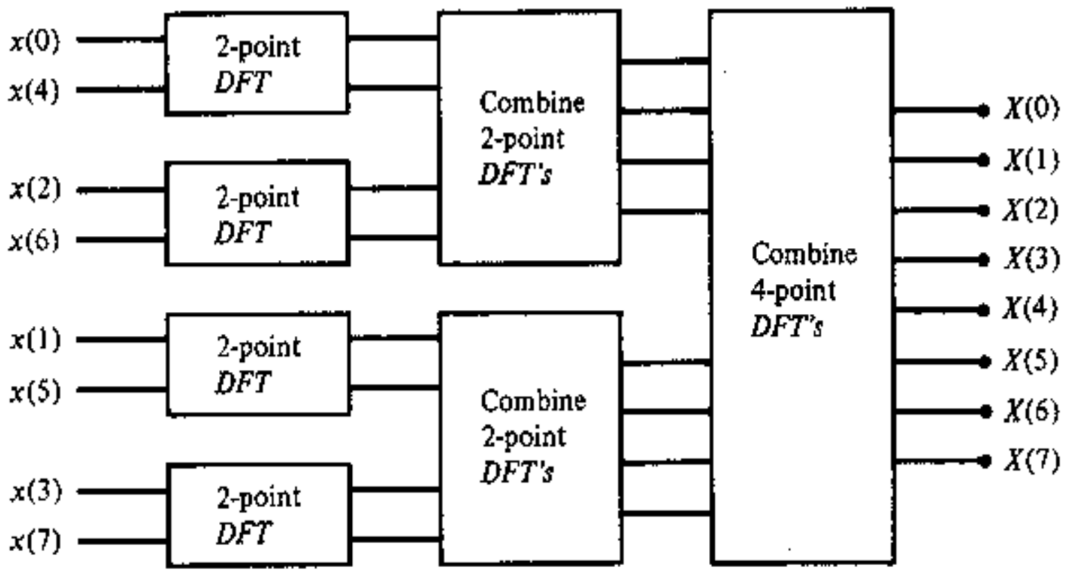


Figure 5.28 Three stages in the computation of an $N = 8$-point DFT.

Mathematically there are two FFT algorithms known as

Decimation-in-Time (DIT) and

Decimation-in-frequency(DIF) FFT algorithms.

These two FFT algorithms that perform the same operations of finding the DFT of a sequence.

Decimation-in-time (DIT) FFT algorithms: DIT algorithm is nothing but separating the input samples into smaller sets.Consider x[n] is an input sequence ,decimating the sequence x[n] by a factor of 2 I.e., odd and even sequence and hence the resulting FFT algorithm is called a radix-2 *decimation-in-time algorithm*.

Basic butterfly diagram of 2-point DFT is shown in figure 5.29,

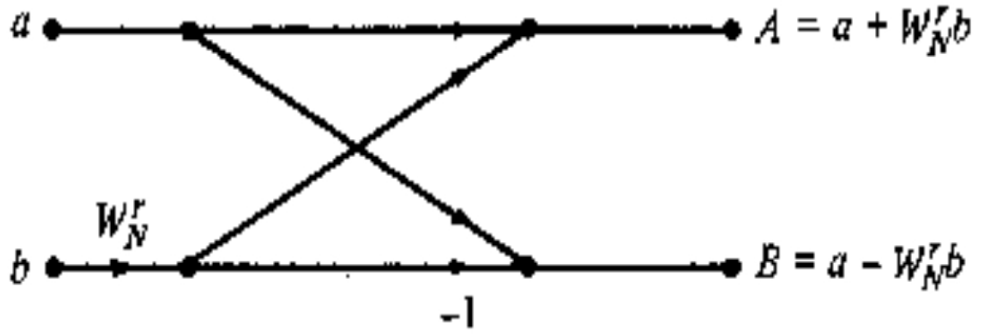


Figure 5.29 Basic butterfly computations in the DIT FFT algorithms

Consider the following example of Figure 5.30 which depicts the computation of N = 8 point DFT. The computation is performed in three stages, beginning with the computations of four two-point DFTs, then two four-point DFTs, and finally, one eight-point DFT. The combination for the smaller DFTs to form the larger DFT is illustrated in Figure 5.30 for N = 8

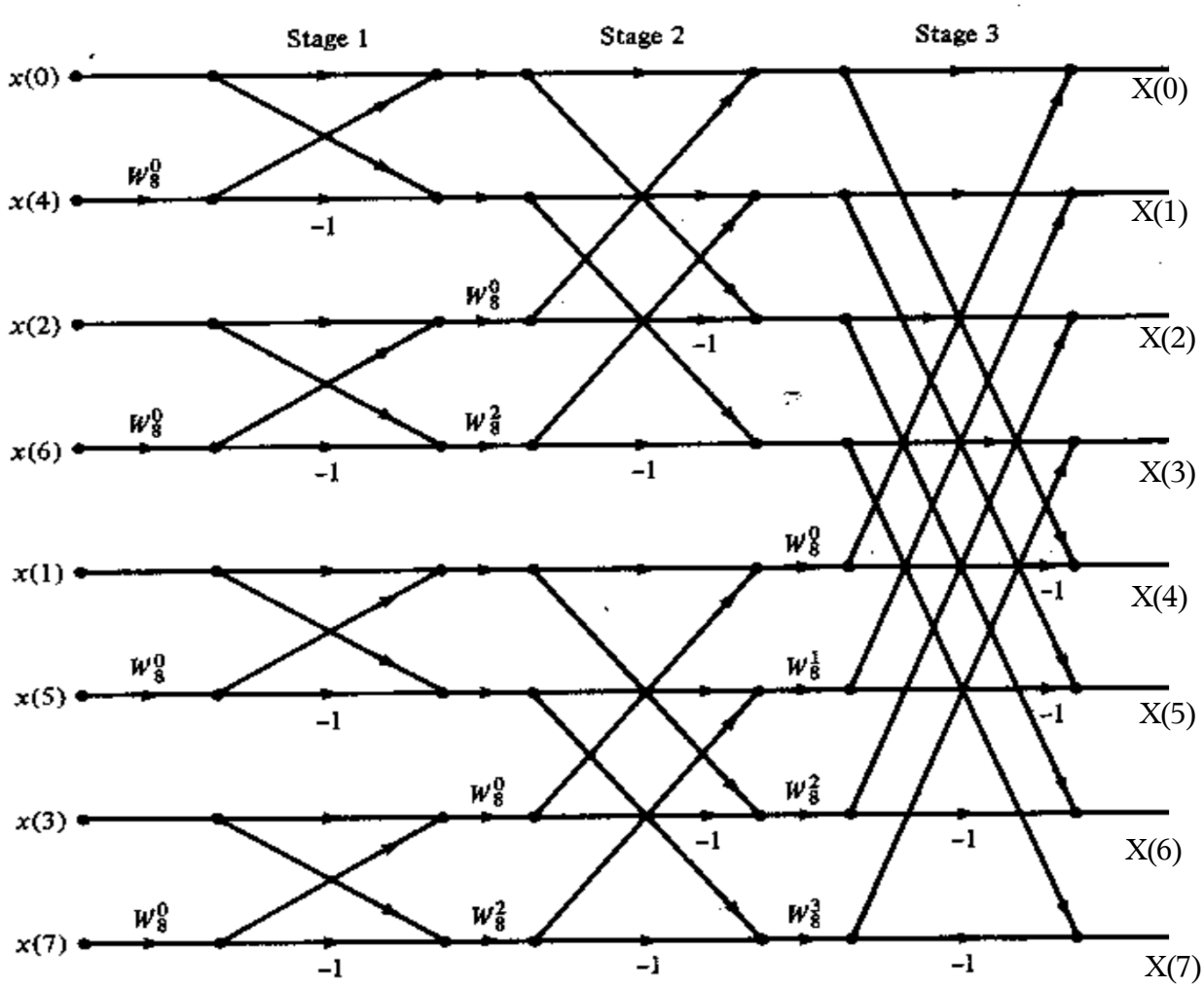


Figure 5.30 Eight point DIT FFT algorithms.

