UNIT-1
Semiconductor PN Junction Diode

Structure of the Unit

1.0 Objectives

1.1 Introduction

1.2 p-n Junction Formation

1.3 Physics of pn junction

1.3.1 Band Diagram

1.3.2 Built in Potential

1.3.3 Electric Field

1.3.4 Electrostatic Potential

1.3.5 Potential Barrier

1.3.6 Width of the space charge region

1.4 Self learning exercise-I

1.5 Biasing of pn junction

1.6 pn junction diode equation

1.7 Self learning exercise-II

1.8 Summary

1.9 Glossary

1.10 Answers to self learning exercises

1.11 Exercise

1.12 Answers to exercise

References and Suggested Readings

1.0 Objectives

The chapter discusses the fundamental physical properties and its volt- ampere characteristics behavior of semiconductor p-n junction diode. We shall also discuss

in this chapter the physical principle and the operation of a pn junction.

The aim of this chapter is to enable the reader to understand the action of charge carrier mechanism in diode.

1.1 Introduction

The semiconductor p-n junction diode is the simplest element of solid-state electronics circuit and begins with it. Its simplicity beliefs its versatile importance in the solid state electronics.

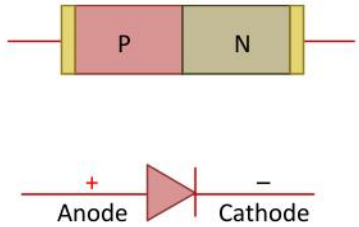


Fig1.1. A pn junction

The p-n junction diode is a two terminal electronic device and acts as switches which can current flow only in one direction. The p-and n-type regions are referred to as anode and cathode respectively. In above fig, arrowhead symbol shows the conventional direction of current flow when the pn – junction is forward-bias. It is the opposite direction in which electrons flow takes place or same direction along the holes flow.

A pn junction can be *step graded* or *linearly graded*. In a step graded (*abrupt*) pn junction, the accepter ion and donor ion concentration in the semiconductor are constant up to the junction. In a linearly graded junction, the impurities concentration varies almost linearly with distance from the junction.

1.2 PN Junction Formation

A pn junction is an electronic device which is formed simply by combining p-type (doped with B, Al) and n- type (doped with P, As, Sb) semiconductors together in close contact atomically. When the p- type material which Contains

NEGATIVELY charged acceptors (immovable[1]) and POSITIVELY charged holes (free) and the n- type material which Contains POSITIVELY charged donors

[1] The donors and acceptors ions are immovable (fixed),don't move (unless you heat up semiconductors, so they cannnot diffuse) because they are part of crystal.

(immovable) and NEGATIVELY charged free electrons are brought to closed contact, due to concentration gradient across the junction majority carrier (holes and electrons) of each type would diffuse across the junction. This shell mean that being free particles, majority **electrons** start diffusing from n-type material into p-material where they become minority carrier and recombine with majority carriers holes. Similarly for **holes,** too**,** start diffusing from p-type material into n-material in opposite direction.

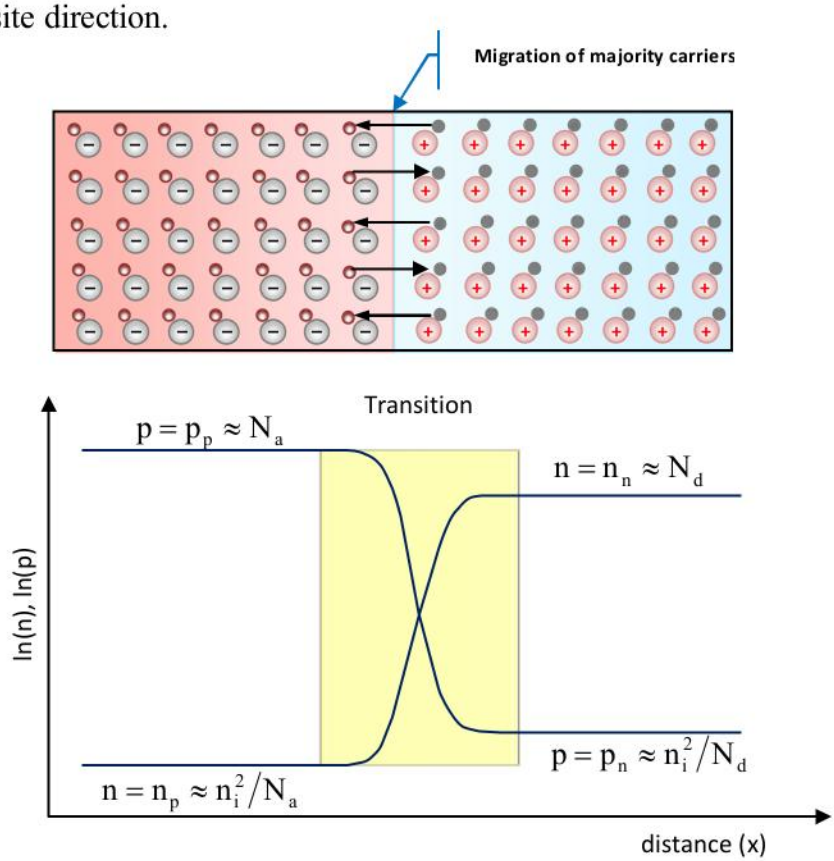


Fig1.2: Diffusion of majority carriers in pn junction

where

$n = n_n \approx N_d$: Concentrations of electrons on n side (majority carriers)

$n_p \approx n_i^2/N_a$: Concentrations of electrons on p side (minority carriers)

$p = p_p \approx N_a$: Concentrations of holes on p side (majority carriers)

$p_n \approx n_i^2/N_d$: Concentrations of holes on n side (minority carriers)

and

$$n_n.p_n = n_i^2 \rightarrow n-\text{type material}$$

$$n_p.p_p = n_i^2 \rightarrow p-\text{type material}$$

In this process, the electrons and holes recombine with each other. As a result, the negative charge (immovable accepter ions) on p- side and positive charge (immovable donor ions) on n- side create across the junction. The un-neutralized ions in the vicinity of the junction are termed as 'space charge regions' (uncovered charges region) and sometimes it is called as *potential barrier*. The uncovered charges generate an electric field barrier across the junction which is directed from n- side to the p- side. The barrier field opposes the further diffusion of electrons and holes through the junction. Sometime the electric field (barrier field) produced would drift the thermally- generated minority carriers in the opposite direction across the junction i.e., total current is sum of diffusion and drift current. Thus when *thermal equilibrium*[2] attained, the drift current flowing in one direction cancels out the diffusion current flowing in opposite direction, producing a net current of zero (see fig.). Mathematically, current density J due to various carriers can be written as

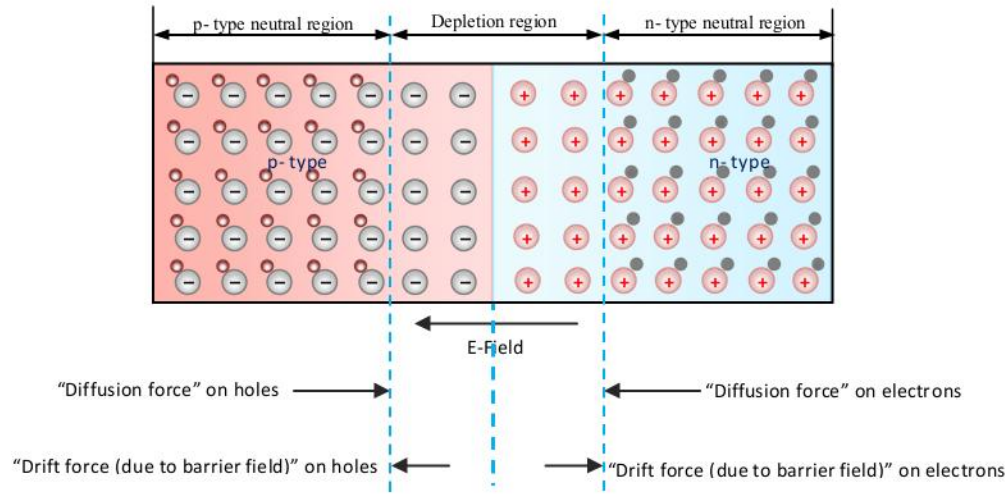


Fig1.3. The force on carriers at thermal equilibrium across a pn junction

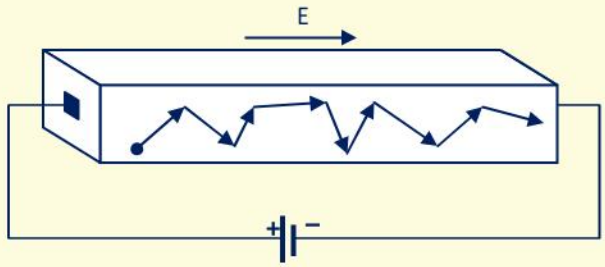
$$J_{holes} = J_{holes}(\text{diffusion}) + J_{holes}(\text{drift}) \qquad (1a)$$

$$J_{electrons} = J_{electrons}(\text{diffusion}) + J_{electrons}(\text{drift}) \qquad (1b)$$

[2] Thermal equilibrium means no extra heat, no applied voltages (zero bias), no light

(a) Drift current

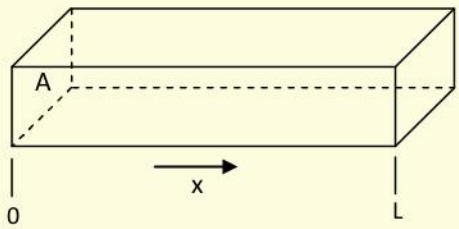
Drift current is due to electric field at the junction; minority carriers. In this process charge particles move under the electric force is called drift.

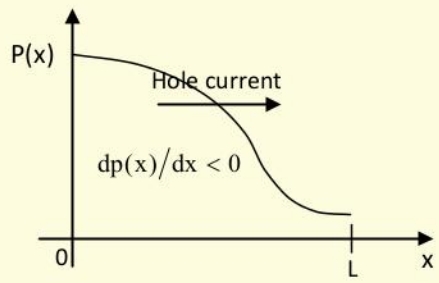


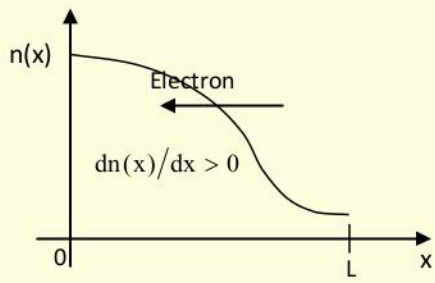
Thus $$J_{tot} = q(\mu_n n + \mu_p p)E$$

(b) Diffusion current

The diffusion current is due to concentration gradient of majority carriers by causing s net transport of carriers from high concentration to regions of lower concentration.







$$J_n(\text{diff}) = qD_n \frac{dn(x)}{dx} \qquad J_p(\text{diff}) = -qD_p \frac{dp(x)}{dx}$$

where D_p and D_n are hole and electron diffusion coefficients

1.3 Physics of PN Junction

1.3.1. Band Diagram

When the P-type material is joined with the N-type material, under thermal equilibrium condition, the Fermi level attains a constant level throughout the diode system, as shown in figure below.

In this situation, the conduction band edge E_{cp} and valence band edge E_{vp} of the p-side will be at higher than the conduction band edge E_{cn} and valence band edge E_{vn} of the n- side respectively. The barrier height is

$$E_B = E_{cp} - E_{cn} = E_{vp} - E_{vn}$$

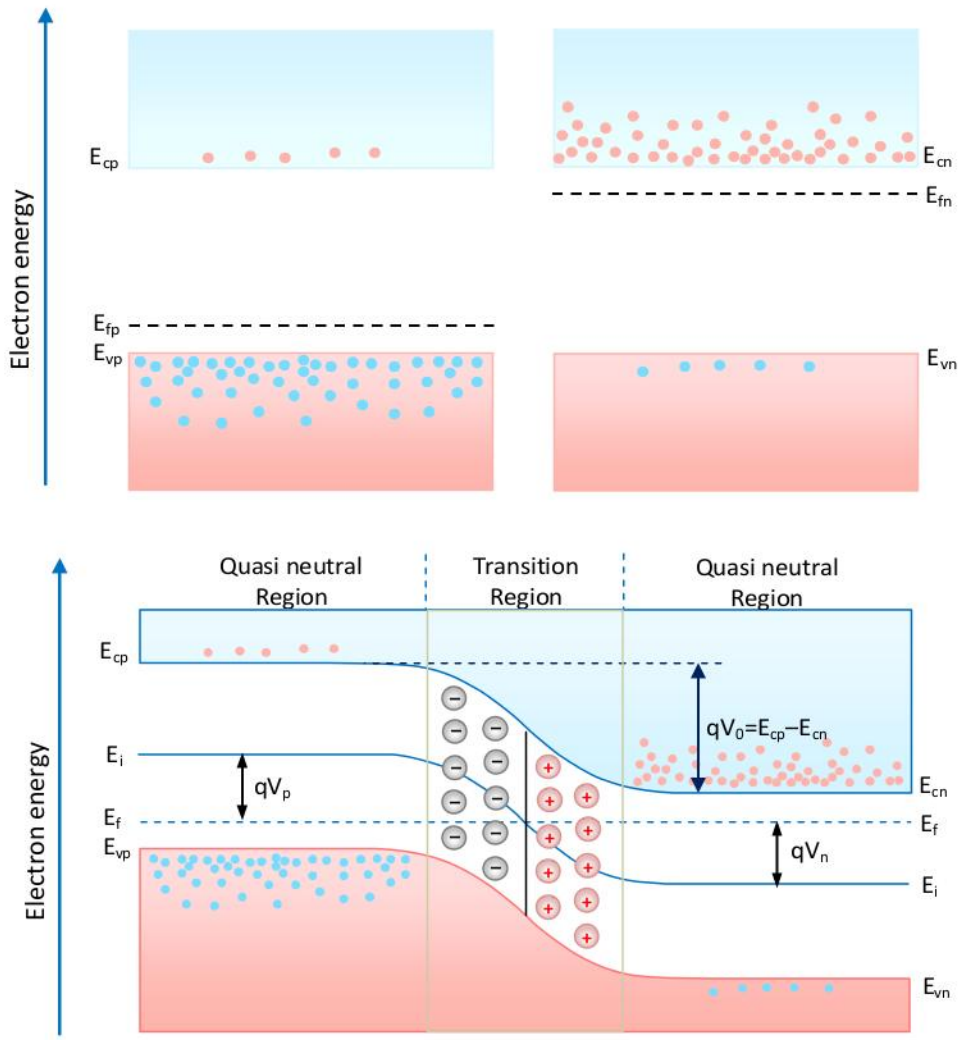


Fig1.3.1. Energy band- diagram at thermal equilibrium (zero bias)

At the thermal equilibrium (zero bias) barrier field stops the diffusion of electrons and holes. At this situation, there is no movement of charge carriers across the junction i.e., the drift current flowing in one direction cancels out the diffusion current flowing in the opposite direction, producing a net current of zero.

$$\left.\begin{aligned} J_{holes} &= J_{holes}(\text{diffusion}) + J_{holes}(\text{drift}) = 0 \\ J_{electrons} &= J_{electrons}(\text{diffusion}) + J_{electrons}(\text{drift}) = 0 \end{aligned}\right\} \quad (1)$$

$$J_p = q\mu_p p E_x - qD_p \frac{dp}{dx} = 0 \quad (2)$$

where

$$E_x = \frac{1}{q}\frac{dE_i}{dx} \qquad \text{and} \qquad D_p = \frac{\mu_p kT}{q} \quad (\text{Einstein's relation})^3$$

putting these values in eqn. (2), we get

$$J_p = \mu_p\left(p\frac{dE_i}{dx} - kT\frac{dp}{dx}\right) = 0 \quad (3)$$

but we know that $p = n_i \exp\left(\frac{E_i - E_f}{kT}\right)$

from which $\frac{dp}{dx} = \frac{p}{kT}\left(\frac{dE_i}{dx} - \frac{dE_f}{dx}\right)$

Finally we get

$$J_p = \mu_p p\left(\frac{dE_f}{dx}\right) = 0 \qquad \text{or} \qquad E_f = \text{constant}$$

which states that *Fermi level is a constant through the diode system* at the thermal equilibrium.

1.3.2. Built in Potential

An electric field is produced which results in a voltage drop across the depletion region is called the barrier voltage or built-in potential (V_{bi}) and also as the junction potential. built in voltage depends on the doping. once again from eqn (2)

$$J_p = q\mu_p p E_x - qD_p \frac{dp}{dx} = 0 \text{ (at thermal equilibrium)}$$

---

[3] Einstein's relationship is the mysterious link between carrier mobility (drift current) and diffusivity (diffusion current) $\frac{D}{\mu} = \frac{kT}{q}$

$$\frac{\mu_p}{D_p} E_x = \frac{1}{p_x}\frac{dp_x}{dx} \text{ (along x-axis)}$$

since the electric field $E_x = -\frac{dv_x}{dx}$ and using Einstein's relationship $\frac{D_p}{\mu_p} = \frac{kT}{q}$

thus
$$-\frac{q}{kT}\frac{dv_x}{dx} = \frac{1}{p_x}\frac{dp_x}{dx}$$

$$-\frac{q}{kT}\int_{V_p}^{V_n} dv = \int_{p_p}^{p_n}\frac{dp}{p}$$

$$-\frac{q}{kT}(v_n - v_p) = \ln p_n - \ln p_p \Rightarrow -\frac{qV_{bi}}{kT} = \ln\frac{p_n}{p_p}$$

It will turn out to be useful to note that

$$\boxed{\frac{p_p}{p_n} = \frac{n_n}{n_p} = e^{(qV_{bi}/kT)}}$$

we note that in particular $p_p \approx N_a$, and $p_n \approx n_i^2/N_d$. Then we can express the barrier voltage as

$$\boxed{V_{bi} = V_t \ln\frac{N_a N_d}{n_i^2}}$$

When external voltage is applied to the pn junction, the potential barrier is raised or lowered from V_{bi}, and the Fermi levels on either side are shifted by energy qV_{ext}.

1.3.3. Electric Field

Space charge density for a step (abrupt transition) pn junction[4]

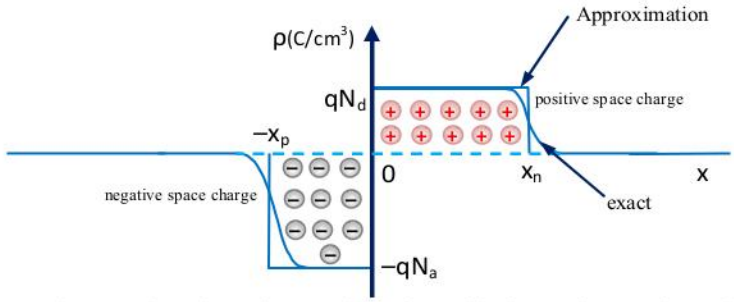


Fig1.3.2. Charge density of a pn junction diode at thermal equilibrium

[4] the charge distribution is very nice and abrupt changes occur at the depletion region (DR) edges. Such a junction is called as an *abrupt junction*.

$$\rho(x)=\begin{cases}0; & x<-x_p & \text{neutral p – region}\\ -qN_a; & -x_p<x<0 & \text{Space charge region}\\ qN_d; & 0<x<x_n & \\ 0; & x>x_n & \text{neutral n – region}\end{cases}$$

The electric field can be determined by integrate the Poisson's equation

$$E(x_2)-E(x_1)=\frac{1}{\varepsilon_s}\int_{x_1}^{x_2}\rho(x)dx$$

Therefore,

$x<-x_p$ $\qquad E(x)=0$

$-x_p<x<0$ $\qquad E(x)-E(-x_p)=\frac{1}{\varepsilon_s}\int_{-x_p}^{x}-qN_a\,dx$

$$E(x)=\left[-\frac{qN_a}{\varepsilon_s}x\right]_{-x_p}^{x}=-\frac{qN_a}{\varepsilon_s}(x+x_p)$$

$0<x<x_n$ $\qquad E(x)=-\frac{qN_d}{\varepsilon_s}(x_n-x)$

$x>x_n$ $\qquad E(x)=0$

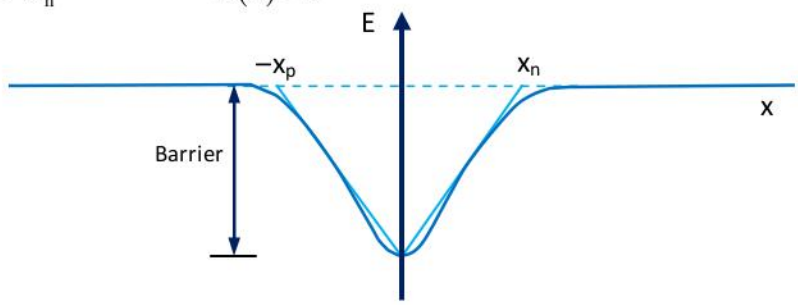


Fig.1.3.3. Electric field across the pn junction diode

Thus the electric field is negative throughout the junction region and varies linearly with x, reaching a maximum at x = 0 as shown in Fig. The electric field must also be continuous at the interface, x = 0, so that

$$N_a x_p = N_d x_n$$

This equation tells us that the width of the depletion region on either side of the junction varies inversely with doping concentration. In other words, the depletion region extends primarily into the side which has the lightest doping level. Note that equation is also another statement of *space-charge neutrality*.

1.3.4. Electrostatic Potential

The potential distribution in a p-n junction is obtained by integrating the electric field.

$$\phi(x_2) - \phi(x_1) = -\int_{x_1}^{x_2} E(x)dx$$

$x < -x_p$ $\qquad V(x) = V_p$

$-x_p < x < 0;$ $\qquad V(x) - V(-x_p) = \int_{-x_p}^{x} -\frac{qN_a}{\varepsilon_s}(x + x_p)dx$

$$= \frac{qN_a}{2\varepsilon_s}(x + x_p)^2$$

$$V(x) = V_p + \frac{qN_a}{2\varepsilon_s}(x + x_p)^2$$

$0 < x < x_n$ $\qquad V(x) = V_n - \frac{qN_d}{2\varepsilon_s}(x - x_n)^2$

$x > x_n$ $\qquad V(x) = V_n$

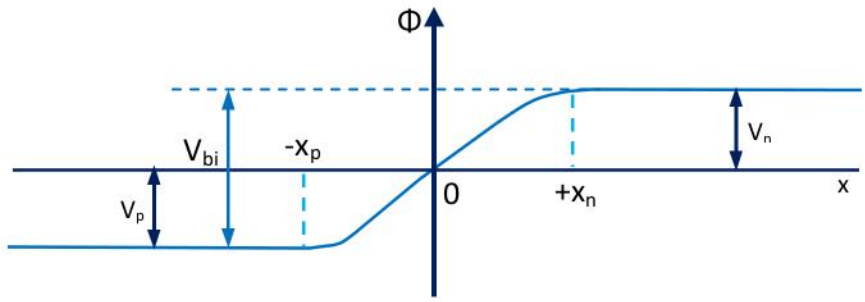


Fig.1.3.4. Electrostatic potential across the pn junction diode

1.3.5. Width of Potential Barrier

Now, by using the continuity of V(x) at x=0 in the above equations for the potential we get

$$V_p + \frac{qN_a}{2\varepsilon_s}(x_p)^2 = V_n - \frac{qN_d}{2\varepsilon_s}(x_n)^2$$

Therefore, $\qquad V_{bi} = V_n - V_n = \frac{q}{2\varepsilon_s}\left(N_d x_n^2 + N_a x_p^2\right)$

By using the space charge neutrality condition, one can get the potential widths as function of doping and built in potential.

$$x_n = \left[\frac{2\varepsilon_s V_{bi}}{q}\left\{\frac{N_a}{N_d(N_a+N_d)}\right\}\right]^{1/2} \qquad x_p = \left[\frac{2\varepsilon_s V_{bi}}{q}\left\{\frac{N_d}{N_a(N_a+N_d)}\right\}\right]^{1/2}$$

1.3.6. Width of the space charge region:

$$W = x_p + x_n = \left[\frac{2\varepsilon_s V_{bi}}{q}\left\{\frac{1}{N_a}+\frac{1}{N_d}\right\}\right]^{1/2}$$

Special Cases

(1) Symmetric junction for which $N_a = N_d$, therefore $x_p = x_n$ and

$$W = \left[\frac{\varepsilon_s V_{bi}}{qN_a}\right]^{1/2}$$

(2) Asymmetric junction for which $N_a > N_d$, therefore $x_p < x_n$ and

(3) Strongly asymmetric junction (p^+n junction) for which $N_a >> N_d$, therefore $x_p << x_n$ and

$$W \approx \left[\frac{2\varepsilon_s V_{bi}}{qN_d}\right]^{1/2}$$

When apply an external voltage across the pn junction then almost all of this voltage drops across the depletion region because the quasi-neutral p and n regions have low resistivity, whereas the depletion region has high resistivity (Think as voltage divider circuit).

1.4 Self Learning Exercise-I

Section A : Very Short Answer Type Questions

Q.1 What is a pn junction diode?

Q.2 Define built in potential.

Q.3 What are step graded and linearly graded pn junction?

Section B : Short Answer Type Questions

Q.4 What diffusion current.

Q.5 Show that Fermi level is constant through the diode system at thermal equilibrium.

Q.6 Describe the depletion region.

1.5 Biasing of PN junction

PN junction can basically work in two modes,

(1) *Forward bias mode* ($V_{applied} > 0$)

In this mode, a positive terminal of battery is connected to p-region and negative terminal is connected to n region (as shown in figure below). As applied voltage (V_{ext}) increases, the potential barrier to carrier diffusion is reduced ($V_{bi}–V_a$) by the applied voltage. Thus, more carriers will diffuse across the junction and large current (exponentially increases) flow through the junction. In this case the junction is said to be forward biased.

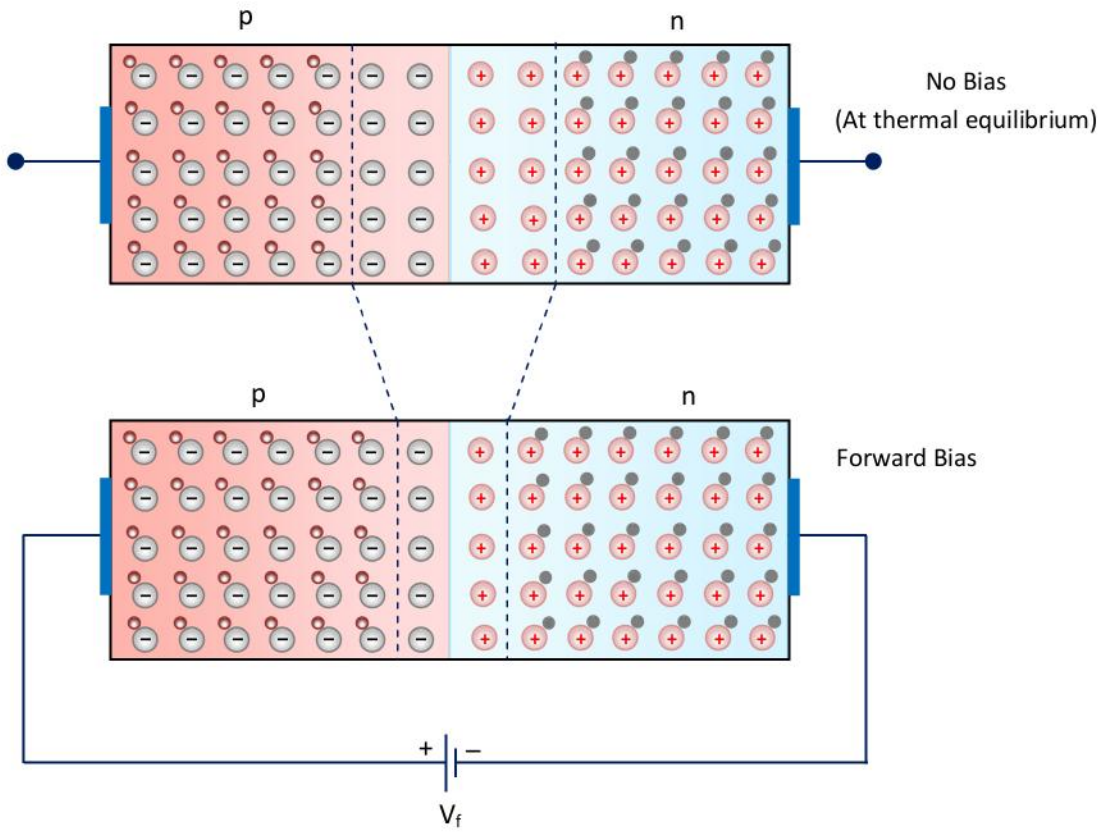


Fig.1.5. Forward bias of a pn junction diode

In process, the carriers that diffuse across the junction region become minority carriers in the quasi-neutral regions and then they recombine with majority carriers, "dying out" with distance. Hence, the width of the depletion region decreases.