Decimation-in-frequency FFT algorithms:are based on structuring the DFT computation by forming smaller subsequences of the input x[n].Alternatively dividing the output sequence X[k] into smaller and smaller subsequence .To derive the algorithm, by splitting the DFT formula into two summations, one of which involves the sum over the first N/2 data points and the second sum involves the last N/2 data points. Basic butterflies of the type shown in Figure 5.31

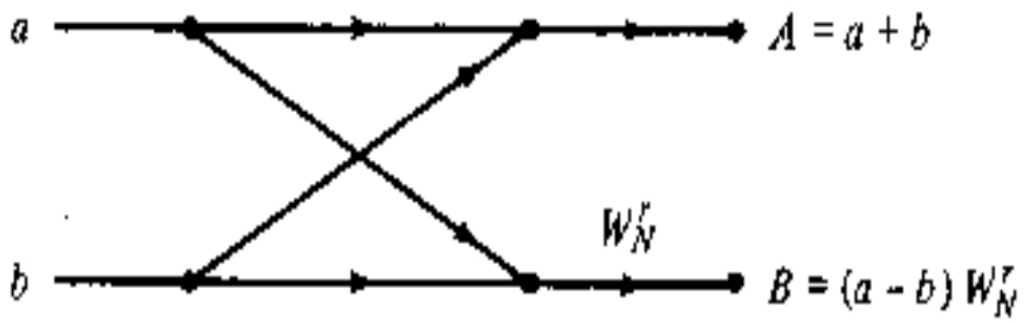


Figure5.31. Basic butterfly computations of 2 point DIF

The eight-point decimation-in-frequency algorithm is given in Figure 5.32

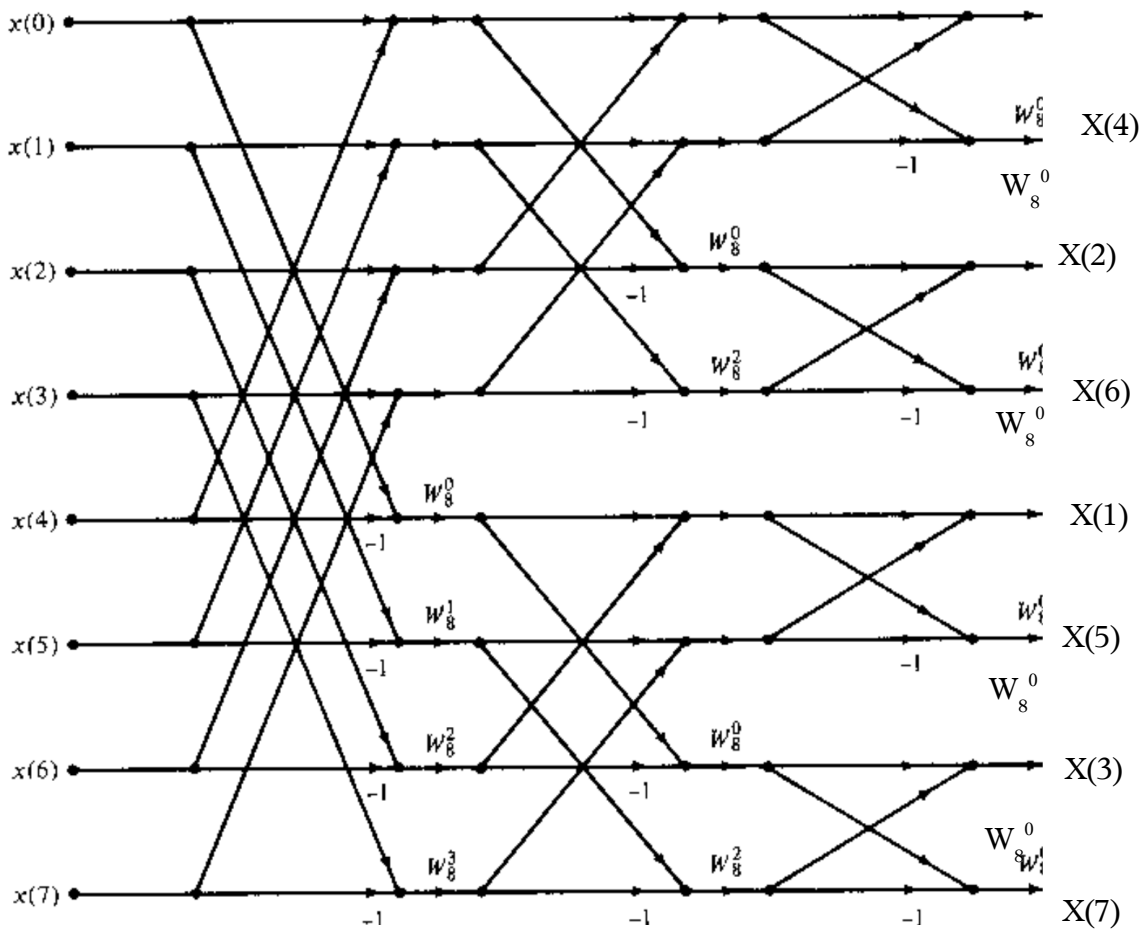


Figure 5.32 N=8 point Decimation-in-frequency FFT algorithm

From this two methods,the bit reversed input required for the DIT algorithm and Bit reversed output gets from DIF algorithm.

The efficient radix-2 FFT algorithm computes the same results with only $(N/2)\log_2 N$ complex multiplications and $N\log_2 N$ complex additions unlike in DFT the number of computations directly involves N^2 complex multiplications and N^2 complex additions. The ratio between a DFT computation and an FFT computation for the same N points is proportional to $N / \log_2(N)$. N complex multiplications=$4N$ real multiplications and N complex additions =$4N$-1 real additions

Note:The "radix" is the size of an FFT decomposition. For radix-2 the computation of the N point DFT can be divided as the *power of 2 I.e.,* $N = 2^v$. To split the N-point data sequence into two $N/2$-point data sequences *x(2n) and x(2n+1)*, corresponding to the even-numbered and odd-numbered samples of $x(n)$, respectively..For radix-4 FFT, the number of data points N in the DFT is a power of 4 (i.e., $N = 4^v$), and

X(0)

to decimate the *N*-point input sequence into four subsequences, $x(4n)$, $x(4n+1)$, $x(4n+2)$, $x(4n+3)$, $n = 0, 1, \ldots, N/4-1$.

Depending on the properties of a given signal corresponding transform can be chosen to convert from time dom ain to frequency domain and it is summarized in table 5.1.

Time domain	Periodic	Non-periodic	
continuous	Fourier series $x(t)=\sum_{k=-\infty}^{\infty} X[k]\ e^{jk\omega 0t}$ $X[k]=1/T \int_{(T)} x(t)\ e^{-jk\omega 0t}\, dt$ x(t) has period T and $\omega 0=2\pi/T$	Fourier transform $X(t)=1/2\pi \int_{\infty} x(j\omega)\ e^{jk\omega t}\, d\omega$ $X[j\omega]=\int_{\infty} x(t)\ e^{-jk\omega t}\, dt$	Non-periodic
discrete	Discrete Time Fourier series $x[n]=\sum_{k=-(N)} X[k]\ e^{jk\Omega_0 n}$ $x[k]= 1/N \sum_{n=(N)} X[n]\ e^{-jk\Omega_0 n}$ x[n]and x[k] have period N $\Omega_0=2\pi/N$	Discrete Time Fourier Transform $x[n]= 1/2\pi \int_{-\pi}^{\pi} X(e^{j\Omega})\ d\Omega$ $X(e^{j\Omega})=\sum_{n=\infty} x[n]\ e^{-j\Omega n}$ $x(e^{j\Omega})$ has period 2π	Periodic
	discrete	Continuous	Frequency domain

Table 5.1 Summarized Fourier representation

5.4.4 Analysis of r andom signals:

The random signal is one whose value cannot be predicted in advance. It can be described in terms of statistical properties such as average power, spectral distribution of average power and the like. Random signals are generally nondeterministic,infinite in duration and infinite energy signals. The concepts can be represented mathematically by random process. Random process constitutes a set of random variables together with probability description of each random variable.

Power spectral density(PSD): A spectrum is a relationship typically represented by a plot of the magnitude or relative value of some parameter against frequency. Power spectrum is a positive real function of a frequency variable associated with a function of time, which has dimensions of power per Hz.

The spectral density is an estimate of power spectral density or power spectrum of a random signal from a sequence of time samples of the signal. The spectral density characterizes the frequency content of the signal. The purpose of estimating the spectral density is to detect any periodicities in the data, by observing peaks at the frequencies corresponding to these periodicities. Fourier method applied to the random power signal result in power spectral density function. Power spectral density can be obtained using the following two approaches.

1.Fourier Transform method

2.Auto correlation method

Power spectral density of the random signal will play the same role as does the transform(such as Fourier transform, Fourier series,Z-transform etc.,) of the non-random signal. Since a signal with nonzero average power is not squarely integrable, the Fourier transform does not exist. So Wiener–khintchine theorem provides a simple alternative. The PSD is the Fourier transform of the autocorrelation function, $R(\tau)$, of the signal if the signal can be treated as a stationary random process.

The general expression of PSD is shown in equation 5.35

$$p=1/2\prod \int_{\infty} s(f)\, df \qquad (5.35)$$

Consider X(t) is a voltage or current signal associated with 1 ohm resistance. Then average power dissipated in that resistance P over a bandwidth of frequency |f1-f2| and it is expressed as in equation (5.36)

$$P = \int_{F_1}^{F_2} S(f)\,df + \int_{-F_2}^{-F_1} S(f)\,df.$$

(5.36)

Now the Power spectral density S(f) can be interpreted as the average power associated with the bandwidth of 1Hz (average power per Hz) The spectral density defined above is referred as two sided spectral density since it exists for both positive and negative values of f. It may also be represented only either positive or negative values of f. This spectral density is known as one sided spectral density.

Mathematically,the PSD s(f) is the Fourier transform of the auto correlation function, R(τ), of the signal. If a given random signal is wide-sense stationary(WSS) random process* then PSD is the Fourier transform of the auto correlation function, R(τ), of the signal and it is expressed as in equation (5.37)

$$S(f) = \int_{-\infty}^{\infty} R(\tau)\, e^{-2\pi i f \tau}\, d\tau.$$

(5.37)

Note:For a wide-sense stationary*(WSS) process harmonics will not vary with respect to time

Properties of PSD:

- ***S*(*f*) >= 0**
- ***S*(−*f*) = *S*(*f*)**

Note: By Wiener khintchine theorem,if the signal is WSS ,then

R(τ)----------FT---------->s(f) (5.38)

`Correlation:

Correlation is a mathematical tool used frequently in signal processing for analyzing functions or series of values, such as time domain signals .Correlation describes the degree of relationship between two or more random variables. The correlation depends upon how rapidly the random variable changes with respect to time .It may also be related with the energy content of the random process and it is distributed with respect to frequency .

Autocorrelation:

Auto correlation is a measure of the dependence between the values of the random signal at different times. The auto correlation of a random signal is simply the correlation process of a signal and its time-shifted version . Auto correlation is useful for finding the presence of a periodic signal which has been buried under noise,or identifying the fundamental frequency within a random signal which doesn't actually contain that frequency component, but implies it with many harmonic frequencies .

If Xi is stationary process with mean μ and standard deviation σ, then the auto correlation Rxx(τ) is expressed as

$$Rxx(k) = \frac{E[(X_i - \mu)(X_{i+k} - \mu)]}{\sigma^2} \tag{5.39}$$

where E is the expected value which is nothing but the probability of collection of all possible time functions Or time average value of any sample function by integrating over a range of possible random variable and k is the time shift.

Cross correlation:

Cross-correlation is a measure of the dependence between two random variables from different random signals. This situation arises when there is more than one random signal being applied to a system or to compare random voltages or currents occurring at different points in the system. It can also be explained as a measure of similarity of two waveforms as a function of a time-lag applied to one of them.

Consider two random processes $X(t)$ and $Y(t)$ .The *cross-correlation function* is given in equation (5.40),

$$R_{X,Y}(t_1, t_2) = E\left[X(t_1)Y(t_2)\right] \quad (5.40)$$

White noise

Noise is the random signal which is usually unwanted that tend to disturb the transmission and processing of signals . The noise analysis can be performed based on its idealised form called as white noise.

White noise is spread over all frequency components with equal proportion in analogy with white light. It has a flat power spectrum for the frequency range $-\infty < f < \infty$. The PSD of a white noise can be expressed as in equation (5.41)

$$S_w(f) = No/2 \quad (5.41)$$

Where No is two sided noise power spectral density and the unit is watts/Hertz. Here the factor ½ indicates that half the power is associated with the positive frequencies and half with negative frequencies. As said above the PSD of white noise is independent of frequency and it is No/2 for all frequency. The fourier transform of autocorrelation function of white noise produces its power spectral density. The autocorrelation function and power spectral density of a white noise forms a Fourier transform pair and is shown in figure 5.33. The auto correlation function of white noise is delta function weighted by a factor No/2 at $\tau=0$.