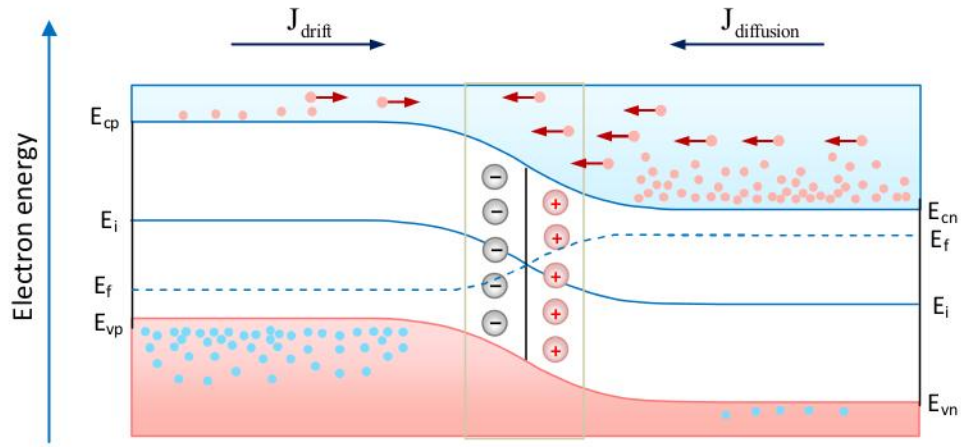


Fig.1.5.1. Energy band diagram of a pn junction under forward bias condition

(2) *Reverse bias mode* ($V_{applied} < 0$)

- If negative terminal of battery is connected to p-region and positive terminal connected to n region (as shown in figure below then, no carriers diffuse across the junction. In this mode depletion region width will increase by ($V_{bi}+V_a$) by the applied voltage and larger built-in electric field across the junction.

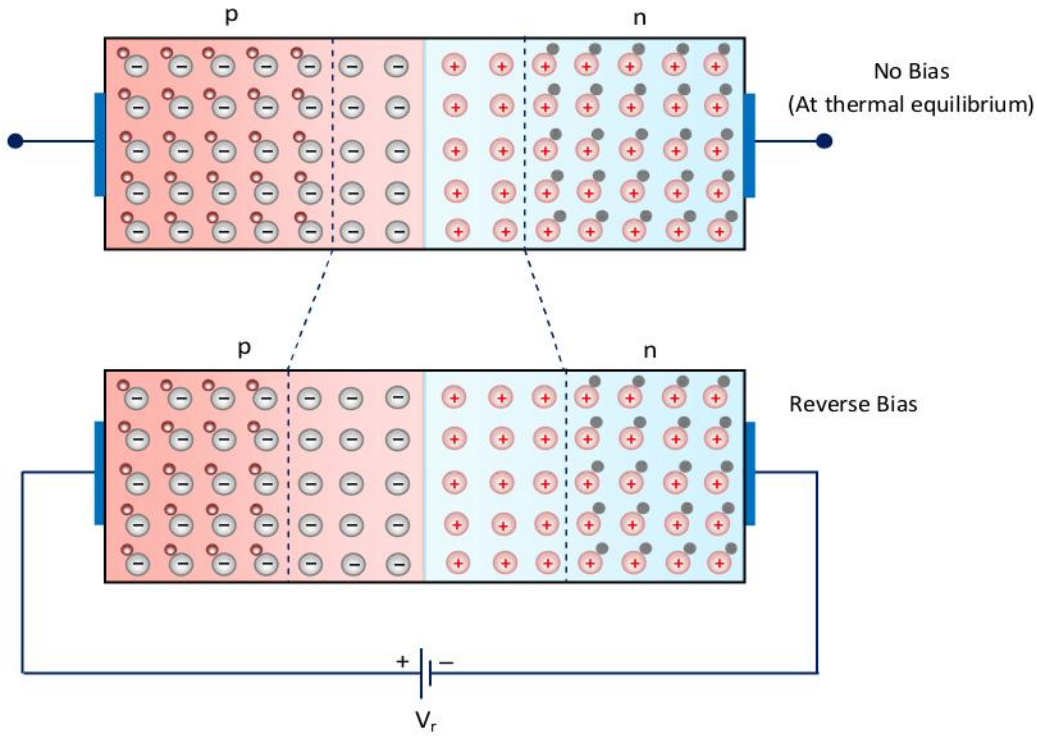


Fig.1.5.2. Backward bias of a pn junction diode

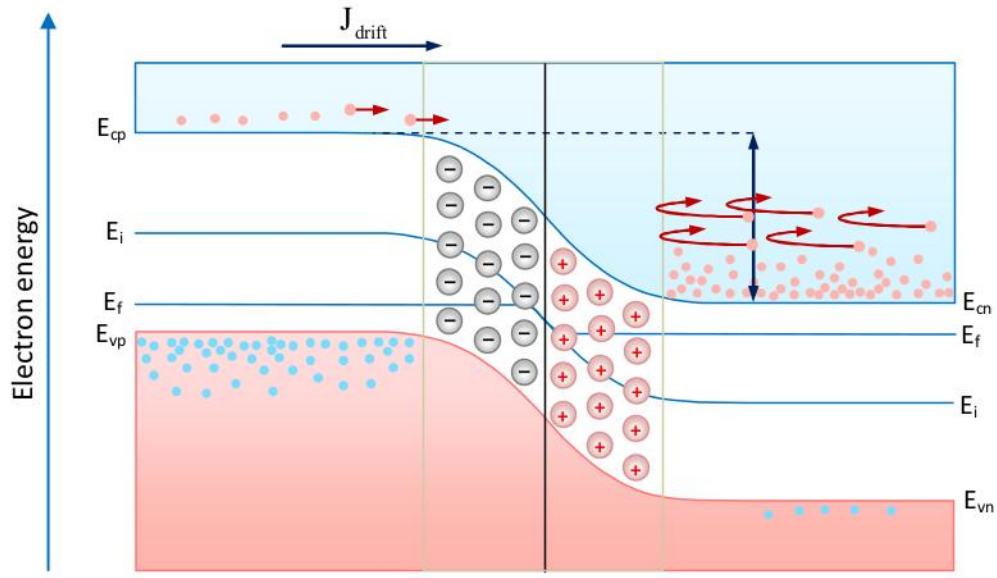


Fig.1.5.3. Energy band diagram of a pn junction under reverse bias condition

1.6 PN Junction Diode Equation

Under forward bias, the barrier to diffusion current is reduced and so more carriers diffuse across junction from where they become majority carrier to minority carrier .Therefore, diffusion currents increase to supply these minority carriers. This process is known as *carrier injection* from one side of the junction to the other side.

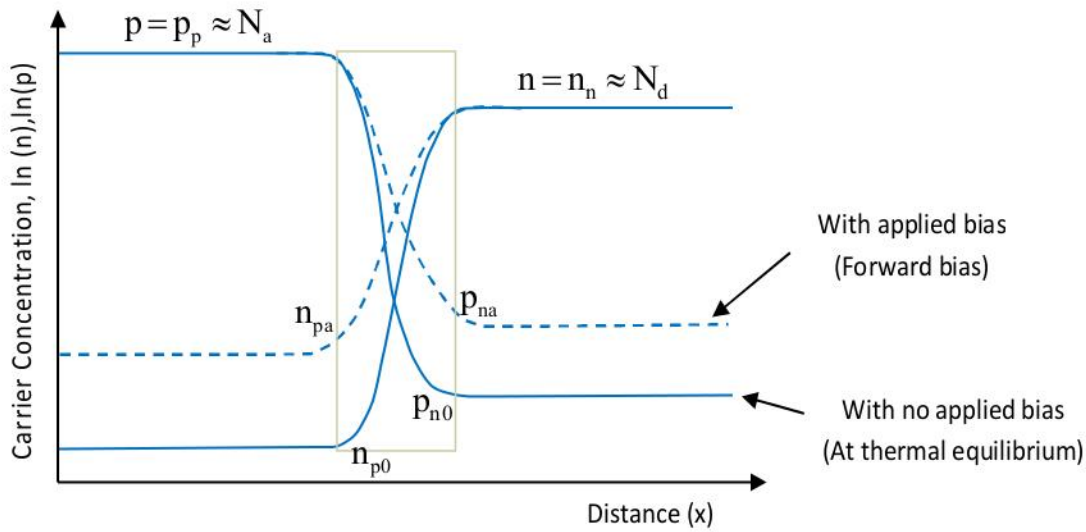
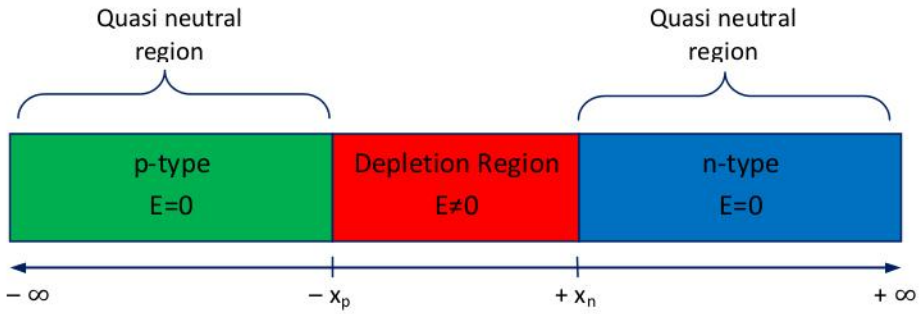


Fig.1.5.4. Carrier concentration under the zero bias and applied bias condition.

Quantitative Analysis



Boundary condition: $\Delta n_p\ (x \to -\infty) = 0$

Boundary condition: $\Delta n_p\ (x \to -x_p) = n_{pa}(-x_p)$

Boundary condition: $\Delta p_n\ (x \to x_n) = n_{p0}$

Boundary condition: $\Delta p_n\ (x \to \infty) = p_{n0}$

Following Assumptions are made for the quantitative analysis of pn junction diode.

(1) Steady state conditions:

(2) Non- degenerate doping:

(3) one- dimensional analysis:

(4) low- level injection: when the excess carrier concentration is small compared to the equilibrium majority carrier concentration.

(5) No generation, optical or otherwise ($G_L = 0$).

(6) In quasi-neutral regions charge density is zero, as is the electric field, this means the current flow is purely diffusive.

Therefore,

(a) Majority carrier concentration under equilibrium condition

$$p_{p0} = p_{n0}\, e^{(qV_{bi}/kT)} \qquad \text{and} \qquad n_{n0} = n_{p0}\, e^{(qV_{bi}/kT)} \qquad (1)$$

(b) In the presence of applied bias condition (non – equilibrium) same mathematics can be used to evaluate carrier concentration under the assumptions of quasi equilibrium.

- For forward bias

$$p_{pa} = p_{na}\, e^{[q(V_{bi}-V_a)/kT]} \qquad \text{and} \qquad n_{n0} = n_{na}\, e^{[q(V_{bi}-V_a)/kT]} \qquad (2)$$

- For reverse bias

$$p_{pa} = p_{na}\, e^{[q(V_{bi}+V_a)/kT]} \qquad \text{and} \qquad n_{n0} = n_{na}\, e^{[q(V_{bi}+V_a)/kT]} \tag{3}$$

The forward bias voltage causes excess[5] minority carriers to be injected across the junction. At low level injection[6]; the number of injected minority carriers is much small than the number of majority carriers. That is the injected or excess minority carriers do not upset the equilibrium majority carrier densities.

$$\left.\begin{aligned} p_{pa} &\approx p_{p0} \\ n_{na} &\approx n_{n0} \end{aligned}\right\} \tag{4}$$

Using Eqns. (1), (2), (3) and (4), we get the minority carrier concentration in the depletion region is given by

$$\left.\begin{aligned} p_{na}(x) &= p_{n0}\, e^{(qV_a/kT)} \\ n_{pa}(x) &= n_{p0}\, e^{(qV_a/kT)} \end{aligned}\right\} \tag{5}$$

where $n_{p0} \approx n_i^2/N_a$ and $p_{n0} \approx n_i^2/N_d$

Also

$$n_{n0}.p_{n0} = n_i^2 \rightarrow n-\text{type material}$$
$$n_{p0}.p_{p0} = n_i^2 \rightarrow p-\text{type material}$$

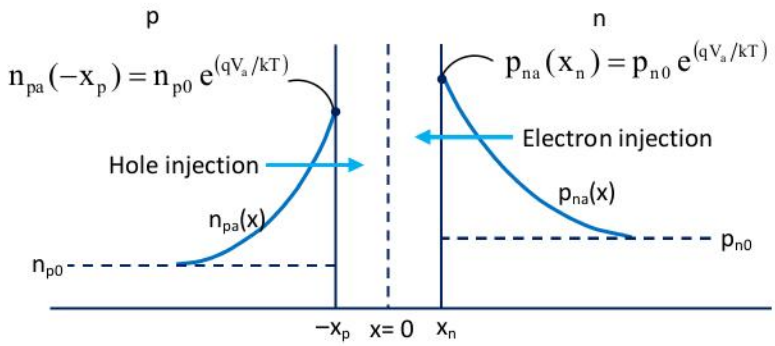


Fig.1.5.5. Excess minority carrier concentration in a pn junction under non equilibrium (applied bias) condition.

Now we can determine the *excess concentration* of minority carriers (holes and electrons) over the equilibrium concentration at the edge of the depletion region.

---

[5] The term excess carrier is used for carriers (electrons and holes) that are excess of their thermal equilibrium.

[6] When the excess (injected) carrier concentration is small compared to the equilibrium majority carrier concentration.

$$\left.\begin{aligned}\Delta p &= p_{na}(x) - p_{n0} \\ \Delta n &= n_{pa}(x) - n_{p0}\end{aligned}\right\} \tag{6}$$

Therefore,

$$\left.\begin{aligned}\Delta p &= p_{n0}\, e^{(qV_a/kT)} - p_{n0} = p_{n0}\left(e^{(qV_a/kT)} - 1\right) \\ \Delta n &= n_{p0}\, e^{(qV_a/kT)} - n_{p0} = n_{p0}\left(e^{(qV_a/kT)} - 1\right)\end{aligned}\right\} \tag{7}$$

Minority carrier distribution

The excess minority carrier hole concentration in the n region can be determined from the continuity equation

$$D_p \frac{d^2(\Delta p_n)}{dx^2} - \mu_p E \frac{d(\Delta p_n)}{dx} + G - \frac{\Delta p_n}{\tau_{p0}} = \frac{d(\Delta p_n)}{dt} \tag{8}$$

where $\Delta p = p_{na}(x_n) - p_{n0}$ is the excess minority hole carrier concentration.

In the n region (for $x>x_n$), we assume steady state $d(\Delta p_n)/dt = 0$, zero generation rate (G=0), and zero drift current. Then continuity equation reduces to

$$D_p \frac{d^2(\Delta p_n)}{dx^2} - \frac{\Delta p_n}{\tau_{p0}} = 0$$

$$\text{or} \qquad \frac{d^2(\Delta p_n)}{dx^2} = \frac{\Delta p_n}{L_p^2} \qquad (x > x_n) \tag{9}$$

where $L_p^2 = D_p \tau_{p0}$,is the hole diffusion length

Similarly, excess minority carrier electron concentration in the p region

$$\frac{d^2(\Delta n_p)}{dx^2} = \frac{\Delta n_p}{L_n^2} \qquad (x < x_p) \tag{10}$$

where $L_n^2 = D_n \tau_{n0}$,is the electron diffusion length

The general solution of the ordinary differential equation is:

$$\Delta p_n(x) = p_{na}(x) - p_{n0} = C_1 \exp\left(-\frac{x}{L_p}\right) + C_2 \exp\left(\frac{x}{L_p}\right) \tag{11}$$

The boundary conditions are

$$x = x_n \qquad p_{na}(x) = p_{n0} \exp\left(\frac{qV_a}{k_B T}\right)$$

$$\text{or} \qquad \Delta p(x) = p_{n0}\, e^{(qV_a/kT)} - p_{n0} = p_{n0}\left(e^{(qV_a/kT)} - 1\right)$$

$$x = \infty \qquad p_{na} = p_{n0} \qquad \text{or} \qquad \Delta p = 0 \tag{12}$$

From above two equation we get $C_2 = 0$ and $C_1 = p_{n0}\left[\exp\left(\frac{qV}{k_B T}\right)-1\right]\exp\left(\frac{x_n}{L_p}\right)$

Substituting the values of constants C_1 and C_2 we get

$$\Delta p_n(x) = p_{na}(x) - p_{n0} = p_{n0}\left[\exp\left(\frac{qV_a}{k_B T}\right)-1\right]\exp\left(\frac{x_n - x}{L_p}\right) \tag{13}$$

and for $(x \le -x_p)$

$$\Delta n_p(x) = n_{pa}(x) - n_{p0} = n_{p0}\left[\exp\left(\frac{qV_a}{k_B T}\right)-1\right]\exp\left(\frac{x_p + x}{L_n}\right) \tag{14}$$

The current density of minority carrier holes in n region along the x direction (at $x=x_n$) by diffusion is given by

$$J_p(x) = -qD_p\frac{d(\Delta p_n)}{dx} \tag{15}$$

Taking the derivative of equation (14) and substituting into equation (15), we get

$$J_p(x) = q\frac{D_p p_{n0}}{L_p}\left[\exp\left(\frac{qV_a}{k_B T}\right)-1\right]\exp\left(\frac{x_n - x}{L_p}\right) \tag{16}$$

The minority carrier hole current density at the edge of the transition region i.e, at $x = x_n$, from above equation can be obtained as:

$$J_p(x_n) = q\frac{D_p p_{n0}}{L_p}\left[\exp\left(\frac{qV_a}{k_B T}\right)-1\right] \tag{17}$$

Similarly, one can get the minority carrier electron current density at the edge of the transition region i.e, at $x = -x_p$, from above equation can be obtained as:

$$J_n(-x_p) = qD_n\frac{d(\Delta n_p)}{dx} \tag{18}$$

we obtain

$$J_n(-x_p) = q\frac{D_n n_{p0}}{L_n}\left[\exp\left(\frac{qV_a}{k_B T}\right)-1\right] \tag{19}$$

The total current density in the pn junction is then

$$J = J_p(x_n) + J_n(-x_p) = \left[q\frac{D_p p_{n0}}{L_p} + q\frac{D_n n_{p0}}{L_n}\right]\left[\exp\left(\frac{qV_a}{k_B T}\right)-1\right] \tag{20}$$

$$J = J_s\left[\exp\left(\frac{qV_a}{k_B T}\right)-1\right]$$

Multiplying by cross-sectional junction area A, then

$$I = AJ_s\left[\exp\left(\frac{qV_a}{k_BT}\right)-1\right] = I_s\left[\exp\left(\frac{qV_a}{k_BT}\right)-1\right]$$

where $$I_s \equiv AJ_s \equiv A\left[q\frac{D_p p_{n0}}{L_p} + q\frac{D_n n_{p0}}{L_n}\right] \qquad (21)$$

This equation is called as the ideal current- voltage relationship of a pn junction.This equation is valid for both forward and reverse biases; just change the sign of V.

In reverse bias, exponential term goes to zero as

$$I = I_s\left[\exp\left(\frac{-qV_a}{k_BT}\right)-1\right] = -I_s$$

The flow of minority carrier produces the reverse saturation current and these are thermally generated (highly temperature dependent). This current increases exponentially with temperature but it is independent of applied reverse voltage i.e., this current is constant as long as tmperature is constant.

This equation is plotted (I-V characteristics of the diode) as shown in figure below.

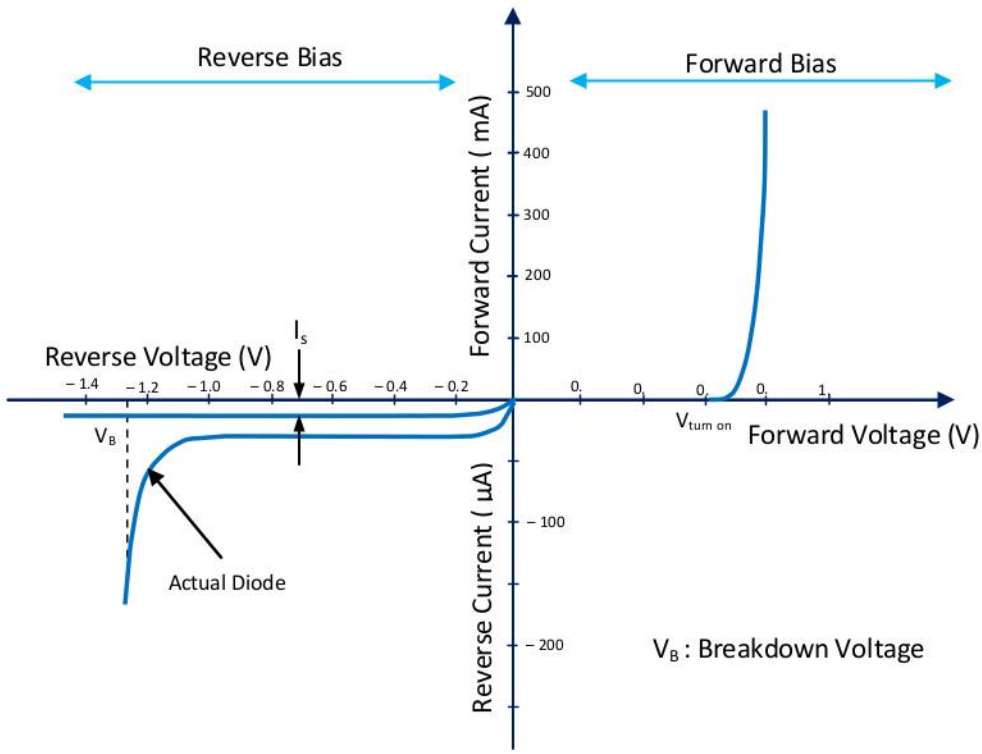


Fig.1.5.6. I-V characteristics of a pn junction diode

The current and voltage relationship of a PN junction is exponential in forward bias condition, and relatively constant in reverse bias condition.

1.7 Self learning Exercise-II

Section A :Very Short Answer Type Questions

Q.1 In which mode, diode is on.

Q.2 What is knee voltage?

Q.3 What is the range of forward bias and reverse bias current in junction diode under normal operation?

Section B : Short Answer Type Questions

Q.4 What is breakdown voltage?

Q.5 What is the knee voltage for silicon and germanium?

Q.6 Name the physical quantities used in diode current equation

$$I = I_s\left[\exp\left(\frac{qV_a}{k_BT}\right) - 1\right]$$

1.8 Summary

In this chapter we have discussed the one of the most important basic semiconductor structure- pn junction diode. We have addressed the physical properties and I-V characteristics behavior of the junction. These include the right from junction formation, barrier potential, electric field, width and its height to operation. We have also seen how the carrier concentration controllable by the external field.

1.9 Glossary

Step Graded pn junction: abrupt change in the accepter ion and donor ion concentration the junction.

Linearly Graded pn junction: In a linearly graded junction, the impurities concentration varies almost linearly with distance from the junction.

Space Charge Regions: The un-neutralized ions in the vicinity of the junction are termed as 'space charge regions' (uncovered charges region) .

1.10 Answers to Self Learning Exercises

Answers to Self Learning Exercise-I

Ans.1: A pn junction is an electronic device which is formed simply by combining p-type (doped with B, Al) and n- type (doped with P, As, Sb) semiconductors material together in close contact atomically.

Ans.2: At the thermal equilibrium (zero bias), an electric field is produced which results in a voltage drop across the depletion region is called the barrier voltage or built-in potential

Ans.3: An abrupt change in charge distribution occurs across the depletion region is called as an abrupt junction whence, in linearly graded junction doping concentration varies almost linearly (gradually) across the junction region.

Ans.4: The diffusion current is due to carrier concentration by causing carriers moves from high concentration to low concentration region.

Ans.5: See section 1.3.1.

Ans.6: Depletion region is a space charge region consisting of negative charge (immovable accepter ions) on p- side and positive charge (immovable donor ions) on n- side across the junction.

Answers to Self Learning Exercise -II

Ans.1: Forward bias condition.

Ans.2: It is defined as forward bias voltage at which the diode just starts conducting

Ans.3: The range of current in forward bias and reverse bias is milli and micro ampere respectively.

Ans.4: It is the reverse bias voltage at which the diode breaks down with rapid increase in reverse current.

Ans.5: 0.7 volts and 0.3volts.

Ans.6: I_s: reverse saturation current

V_a: applied voltage

K_B: Boltzmann constant

T: Temperature

1.11 Exercise

Section-A: Very Short Answer Type Questions

Q.1 In which mode a current flow in pn junction.

Q.2 Is diode Ohmic device?

Q.3 Can diode act like a switch? How?

Q.4 Is it possible to fabricate a diode by metal? Why?

Q.5 What indicate the arrow in the diode symbol?

Section-B: Short Answer Type Questions

Q.6 Draw the energy level diagram of a pn junction at thermal equilibrium (zero bias).

Q.7 Determine the value of thermal voltage.

Q.8 What is the effect of temperature on the reverse saturation current?

Q.9 What is the relationship between depletion width and the concentration of impurities?

Q.10 Define forward bias.

Section-C: Long Answer Type Questions

Q.11 Draw the I-V characteristics of the pn junction diode and explain.

Q.12 Derive the pn junction diode current equation.

Q.13 Explain the formation of a potential barrier in a p-n junction diode.

Q.14 Explain the all current components in a pn junction and how are they generated?

Q.15 Consider a silicon pn junction at T=300K with doping densities $N_d=10^{15}$ cm^{-3} and $N_a=10^{17}$ cm^{-3}. Calculate the built in potential for the pn junction .Assume that $n_i=1.5\times10^{10}$ cm^{-3}.

1.12 Answers to Exercise

Ans.1: Forward Bias

Ans.2: No, because its I-V characteristics is nonlinear.

Ans.3: Yes, diode can act like a switch. In diode, the current flows only in the forward bias condition and there is no current in reverse bias.

Ans.4: No.

Ans.5: Direction of conventional current.

Ans.6: See figure1.3.1.

Ans.7: $V_T = \frac{k_B T}{e} = \frac{(1.38\times10^{-23}\,J/K)(300K)}{(1.6\times10^{-19}C)}$

$= 0.0259volt \approx 0.026volt$

Ans.8: In the reverse bias condition, the reverse saturation current of Si and Ge diodes become doubles for every 10° C rise in temperature.

Ans.9: $W = x_p + x_n = \left[\frac{2\varepsilon_s V_{bi}}{q}\left\{\frac{1}{N_a}+\frac{1}{N_d}\right\}\right]^{1/2}$

Ans.10: A positive terminal of battery is connected to p-region and negative terminal is connected to n region. In this mode the junction is said to be forward biased.

Ans.11: See figure 1.5.6.

Ans.12: See section 1.6.

Ans.13: The negative charge (immovable accepter ions) on p- side and positive charge (immovable donor ions) on n- side create across the junction. The un-neutralized ions in the vicinity of the junction are termed as 'space charge regions' (uncovered charges region) and sometimes it is called as *potential barrier*.

Ans.14: Drift and diffusion current by carriers (electrons and holes).

Ans.15: The built in potential for the diode

$$V_{bi} = V_t \ln\frac{N_a N_d}{n_i^2}$$

$$V_{bi} = (0.026V)ln\left[\frac{10^{15}\times10^{17}}{(1.5\times10^{10})^2}\right]$$

$$= 0.026\times26.82 = 0.69Volt$$

References and Suggested Readings

1. Jasprit Singh, "Semiconductor Devices Basic Principles". John Wiley & Sons (Asia) Pvt. Ltd. (2001).
2. M.S. Tyagi, "Introduction to Semiconductor Materials and Devices". John Wiley & Sons (Asia) Pvt. Ltd. (2000).
3. S.M.Sze, "Physics of Semiconductor Devices 2nd Edition". John Wiley & Sons (Asia) Pvt. Ltd. (2002).
4. S.O.Kasap, "Principles of Electronic Materials and Devices SECOND EDITION". Tata Mc Graw – Hill NEW DELHI (2002).
5. Donald A Neamen, "Semiconductor Physics and Devices Basic Principles Third Edition". Tata Mc Graw – Hill NEW DELHI (2002).
6. Michael Shur, "Physics of Semiconductor Devices 2nd Edition". Prentice Hall of India New Delhi (2002).
7. Ben. G. Streetman, Sanjay Banerjee, "Solid State Electronic Devices First Edition". Prentice Hall of India New Delhi (2002).

UNIT-2
Applications of PN Junction Diodes

Structure of the Unit

2.0 Objective

2.1 Introduction

2.2 Avalanche and Zener breakdown

2.2.1 Avalanche Breakdown

2.2.2 Zener Breakdown

2.3 Self Learning Exercises-I

2.4 Junction and Diffusion capacitance

2.5 Diodes circuits: load line concept

2.6 Illustrative Examples

2.7 Clipping and Clamping Circuits

2.7.1 Series and Shunt clipper circuits (without threshold voltage)

2.7.2 Series and Shunt clipper circuits (with threshold voltage)

2.7.3 Negative and positive clamping (without bias)

2.7.4 Biased clamping Circuit

2.8 Self Learning Exercises-II

2.9 Summary

2.10 Glossary

2.11 Answers to self learning exercises

2.12 Exercise

2.13 Answers to exercise

References and Suggested Readings

2.0 Objectives

In previous chapter we studied in detail about the physics of pn junction diode and characteristics. in this chapter student will able to understand the breakdown

mechanism in ordinary diode and zener diode as well. The chapter has covered the application of diode as wave shaping circuit like clipping and clamping circuits. Topic also discusses the certain basic ideas about the normal operation of the diode with the help of load line analysis.