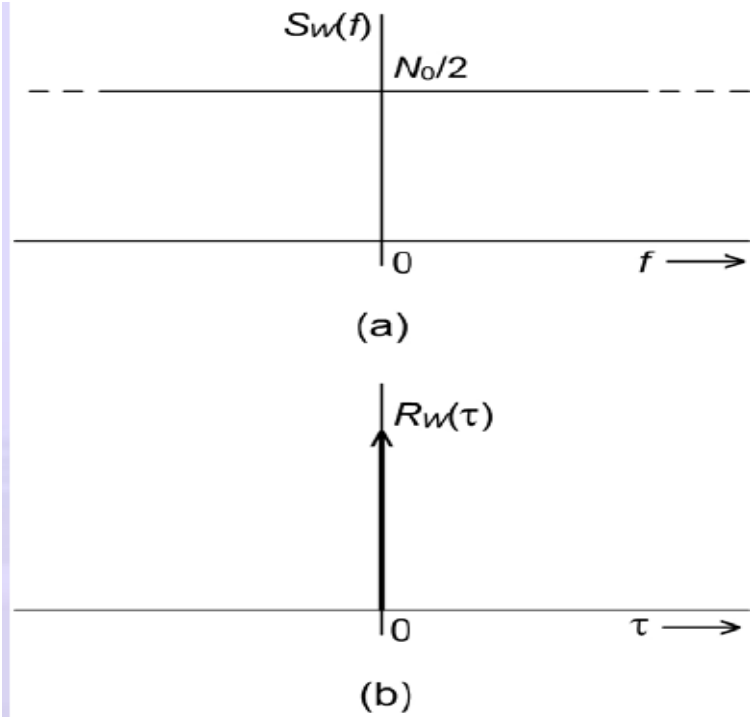


Figure 5.33 characterisation of white noise: a) Power spectral density

b) Auto correlation function

Gaussian noise:

Gaussian noise is defined as the noise with a Gaussian amplitude distribution. Gaussian noise is really like white noise i.e.,it has equal energy at all frequencies but the amplitude distribution is different. Gaussian noise has near uniform amplitudes close to zero, and non uniform amplitudes at extreme values whereas white noise is having equal probability at all values.

Additive white Gaussian noise (AWGN) is modelled by a linear addition of wideband or white noise with a constant spectral density and a Gaussian distribution of amplitude.It is commonly used to simulate background noise of the channel, in addition to multipath, terrain blocking, interference, ground clutter and self interference that modern radio systems encounter in terrestrial operation.

The random variable x results in Gaussian distribution if its probability density function has the form as in equation (5.42),

$$p_X(x)=1/\ \sigma\sqrt{\ }(2\prod).\ e^{-(x-\ \mu)\hat{}2/2\ \sigma\hat{}\ 2} \qquad (5.42)$$

where $\sigma > 0$ is the standard deviation, μ is the expected value.

5.5 Logarithmic Unit of measurement-dB , dBm, dBc conversion:

dB"decibel" : is a logarithmic unit of measurement that expresses the magnitude of a physical quantity relative to a specific reference level.

The decibel measurements are used to represent a very large or small ratios by a convenient number,For example :-70 dB =10 log(1/10,000,000) ,70 dB =10 log(10,000,000) .

In dealing with gain of an electronic circuit it is common practice to express the gain in terms of decibels. When referring the measurements of power ,if two powers Pin and Pout are power flowing into the circuit and flowing out of the circuit ,then the power gain is expressed as in equation 5.43

$$Gp = 10*\log_{10}(P2/P1) \quad (dB) \tag{5.43}$$

Also the voltage gain can be derived as,

$$Gp = 10*\log_{10}(P2/P1) \quad (dB)$$

$$Gp=10 *\log_{10}(V2^2R/V1^2R)$$

$$Gv= 20*\log_{10}(V2/V1) \quad (dB) \tag{5.44}$$

Similarly the current gain , $Gi= 20*\log10(I2/I1) \quad (dB)$ (5.45)

dBm:Absolute power levels are expressed in dBm and generally refers to input and output power levels individually. The 'm' refers to the reference level used, in this case mW (milliWatts).

$$Gp(dBm) = 10*\log_{10}(P/1\ mw) \quad (dBm) \tag{5.46}$$

where P is the power .For example: the value of 0 dBm indicates power of 1 mW.

dBc:Power in dBc refers the power levels with reference to the main carrier power .Typically it is used to describe spurs, noise, channel crosstalk and intermodulated signals which may interfere with the carrier.

$$Gp(dBc) = 10*\log_{10}(P2 / Pc) \quad (dBc) \tag{5.47}$$

Common power gain values:

3dB increment for factor of 2 for power gain

Let the power gain be 2,then the decibel gain is given by,

For power gain is 2 $Gp= 10\log_{10} 2=3.01dB$

For power gain is 4 $Gp= 10\log_{10} 4=6.02dB$

For power gain is 8 Gp= $10\log_{10} 8$=9.03dB

For power gain is 16 Gp=$10\log_{10} 16$=12.04dB

10dB increment for factor of 2 for power gain

Let the power gain be 10,then the decibell gain is given by,

For power gain is 10 Gp=$10\log_{10} 10$=10dB

For power gain is 100 Gp=$10\log_{10} 100$=20dB

For power gain is 1000 Gp=$10\log_{10} 1000$=30dB

For power gain is 10000 Gp=$10\log_{10} 10000$=40dB

Common voltage gain values:

6dB increment for factor of 2 for voltage gain

Let the voltage gain be 2,then the decibell gain is given by,

For **voltage** gain is 2 Gv=$20 \log_{10} 2$=6 dB

if **voltage** gain is 4 Gv=$20 \log_{10} 4$=12.00 dB

if **voltage** gain is 8 Gv=$20 \log_{10} 8$=18.00 dB

20 dB increment for factor of 10 for voltage gain

For **voltage** gain is 10 Gv=$20\log_{10} 10$=20dB

For **voltage** gain is 100 Gv=$20\log_{10} 100$=40dB

For **voltage** gain is 1000 Gv=$20\log_{10} 1000$=60dB

Calculating dBm with known Vrms :In a 50 ohm system if you know the Vrms value of the output that is being measured,it can be easily converted using the below formula,

PdBm=10 log(pout/1mW)

PdBm=10 log((Vrmsout^2/50)/1mW)

PdBm=10 log((Vrmsout^2/.05)

PdBm=10 log((Vrmsout^2*20))

calculating dBm with known Vpk: In a 50 ohm system if you know the Vpk value of the output that is being measured,it can be easily converted using the below formula,

PdBm=10 log((Vrmsout^2*20))

PdBm=10 log((Vpk/$\sqrt{2}$)^2*20))

PdBm=10 log((Vpk^2 /2*20))

PdBm=10 log((Vpk^2 *10))

PdBm=10 log Vpk^2 *+10

PdBm=20 logVpk *10

Calculating Vrms with known dBm:

$$V_{rms} = \frac{\sqrt{10^{\left(\frac{P_{dBm}}{10}\right)}}}{\sqrt{10} \times \sqrt{2}}$$

$$V_{rms} = \sqrt{\frac{10^{\left(\frac{P_{dBm}}{10}\right)}}{10 \times 2}}$$

Calculating VPk with known dBm:

$$V_{pk} = \sqrt{\frac{10^{\left(\frac{P_{dBm}}{10}\right)}}{10}}$$

5.6 Introduction to communication Theory:

The purpose of communication system is to transmit the information bearing(baseband) signal from transmitter to receiver through a channel. The proper utilization of communication channel requires a shift of a range of base band frequencies into other frequency ranges(carrier) suitable for transmission. The shift of the range of frequencies is accomplished by *modulation* process.