2.1 Introduction

In this section, the junction breakdown phenomena in a pn junction diode are discussed. The junction breakdown phenomena commonly observed in a p-n junction diode under large reverse bias.

If the reverse – bias voltage is raised above a threshold point, the pn junction will breakdown and a large current will start to flow rapidly. There are two physical mechanisms are responsible for this behavior; The Zener breakdown and avalanche breakdown.

2.2 Avalanche and Zener Breakdown

As studied previously, When the diode is reverse biased (i.e., n side more positive than p side) the value of reverse current suddenly rises, after some specified value of the reverse bias voltage. This situation is called *reverse breakdown*

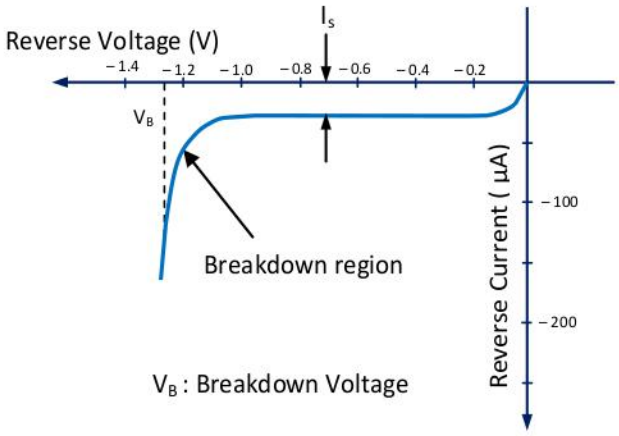


Fig.1 *Breakdown region in a pn junction diode*

In this situation the junction breaks and suddenly a very large current starts to flow. This breakdown mechanism can be understood by two different ways: the avalanche breakdown and the zener breakdown.

If the breakdown voltage V_{BR} is less the $4E_G/q$, which is near about to 5V for silicon diode, zener effect would occur. If the breakdown voltage V_{BR} is greater than $6E_G/q$, which is 7V for silicon diode, avalanche breakdown would occur.

2.2.1. Avalanche Breakdown or Impact Ionization

This occurs in ordinary (lightly doped diode, which has doping concentration less than $10^{15}cm^{-3}$) pn junction diode. *Avalanche breakdown* is a high-field effect that occurs when the *electrostatic field strength* associated with the p-n junction is strong enough to pull electrons out of the valence band within the depletion region.

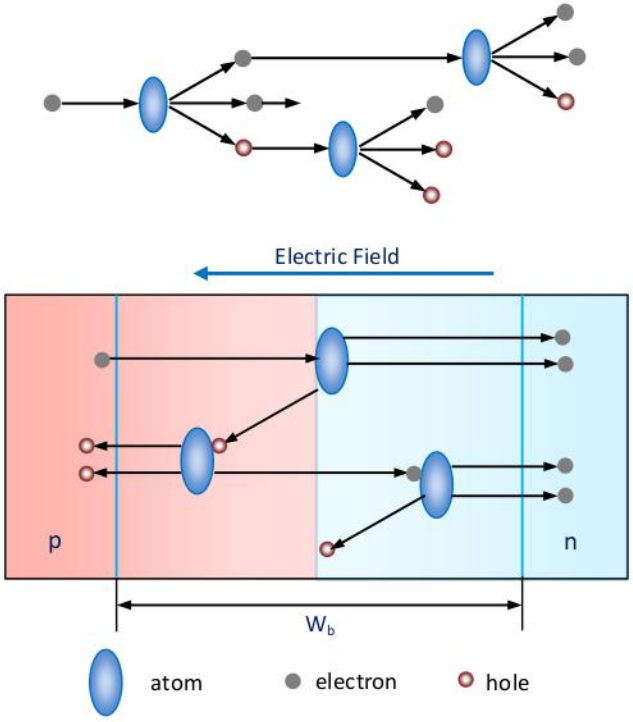


Fig.2 *Creation of electron – hole pair by impact ionization*

It is a result of electrons or holes colliding with the fixed ions inside the depletion region through the impact ionization[7].When applying a high reverse bias voltage to the diode, carriers gain kinetic energy and generate additional electron-hole pairs through impact ionization. This process is called avalanche multiplication. It is a Non-destructive but increases power consumption

Avalanche Breakdown Mechanism

In reverse biased condition, the barrier potential increases. As the result, lesser chance for majority carrier to diffuse across the junction and more easily for the minority carrier to drift across the junction would be happened.

If the reverse bias voltage is very high, then the depletion region widens, and the

[7] Impact Ionization: Electron-hole pairs are created on collision with immobile atoms.

electric field is quite strong. The minority charge carriers get accelerated in this depletion region, and gain kinetic energy. Once the field is strong enough (larger than a critical value E_{CR}), then the junction will "breakdown".

$$E_{CR} = \sqrt{\frac{2qN(V_{bi} + V_{BR})}{\varepsilon_s}}$$

They have enough energy to knock a valance electron free from its host atom during a collision (creating a hole and another conduction electron). The beauty is one energetic electron (or hole) produces two more charge carriers.

The field then separates the electron and hole of this newly created EHP and we now have three mobile carriers instead of one (the increase is exponential).

As a result, more carriers generated by this collision thereby initiating a chain reaction. This process is called avalanche multiplication.

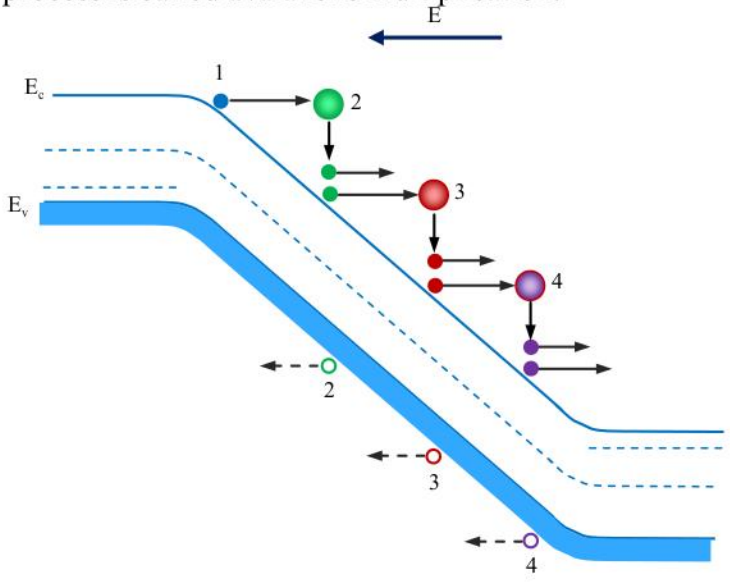
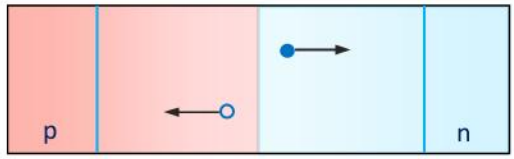


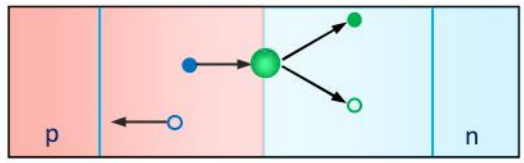
Fig.3 *Avalanche breakdown mechanism*

Once the process starts, the number of multiplication that can occur from a single collision increases rapidly with further increase in the reverse-bias, so the terminal current increases rapidly, and we say that the junction breaks down. This is called *avalanche breakdown.* The problem is, due to this process (Impact Ionization), heat is generated, and this heat could destroy the diode.

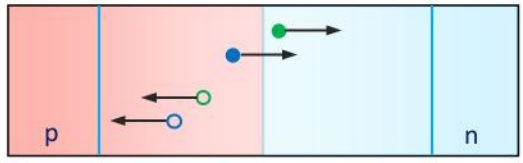
1. Minority carriers (Electrons from the p-side and holes from the n-side) get accelerated in the depletion regions by strong electric field.



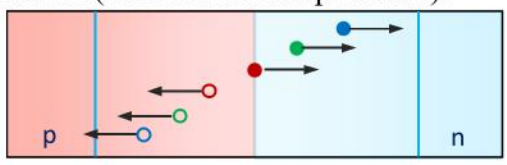
2. At collision by fixed atom, electron produces new electron – hole (e-h) pairs



3. These secondary e-h pairs acquire high energy and may in turn produce more e-h pairs. New created electron hole pairs will enhance the reverse current.



4. Similarly, new created electron-hole pairs will also create more electron hole pairs by impact ionization (avalanche multiplication).



The efficiency of the avalanche effect is characterized by a so-called multiplication factor M that depends on the reverse voltage. It can be determined by an empirical relation.

$$M = \frac{1}{1-(V/V_{br})^n}$$

where, n is between 3-6 depending on the material. Critical voltage for breakdown increases with bang gap. Peak field in depletion region increases with increased doping, thus V_{br} decreases as doping increases.

2.2.2 Zener Breakdown or Field Ionization

Zener breakdown is a type of reverse breakdown that occurs in a zener diode at low reverse voltage. The breakdown voltage is controlled by the doping level. The n-type and p-type materials of a zener diode are heavily doped (greater than $10^{17}cm^{-3}$.), resulting in a very narrow depletion region. Therefore, the resulting electric field at the junction imparts a very large force on a bound electron, enough

to pull from their valence bands (dislodge from its covalent bond). The breaking of the covalent bonds produces a large number of EHP (electron–hole pairs). Consequently a large reverse current produces at a low reverse voltage (V_R). This type of breakdown phenomena is known as *Zener breakdown.*

We recall that the width of the depletion region decreases as the doping concentration increases.

$$x_d = x_n + x_p = \left[\frac{2\varepsilon_s}{q}\left(\frac{1}{N_a}+\frac{1}{N_d}\right)V_{bi}\right]^{1/2}$$

Thus the depletion region can be narrow for high doping levels resulting in the energy bands across the junction region being bent more steeply. When the width of the depletion layer is small enough, tunneling of electrons from valence band to the conduction band can occur. This type of tunneling of electrons through the band gap is called Zener tunneling. Tunneling of electron leads a reverse current.

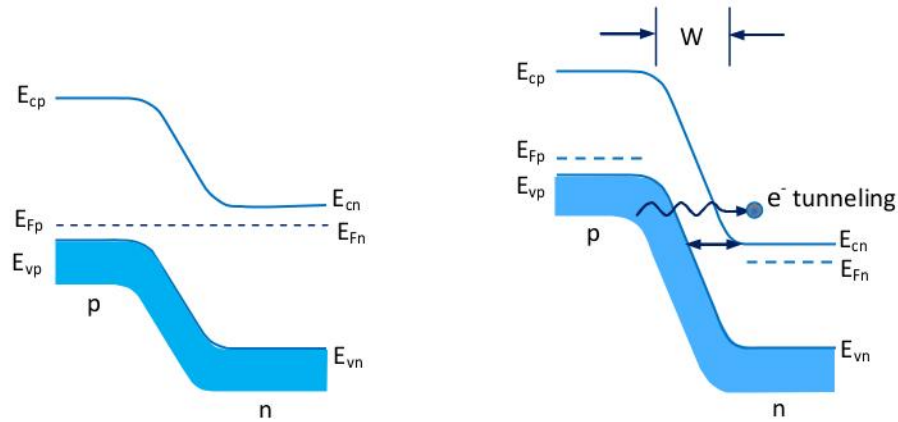


Fig.4 Zener breakdown tunneling

Comparison between Zener and avalanche breakdown

Avalanche breakdown (ABD)	Zener breakdown(ZBD)
1. Ordinary p-n junction (low doped $< 10^{15}cm^{-3}$) with sufficient reverse bias	1. Heavily – doped $> 10^{17}cm^{-3}$ p^{++}-n^{++} junction diode with reverse bias required
2. Dominant above ~ 6 volts	2. Dominant below ~ 6 volts
3. Impact ionization	3. Field ionization
4. Large depletion region width	4. Small depletion region width

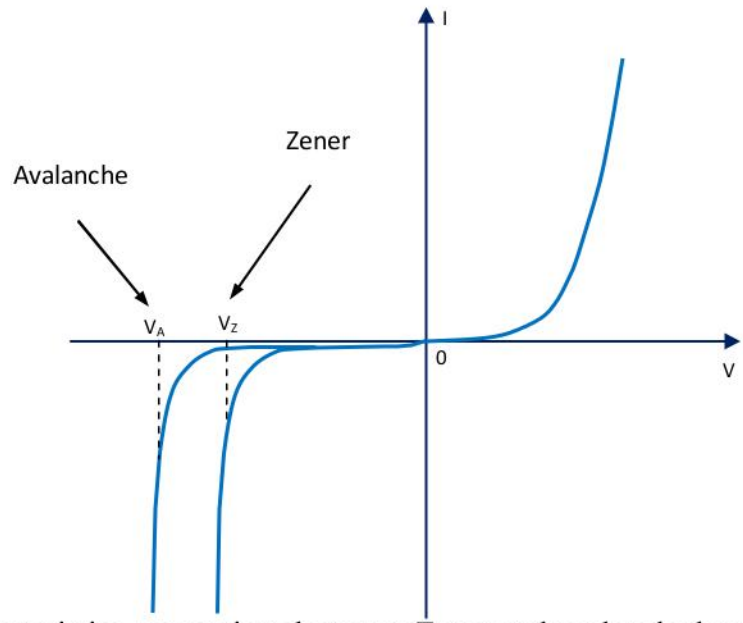


Fig.5 I–V characteristics comparison between Zener and avalanche breakdown

2.3 Self Learning Exercises-I

Q1 Why an ordinary diode suffers avalanche breakdown rather than zener breakdown?

Q2 Is the reverse breakdown means destruction of diode?

Q3 What do you mean by reverse breakdown in diode?

2.4 Junction and Diffusion Capacitance

Basically, there are two types of differential capacitance are associated with a p-n junction diode.

(a) Junction capacitance (depletion capacitance) and

(b) Diffusion capacitance (charge storage)

(1) Junction capacitance (C_j):

Junction capacitance is also called *transition region capacitance* or *depletion layer capacitance*. The junction capacitance is formed by the majority charges in the pn junction. The depletion capacitance can be treated as a plate capacitor, where the distance between the plates corresponds to the width of the depletion region. Only majority carrier contributes to the depletion capacitance.

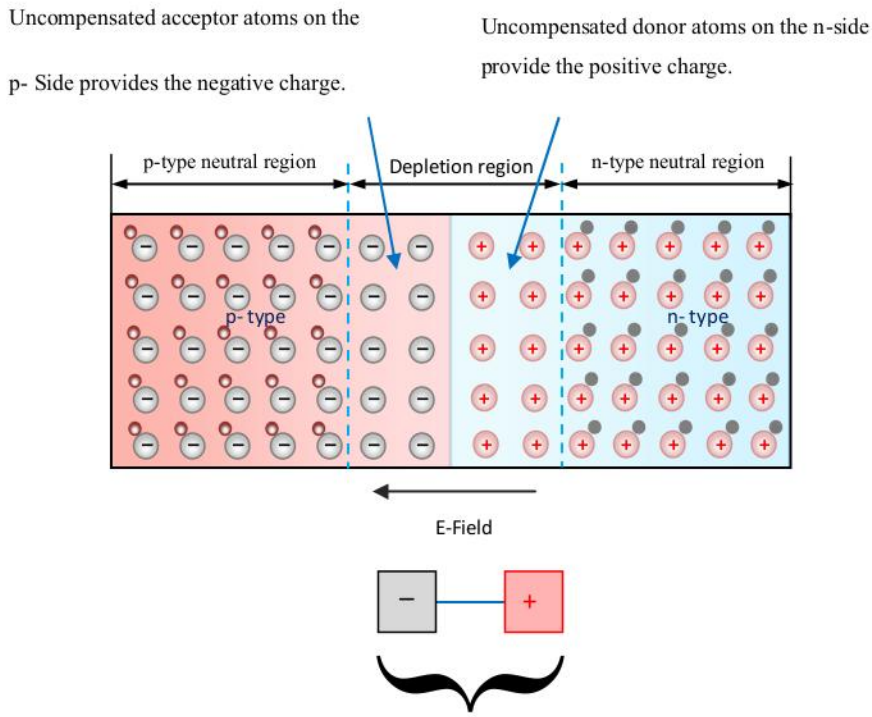


Fig.6 *Formation of charge dipole*

This capacitance dominates the under reverse bias condition. The depletion layer capacitance per unit area is defined as the ratio of the change in stored charge to the associated change in potential:

$$C_{dep} = \frac{dQ}{dV}$$

where dQ is incremental charge per unit area as for an incremental charge per unit voltage. The capacitance of symmetric and abrupt pn-junction can be calculated as

$$C_{dep} = A\frac{\varepsilon_s}{w} = A\frac{\varepsilon_s}{x_p + x_n}$$

where A is area of the diode, w is the width of the depletion region and ε_s is the dielectric constant of the semiconductor. The width of the depletion region is the sum of the depletion region in the n- and the p-region. The width of the depletion region can be obtained as

The width of the space charge region in the n-type region

$$x_n = \sqrt{\frac{2\varepsilon_s}{q}(V_{bi} - V)\frac{N_d}{N_a}\left(\frac{1}{N_a + N_d}\right)}$$

Similarly, for the p-type region

$$x_p = \sqrt{\frac{2\varepsilon_s}{q}(V_{bi} - V)\frac{N_a}{N_d}\left(\frac{1}{N_a + N_d}\right)}$$

Total width of space charge region $w = x_n + x_p$

$$w = \sqrt{\frac{2\varepsilon_s}{q}(V_{bi} - V)\left(\frac{N_a + N_d}{N_a N_d}\right)}$$

Now one can get the following expression of the depletion capacitance

$$C_{dep} = \left|\frac{dQ}{dV}\right| = A\sqrt{\frac{q\varepsilon_s N_a N_d}{2(N_a + N_d)(V_{bi} - V)}}$$

The depletion capacitance is usually written in the following form

$$C_{dep} = A\sqrt{\frac{q\varepsilon_s N_a N_d}{2(N_a + N_d)V_{bi}}}\frac{1}{\sqrt{1 - V/V_{bi}}} = \frac{C_j}{\sqrt{1 - V/V_{bi}}}$$

where C_j is the junction capacitor, which is dependent of the applied voltage. Therefore, the depletion capacitance of an abrupt pn-junction is proportional to $\frac{1}{\sqrt{1 - V/V_{bi}}}$.

(2) Diffusion Capacitance (Charge Storage Capacitance)

Diffusion capacitance is caused by the diffusion of carriers across the junction. It arises from the injected charge stored near the junction outside the space charge region (transition region) caused only by the minority carriers. By result that the voltage lagging behind the current due to this charge storage effects. It Dominants when the junction is forward biased. It is defined as the rate of change of injected charge with voltage called diffusion or charge storage capacitance (C_d).

In the figure7, Δn and Δp are the *excess concentration* of minority carriers (holes and electrons) over the equilibrium concentration at the edge of the depletion region. Q_n and Q_p are Total injected electrons and holes per unit area respectively.

The diffusion capacitance can be calculated by solving the Diffusion equation of excess charge by applying forward bias. The diffusion capacitance is caused by an increase of the minority carrier concentration in the neutral and the depletion region. The increase of the minority carrier concentration is caused by the applied

forward bias.

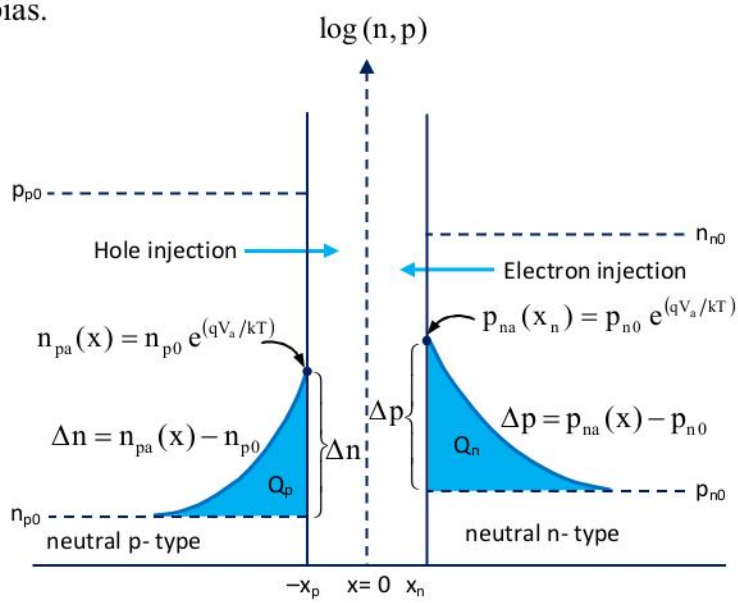


Figure 7

Forward biased junction stores an excessive charge ΔQ:

$$\Delta Q_p = q\,A\int_0^\infty \Delta p_n(x)\,dx = qAL_p p_{n0}(e^{qV/kT}-1)$$

$$\Delta Q_n = q\,A\int_0^\infty \Delta n_p(x)\,dx = qAL_p n_{p0}(e^{qV/kT}-1)$$

Therefore associated differential capacitance (diffusion capacitance)

$$C_{diffusion} = \frac{dQ_{stored}}{dV}$$

then $\dfrac{d\Delta Q_p}{dV} = qAp_{n0}L_p\dfrac{q}{kT}e^{qV/kT}$ and $\dfrac{d\Delta Q_n}{dV} = qAn_{p0}L_n\dfrac{q}{kT}e^{qV/kT}$

$$C_{diffusion}(\text{total}) = \frac{d\Delta Q_p}{dV} + \frac{d\Delta Q_n}{dV}$$

$$C_{diffusion} = qAp_{n0}L_p\frac{q}{kT}e^{qV/kT} + qAn_{p0}L_n\frac{q}{kT}e^{qV/kT}$$

Compare this to the diode current equation

$$I = I_p + I_n = qAp_{n0}\left(e^{qV/kT}-1\right)\frac{L_p}{\tau_p} + qAn_{p0}\left(e^{qV/kT}-1\right)\frac{L_n}{\tau_n}$$

therefore, assuming for simplicity $\tau_p \approx \tau_n$ and $V >> kT/q$

$$C_{diffusion} = I\tau\frac{q}{kT} = \frac{I\tau}{V_{Th}}$$

where, $V_{Th} = kT/q = 0.026$ Volt at 300 K

$$C_{diffusion} = \begin{cases} \frac{I\tau}{V_{Th}} & \text{Forward bias} \\ 0 & \text{Reverse bias} \end{cases}$$

In figure, pn junction capacitance as a function of applied voltage is shown as

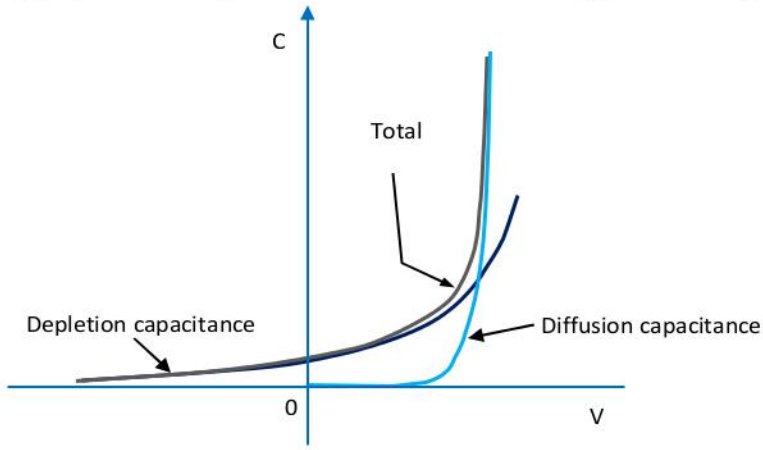


Fig.8 *pn junction capacitance vs. applied voltage*

2.5 Diodes Circuits : Load Line Concept

Load line (A straight line between two points) is a graphical approach to analyze the operational properties of a diode. It relates circuit parameters through standard linear analysis techniques. The load of a diode circuit determines the point or the region of operation of a diode (or device). A line is drawn on the characteristic of the device. The intersection point gives the point (called Q-point) of operation which decides the normal operation of device in a circuit. This kind of an analysis is known as the *load-line analysis.*

DC Analysis of Diode Circuits

Load line is a straight line curve between the current (I) and voltage drop across the diode of possible values of current flow in the circuit. The basic diode circuit consists of a voltage source in series with a resistor and a diode.

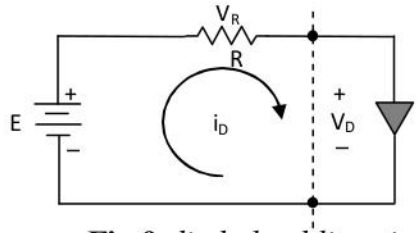


Fig.9 *diode load line circuit*

Applying Kirchhoff's voltage law (KVL) of circuit, we obtain $E - V_D - V_R = 0$

$$E = V_D + I_D R \quad (1)$$

from the characteristics equation of diode

$$I = I_s \left[\exp\left(\frac{V_D}{\eta V_T} \right) - 1 \right] \quad (2)$$

$$E = V_D + I_s \left[\exp\left(\frac{V_D}{\eta V_T} \right) - 1 \right] R$$

It is a *transcendental* equation for V_D. There is no simple analytical solution to this equation. Therefore graphical method (load line analysis) is an empirical solution method for solving the diode equation.

Rearranging the eqn. 1 in terms of I_D

$$I_D = \frac{E}{R} - \frac{V_D}{R} \quad (3)$$

which is an equation of straight line (y=b+mx). We call this straight line the load line.

If $V_D = 0$, we can calculate I_D and plot the magnitude of I_D on the vertical axis.

$$I_D = \left. \frac{E}{R} \right|_{V_D = 0}$$

If $I_D = 0$, we can calculate V_D and plot the magnitude of V_D on the horizontal axis.

$$V_D = E \big|_{I_D = 0}$$

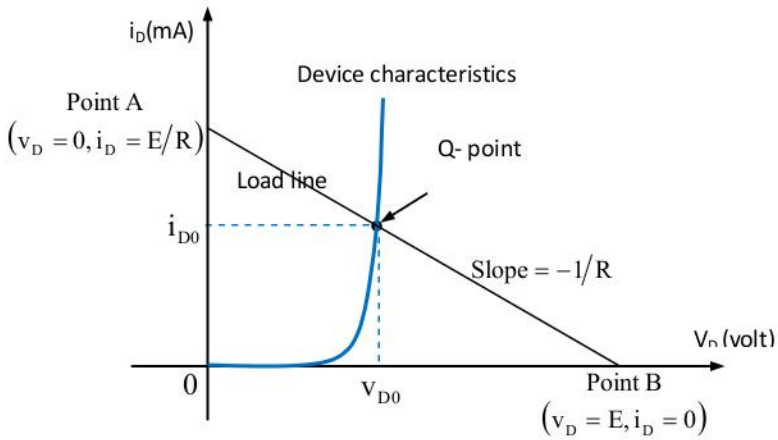


Fig.10 *Q-point on load line*

Intersecting point Q is simultaneous solution of eqn.1 and 2. This graphical method is an impractical solution method which helpful for a qualitative understanding of

these circuit. For example, what happens when

(a) E increases (voltage varies and the resistance is constant)

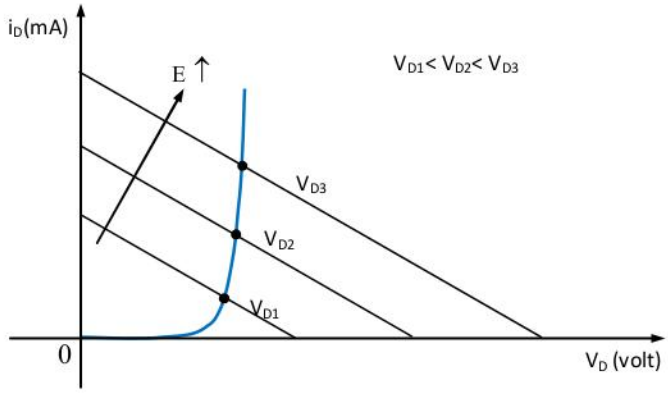


Fig.11 variation of Q-point with increasing source voltage

If the source voltage changes the diode current will also change. With R fixed the load line slope is unchanged but the V_D intercepts shifts along the axis.

(b) R increases (voltage is constant and resistance varies)

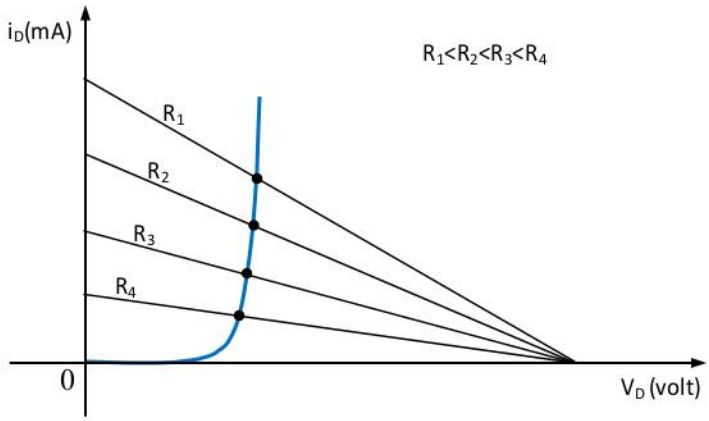
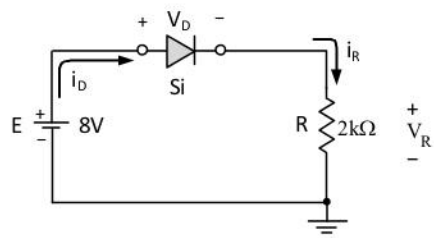


Fig.12 variation of Q-point with increasing R

2.6 Illustrative Examples

1. Plot the load line and find the Q-point for a diode with forward and reverse bias in the circuit. If V = 5 V and R = 10 kΩ. Use the i-v characteristic in Figure below (for silicon diode V_D=0.7V).



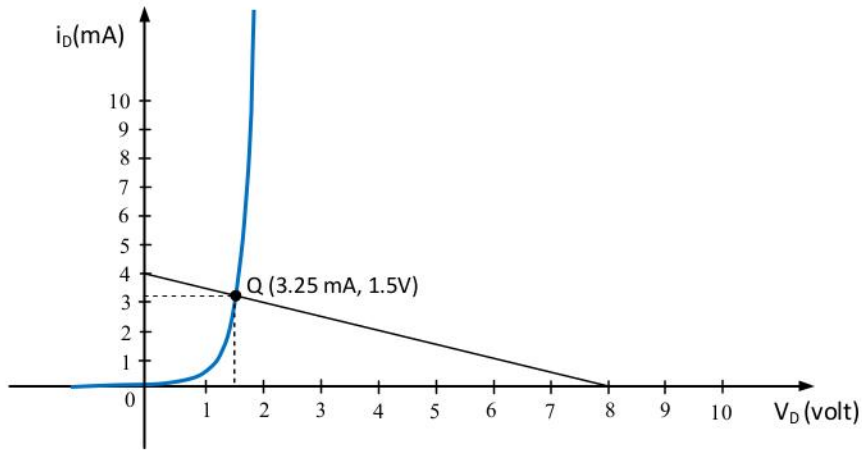
Sol. The applied voltage develops a current in the clockwise direction and diode is in the ON state.Applying Kirchhoff's voltage law (KVL) of circuit, we obtain $E = V_D + I_D R$

If $V_D = 0$, we can calculate I_D and plot the magnitude of I_D on the vertical axis.

$$I_D = \frac{E}{R}\Big|_{V_D=0} = \frac{8V}{2k\Omega} = 4mA$$

If $I_D = 0$, we can calculate V_D and plot the magnitude of V_D on the horizontal axis

$$V_D = E\big|_{I_D=0} = 8Volt$$



2.7 Clipping and Clamping Circuits

The circuit with which the waveform is shaped either by removing (or clipping) or shifting (or clamping) a certain portion of the applied wave are known as a wave shaped circuit.

The basic components required for these circuits are diode, resistor, capacitor and battery. These circuits are extensively used in radar, digital and other important electronic systems.

CLIPPERS

Clipping circuits (also known as limiters, wave shapers), are used to shape the waveform by removing (or clipping) a certain portion of the applied waveform that is above or below some specified reference level. Clipping circuits comprise of resistor, battery, and diode. In order to fix the clipping level to the desired amount, a dc battery must also be required in circuit. In this section, we shall focus only to diode clippers.

As we know, diodes can be used as switches depending on the biasing type, reverse of forward.



Fig. *on-off condition of diode*

In the forward direction, it behaves like a perfect conductor with zero voltage drop (short circuit) while in the reverse direction, it acts as open circuit.

Broadly, clippers are classified into two main categories:

(a) Series clipper circuits (b) Shunt clipper circuits

Further it can also be classified into:

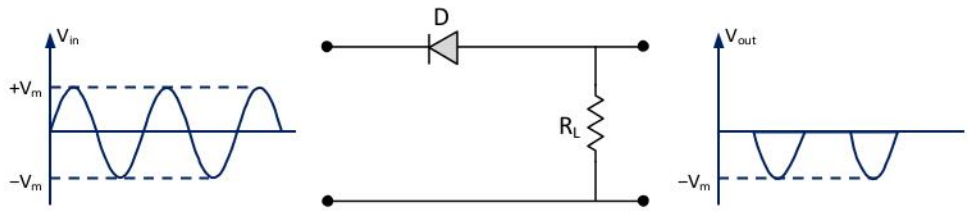
1. Positive and negative clippers.
2. Biased clippers.
3. Combinational clippers.

In series clippers diode is connected in series with the load while in parallel (shunt) clippers it is parallel as shown in fig.(x). Clipping the positive or negative portion of the waveform depends on *the orientation (i.e., forward or reverse bias) of the diode* and accordingly the diode clippers may be positive or negative.

2.7.1. Series and Shunt clipper circuits (without threshold voltage)

(a) Positive series clipping

As shown in the figure below, the diode is connected in series with the load. During the positive half cycle of the input waveform, the diode 'D' is reverse biased, which maintains the output voltage at 0 Volts. This causing the positive half cycle to be clipped off or remove. During the negative half cycle of the input, the diode is forward biased and so the negative half cycle appears across the output.



(b) Negative series clipping

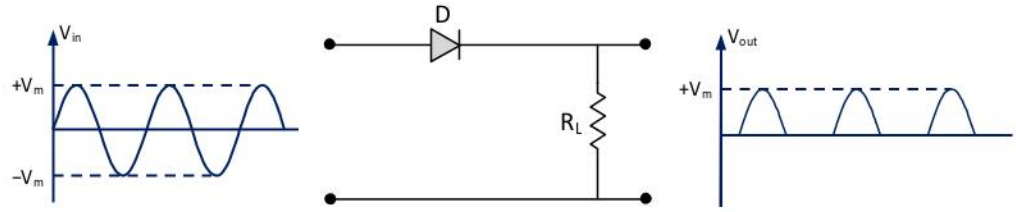
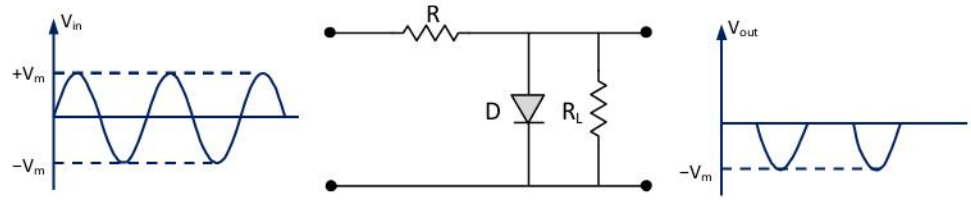


Fig. Zero threshold series clippers

(c) Positive shunt clipping

In Figure, the diode is connected in parallel with the load.



During the positive half cycle, the diode 'D' is forward biased and the diode acts as a closed switch. This causes the diode to conduct well. This causes the voltage drop across the diode or across the load resistance R_L to be zero. Thus output voltage during the positive half cycles is zero (i.e., clips the positive portion of the signal), as shown in the output waveform. During the negative half cycles of the input signal voltage, the diode D is reverse biased and behaves as an open switch. Consequently the entire input voltage appears across the diode or across the load resistance R_L.

(d) Negative shunt clipping

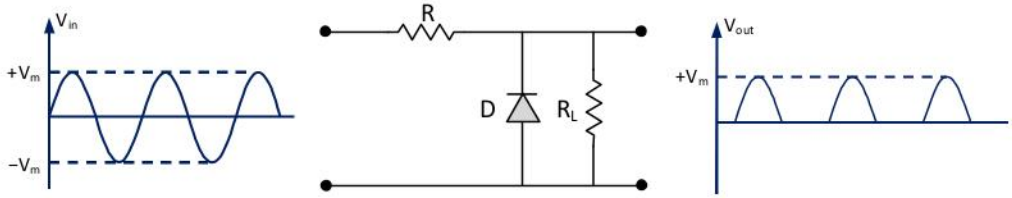


Fig. Zero threshold shunt clippers

2.7.2. Series and Shunt clipper circuits (with threshold voltage)

In the threshold series clipping, the voltage source of V_B (positive or negative) is applied between output terminals as in Fig. 2.4.Output is depending on the diode connection (forward or reverse), the values smaller (Fig.2.4.a) or greater (Fig.2.4.b) than V_B is clipped and assigned as V_B.

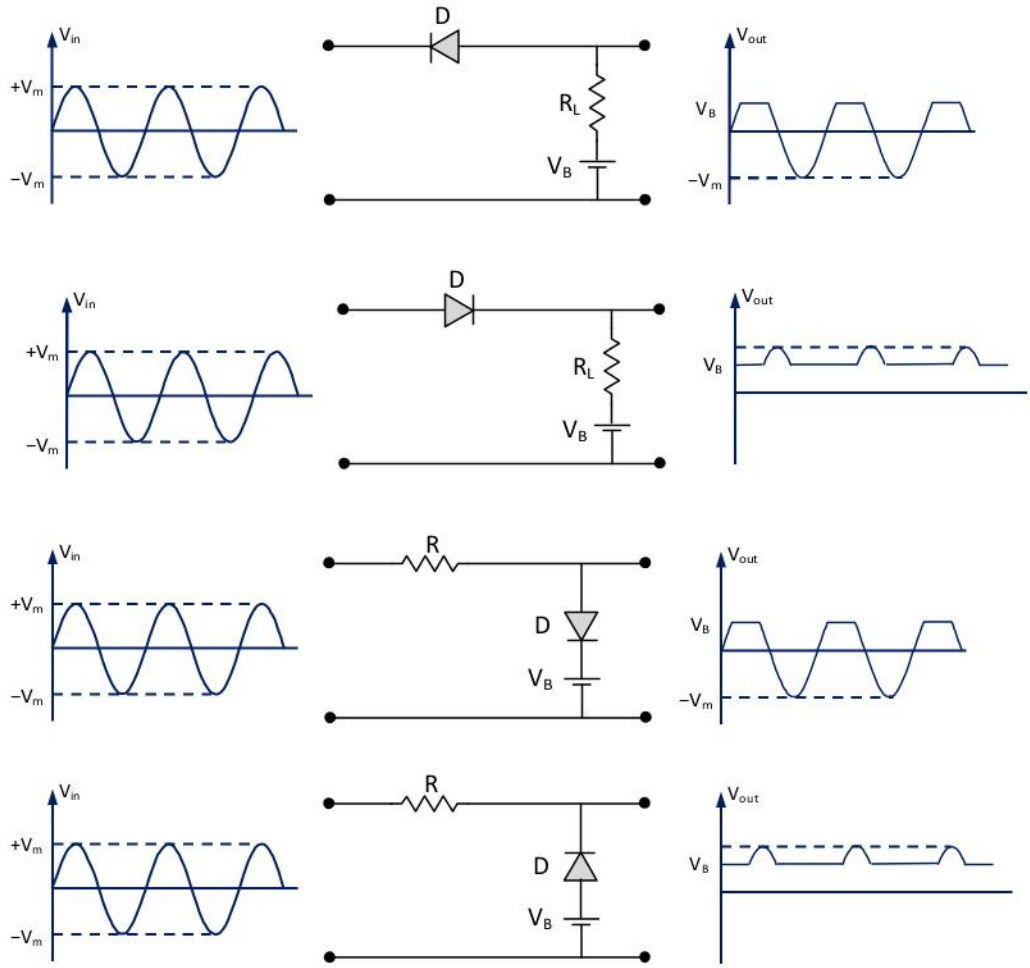


Fig. Series and Shunt clipper circuits (with threshold voltage)

CLAMPER CIRCUITS

A clamper is a network consisting of a diode, a resistor, and capacitor that shifts an input signal by a desired value without changing the appearance of the applied signal.

2.7.3 Negative and positive clamping without bias

(a) Negative Clamper

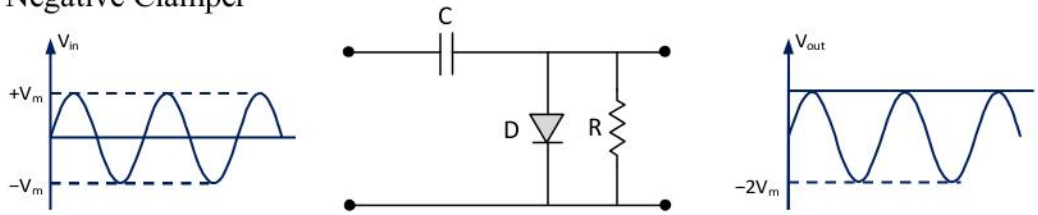


Fig. A negative clamping circuit

During the positive cycle of input signal, the diode is in forward bias mode. In this situation diode is short circuited and acts like a closed switch (i.e. "on" state). No voltage drop will be across the diode (V_0=0) because the current is shorted through the diode and capacitor will rapidly be charged up to voltage V_m due to small time constant.

During negative cycle of input signal, the diode is in reverse bias. In this situation diode is open circuited and acts like a open switch (i.e. "off" state). The voltage across R will be approximately

$$V_{out} = V_{in}(dc) + V_{cap} = -V_m + (-V_m) = -2V_m$$

(b) Positive clamper

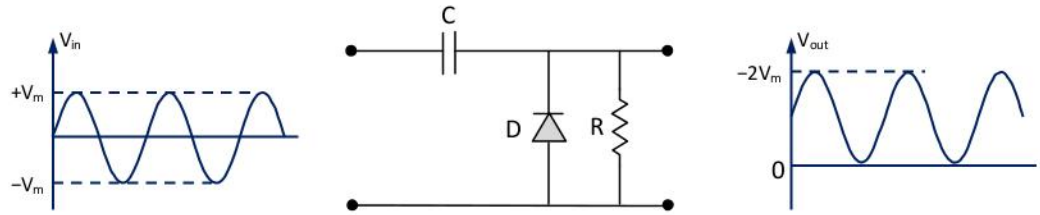


Fig. A positive clamping circuit

2.7.4. Biased clamping Circuit

A positive bias clamper circuit is shown in Fig. The bias voltage V_B is used to shift the negative clamped waveform to a voltage level V_B above the zero reference level.

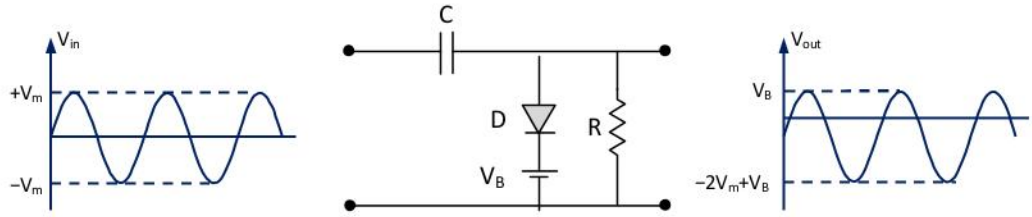


Fig. A positive bias clamping circuit

2.8 Self Learning Exercises-II

Q.1 Explain the concept of load line of diode.

Q.2 On what factor the Slope of load line depend?

Q.3 Define clipping circuits.

Q.4 Define clamping circuits.

2.9 Summary

Avalanche breakdown is a high-field effect that is result of electrons or holes colliding with the fixed ions inside the depletion region through the impact ionization.

Zener breakdown occurs in heavily doped ($> 10^{17}cm^{-3}$) junction diode with resulting in a very narrow depletion region and tunneling of electrons from valence band to the conduction band occurs. Tunneling of electron leads a reverse current.

The intersection of the load line and the diode graph gives the Q point. When the source voltages change, different load lines are obtained with different Q points.

The circuit with which the waveform is shaped either by removing (or clipping) or shifting (or clamping) a certain portion of the applied wave are known as a wave shaped circuit.

2.10 Glossary

Clamper: It is a network consisting of a diode, a resistor, and capacitor that shifts an input signal by a desired value without changing the appearance of the applied signal.

Depletion Region : region where electrons and holes recombime.

2.11 Answers to Self Learning Exercises

Answers to Self Learning Exercise -I

Ans.1: See avalanche and zener diode

Ans.2: No, the reverse breakdown does not mean the destruction of the diodes. Diodes can be operated in and out of the break-down region with no irreversible changes, until it is not exceeded the maximum power dissipation limit.

Ans.3: A reverse breakdown is a property of diode under which the voltage is almost constant with increasing the reverse current (i.e., voltage is nearly independent of the current). Zener diode is specially designed for this mode of operation.

Answers to Self Learning Exercise -II

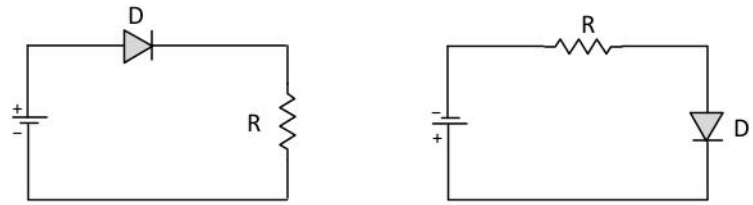
Ans.1: See load line

Ans.2: Resistor

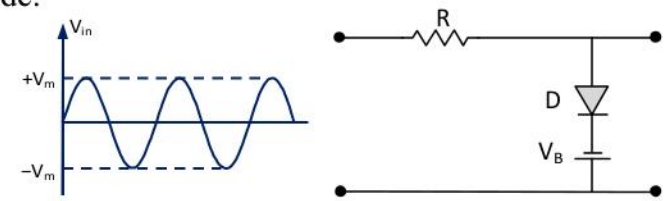
Ans.3: The circuit with which the waveform is shaped by removing (or clipping) a certain portion of the input wave is known as a *clipping circuit.*

2.12 Exercise

Q.1 Determine the diode states for the circuits shown below.



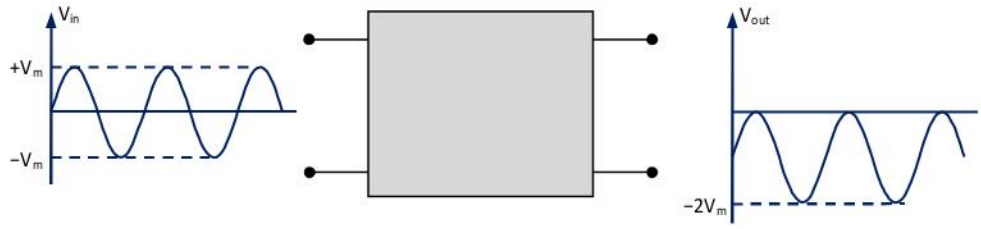
Q.2 Sketch the output waveform for this clipper circuit, assuming with an ideal diode.



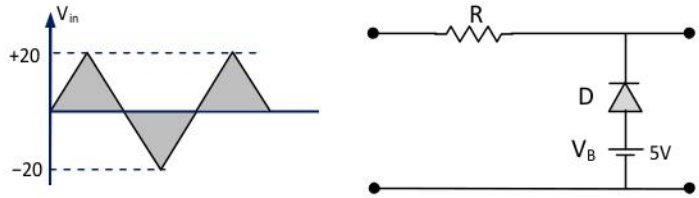
Q.3 Design a clipper circuit that removes the positive portion of the input waveform, leaving only the negative half-cycles to appear on the output:



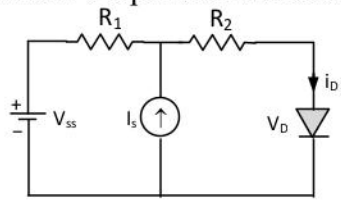
Q.4 Design a clamper circuit that biases the input waveform so it lies completely below (negative) the zero line:



Q.5 Determine the output waveform for the following circuit.



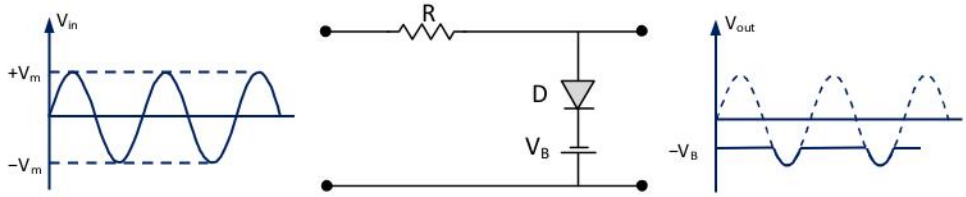
Q.6 Obtain the Thevenin's equivalent circuit for determining the load line analysis.



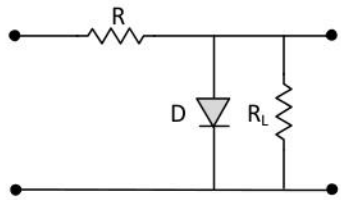
2.13 Answers to Exercise

Ans.1: (a) D is on (forward bias) (b) D is off (reverse bias).

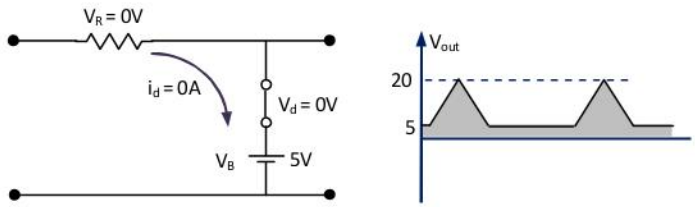
Ans.2.



Ans.3:



Ans.5:

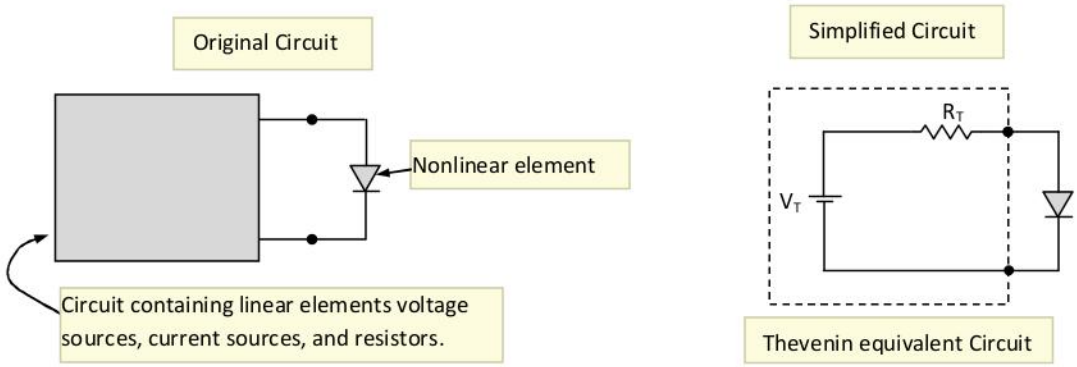


The transition level will be at $V_i = 0$ since $i_d = 0$

When Vi more than V=5 V , V_o will follow V_i

When Vi less than V=5V , V_o will stay at V=5V

Ans.6:



References and Suggested Readings

1. Jasprit Singh, "Semiconductor Devices Basic Principles". John Wiley & Sons (Asia) Pvt. Ltd. (2001).
2. M.S. Tyagi, "Introduction to Semiconductor Materials and Devices". John Wiley & Sons (Asia) Pvt. Ltd. (2000).
3. S.M.Sze, "Physics of Semiconductor Devices 2nd Edition". John Wiley & Sons (Asia) Pvt. Ltd. (2002).
4. S.O.Kasap, "Principles of Electronic Materials and Devices SECOND EDITION". Tata Mc Graw – Hill NEW DELHI (2002).
5. Donald A Neamen, "Semiconductor Physics and Devices Basic Principles Third Edition". Tata Mc Graw – Hill NEW DELHI (2002).
6. Michael Shur, "Physics of Semiconductor Devices 2nd Edition". Prentice Hall of India New Delhi (2002).
7. Ben. G. Streetman, Sanjay Banerjee, "Solid State Electronic Devices First Edition". Prentice Hall of India New Delhi (2002).
8. Boylestad, Nashelsky, " Electronic Devices and circuit theory 6th edition". Prentice Hall of India New Delhi (2001).

UNIT-3
Semiconductor Diode Devices

Structure of the Unit

3.0 Objectives

3.1 Introduction

3.2 Varactor Diode

3.2.1 Illustrative Examples

3.3 Tunnel Diode

3.4 Self Learning Exercises-I

3.5 Metal- Semiconductor Contacts

3.6 Schottky Diode

3.7 Light Emitting Diode (Led)

3.8 Solar Cell

3.9 Photodiode (Photodetector)

3.9.1 Pin Photodetector

3.10 Gunn Diode

3.11 Self Learning Exercise-II

3.12 Summary

3.13 Glossary

3.14 Answers to Self Learning Exercises

3.15 Exercise

3.16 Answers to Exercise

References and Suggested Readings

3.0 Objectives

In this chapter students will enable to understand the knowledge of several semiconductor diodes, their structure, principles of operation and their working mechanism, characteristics and essential parameters; learning to measure I-V

characteristics, also called V-I characteristics and important applications in semiconductor electronics.

3.1 Introduction

Chapter introduces several devices that are formed by joining two different materials together. The p-n junction (a junction between a p-type and n-type semiconductor materials) and metal semiconductor junction are basic building of various semiconductor diodes. It acts as a diode, which can serve in electronics as a rectifying diode, varactor diode, Schottky diode etc. in optoelectronics as a light emitting diode (LED), photodetector, or solar cell and in microwave electronic devices as a tunnel diode, Gunn diode.

These junctions are analyzed in different modes of operation (the forward bias and reverse-bias conditions). The concepts of the depletion region and mechanism of carrier injection has particular importance. Semiconductor optoelectronic devices are also presented widely because of their rising importance for renewable energy generation. The metal–semiconductor junction can be a rectifying junction or an ohmic contact. These contacts have great importance in the fast switching operation of the devices.

3.2 Varactor Diode

A varactor diode is a special purpose pn junction diode that operates in reverse bias. The word varactor comes from the term “variable reactor”. It is basically a voltage variable capacitor in which the capacitance of the pn junction is controlled by the amount of reverse bias voltage. Fig. shows the schematic symbol of a varactor diode.

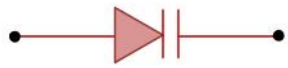


Fig.1 Schematic symbol of a varcator diode

One property of p-n junction is that the variation in reverse bias voltage change the width of the depletion region and hence the depletion capacitance i.e., this produce a voltage dependent capacitor. This special property is utilized in the application.

A reversed-biased p-n junction diode has two conducting regions separated by an insulating depletion region acts as a variable capacitor and its capacitance is found as

$$C = \frac{A\varepsilon}{d} \qquad (1)$$

where, C = the total junction capacitance.

A = the plate area.

ε = the dielectric constant (permittivity).

d = the width of the depletion region (plate separation).

When the diode is reverse biased, we find the depletion layer widens, and the capacitance across the pn junction decreases.