Modulation: ***is the process of changing some characteristics of a carrier signal like amplitude, frequency or phase in accordance with the instantaneous values of the information signal*** . Modulation is performed at the transmitting end. At the receiving end,the original baseband signal is restored by the process called d*emodulation*. The various modulation techniques used in communication systems are discussed below ,

5.6.1 Analog modulation

In analog modulation, the modulation is performed continuously in response to the analog information signal.

Amplitude modulation:

It is defined as the process in which amplitude of the carrier signal is varied in accordance with

the instantaneous values of baseband signal. Figure 5.50 shows the Amplitude modulated signal .Let us represent the modulating wave Eam(t) and it is given as

$$em=Em \sin \omega_{mt} \quad (5.48)$$

,and the carrier signal can be represented as ec as,

$$ec=Ec \sin \omega_c t. \quad (5.49)$$

where Ec is the maximum amplitude of the carrier signal

Em is the maximum amplitude of the modulating signal

ω_m is the frequency of the modulating signal

ω_c is the frequency of the carrier signal

Therefore the instaneous value of the amplitude modulated wave can be given as,

$$eAM=(Ec+Em \sin \omega_m t) \sin \omega_c t \quad (5.50)$$

This is an equation of AM wave.It can also be expressed as in equation (5.51),

$$eAM=Ec \sin (\omega_c t)+(Ec * m/2)[\cos((\omega_c - \omega_m)t)-\cos((\omega_c + \omega_m)t)] \quad (5.51)$$

Modulation Index(m): is the quantity which indicates the percentage change in the modulated variable around its unmodulated level.It is a measure of quality of modulation process. Here in AM the modulation index is defined as the ratio of the modulating signal amplitude to the carrier amplitude.

It can be expressed as in equation (5.52)

$$m=Em/Ec \quad \text{where } 0<m<1 \quad (5.52)$$

Value of Em must be less than Ec to avoid any distortion in the modulated signal.

If the modulation index is zero (m = 0) the modulated signal is simply a constant amplitude carrier.

If the modulation index is 1 (m= 1),then resultant signal has maximum or 100% amplitude modulation.

When the modulation index is greater than one ,the it is called over modulation which causes loss of information.

The simple AM signal is as shown in figure 5.34,This is called time domain representation of AM

signal.

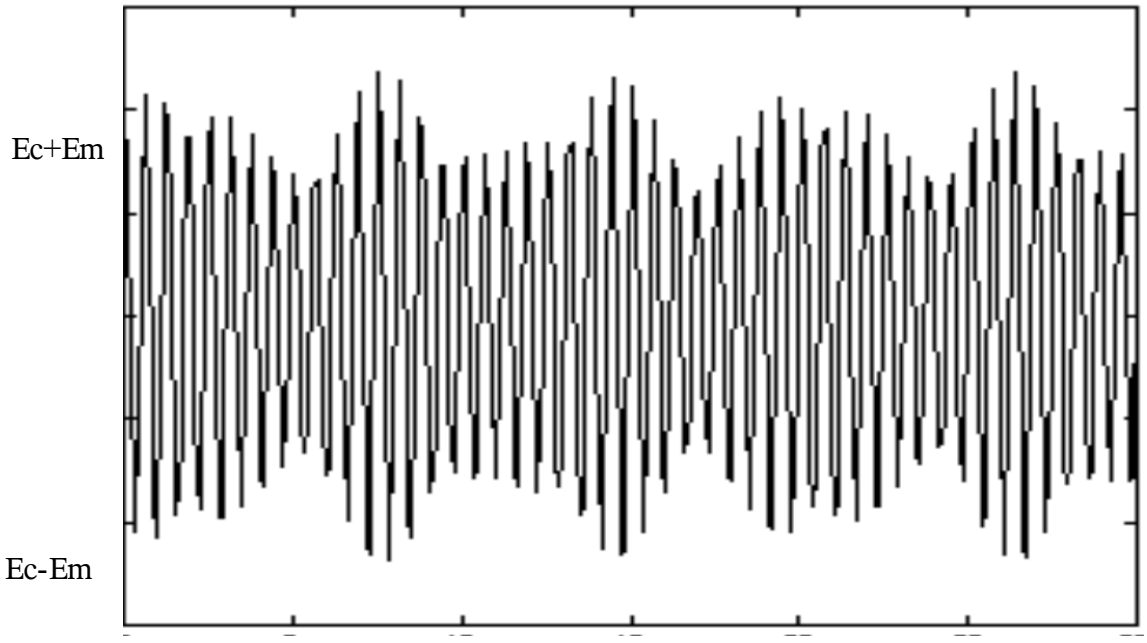


Figure 5.34 AM signal

Amplitude modulation produces a signal with power concentrated at the carrier frequency(ωc) and in two adjacent sidebands. Each sideband(upper (ωc+ωm) and lower(ωc-ωm)) is equal in bandwidth to that of the modulating signal and it is a mirror image of the other. Therefore the AM spectrum is named as double-sideband full-carrier(DSBFC) spectrum**.**

Amplitude modulation is inefficient in terms of power consumption . Because atleast two-thirds of the power is concentrated in the carrier signal, which carries no useful information. The remaining power is shared between two identical sidebands which contain identical information.

Double-Sideband Suppressed-Carrier (DSBSC): To increase transmitter efficiency, the carrier component can be removed (suppressed) from the AM signal. This produces a DSBSC signal. A suppressed-carrier in amplitude modulation scheme is three times more power-efficient than DSBFC.

Single-Sideband Suppressed-Carrier SSBSC: In this scheme carrier component and one of the side bands are removed. Hence maximum efficiency is achieved but at the cost of increased transmitter and receiver complexity .Single-sideband modulation is widely used in point to point communication due to its efficient use of both power and bandwidth.

Vestigial Sideband (VSB): In this transmission one of the sideband is partially suppressed and a portion(vestige) of other sideband is added to compensate for this removal. This transmission is widely

used in television broadcasting.

Comparison of various AM :

The DSBSC, SSBSC, VSB AM modulation techniques are compared as shown in below table 5.2

sl.no	Parameter	DSBSC	SSBSC	VSB
1	Method of transmission	Only side bands	Only one side bands	Partially suppressed one side band and other sideband
2	Bandwidth	2Fm	Fm	Fv+Fm
3.	Power saving	66%	83%	75%
4.	Applications	Carrier telephony	Police wireless	Television

Table 5.2 Comparison of AM

Frequency Modulation(FM): The frequency of the carrier wave is varied in accordance with instantaneous values of the message signal.The amplitude and phase of the carrier remains constant

The modulated signal y(t) is given in equation (5.53),.

$$y(t)=Ac \cos (2\prod fc(t) + m \sin(2\prod fm(t)) \quad (5.53)$$

where m is modulation index ,A is the amplitude of the modulated signal, fm is the highest modulating frequency and fc is the carrier frequency.

Modulation Index :

Modulation index of FM is defined as in equation (5.54)

$$m=\Delta f/fm \quad (5.54)$$

where **Δf** is the frequency deviation viz the maximum instantaneous difference between an FM modulated frequency and the carrier frequency. It indicates the largest swing or deviation in frequency on either side of fc. This value depends upon the magnitude of the modulating signal , but not upon the modulating signal frequency fm,

The effective bandwidth of the FM is given by $2(\Delta f+Fm)$

If M<<1,the modulation is called narrowband FM,and its bandwidth is approximately 2Fm .