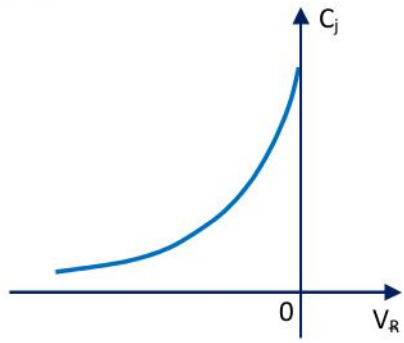


Fig.2: Voltage-Dependent Capacitance

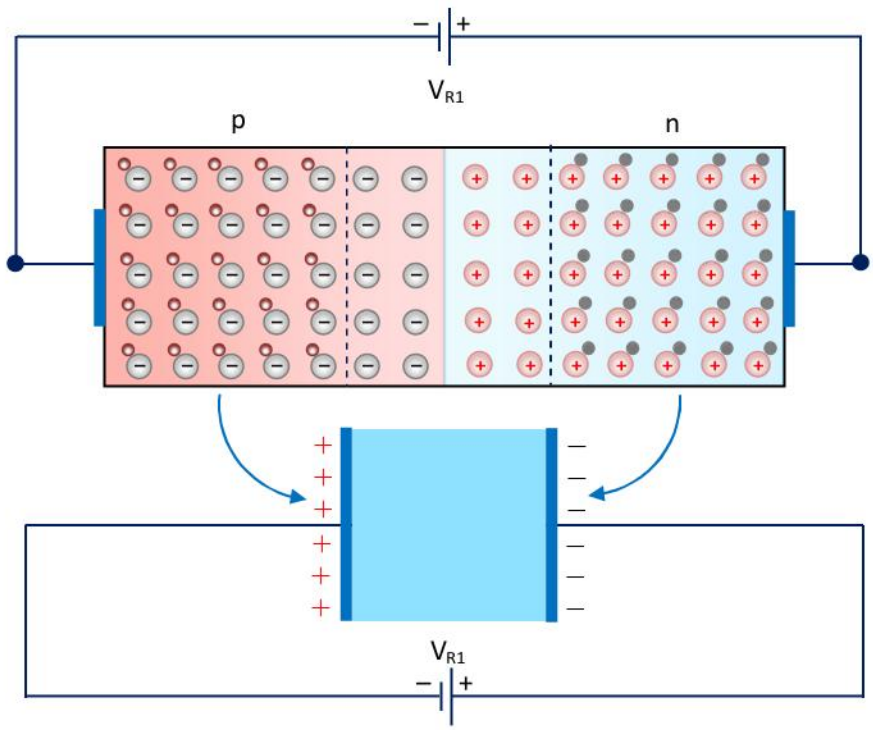


Fig.3a :Variation of depletion region width with changing of reverse bias voltage

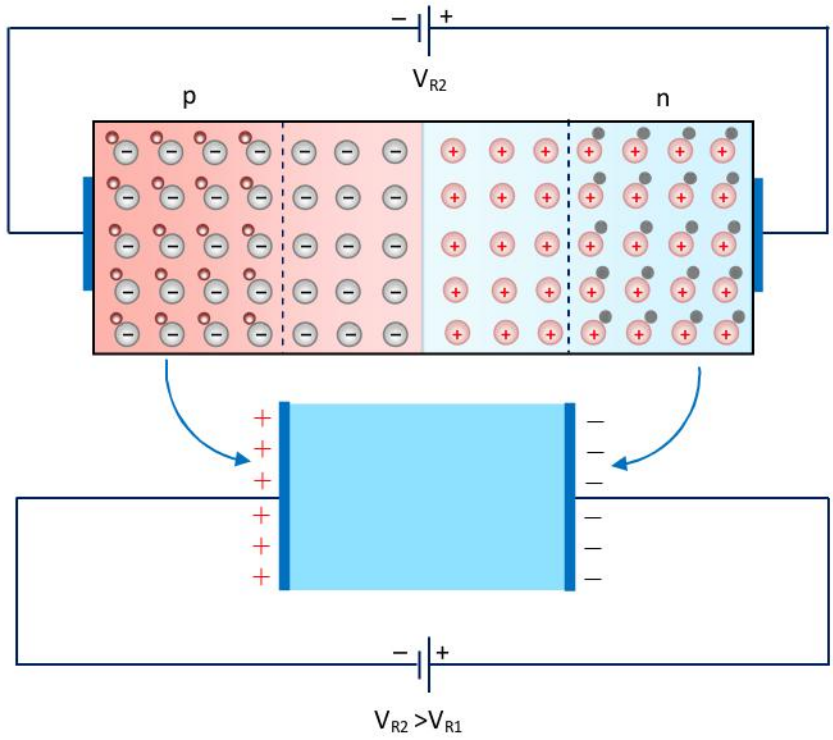


Fig.3b: Variation of depletion region width with changing of reverse bias voltage

Changes in reverse voltage lead to changes in depletion layer width and charge. This leads to a capacitance that can be calculated from the voltage-dependent capacitance is

$$C_j = \frac{C_{j0}}{(V_a + V_0)^n} \qquad (2)$$

where C_{j0} is the zero bias junction capacitance per unit area.

V_0 junction contact potential (≈ 0.7 volt)

n capacitance exponent depends on the doping profile

V_a the applied voltage is positive for reverse bias and negative for forward bias

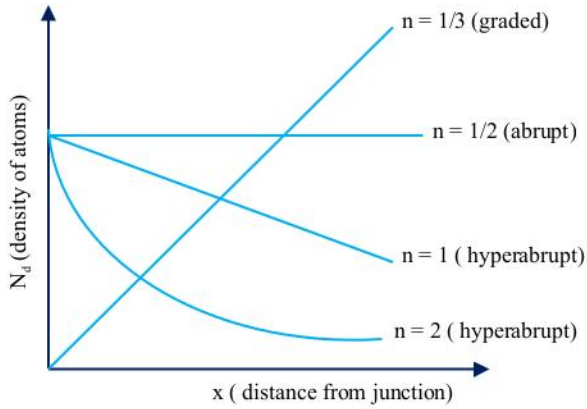


Fig.4

The values of capacitance of varactor diodes are in the picofarad (10^{-12} F) range.

APPLICATIONS

They are commonly used in communication systems.

•Automatic tuning capacitor

•Microwave frequency multiplication

•Active filters

•Voltage Controlled Oscillators

•RF Filters

3.2.1 Illustrative Example

Example1: A junction pn junction diode has the parameters C_{J0} = 15pF, V_0 = 0.9 V, and n = 0.5. Calculate the change in junction capacitance if the reverse bias voltage is changed from 5V to 10V.

Sol. The junction capacitance is given by Eq. (2).

(a) For V= 5volts

$$C_j = \frac{C_{j0}}{(V_a + V_0)^n} = \frac{15\times10^{-12}}{(5+0.9)^{0.5}} = 6.18\,\text{pF}$$

(b) For V= 10volts

$$C_j = \frac{C_{j0}}{(V_a + V_0)^n} = \frac{15\times10^{-12}}{(10+0.9)^{0.5}} = 4.54\,\text{pF}$$

Thus the capacitance changes by 6.18 – 4.54 = 1.64 pF.

3.3 Tunnel Diode

The tunnel diode was invented in August 1957 by Japanese physicist Leo Esaki. It is made of heavily doped p-n junction (10^{19} to 10^{20} atoms/cm^3 or Impurity concentration is 1 part in 10^3 as compared to 1 part in 10^8 in an ordinary p-n junction diode). The heavy doping results in a broken bandgap, where Fermi level goes above the conduction band for n- type and below valence band for p- type material (degenerate materials)[8]. A heavily doped p-n junction leads to extremely

[8] Degenerate semiconductor

narrow depletion layer (~ 10 nm) which differs from the normal rectifying diode in ways like I-V characteristics and physical mechanism of its operation.

It is usually made up of semiconducting materials like germanium (Ge), gallium arsenide (GaAs) and silicon (Si). They can be used as high frequency oscillators, amplifiers, frequency converters and detectors.[1] Fig. shows the schematic symbol of a tunnel diode.



Fig.5 Schematic symbol of a Tunnel diode

Tunnel diode is the p-n junction device that exhibits negative resistance between two values of forward voltage. That means when the voltage is increased the current through it decreases.

Theory

The tunnel diode is basically a heavily doped (1000 times as heavily as a conventional diode) p-n junction diode. By the result of heavy doping a large number of majority carriers are available for the recombination. Most of them are not used during initial recombination and produces a narrow depletion (100 times narrower than an ordinary diode) layer.

Operation of tunnel diode depends on the tunneling effect. Tunnel diode is named on the basis of tunneling effect. It is a *quantum mechanical phenomenon* which Occurs when an electron passes through a potential barrier without having enough energy to do so. It is caused by the wave nature of electrons.

Tunneling

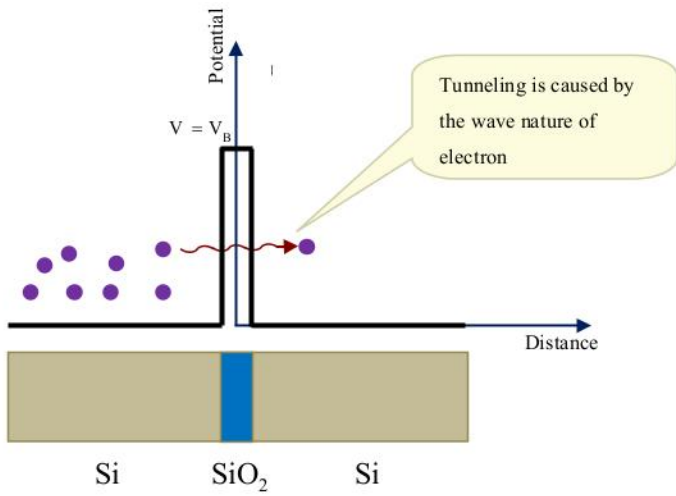


Fig.6 Tunneling of electron

In tunneling, the valence electrons moves from the valence energy band to the conduction band with little or no applied forward voltage. The particle energy does not change during the tunneling.

I-V Characteristics

The current-voltage characteristic of tunnel diode is represented in Figure below. In this sketch I_p and U_p are the peak, and I_v and U_v are the valley values for the current and voltage respectively.

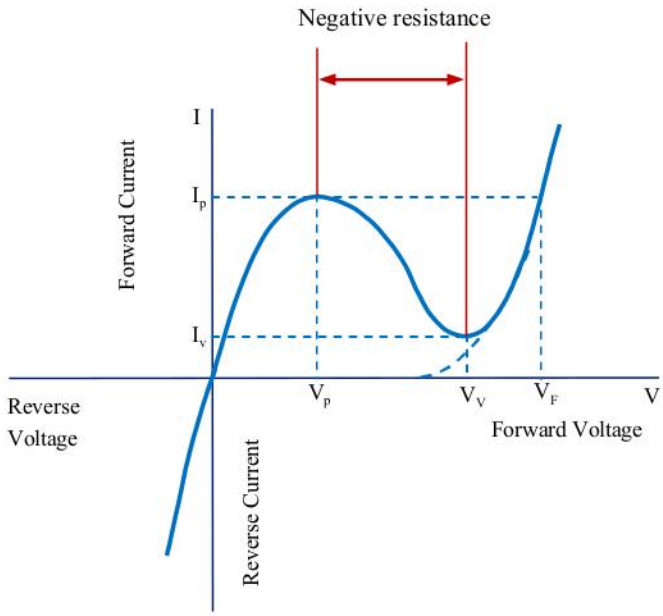


Fig.7 I-V characteristics of tunnel diode

As the forward voltage is start increased, the tunnel diode current is increased from zero, causing electrons tunnel from n- region to the p- region. As the forward voltage increases the diode current also increases to its peak value.

As the voltage is increased beyond the V_p (valley point) the diode current starts decrees and diode exhibits negative resistance. When diode is operated in the negative region it can be used as an oscillator.

The form of this type I-V characteristic can be qualitatively explained by considering the tunneling processes that take place in a thin p-n junction.

Tunnel Diode Operation

- *Forward Bias Operation*

(a) Step1- At zero bias (at thermal equilibrium): At zero bias there is no current flow as shown in figure. All energy states are filled below Fermi level on both sides of the junction.

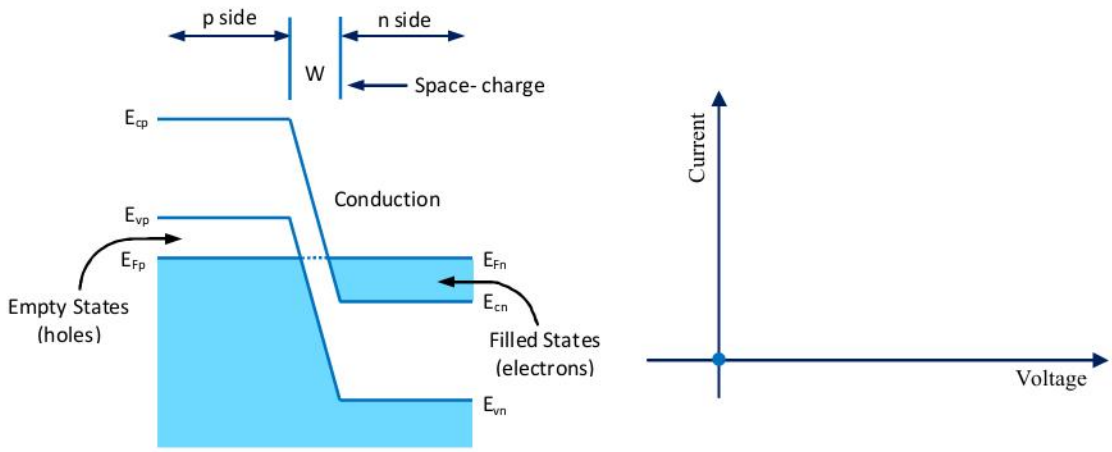


Fig.8 Energy- band diagram and I-V characteristics at zero bias equilibrium.

(b) Step2- Under small forward bias is applied: A small forward bias is applied. Potential barrier is still very high – no forward current is predicted across the junction. Electrons in the conduction band of the n side will directly tunnel to the empty states of the valence band in p- side across the small depletion region, see fig. This will produce a small forward bias tunnel current.

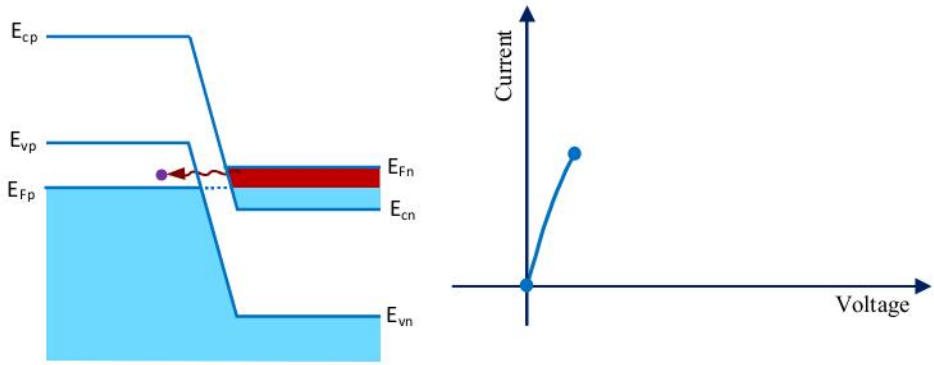


Fig.9 Direct tunneling current starts

(c) Step3- Further increasing forward voltage, the energy of the electrons in the n-region is equal to that of the empty states (holes) in the valence band of p-region. Hence, the maximum number of electrons in the n region will tunnel opposite to the empty states in the p region. This will produce maximum tunneling current.

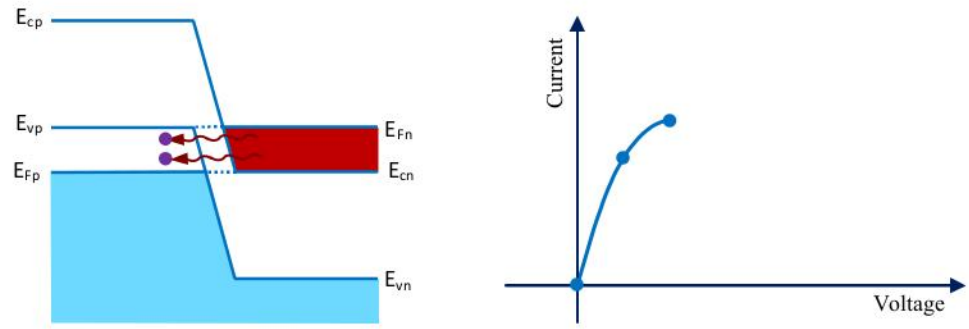


Fig.10 Maximum Direct tunneling current

(d) Step4- As the forward bias voltage increases, the number of electrons in the conduction band (n- side) that are directly opposite to the empty states in the valance band (p-side) starts decrease. Therefore direct tunneling current will start to decline.

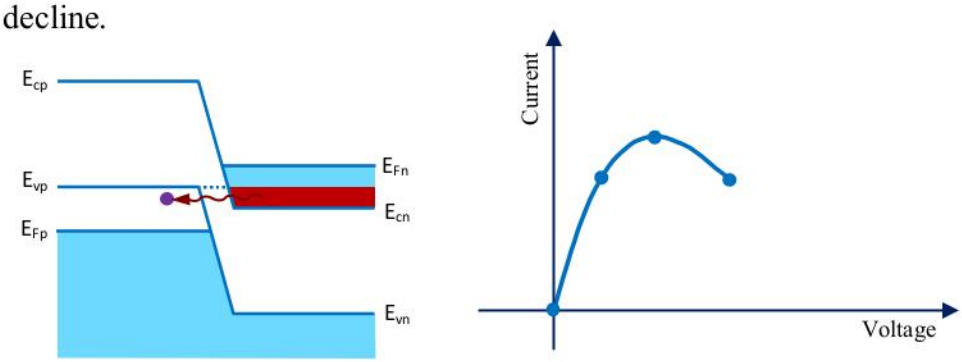


Fig.11 Direct tunneling current starts

(e) Step5- As the forward bias voltage continues to increase; it will approach to its minimum value.

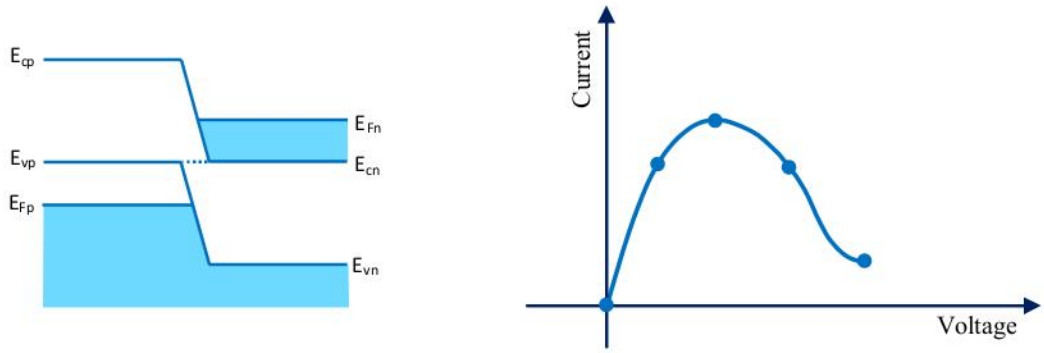


Fig.12 Minimum direct tunneling current

(f)Step6- with further voltage increase, the direct tunneling current tends to zero and height of the valley potential is reduced to a level that allows conventional

(thermal diffusion) current to flow over the barrier. The forward current starts to increase due to electron-hole injection via thermal diffusion mechanism.

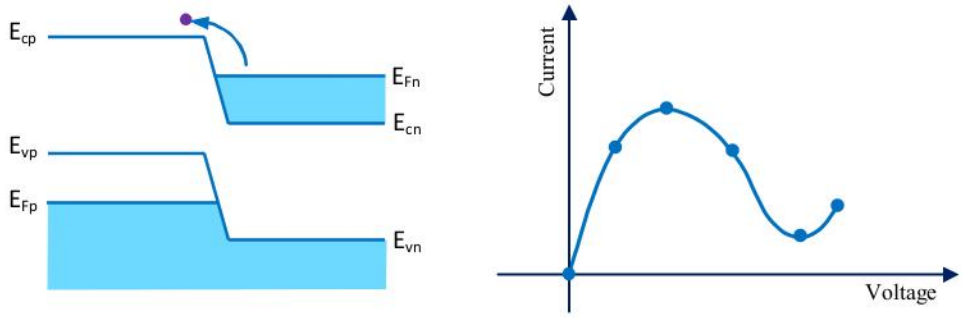


Fig.13 Diffusion current starts

(g) As more forward voltage is applied, until the height of the barrier is reduced to a level that allows conventional current to flow over the barrier and I-V characteristics is similar to that of an ordinary pn junction diode.

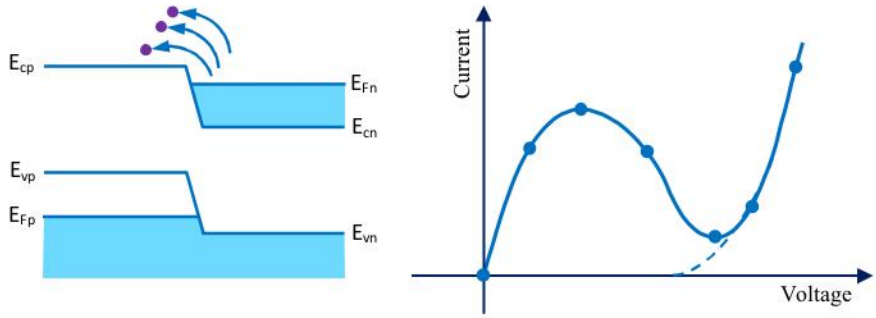


Fig.14 increase in diffusion current

● *Reverse Bias Operation*

In this case, the valence band of the p- side tunnel directly towards the empty states in the conduction band of the n side producing large tunneling current across the depletion region which increases with the reverse bias voltage.

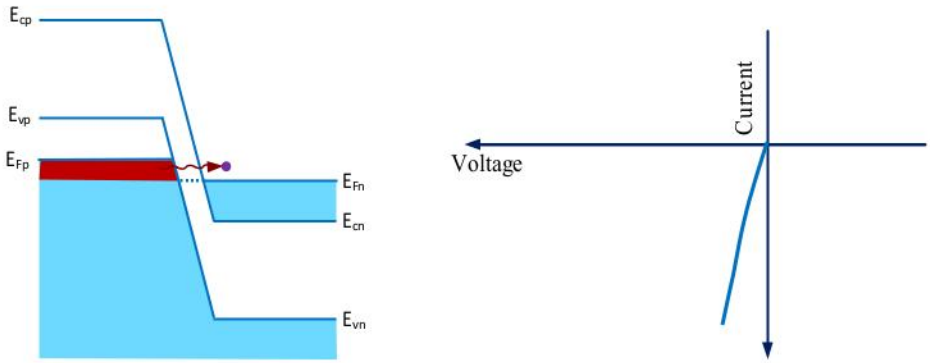


Fig.15 Reverse bias characteristics

Therefore under reverse bias case, there will be a net conventional current flow from n- side to p- side. The reverse I-V characteristic is similar to Zener diode with nearly zero breakdown voltage.

3.4 Self Learning Exercise-I

Q.1 What is tunnel diode?

Q.2 Plot the I-V characteristics of Tunnel diode.

Q.3 How does the capacitance affect the depletion layer in varactor diode?

3.5 Metal- Semiconductor Contacts

Metal- semiconductor contacts have great importance since they are essentially present in every semiconductor device for use in an electronic circuit. There are two types of metal- semiconductor contacts are broadly used in the fabrication of semiconductor devices and IC. They are the ohmic and Schottky contacts. When a metal and semiconductor are brought together into contact, there are two types of contacts formed depending on their work function and their relationships are following:

	P-type	**N-type**
$\Phi m < \Phi semi$	Schottky(rectifying)	ohmic
$\Phi m > \Phi semi$	ohmic	Schottky(rectifying)

(1)*Ohmic Contact*

Ohmic contact is a junction which has linear current-voltage relationship and electrically behaves like a simple homogeneous metal bar. Ohmic contact is used to carry current into and out of the semiconductor device. It means an unimpeded (like a short wire) transfer of majority carriers take place between the materials (from metal to semiconductor or semiconductor to metal) i.e. contacts do not limit the current. It is also defined as the current flow regardless of the polarity of the bias (i.e., current flow in the both direction).

(2) *Schottky Contact*

A Schottky or rectifying contact is defined as one in which a forward bias drives a large current but a reverse bias results in a small current i.e. I-V characteristics

similar to that of PN junction.

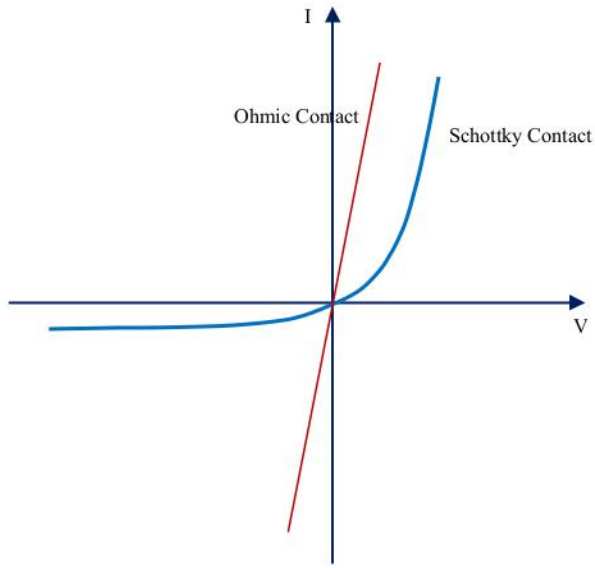


Fig.16 I-V characteristics of ohmic and rectifying contacts

3.6 Schottky Diode

A Schottky diode (named after German physicist Walter H. Schottky; also known as *hot carrier diode*) is a special type of metal-semiconductor junction diode with a very low forward-voltage drop and a very fast switching action. When current flows through diode, there is a small forward voltage drop across the diode terminals.

A conventional diode has forward voltage drops between 0.7-1.7 volts, while a Schottky diode voltage drop is between approximately 0.15-0.45 volts. This lower forward bias voltage drop making these diodes for use in power rectification applications.

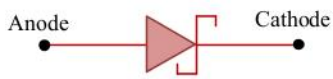


Fig.17 Schematic symbol of Schottky diode

A Schottky barrier diode is a rectifying metal – semiconductor junction formed by plating, evaporating or sputtering of metal (e.g. aluminum, molybdenum, platinum, chromium or tungsten) onto n-type or p-type semiconductor material(instead of a semiconductor-semiconductor junction as in conventional diodes). Generally, n-type silicon and n-type GaAs are used for the fabrication of Schottky diodes. The

metal side acts as the anode and N-type semiconductor acts as the cathode of the diode. This Schottky barrier results in both very fast switching and low forward voltage drop.

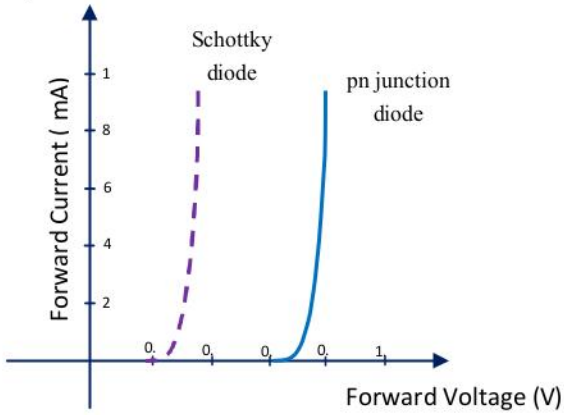


Fig.18: Comparison between forward bias voltage between Schottky and pn junction diode

The most important difference between the p-n and Schottky diode is reverse recovery time. The Schottky barrier diode is much faster than a conventional pn junction diode. The basic reason that Schottky diodes are majority carrier devices that can be switched rapidly from forward (turn on) to reverse bias (turn-off) without minority carrier storage problems that slow down the switching speed of diode. Because of this characteristic they *have a faster "reverse recovery time" than p-n junction diodes* which makes it much faster than a conventional junction diode. The properties of conventional diodes are determined by minority carriers. A pn junction diode has the reverse recovery time in order of 100 ns and Schottky diodes do not have a recovery time, as there is nothing to recover from (i.e. no charge carrier depletion region at the junction) junction region.

3.7 Light Emitting Diode (LED)

A Light Emitting Diode (LED) is a simple pn junction semiconductor diode that converts electrical energy into light energy of a certain wavelength (color). The phenomenon of spontaneous emission of light takes place when it is connected in a circuit and operated in a forward biased direction.

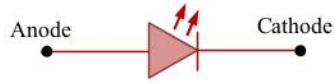


Fig.19

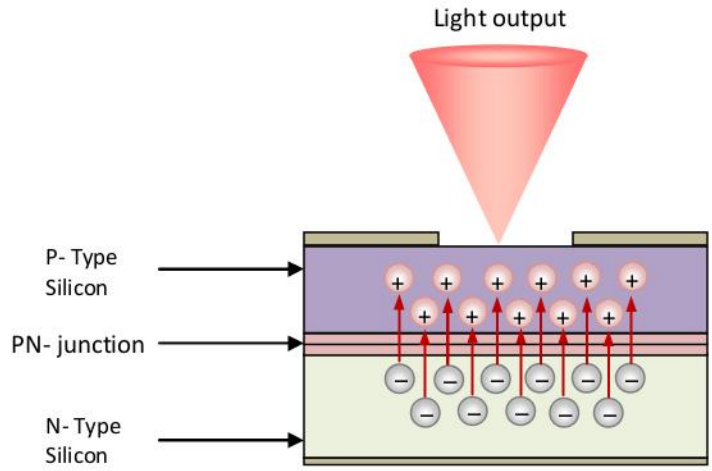


Fig.20 Light output form LED junction

When sufficient electrical energy (threshold voltage) supplied to the LED and current starts to flow (i.e., p- side of the junction is connected to the positive terminal of a battery and the n side to the negative terminal), electrons gain enough energy to cross over the junction and enter into the p region (i.e., electrons are attracted by positive charge due to columbic forces). Afterward they recombine with holes within the device and lose their energy by emitting photons of light. As the p region is lightly doped, only few numbers of holes cross over to the n region than the more numbers of electrons diffuse to the n region. Such process is termed as radiative recombination. *Recombination* takes place in the depletion region (active region) and photons are emitted in random direction. This effect is called *electroluminescence* (emission of light when the device is supplied with some electrical energy).

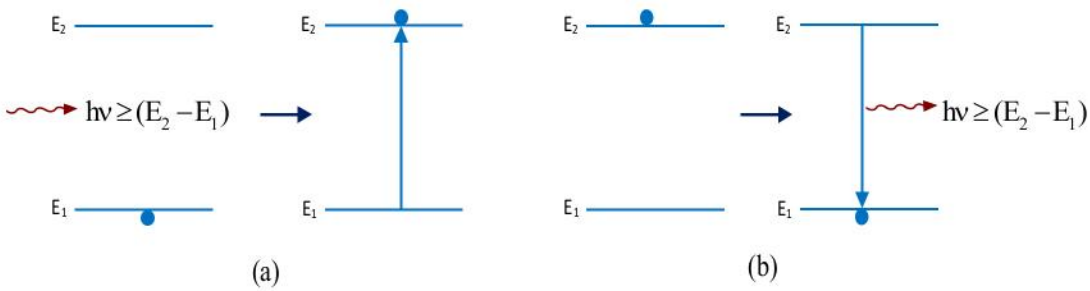
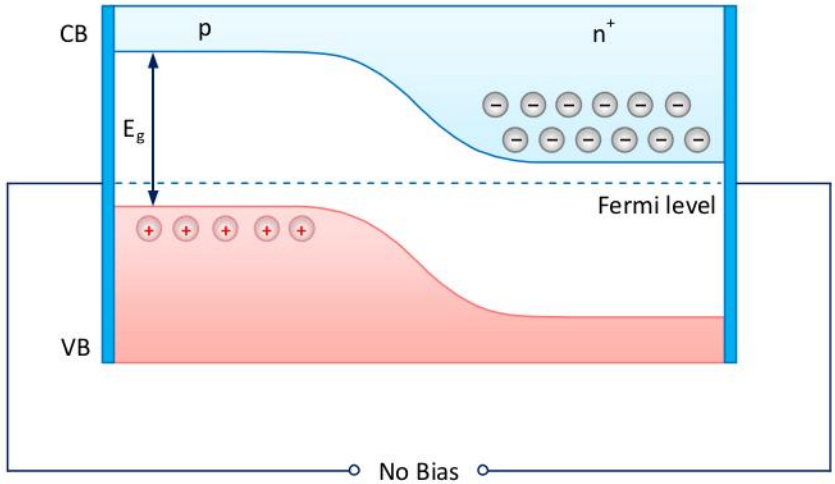


Fig.21 Schematic diagram of the basic process of (a) photon absorption and (b) Photon emission.

Energy band diagram

(A) Under zero bias (no bias) condition



(B) Under forward bias condition

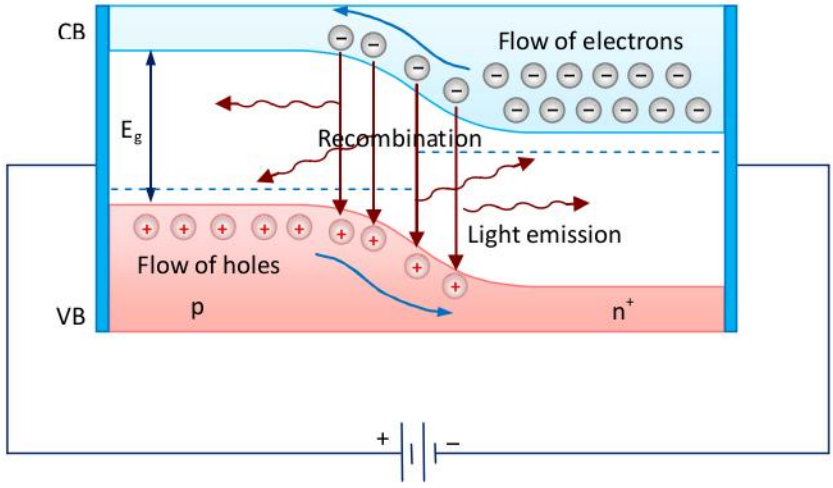


Fig.22 Energy band diagram of LED under the forward bias and Reverse bias.

When apply the certain voltage across the device (Under forward condition), the majority carriers from both sides (electrons in n-type and holes) cross the depletion region and entering the other side.

Upon entering, the majority carriers become minority carriers (for example, the electrons as minority carries in the p- region and holes as minority carries in n-region) and they recombine with that side of majority carries. The recombination process causes light to be emitted. The wavelength of emitted photons is decided

by the bandgap energy of the semiconductor materials.

$$E_g = h\nu = \frac{hc}{\lambda} = eV_{th}$$

where h = Planck's constant

λ = wavelength of emitted photons

V_{th}= Threshold voltage

c = velocity of light

e = electron charge

Above the threshold voltage, the excess electrical energy create more photons- the LED gets brighter. Given this, by measuring the threshold voltage V_{th} we can determine the color (wavelength) of emitted light.

$$\lambda = \frac{hc}{eV_{th}}$$

LED Semiconductor Materials

The colors of emitted light by some common LED semiconductor material are listed in table below:

Table.1. Approximate threshold voltages of LED of different wavelength

LED Materials	V_{th} (volt)	LED colour
GaAs	1.5	Infrared
GaP	2.4	Green
GaN	4.1	White
AlGaAs	1.8	Red

LED construction

A semiconductor chip is encapsulated by special type of transparent plastic (called as epoxy lens of LED) material inside light emitting diode. In LED, the semiconductor chip is mounted in a reflected cup in order to improve the light output. The transparent plastic case encloses the LED frame which provides high resistance against moisture (environmental effect) and mechanical strength.

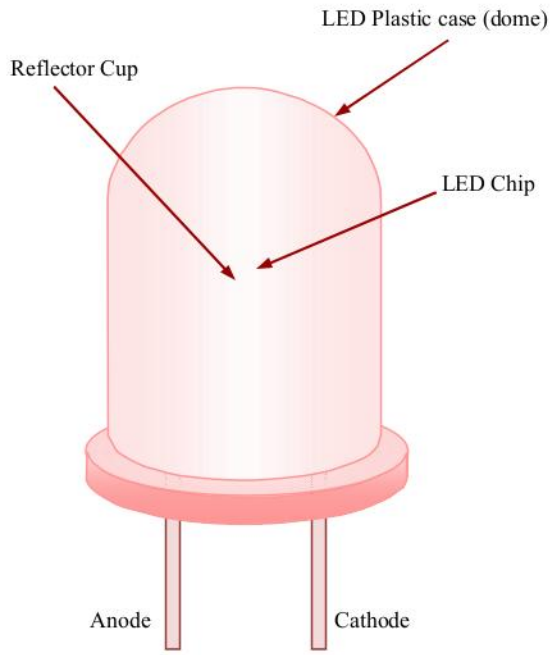


Fig.23 LED construction

3.8 Solar Cell

A solar cell is a semiconductor device that converts sunlight (solar light) from the sun into electricity based on the principles of photovoltaic effect. In general a solar cell that includes both solar and non-solar sources of light (such as photons from incandescent bulb) is termed as photovoltaic cell.

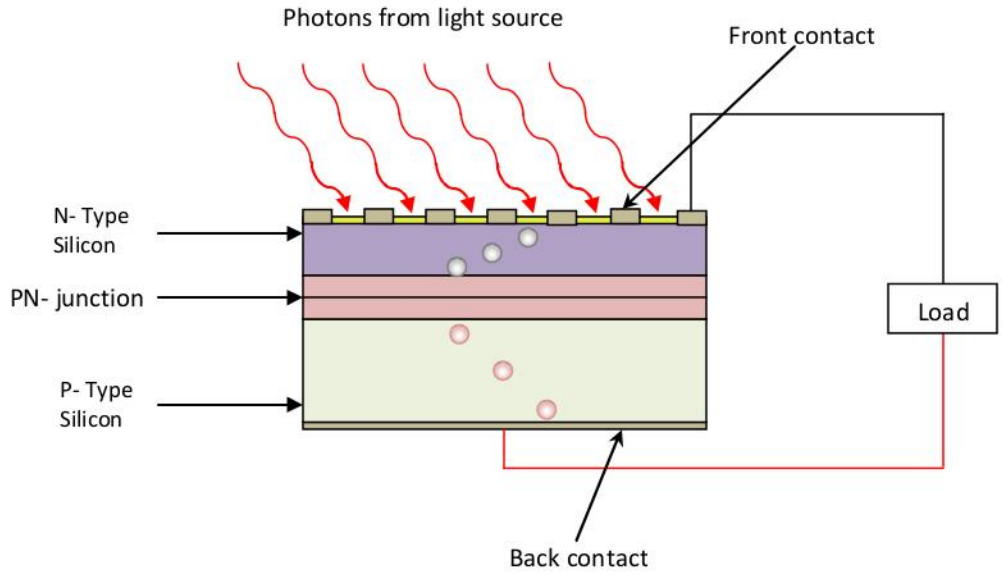


Fig.24 Solar cell structure

When light strikes on the photovoltaic cell then it is absorbed by semiconducting materials. This absorption of light creates charge carriers (electron-hole pairs) generating electrical power.

3.9 Photodiode (Photodetector)

Photodiode (it is also called as photodetector or photosensor) is a semiconductor device that produces current or voltage when it is illuminated by light. The generated current is proportional to the intensity of the incident light. In process, incident photons generate free (mobile) electron-hole pairs in the active region and Transport of the free electrons and holes upon an electric field results in a current. These charge carriers are known as photocarriers, since they are generated by a photon.



Fig.25 Schematic symbol of photodiode

An incident photon is able to generate to e-h pair if it has an energy that is greater than or equal to the bandgap energy. A particular semiconductor material absorbs only over a limited wavelength range. The upper wavelength λ_c (cutoff or threshold) is determined by the band-gap energy E_g of the material.

$$\lambda = \frac{hc}{E_g} = \frac{1.24}{E_g(eV)}$$

The device structure consists of a silicon or germanium pn junction in which reverse current increases when the junction is exposed to light.

Operating Principle

When a photodiode is illuminated by light, the valence band electrons are excited to the conduction band leaving holes in their place in the valence band [Figure]. These e-h pairs occur throughout the active region (the p- layer, depletion region and N-layer).

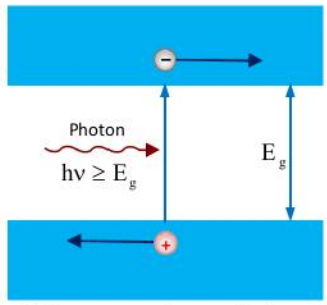


Fig.26 generation of photocarriers

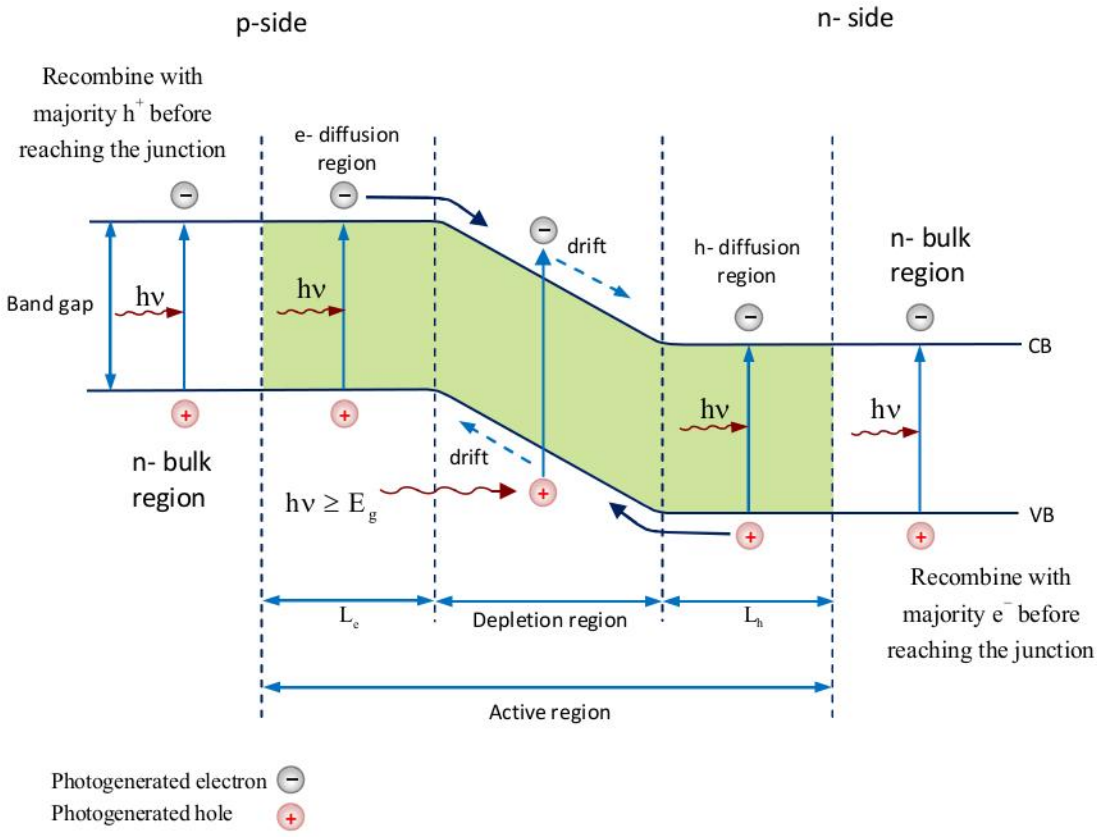


Fig.27 Energy band diagram of pn photodiode

In depletion region, the photogenerated e-h pairs are subjected to the local electric field within the region. The local electric field sweeps the photogenerated-e^- to the n side and the photogenerated- h^+ to the p side i.e. e-h carriers drift in opposite directions. This transport process induces an electric current in the external circuit. A *drift current,* that flows in the reverse direction from the n- side (cathode) to the- p side (anode).

At the edges of the depletion layer diffusion region occurs. The photogenrated minority carriers (hole in the n- side and electron in the p- side) reach the depletion region by diffusion and then be swept to the other side by the local field. A diffusion current, that also flows in the reverse direction.

In the p or n bulk region, essentially no current is generated because there is no local field is present which can separate the photogenrated minority carriers. Therefore, photogenerated minority carriers in bulk region cannot diffuse to the depletion region before combining with the majority carriers.

I-V Characteristics Of Photodiode

Recall the current-voltage (I-V) characteristic of the junction is given by the diode equation:

$$I = I_s\left[\exp\left(\frac{qV_a}{k_B T}\right) - 1\right]$$

I_s is the "saturation current" representing *thermal-generated* free carriers which flow through the junction (*dark current*).

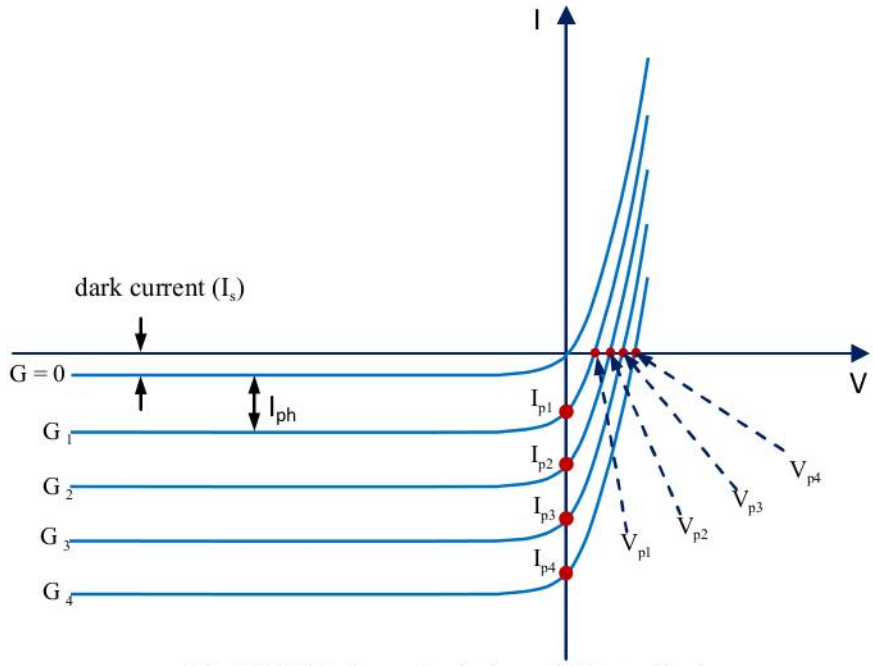


Fig.28 I-V characteristics of Photodiode

The photodiode therefore has an I-V characteristic (under illumination):

$$I = I_s\left[\exp\left(\frac{qV_a}{k_B T}\right) - 1\right] - I_{ph}$$

This is the usual I-V curve of a p-n junction with an added photocurrent $-I_p$ and proportional to the photon flux (light intensity).

The *short-circuit current* (V = 0) is the *photocurrent* I_p and The *open-circuit voltage* (I = 0) is the *photovoltage* V_p.

3.9.1 Pin Photodetector

A PIN photodiode is a semiconductor device with an intrinsic (usually lightly doped p-or n- region) layer that is sandwiched between the p-type and n- type

layer. The concept of PIN diode design is to obtain the wide depletion region (the drift region) which is achieved by intrinsic layer.

As a photodetector, the PIN diode has main advantages:

(1) Wide depletion region provides large drift region in which photons can create more photocarriers (e-h pairs) that leads to improved efficiency.

(2) Large width of the depletion layer decreases the junction capacitance and thereby the time constant RC which means fast response of device.

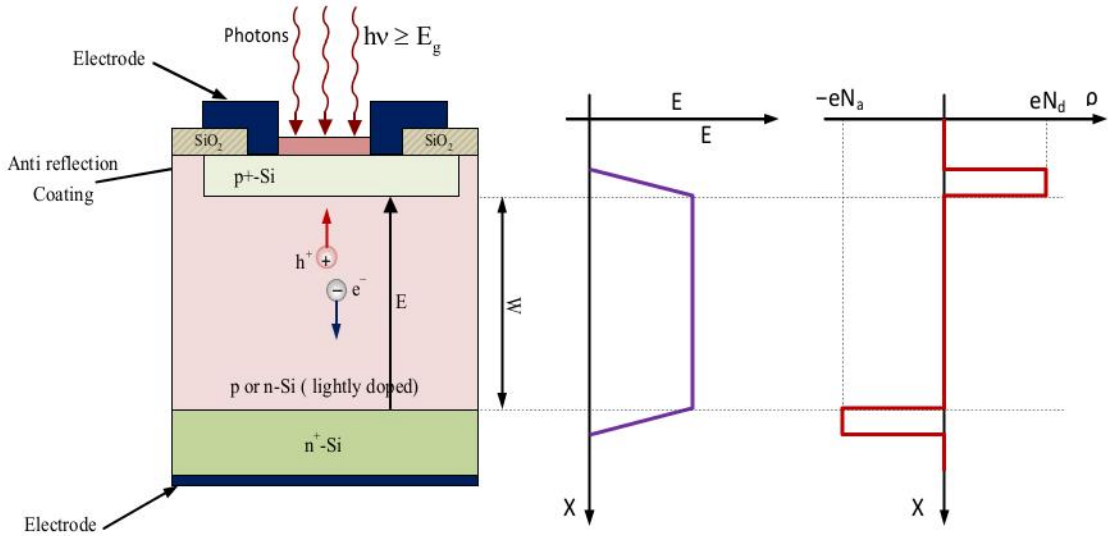


Fig.29 PIN photodetector Structure, electric field and charge density at thermal equilibrium.

3.10 Gunn Diode

Gunn diode (named after Physicist John Battiscombe Gunn) is a microwave device that is characterized by "Negative Differential Resistance" (or "Negative Differential Mobility" or "Negative Differential Conductivity") region (NDR).

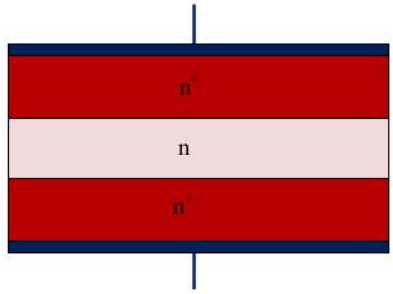


Fig.30 Basic structure of Gunn diode

It is also known as transferred electron devices (TED) is widely used in microwave RF applications for frequencies between 1 and 100 GHz**.** Typically, a Gunn diode consists of a uniformly doped n-type III-V material (e.g., GaAs, InP) sandwiched between heavily doped (n-type) regions at each terminal.

In certain semiconductors (such as GaAs, InP, GaN), the conduction electron can be transferred from a state of high mobility to a state of low mobility by influence of strong electric field i.e., transferred electron mechanism.

Gunn Diode Operation

These are semiconductors having a many closely spaced energy valley in E-K diagram as shown in Fig. for GaAs.

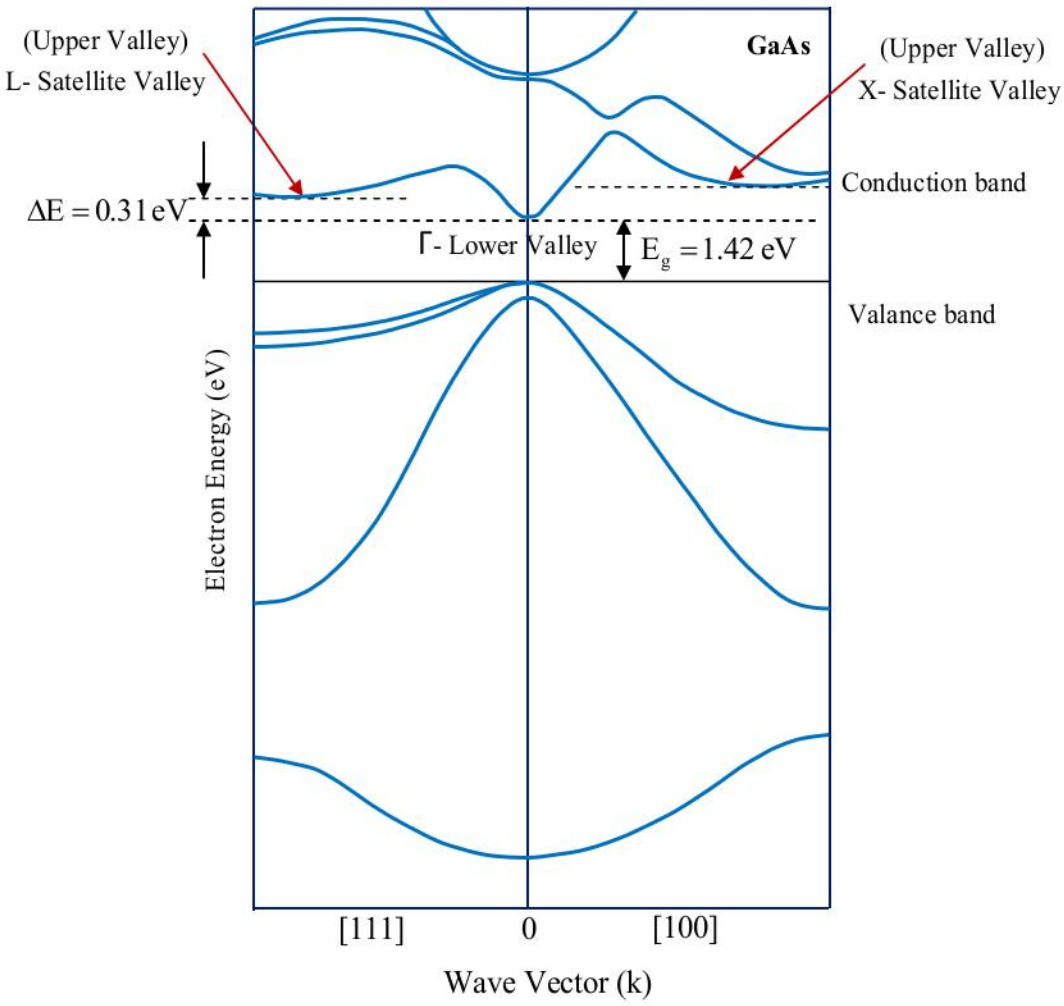


Fig.31 E-K diagram for GaAs

When a dc bias voltage is applied across the Gunn diode, an electric field is established across it. At low **E**-field $(E > E_0)$, most of the electrons will be located in the lower energy central valley Γ. At higher **E**-field $(E_0 < E < E_{th})$, most of the electrons will be transferred in to the high-energy satellite L valley where these electrons have a lower mobility (or drift velocity) because of the increase in their effective mass. The mobility can be defined as negative $((dv_d/dE) < 0)$ when the electric field is above the threshold value $(E > E_{th})$ and diode exhibits the negative differential resistance (NDR). As the field is continually increased beyond the saturation field $(E > E_s)$, the drift velocity of electron will be saturate. This is called the *transferred electron effect* and the device is also called 'Transfer Electron Device (TED) or Gunn diode

Two Valley Model

(1) The energy difference between two valleys must be several times larger than the thermal energy ($kT \approx 0.0026eV$) but smaller than the bandgap energy (eV).

(2) Electron in lower valley must have a higher mobility (or smaller effective mass) than that of upper valley.

The effective mass of an electron is given by

$$m^* = \frac{\hbar^2}{d^2E/dk^2}$$

where d^2E/dk^2 Rate of change of the curve slope

Since the lower valley has a sharp curvature (slope is more shaper) than in the upper valley, thus electron effective mass in lower valley is less than that in upper valley. Hence, the mobility of electron in lower valley is high due to the low effective mass.

$$\mu_n = \frac{e\tau}{m_n^*}$$

Table. Effective mass and mobility In lower and upper valleys for n- GaAs

Valley	Effective mass (m^*)	Mobility (μ_n)
Lower (Γ-central)	$0.068\ m_0$	8000
Upper(L- satellite)	$1.22\ m_0$	180

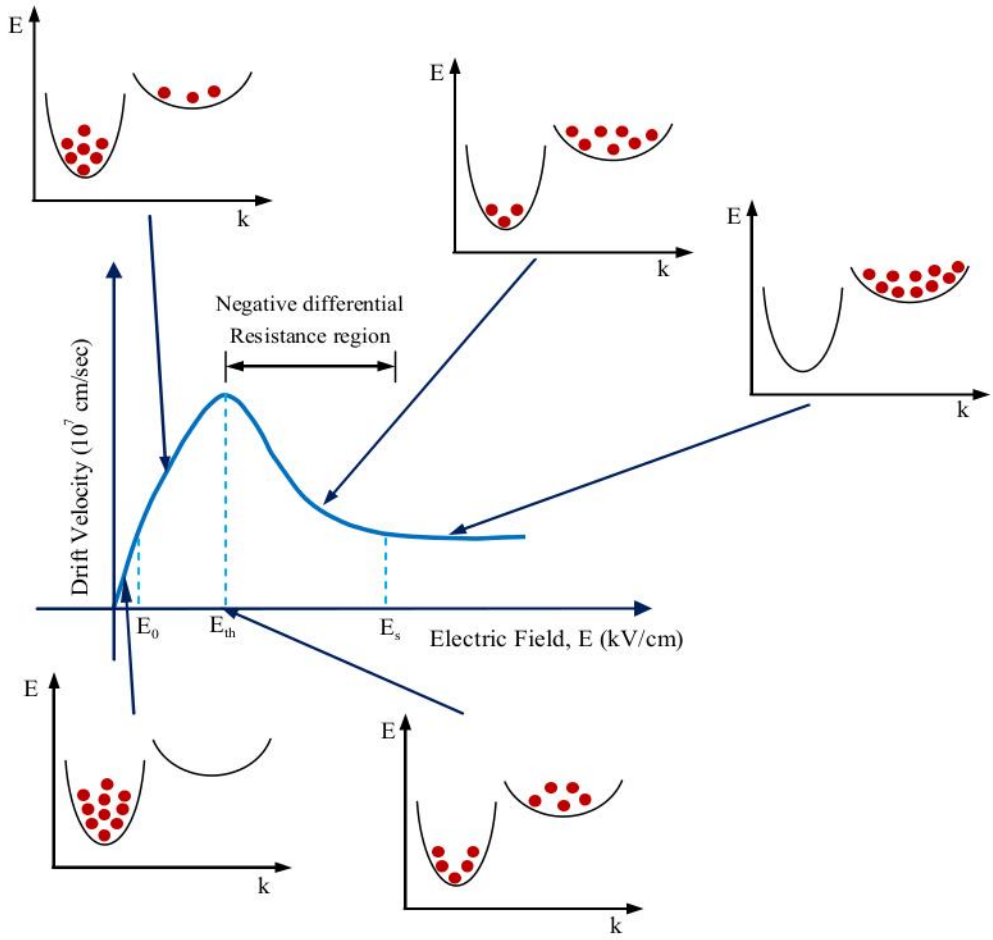


Fig.32 Transfer electron mechanism (TEM)

3.11 Self Learning Exercise-II

Q.1 What does LED stand for?

Q.2 What is relationship between bandgap and wavelength for a LED?

Q.3 What is ohmic contact?

3.12 Summary

In this chapter we have addressed the different varieties of one of the most important semiconductor structure pn junction diode. we discussed the numerous applications of pn junction diode as microwave devices (tunnel diode and Gunn diode), and optoelectronics devices(light emitting diode, solar cell photodiode). We discuss the basic structure and their operation in briefly. We have also seen how the different IV characteristics of different semiconductor devices.

3.13 Glossary

LED: Light emitting diode

Tunnel diode : It is the p-n junction device that exhibits negative resistance between two values of forward voltage.

Schottky: German physicist Walter H. Schottky

TED: Transfer Electron Device (TED) or Gunn diode

3.14 Answers to Self Learning Exercises

Answers to Self Learning Exercise-I

Ans.1: The tunnel diode is basically a heavily doped (1000 times as heavily as a conventional diode) p-n junction diode. Operation of tunnel diode depends on the tunneling effect.

Ans.2: See I-V characteristics of tunnel diode

Ans.3: One property of p-n junction is that the variation in reverse bias voltage change the width of the depletion region and hence the depletion capacitance i.e., this produce a voltage dependent capacitor. This special property is utilized in the varactor diode.

Answers To Self Learning Exercise -II

Ans.1: Light emitting diode.