If M>>1 the modulation is called wideband FM and its bandwidth is approximately $2\Delta f$

Here the simple FM signal is as shown in figure 5.35 ,

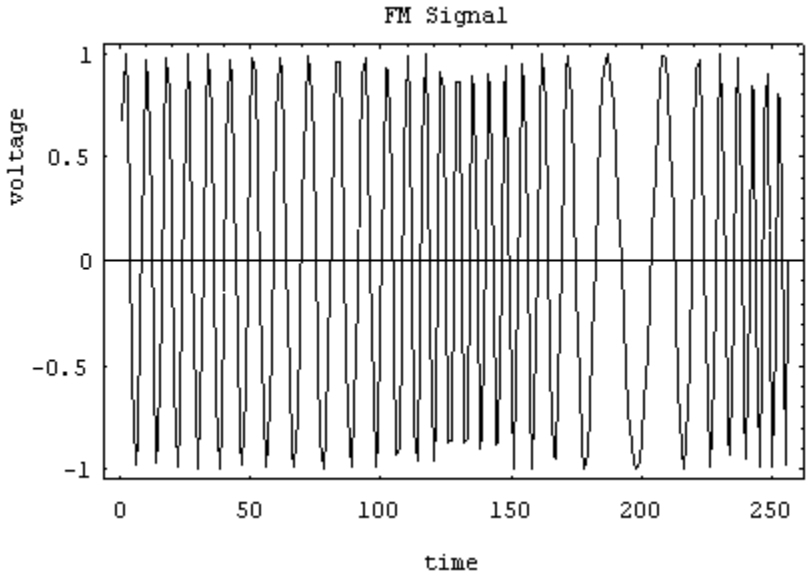


Figure 5.35 FM signal

Phase modulation: In phase modulation the phase of the carrier wave is varied in accordance with instantaneous values of message signal. It produces an infinite number of sideband frequencies. The spacing between these sidebands will be equal to the frequency of the modulating signal. However, the only difference is distribution of power in PM sidebands is not similar to that in FM sidebands. Amount of phase shift of phase modulation is proportional to the amplitude of the modulating signal. Rate of phase shift of Phase modulation is proportional to the frequency of the modulating signal. Main advantages of PM are improved signal to noise ratio and less radiated power.

The modulated signal y(t) is given in equation (5.55),

$$y(t) = A_c \sin\left(\omega_c t + m(t) + \phi_c\right),$$

(5.55)

where Φ_c is the maximum value of carrier phase , Ac is the amplitude of the carrier, $\omega c$is the angular frequency of the carrier.

Modulation Index: It relates to the variations in the phase of the carrier signal and is defined as in equation (5.56)

$$m = \Delta\varphi c \qquad (5.56)$$

where Δφc is the peak phase deviation.

For small amplitude signals, Bandwidth of PM is similar to amplitude modulation (AM) and its value is 2fm . For a single large sinusoidal signal, the BW is similar to FM, and its bandwidth is approximately 2(m+1)fm. The simple PM signal is shown in figure 5.36.

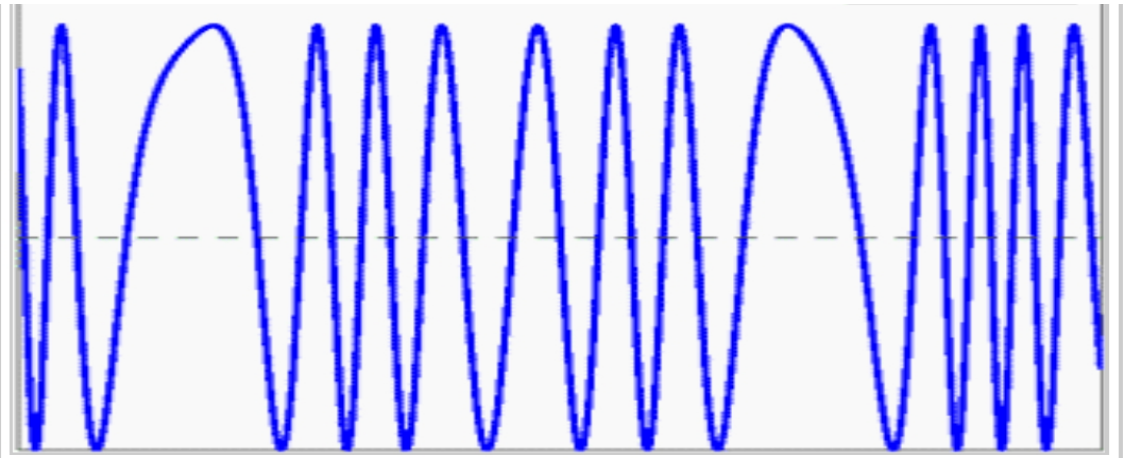


Figure 5.36 PM signal

Comparison of Analog modulation:

The following table 5.3 shows the comparison of AM,FM and PM

Sl.no	AM	FM	PM
1.	Amplitude of carrier is varied,Frequency and Phase are constant	Frequency of carrier is varied,Amplitude and Phase are constant	phase of carrier is varied,Frequency and Amplitude are constant
2.	Modulation index M=Em/Ec	M=(Δf)/Fm	M= Δφ
3.	Total bandwidth Bt=2*Bm where Bm is message signal Bandwidth.	Total bandwidth Bt=2(Δf)+Fm)	Total bandwidth Bt=2(M+1)Fm
4	No amplitude deviation	Maximum frequency deviation is depends upon amplitude of modulating voltage and modulating frequency.	Maximum phase deviation is depends upon the amplitude of modulating voltage.

5	Depth of the modulation have limitation. it cannot be increased above 1.	Depth of the modulation have no limitation. modulation index can be increased by reducing modulation frequency.	Modulation Index remains same if modulating frequency is changed.

Table 5.3 comparison of analog modulation

5.6.2 Digital Modulation

In digital modulation, an analog carrier signal is modulated by a digital bit stream viz baseband or information signal. The number of distinct changes in the carrier signal characteristics is equal to the number of distinct symbols(M) in the modulating bit stream.For example consider a bit stream 001101101011.....If M=4 then the number of bits to represent each symbol should be 2(2^2=4).Hence the symbols are 00,01,10,11 .In order to modulate these four symbols a carrier with four different values in its frequency or phase or amplitude is required.

The advantages of digital modulation are higher information handling capacity, compatibility with digital data services, enhanced data security and better fidelity .The various digital modulation techinques are shown in figure XXX and some of them are explained below,

Amplitude-shift keying (ASK): It is a form of modulation that converts digital data as variations in the amplitude of a carrier wave. The amplitude(not frequency or phase) of an analog carrier signal varies in accordance with the modulating bit stream. The level of amplitude can be used to represent binary symbols logic 0s and 1s. Mathematically ,digital amplitude modulation (ASK) of a binary signal can be represented as in equation (5.57),

$$\mathbf{V_{ASK}(t)=[1+V_m(t)](A_c/2\ \cos \omega_c(t))} \qquad (5.57)$$

Where Vm(t) is the modulating binary signal in volts,Ac is the amplitude of unmodulated carrier signal in volts.,ωc is the carrier angular frequency in radians per second.