Ans.2: $E_g = h\nu = \frac{hc}{\lambda}$

Ans.3: Ohmic contact is a metal semiconductor junction which has linear current-voltage relationship.

3.15 Exercise

Section A:Very Short Answer Type Questions

Q.1 Draw the schematic symbol of varactor diode, tunnel diode, Schottky diode, Light emitting diode, photodiode and Gunn diode.

Q.2 What is tunneling?

Q.3 What is Gunn diode?

Q.4 What is Schottky diode?

Q.5 What is varactor diode?

Section-B : Short Answer Type Questions

Q.6 Draw the I-V characteristics of ohmic and Schottky contacts.

Q.7 Explain operation of the Tunnel diode.

Q.8 Draw the energy band diagram of LED.

Q.9 What are ohmic and Schottky contacts?

Q.10 What are the semiconductor elements that exhibit Gunn effect?

Section C :Long Answer Type Questions

Q.11 Describe the operating principle of tunnel diode using I-V characteristic curve.

Q.12 Describe how a negative mobility can possibly happen for a GUNN diode.

Q.13 Explain operation of the Gunn diode.

Q.14 Does LED emitted coherent or non-coherent light? Explain.

Q.15 Explain structure and energy band diagram of PIN photodetector?

3.16 Answers to Exercise

Ans.2: *Tunneling* is a *quantum mechanical phenomenon* which Occurs when an electron passes through a potential barrier without having enough energy to do so caused by the wave nature of electron.

Ans.5: It is basically a voltage variable capacitor in which the capacitance of the pn junction is controlled by the amount of reverse bias voltage.

Ans.9: Ohmic contact is a junction which has linear current-voltage relationship and electrically behaves like a simple homogeneous metal bar in which current can flow in both directions. Schottky or rectifying contact is defined as one in which current flows in one direction i.e. I-V characteristics similar to that of PN junction.

Ans.10: Gallium arsenide (GaAs), Indium phosphide (InP) and cadmium telluride (CdTe)

Ans.14: Incoherent Light

References and Suggested Readings

1. Jasprit Singh, "Semiconductor Devices Basic Principles". John Wiley & Sons (Asia) Pvt. Ltd. (2001).
2. M.S. Tyagi, "Introduction to Semiconductor Materials and Devices". John Wiley & Sons (Asia) Pvt. Ltd. (2000).
3. S.M.Sze, "Physics of Semiconductor Devices 2nd Edition". John Wiley & Sons (Asia) Pvt. Ltd. (2002).
4. S.O.Kasap, "Principles of Electronic Materials and Devices SECOND EDITION". Tata Mc Graw – Hill NEW DELHI (2002).
5. Donald A Neamen, "Semiconductor Physics and Devices Basic Principles Third Edition". Tata Mc Graw – Hill NEW DELHI (2002).
6. Michael Shur, "Physics of Semiconductor Devices 2nd Edition". Prentice Hall of India New Delhi (2002).
7. Ben. G. Streetman, Sanjay Banerjee, "Solid State Electronic Devices First Edition". Prentice Hall of India New Delhi (2002).
8. Boylestad, Nashelsky, " Electronic Devices and circuit theory 6th edition". Prentice Hall of India New Delhi (2001).
9. Samuel Y. Liao, "Microwave Devices and Circuits", Pearson Education (2004).

UNIT-4
Bipolar Junction Transistor

Structure of the Unit

4.0 Objective

4.1 Introduction

4.2 Transistor structure and symbols

4.2.1 Physical structure

4.2.2 Two-diode Model

4.3 Transistor Currents and voltages

4.4 Basic Transistor Operation

4.4.1 BJT Conditions

4.5 Self Learning Exercises-I

4.6 Transistor Circuit Configurations

4.6.1 DC characteristics

4.6.2 Common-base (CB) configuration

4.6.3 Common-emitter (CE) configuration

4.6.4 Common-collector (CC) configuration

4.7 Self Learning Exercises-II

4.8 Summary

4.9 Glossary

4.10 Answers to self learning exercises

4.11 Exercises

4.12 Answers to exercise

References and Suggested Readings

4.0 Objectives

This chapter deals with the basic concepts concerning p-type and n-type

semiconductors are applied to analysis as the pnp and npn transistor. The physical structure, operation, gain, and impedance characteristics of typical transistors are considered. The chapter also discusses as specific application of four- terminal network analysis to the transistor with their input/output characteristics and explores its use in amplifier and fundamental switching circuits. The comparison between three main configurations CBC, CEC, and CCC are discussed in briefly.

4.1 Introduction

In 1947, Bardeen, Brattain, and Shockley invented the first (bipolar) transistor and awarded the Nobel Prize in Physics in 1956. The transistor is one of the most important and widely used semiconductor device which is main building block "element" of electronics. It comes in two main types; the bipolar junction transistor (BJT) and the field effect transistor (FET) and we will focus only the BJT. The operation of bipolar junction transistor depends on the flow of majority and minority carriers. This chapter explains how the transistor works, as well as the performance parameters of a transistor. Here we will discuss the system characteristics of the BJT configuration and explore its main use as an amplifier and switching circuits.

A Bipolar Junction Transistor (BJT) is a three layer / terminal semiconductor device containing three differently doped regions. The word transistor comes from (transfer+ resistor).These three differently doped semiconductor regions are called as the emitter, base and collector region respectively. Each semiconductor region is connected to a terminal, which is labelled as emitter (E), base (B) and collector (C). Both electrons and holes participate in the conduction process, which is the reason for the name *bipolar*. The regions of transistor are made of *n*–type or *p*–type material in the order of pnp or *npn.*

The **emitter** is a medium sized, heavily doped ($\sim 10^{19}$ cm^{-3}) region whose main purpose is to supply majority charge carriers (For example, electrons for n-type, holes for p-type) through the base and into the collector region of the transistor.

The **base** is a very thin, lightly doped ($\sim 10^{16}$ cm^{-3}) region which is sandwiched between the emitter and collector. Its main function is to control the flow of the majority carriers coming from the emitter to the collector.

The **collector** is thick, moderately (medium) doped (~10^{17} cm^{-3}) region is designed to collect the majority carriers coming from the emitter and passing through the base.

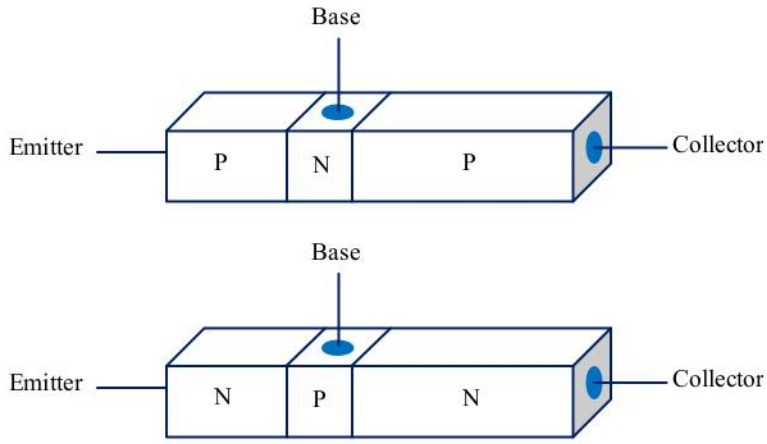


Figure:1

4.2 Transistor Structure and Symbols

4.2.1 Physical Structure

The lightly doped base layer is sandwiched between the collector and the emitter. The collector is moderately doped and the emitter is heavily doped. The base region is much thinner than the either the collector or emitter regions. Typical base widths are about 10^{-6} m. The collector region is usually thicker than the emitter because the largest amount of heat can be dissipated through it.

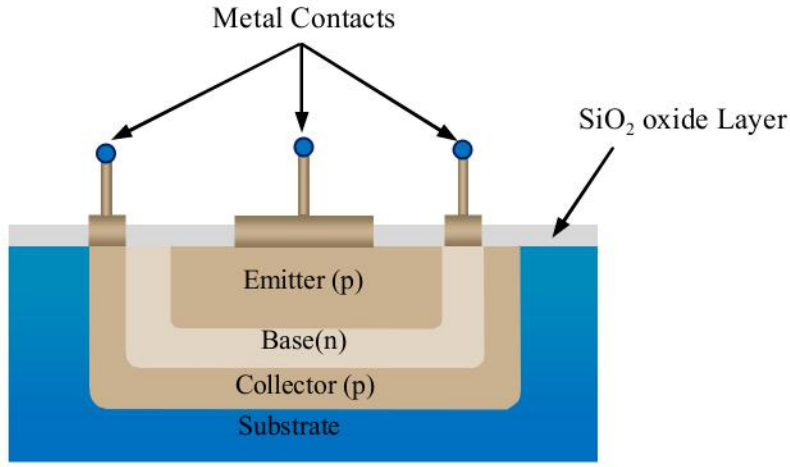


Figure 2. Physical Structure of BJT

The majority of current in a pnp transistor is carried by holes (positively charged particles) which move in the same direction as conventional current flow. In an npn transistor, the majority of current is carried by electrons. Transistor circuit symbols are shown below

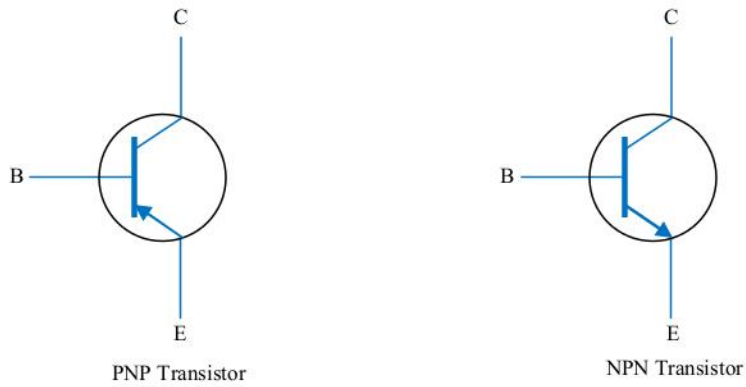


Figure 3: Schematic symbol of PNP and NPN transistor

The arrow on the emitter indicates the direction of conventional current flow under normal bias conditions. The direction of the arrow on the emitter distinguishes the NPN from the PNP transistor. If the arrow points in, the transistor is a PNP. On the other hand if the arrow points out, the transistor is an NPN. Transistor is a current controlled device in which base controls how much current flows between the collector and emitter (i.e., the small base current controls the larger collector current.), just like the gate controls the flow of water in the pipe.

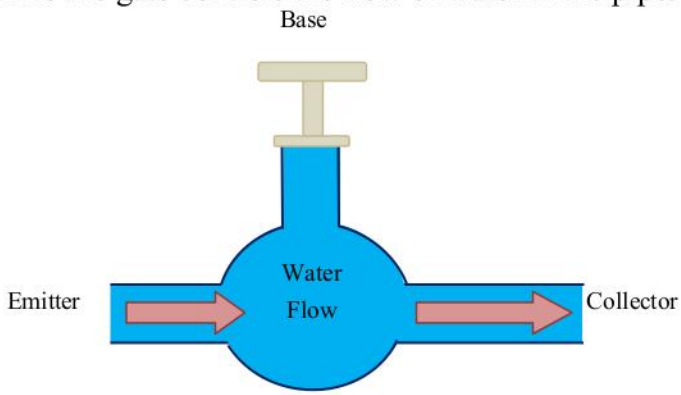


Figure 4: Water tap analogy

A transistor may be thought of as an electronic tap that is able to control a water flow with only small variations of the handle. The handle in the case of a transistor is called the base terminal and the in and out flow of water (current) through the pipe are called the emitter and collector respectively. A small change in base

voltage of the transistor results in wide change to the flow of current through the transistor.

4.2.2 Two-Diode Model

A Bipolar junction transistor (BJT) can be viewed as two pn junctions connected back to back to form np-pn or pn-np structures. Figure (5) illustrates the structure showing two p-n junctions.

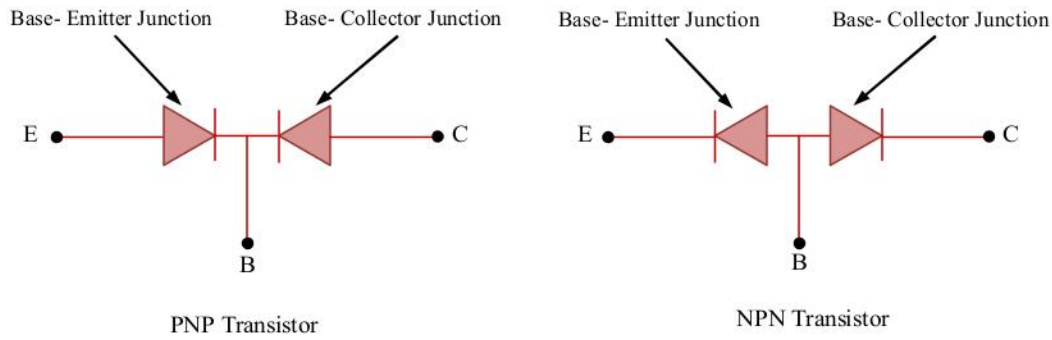


Figure 5: Two Diode Model

4.3 Transistor Currents and Voltages

There are three current parameters i_C, i_B, and i_E define the state of the bipolar junction transistor. However, because BJT has three terminals, KVL and KCL should hold for these terminals i.e.

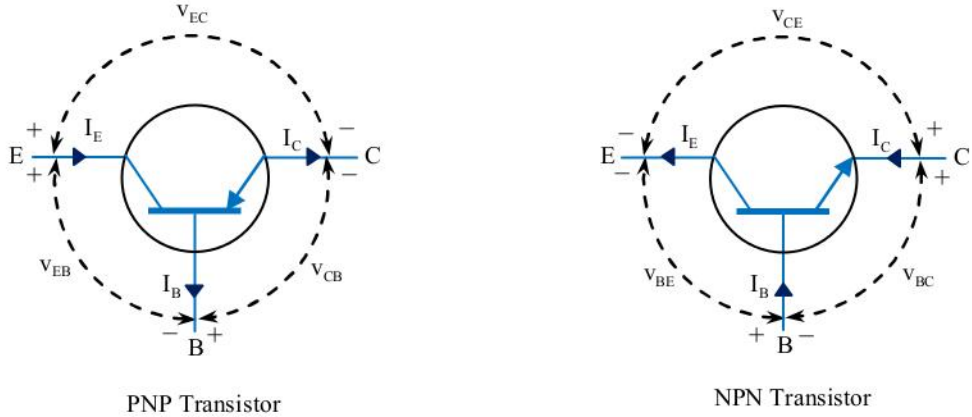


Figure:6

These diagrams show that the emitter current (I_E) is always the sum of the collector current (I_C) and the base current (IB), expressed as

$$I_E = I_C + I_B$$

There are three different types of voltages are involved in the description of transistors and transistor circuits. These are:

1] When the double subscripts are the same, it is a supply voltage (source).

Transistor supply voltages: V_{EE}, V_{BB}, V_{CC}

2] Single subscripts are used for node voltages with ground serving as the reference.

Transistor terminal voltages: V_E, V_B, V_C.

3] When two subscripts are different, this shows the voltage measured between two terminals.

Voltages between two points: V_{BE}, V_{CE}, V_{CB}.

From KVL only we find:

$$V_{CE}=V_{BE}+V_{CB} \qquad \text{(npn)}$$

$$V_{EC}=V_{EB}+V_{BC} \qquad \text{(pnp)}$$

where

$$V_{CE} = V_C - V_E \qquad V_{BE} = V_B - V_E \qquad V_{CB} = V_C - V_B$$

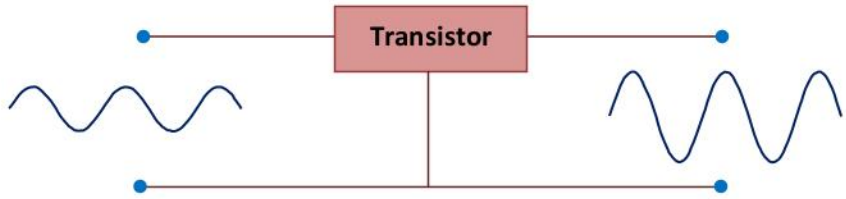
$$V_{EC} = V_E - V_C \qquad V_{EB} = V_E - V_B \qquad V_{BC} = V_B - V_C$$

4.4 Basic Transistor Operation

There are two main uses of transistor as switches (memory, microprocessor etc.) and an amplifier (Sensor, radio etc.) makes it useful in electronics. In this section, we will focus on the basic operating modes of a BJT. The proper operation of transistor depends on the bias condition (forward or reverse) of the two pn junctions i.e. it is essential to apply voltages of correct polarity across two pn junctions. Four different modes of operation of the BJT are possible. These are

Mode of operation	**E-B Junction**	**C-B Junction**	**Applications**
Normal Active	Forward	Reverse	Amplifier
Saturation	Reverse	Forward	Switching application in digital electronics
Cut-off	Reverse	Reverse	
Inverse Active	Forward	Forward	Performance degradation

•*Active*: Is useful for amplifiers. Most common mode



•*Saturation*: Equivalent to an on state when transistor is used as a closed switch.



•*Cut -off*: Equivalent to an off state when transistor is used as an opened switch.



•*Inverse*: Rarely if ever used.

4.4.1 BJT Conditions

For proper working of BJT, the majority carriers diffuse from emitter into the base and then go on to the collector. For this, the emitter-base junction of the bipolar junction transistor is always forward biased (or turns on).

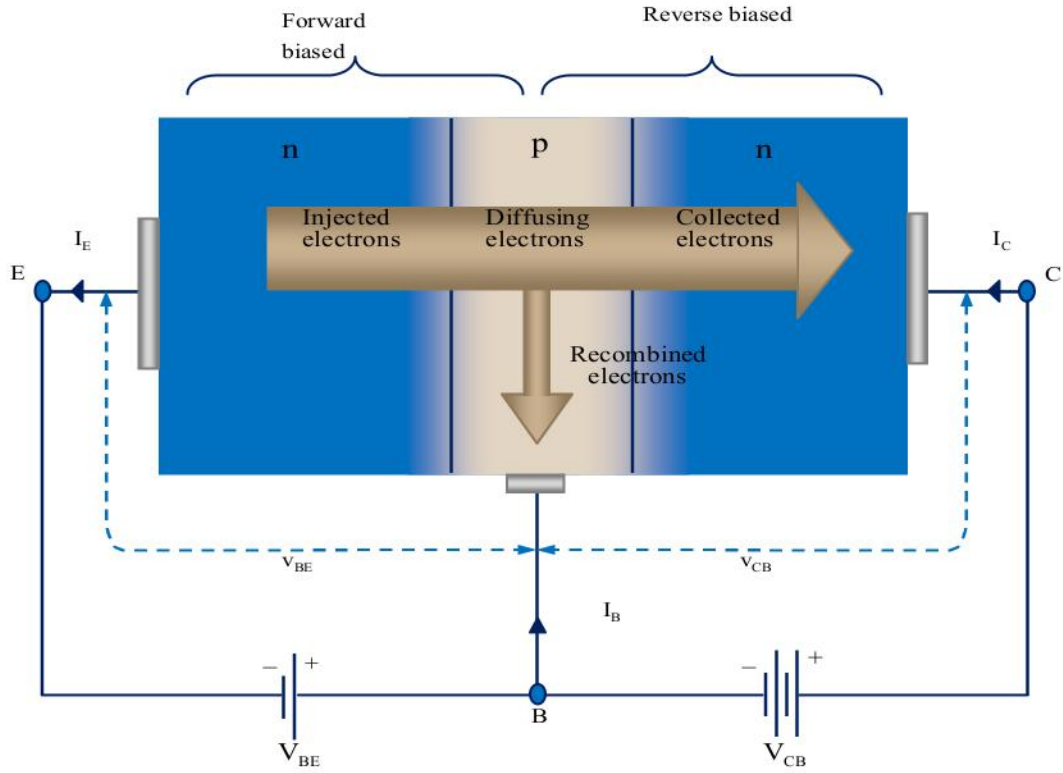


Figure:7

The collector-to-base is always reverse biased (or turns off) is referred as transistor biasing. Figure7 shows the current components of a bipolar junction transistor under normal bias.

With the forward biased of EB junction electrons (majority carriers for npn) to diffuse from emitter into base. As the base region is very thin and lightly doped, the electrons carriers injected from the emitter into the base region diffuse to the edge of the depletion region of CB junction and are captured by the high electric field of the reverse- biased CB junction. A small fraction of these electrons injected into the base from the emitter do recombine with the holes (majority carriers) in base region.

There are three current types flowing in bipolar junction transistor. There are emitter current (I_E), collector current (I_C), and base current (I_B). the emitter current is always the sum of the collector current and the base current, expressed as

$$I_E = I_C + I_B$$

It means that a small part (about 1-2%) of emitter current goes to supply base current due to recombination and the remaining major part (98-99%) goes to supply collector current. This statement is true regardless of *transistor type* or *transistor configuration*.

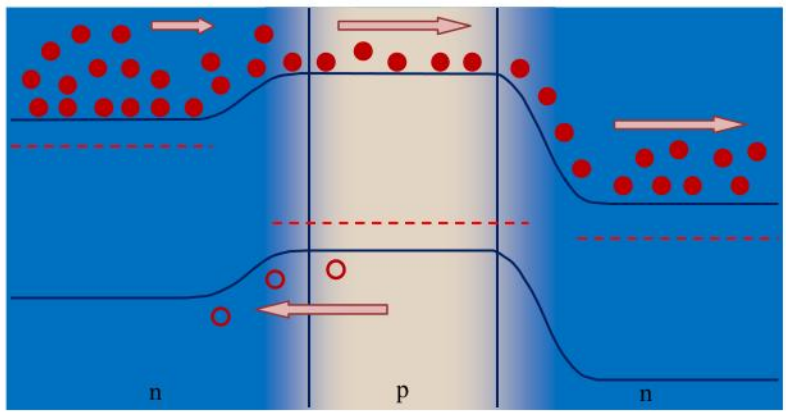


Figure 8: Energy band diagram of npn transistor under normal biasing

4.5 Self Learning Exercise-I

Q.1 Draw the circuit symbol of pnp and npn transistor.

Q.2 What are the names of three terminals of a BJT transistor?

Q.3 Describe the impurity profile of three terminal of transistor.

Q.4 Sketch the diagram of physical structure of transistor?

Q.5 In what modes transistor acts like a switch?

4.6 Transistor Circuit Configurations

Basically, there are basically three types of circuit connections (called configurations) for operating a transistor with one terminal being common to both the input and output. These are

1] Common-base configuration (CBC)

2] Common-emitter configuration (CEC)

3] Common-collector configuration (CCC)

4.6.1 DC Characteristics

Four types of characteristics can be defined for each of the transistor operating configurations. These characteristics define relationships among DC currents and DC voltages at the transistor input and output.

(a) Input characteristics V_{IN} vs. I_{IN} for V_{OUT} = const.

(b) Output characteristics I_{OUT} vs. V_{OUT} for I_{IN} = const.

4.6.2 Common-Base (CB) Configuration

In common base configuration circuit (see figure), the base is grounded and it is used as the common terminal for both input and output. Emitter is used as an input terminal where as collector is the output terminal.

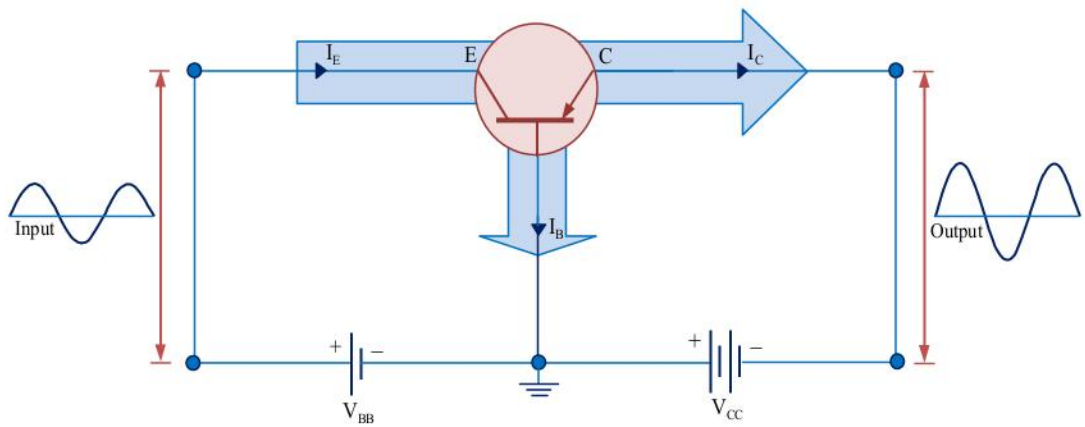


Figure9(a): Common base configuration circuit for pnp transistor

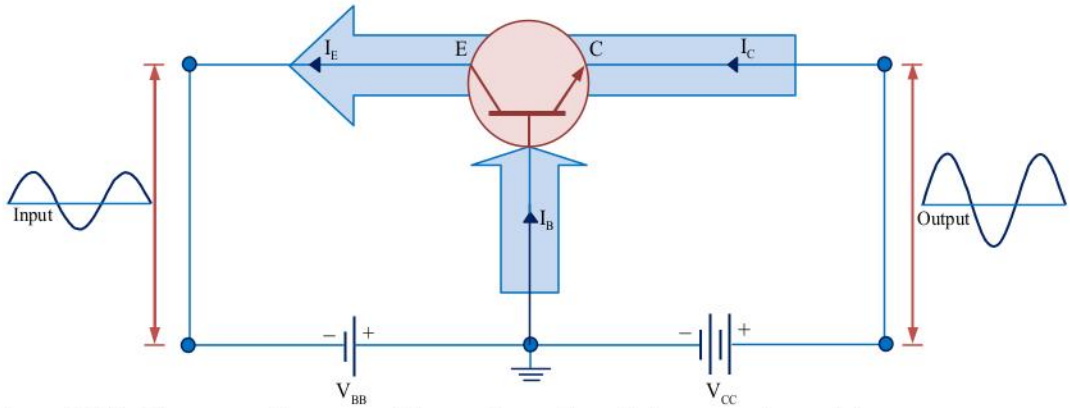


Figure9(b) Common base configuration circuit for npn transistor

●Input characteristics

It is defined as the characteristic curve drawn between input voltage to input current whereas output voltage is kept constant is shown in figure. To determine input characteristics for CBC, the emitter current I_E (input current) is increased by increasing emitter- base voltage (input voltage) V_{BE} with the collector-base voltage (output voltage) is kept constant.

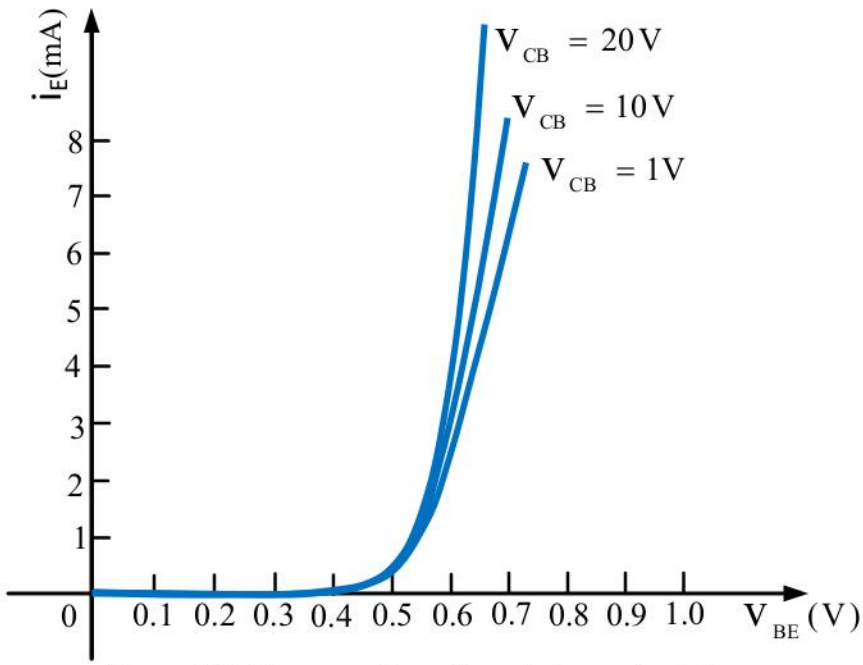


Figure10: Common base input characteristics

●Output characteristics

The output characteristics are obtained between output voltage to output current whereas input current is kept constant. To determine output characteristics, the output voltage V_{CB} is adjusted in suitable steps and the corresponding levels of collector current I_C are recorded at the fixed value of emitter current I_E. This is repeated for next convenient higher fixed values of I_E.