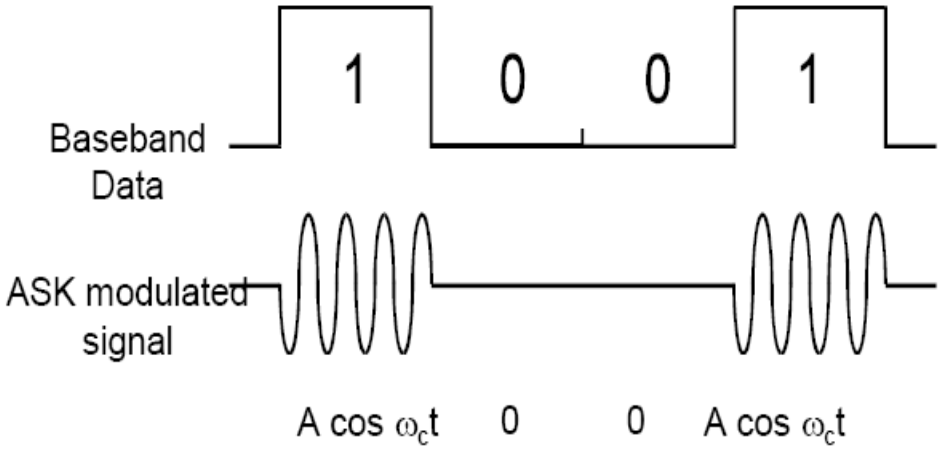


Figure 5.37 ASK modulated signal

As shown in figure 5.37 the amplitude modulated signal VASK(t) is either AcosWc(t) or 0 i.e.,the carrier is either "ON" or "OFF" so that the digital amplitude modulation is also called as "on-off keying"(OOK).

Bit error rate (BER): is defined as the number of bit errors that occur per second. This measurement is one of the prime considerations in determining signal quality. The BER is an indication of how often data has to be retransmitted because of an error. Too high a BER may indicate that a lower data rate would actually improve overall transmission rate .The higher the data transmission rate the greater the standard of modulation.

The bit error rate (Probability of error, Pe) of ASK can be expressed as in equation (5.58),

$$P_e = \left(1 - \frac{1}{L}\right) \operatorname{erfc}\left(\frac{Ag(0)}{\sqrt{2}(L-1)\sigma_N}\right)$$

(5.58)

where erfc() is the complem,entary error function,L is the number of Levels, σN be the variance of Gaussian function,A is the maximum amplitude,g(t) is a nyquist function.

From this equation, the probability of an error decreases if the maximum amplitude of the transmitted signal or the amplification of the system becomes greater. Similarly it increases if the number of levels or the power of noise becomes greater. This relationship is valid when there is no intersymbol interference(ISI) .

Note:In communication, InterSymbol Interference (ISI) is a form of distortion of a signal in which one symbol interferes with subsequent symbols. This is an unwanted phenomenon thus making the communication less reliable. ISI is usually caused by multipath propagation

Frequency Shift Keying(FSK):

It is a frequency modulation scheme in which digital information is transmitted through discrete frequency changes of a carrier wave. The frequency of the carrier is changed as a function of the modulating signal (data) which is being transmitted. Amplitude and phase remain unchanged.

The general expression for FSK signal is defined in equation (5.59),

$$\mathbf{Vfsk(t)=Vc\ cos\{2 \prod [Fc + Vm(t)\ \Delta f]t\}} \qquad (5.59)$$

where Vc=peak carrier amplitude(volts)

fc=centre carrier frequency

Δf= peak frequency deviation

Vm(t)=input modulating signal

The simplest FSK is **binary FSK (BFSK).** BFSK implies using a couple of discrete frequencies to transmit binary (0s and 1s) information. With this scheme, the "1" is called the mark frequency(f1) and the "0" is called the space frequency(f0).The mark and the space frequency are separated from the carrier frequency by the frequency deviation(Fc+-**Δf**]. The time domain of an FSK modulated carrier is illustrated in the figures .

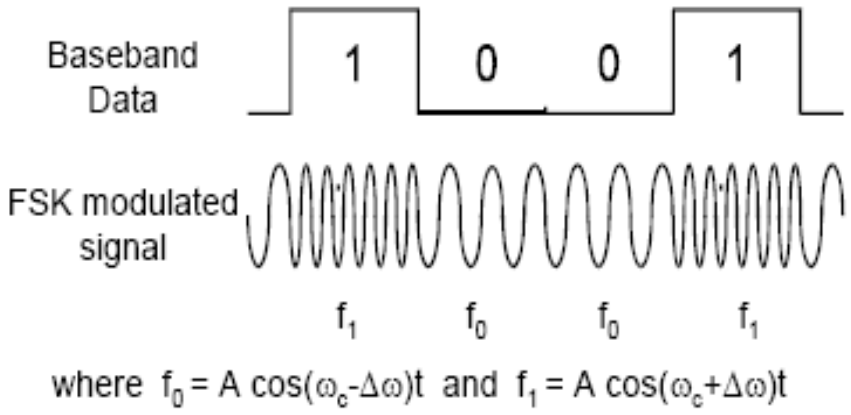


Figure 5.38 FSK Modulated signal

*Bandwidth of FSK:*The output of the FSK modulator is related to the input .The peak frequency deviation is given as

$$\Delta f = |f1-f0)\ /2| \tag{5.60}$$

the minimum bandwidth of FSK signal can be approximated as ,

$$B=2(\ \Delta f + fb) \text{ where fb is the bit rate} \tag{5.61}$$

The bit error rate (Pe) of BFSK can be expressed as in equation (5.62),

$$P_e=1/2\ \text{erfc}\ (\ (E_b/2N_0\ )\text{^}(1/2) \tag{5.62}$$

Note: In binary FSK the input and the output rate of change are equal. The rate of change at the input to the modulator is called bit rate (bits/sec),and the rate of change at the output of the modulator is called baud rate.

Constellation diagram: or signal space diagram contains points that displays the signal as a two-dimensional scatter diagram in the complex plane at symbol sampling instants. In a more abstract sense, it represents the possible symbols that may be selected by a given modulation scheme as points in the complex plane. Measured constellation diagrams can be used to recognize the type of interference and distortion in a signal.

Figure 5.39 shows the constellation diagram of BFSK where e is the energy if the signal Vfsk(t).

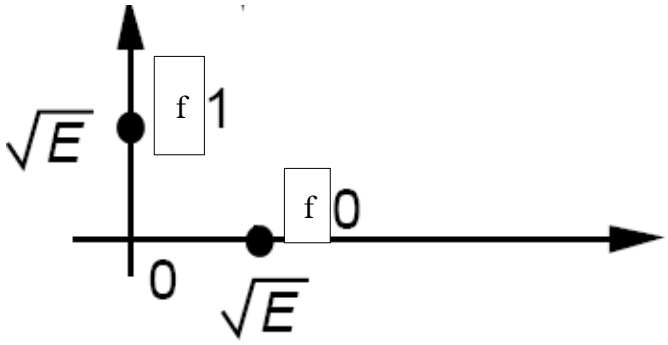


Figure 5.39 shows the constellation diagram of BFSK

Phase-shift keying (PSK): is a scheme in which the phase of the carrier signal is altered according to the digital data input. PSK uses a finite number of phases, each of which is assigned to a specific symbol. The general expression for a PSK modulated signal vpsk(t) is expressed as in equation (5.63) ,

$$Vpsk(t)=Ac\sqrt{(2/T)}\ [\cos(2\prod fc(t)+2\prod i/M)] \qquad (5.63)$$

where Ac is carrier signal amplitude,fc is carrier signal frequency,i is the variable varies from 0 to M-1 ,T is the symbol period and M is the number of phases used. The PSK modulated signal is shown in the figure,

Note:. Symbol rate is measured in symbols-per-second or baud rate (Bd). The symbol rate is the bit rate divided by the number of bits transmitted in each symbol The symbol rate is distinct from the bit rate because one symbol may carry more than one bit of information. For example with 4 phases, QPSK can encode two bits per symbol.

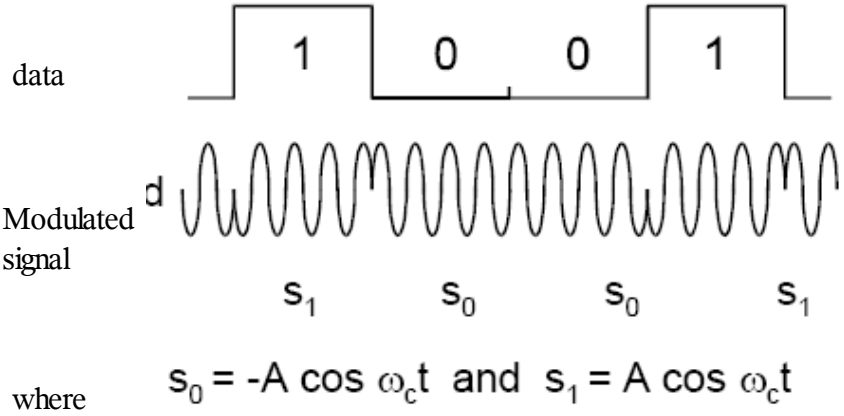


Figure 5.40 Modulated signal of PSK

The two common PSK methods are Binary phase-shift keying (BPSK) which uses two phases, and Quadrature phase-shift keying (QPSK) which uses four phases for modulation.

Although any number of phases may be used ,since the data to be conveyed are usually digital , the PSK scheme is usually designed with the number of constellation points being a power of 2.

BPSK*:* **(**Phase Reversal Keying) is the simplest form of PSK. It uses two phases which are separated by 180° and so can also be termed 2-PSK.It can modulate at 1 bit/symbol and so it is unsuitable for high data-rate applications when bandwidth is limited.

The baud rate(symbol rate) of BPSK is equal to the bit rate and is a measure of bandwidth of BPSK. The widest bandwidth occurs when the input binary data are an alternating 1/0 sequence.The fundamental frequency of an alternative 1/0 bit sequence is equal to one-half of the bit rate(fb/2).The minimum double sided nyquist bandwidth (B) can be defined in equation (5.64),

$$B = 2\ fb/2 = fb \tag{5.64}$$

Constellation diagram of BPSK is shown in figure:

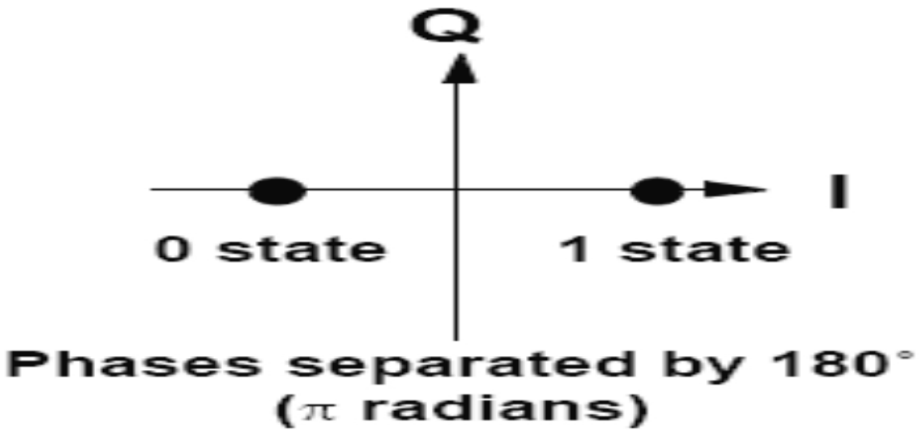


Figure 5.41 Constellation diagram of BPSK

The Probability of bit-error (Pe) of BPSK can be expressed in equation (5.65),

.

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \text{or } P_b = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E_b}{N_o}}\right)$$

(5.65)

where Eb = Energy-per-bit

N0 / 2 = Noise power spectral density (W/Hz)

QPSK: It uses four phases which are separated by 90°. It can modulate at the rate of 2 bits/symbol . The binary inputs are combined into two bits per symbol (00,01,10,11)and each symbol generates one of the four possible output phases .Therefore the rate of change of output(baud rate) is one-half of the input bit rate. The minimum bandwidth(f_N) of QPSK is fb/2 .The constellation diagram is shown in the figure5.42

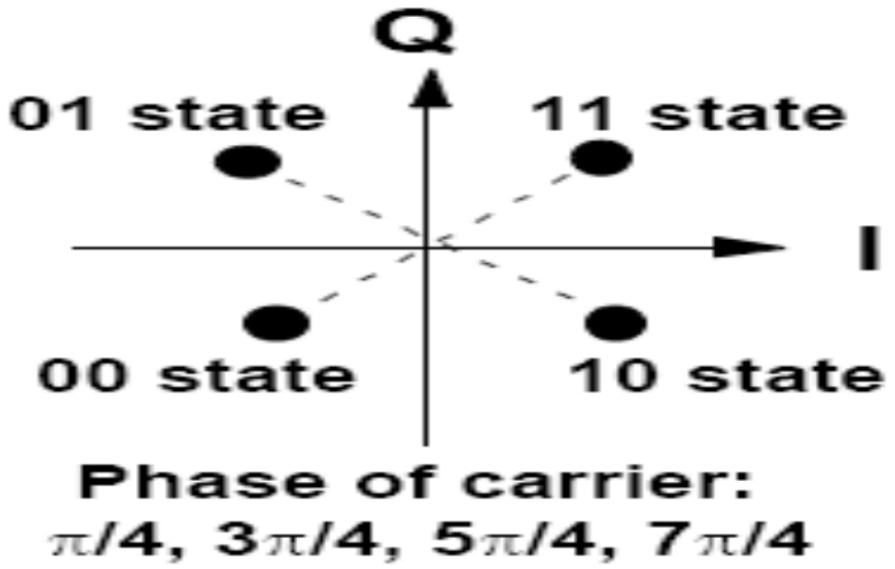


Figure 5.42 Constellation diagram

where I and Q are inphase and quadrature phase components.

The bit error rate (Pb) of QPSK can be defined in equation (5.66),

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \text{or } P_b = \frac{1}{2}\text{erfc}(\sqrt{\frac{E_b}{N_o}})$$

(5.66)

Comparison of digital modulation:

Table 5.4 shows that the comparison of Digital modulation Where fb is the bit rate.

Type	Bandwidth	Error performance
OOK	fb	erfc√(Eb/No)
BPSK	fb	erfc√(2Eb/No)
FSK	2 Δ f+fb	erfc√(Eb/No)
QPSK	fb/2	erfc√(2 Eb/No)
MSK	21.5 fb	erfc√(2Eb/No)

Table 5.4 comparison of Digital modulation