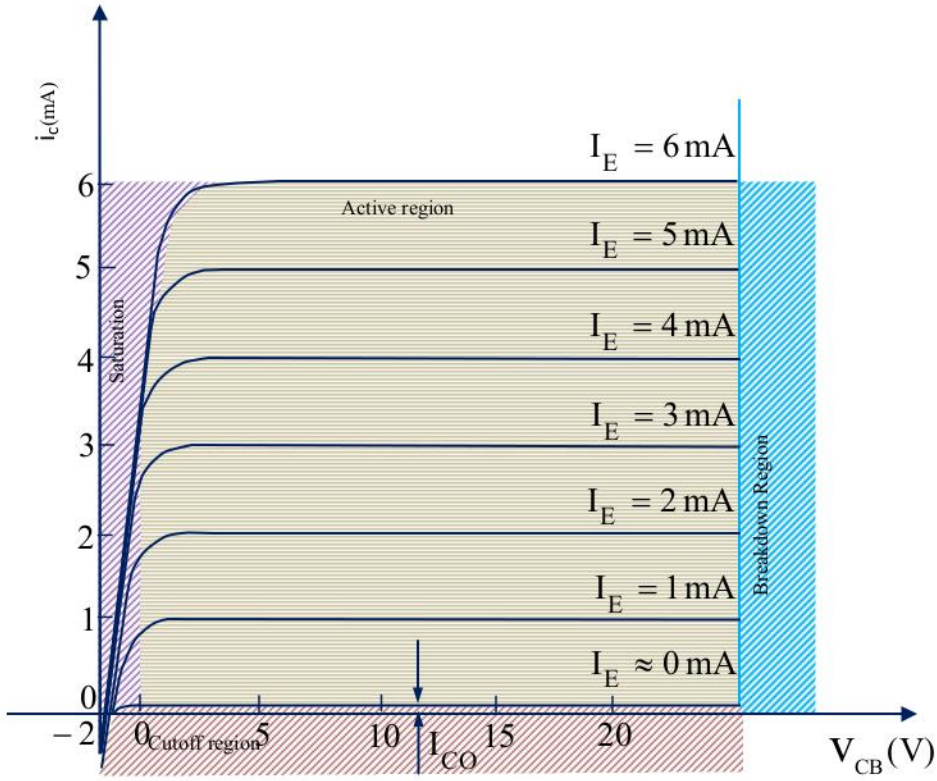


Figure11: Common base V-I (output) Characteristics curves

The ratio of collector current to emitter current is called α , which is also named as *h*-parameter h_{FB}. This parameter is commonly known as *common base gain*.

$$\alpha = \frac{I_C}{I_E}$$

and $$I_C = \alpha I_E + I_{CBO}$$

The typical value of α ranges from 0.95 to 0.99(recombination in the base reduces its value). α is the measure of quality of a transistor, for a good transistor, its value is closed to one.

α is the fraction of majority carriers(electrons for npn and holes for pnp transistor) that diffuse across the narrow Base region. 1- α is the fraction of majority carriers

that recombine with minority carriers in the Base region to create base current.

4.6.3 Common-Emitter (CE) Configuration

In common emitter configuration circuit (see figure), emitter is grounded and it is used as the common or reference to both input and output terminals. Base is used as an input terminal whereas collector is the output terminal. this configuration is almost used to amplifier designing due to high current and voltage gain.

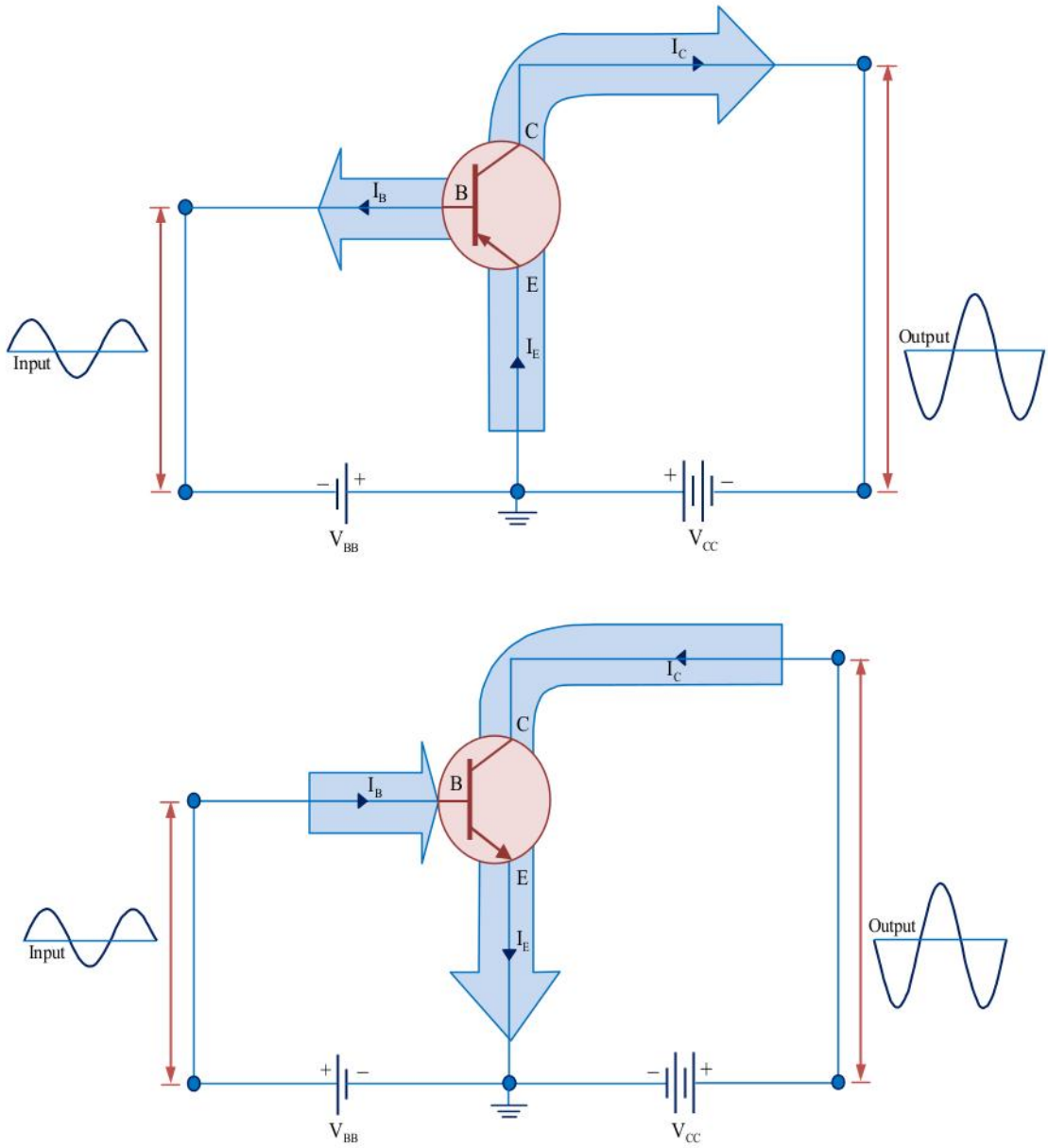


Figure12: Common emitter configuration for (a) pnp and (b) npn transistor

Two set of V-I characteristics are necessary to describe the behavior for CEC, input (base terminal) and output (collector terminal) parameter as shown in figure.

- Input characteristics

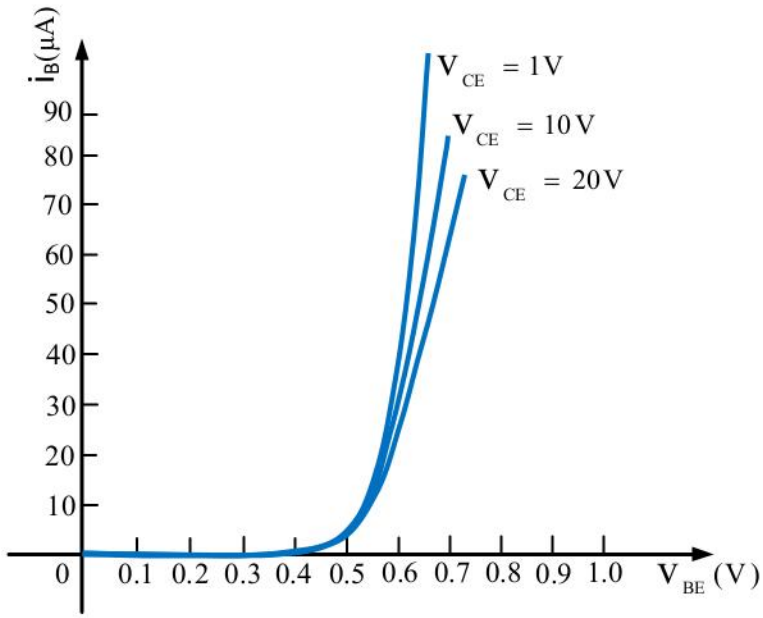


Figure13: Common Emitter Input Characteristics

- **Output Or Collector Characteristics**

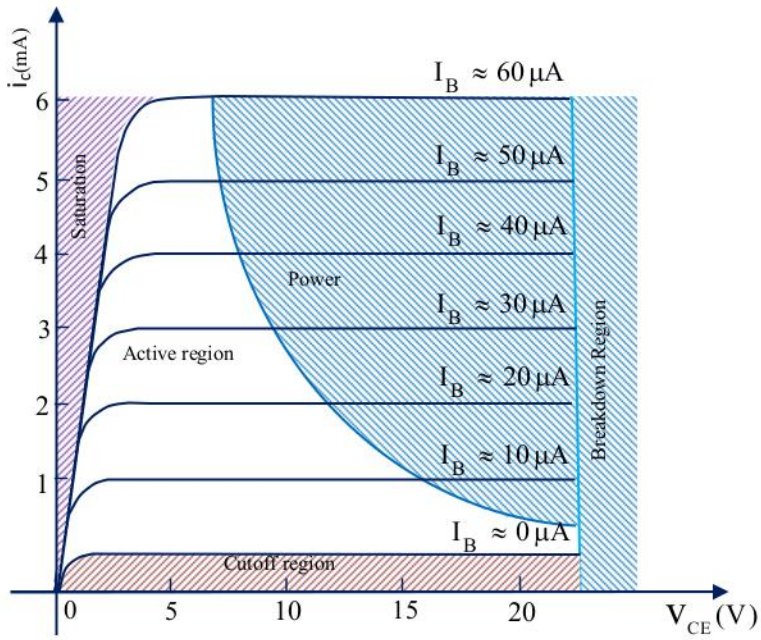


Figure14: Common Emitter Output Characteristics

The ratio of collector current to base current is β which also denoted as *h* parameter- h_{FE}. This parameter is commonly known as *common emitter gain*.

$$\beta = \frac{I_C}{I_B}$$

and $$I_C = \beta I_B + (\beta + 1) I_{CEO}$$

$$I_C = \beta I_B \qquad \text{(Neglecting } I_{CEO}\text{, very small)}$$

The typical range value of β is between 20 to 500.

Relationship between α and β

we know that $I_C = \alpha I_E$ and $I_C = \beta I_B$, we get

$$\frac{\beta}{\alpha} = \frac{I_E}{I_B} = \frac{I_C + I_B}{I_B} \quad \Rightarrow \quad \frac{\beta}{\alpha} = \frac{I_C}{I_B} + 1 = \beta + 1$$

which yields

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

In practice the most important transistor characteristic is the output characteristic which is used to determine to the transistor operating regions (see Fig.). These regions of BJT operation are as follows:

Cut-off region: In this region, the transistor is off and acts like an open switch. There is no conduction between the collector and the emitter. (I_B = 0 therefore I_C = 0). The region lying below the I_B = 0 curve; in this region, both transistor junctions (base –emitter and emitter-collector) must reverse biased.

Saturation region: In this region, the transistor is on and acts like a closed switch (short circuit). The collector current varies very little with a change in the base current in the saturation region. The V_{CE} is small, a few tenths of volt. Both transistor junctions are forward biased.

Active region: In this region the transistor can be an amplifier and lying between the saturation and the cut-off regions with the flat fragments of the output characteristics. The base-emitter junction is forward biased and the base-collector junction is reversed biased.The collector current is proportional to the base current ($I_C = \beta I_B$).

4.6.4 Common-Collector (CC) Configuration

In common collector configuration (CCC) circuit, the collector is grounded and it is used as the common terminal for both input and output. Base is used as an input terminal whereas emitter is the output terminal.

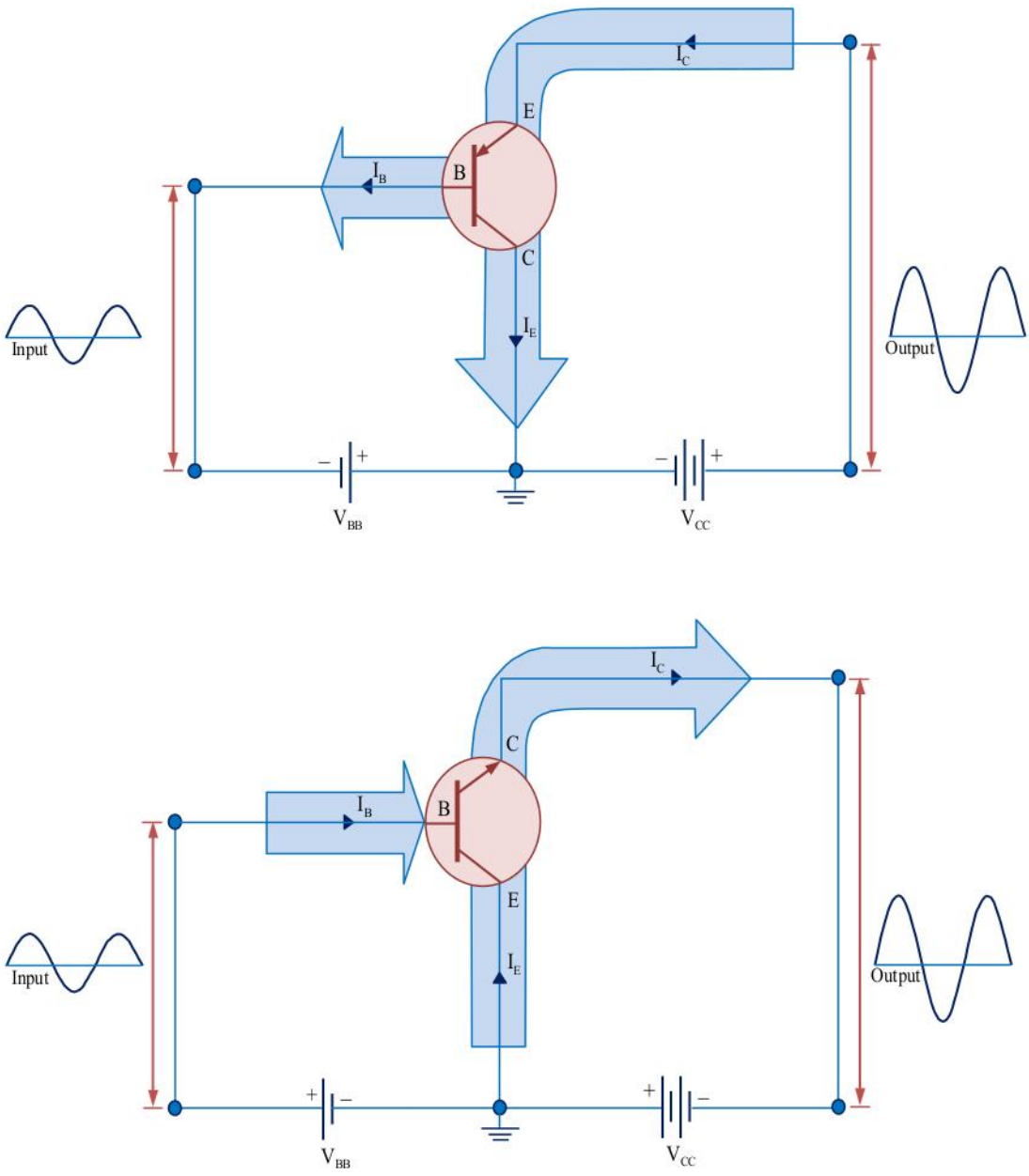


Figure15: Common collector configuration for (a) pnp and (b) npn transistor

●Input Characteristics

To obtain the input characteristics, the emitter - collector voltage V_{EC} is kept constant and base current I_B is increased with increasing V_{BC}. This is repeated for higher fixed values of V_{CE}.

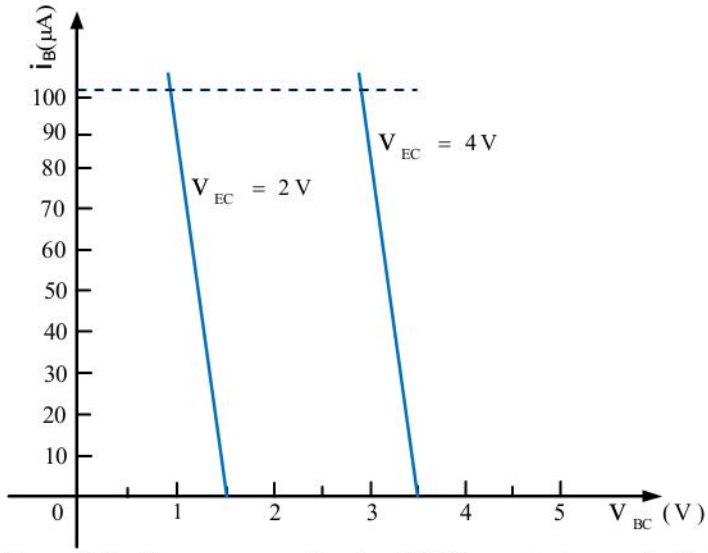


Figure16: Common collector V-I input characteristics

●Output Characteristics

Similarly output characteristics can be determined that show how the collector current, I_C, varies with the collector-to emitter voltage, V_{CE}, for specified values of base current, I_B.

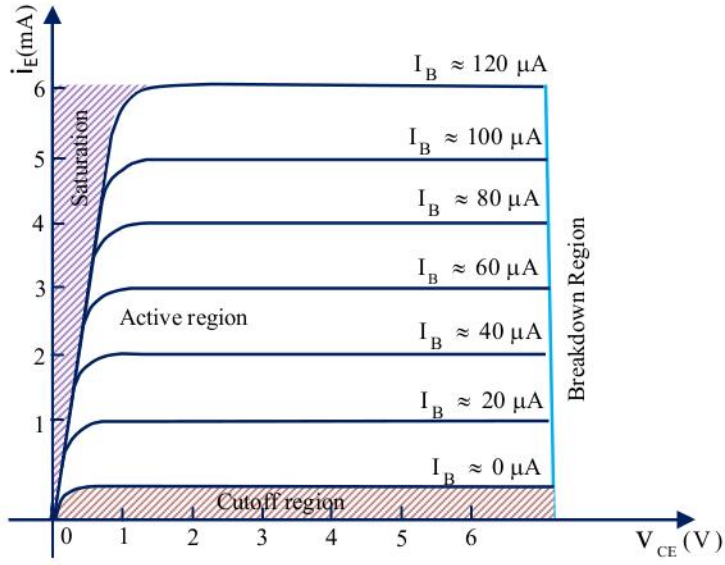


Figure17: Common collector V-I output characteristics

The current gain for CCC is given as

$$\text{Currentgain}(\gamma) = \frac{I_E}{I_B} = \frac{I_C + I_B}{I_B}$$

$$\Rightarrow \frac{I_C}{I_B} + 1 = \beta + 1$$

Table: Comparison between CBC, CBE and CCC configuration

Circuit Configuration	**Common Base**	**Common Emitter**	**Common Collector**
Current Gain	Low (0.99)	Medium (50)	Highest (60)
Voltage gain	Highest (200 to 2000)	High (100 to 500)	Low (Less than 1)
Power gain	Medium (200 to 1000)	Highest (200 to 20,000)	Low (20 to 80)
Input Impedance (R_{in})	Very Low (15Ω to150 Ω)	Medium (1k Ω to 5k Ω)	High (2kΩ to 500k Ω)
Output Impedance (R_{out})	Very High (250Ω to1M Ω)	High (40 Ω to 60k Ω)	Very Low (25 Ω to 1k Ω)
Phase Shift	0° (No change)	180° (Phase Reversal)	0° (No change)

4.7 Self Learning Exercise-II

Q.1 In what region transistor is operated as an amplifier?

Q.2 What are the formula for Alfa (α) and beta (β)?

Q3. Which transistor configuration provides a phase reversal between the input and output signals?

Q.4 Explain the conditions in cutoff region.

Q.5 What are the three transistor configurations?

4.8 Summary

In this chapter we have discussed the one of the most important semiconductor deice the BJT (bipolar junction transistor). The BJT (bipolar junction transistor) is constructed with three regions: base, collector, and emitter.

The BJT can be liked two back to back connected pn junctions, the base-emitter junction and the base-collector junction. The base region is very thin and lightly doped compared to the collector and emitter regions.

There are three currents in the transistor are the base current (I_B), emitter current (I_E), and collector current (I_C) in which I_B is very small compared to I_C and I_E.

To operate as an amplifier, the base-emitter junction must be forward-biased and the base-collector junction must be reverse-biased.

A transistor can be operated as an electronic switch in cutoff and saturation region.

Current gain (amplification factor) is the ratio of output current to input current.

4.9 Glossary

BJT: Bipolar Junction Transistor

α : The ratio of collector current to emitter current is called α

β : The ratio of collector current to base current is β which also denoted as *h* parameter- h_{FE}. This parameter is commonly known as *common emitter gain*.

4.10 Answer to Self Learning Exercises

Answers to Self Learning Exercise -I

Ans.1: see figure

Ans.2: emitter, base and collector

Ans.3: emitter-heavily doped, base- medium and collector- low

Ans.4: see figure

Ans.5: saturation and cutoff region

Answers to Self Learning Exercise -II

Ans.1: Active region

Ans.2: $\alpha = \frac{I_C}{I_E}$ and $\beta = \frac{I_C}{I_B}$

Ans.3: Common emitter configuration

Ans.4: Both transistor junctions (base–emitter and emitter-collector) must reverse biased.

Ans.5: Common-base configuration (CBC), Common-emitter configuration (CEC) and Common-collector configuration (CCC)

4.11 Exercise

Section A :Very Short Answer Type Questions

Q.1 What is the majority carriers in the emitter region of an *npn* transistor called?

Q.2 What is the formula for GAMMA (γ) in CCC?

Q.3 Which region of BJT is made thin and is lightly doped?

Q.4 What kind of biasing is required for the normal operation to the transistor when it is used as an amplifier?

Q.5 Under what condition a transistor acts like an open switch?

Section B: Short Answer Type Questions

Q.6 Describe BJT switching operation

Q.7 How is voltage gain defined?

Q.8 Which transistor configuration has a current gain of less than 1?

Q.9 What bias conditions must exist for a transistor to operate as an amplifier?

Q.10 Define current amplification factor of transistor in CE mode.

Section C :Long Answer Type Questions

Q.11 Explain the cause of a thin, lightly doped base region.

Q.12 In a transistor, the base current is 1 percent of the 40 mA emitter current. Determine the collector current.

Q.13 In a common emitter configuration (CEC) circuit, the collector current is 99 percent of 30mA emitter current. Determine the current amplification ratio.

Q.14 Explain phase reversal between the input and output signals in CEC configuration?

4.12 Answers to Exercise

Ans.1: electrons

Ans.2: $\gamma = \beta + 1$

Ans.3: Base.

Ans.4: Emitter-base junction is forward biased and collector-base junction is reverse biased.

Ans.5: When the transistor is in cut off state it acts as an open switch.

Ans.6: See CEC.

Ans.7: Voltage gain is defined as the ratio of the output voltage to input voltage.

Ans.8: Common base configuration (CCC).

Ans.9: base- emitter junction is forward bias and collector- base junction is reversed bias.

Ans.10: The current gain (β) is defined as the ratio of change in collector current to corresponding change in base current at constant V_{CE}.

Ans.11: large number of majority carriers from emitter can diffuse across the base and be swept into collector.

Ans.12: I_B=0.4mA and I_E=40mA therefore,

$I_C = I_E - I_B = (40-0.4) = 39.6$ mA

Ans.13: $I_C = \frac{30 \times 99}{100} = 29.7$mA and I_E=30mA therefore,

$I_B = I_E - I_C = (30-29.7) = 0.3$ mA

$\beta = \frac{I_C}{I_B} = \frac{29.7}{0.3} = 99$

Ans.14: In a common-emitter configuration (CEC), the input signal is applied between the base – emitter junction , during the positive cycle of input signal, the base to swing positive with respect to the emitter. This action decreases forward bias which reduces collector current (I_C) and increases

collector voltage which makes V_C more negative. During the negative cycle of the input signal, the base became more negative with respect to the emitter. This increases forward bias and allows more current carriers to be released from the emitter, which results in an increase in collector current and a decrease the collector voltage which makes the V_C more positive.

References and Suggested Readings

1. Jasprit Singh, "Semiconductor Devices Basic Principles". John Wiley & Sons (Asia) Pvt. Ltd. (2001).
2. M.S. Tyagi, "Introduction to Semiconductor Materials and Devices". John Wiley & Sons (Asia) Pvt. Ltd. (2000).
3. S.M.Sze, "Physics of Semiconductor Devices 2nd Edition". John Wiley & Sons (Asia) Pvt. Ltd. (2002).
4. S.O.Kasap, "Principles of Electronic Materials and Devices Second Edition". Tata Mc Graw – Hill New Delhi (2002).
5. Donald A Neamen, "Semiconductor Physics and Devices Basic Principles Third Edition". Tata Mc Graw – Hill New Delhi (2002).
6. Michael Shur, "Physics of Semiconductor Devices 2nd Edition". Prentice Hall of India New Delhi (2002).
7. Ben. G. Streetman, Sanjay Banerjee, "Solid State Electronic Devices First Edition". Prentice Hall of India New Delhi (2002).
8. Boylestad, Nashelsky, " Electronic Devices and Circuit Theory 6th Edition". Prentice Hall of India New Delhi (2001).

UNIT-5
Transistor Hybrid Model

Structure of the Unit

5.0 Objectives

5.1 Introduction of h-parameters

5.2 CB, CE and CC Configurations of the transistor and their equivalent circuits.

5.3 Analysis of a transistor Amplifier using h-parameters.

5.4 Transistor Amplifier circuits in different configurations and their analyses

5.5 Self Learning Exercise-I

5.6 Emitter follower

5.7 Miller's theorem and it dual

5.8 Self Learning Exercise – II

5.9 Summary

5.10 Glossary

5.11 Answers the self learning exercises

5.12 Exercise

References and Suggested Readings

5.0 Objectives

This unit introduces you to the two-port devices and hybrid model of transistor. After going through this unit you will h parameters. CB, CE and CC configuration of the transistors and their equivalent circuits analysis of a transistor amplifier using h-parameters, transistor amplifier circuits in different configurations and their analysis and Miller theorem and its dual.

5.1 Introduction of h- Parameters

The study of four terminal or two port network is usually done through the quantities measured at the input and output terminals. Out of the four quantities

related with a four terminal network, the input current I_1 input voltage V_1, output current I_2 and output voltage V_2 two are treated as independent variables and the remaining two are expressed in terms of these independent variables. The choice of independent variables is not arbitrary but based on certain logic. In transistor circuits the input current I_1 and output voltage V_2 control the behavior of the circuit. Hence the choice of these quantities as independent variables is more appropriate. Accordingly

$$V_1 = h_{11}I_1 + h_{12}V_2 \quad (1)$$

$$I_2 = h_{21}I_1 + h_{22}V_2 \quad (2)$$

The quantities h_{11}, h_{12}, h_{21} and h_{22} appearing in relations (1) and (2) are called h-parameters or hybrid parameters.

$h_{11} = \frac{V_1}{I_1}$ when $V_2 = 0$, i.e., it represents input impedance when output terminals are short circuited. The unit of h_{11} will be Ω.

$h_{12} = \frac{V_1}{V_2}$ When $I_1 = 0$, i.e., it represents the fraction of output voltage at the input when the input is open. It is also called reverse voltage ratio. It is a dimensionless quantity.

$h_{21} = \frac{I_2}{I_1}$ When $V_2 = 0$, i.e., it represents forward current gain when the output is short circuited. It is also a dimensionless quantity.

$h_{22} = \frac{I_2}{V_2}$ When $I_1 = 0$, i.e., It represents output admittance when input terminals are open. It unit is mho.

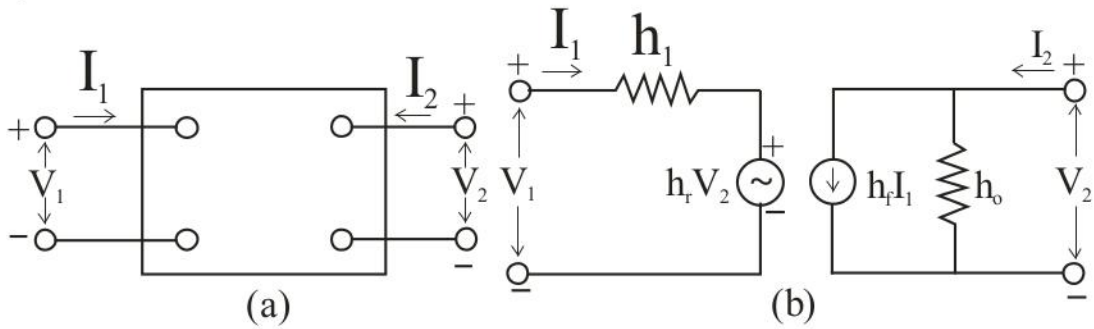


Figure: 5.1

h-parameters are also called hybrid parameters because in defining these both open

and short circuit configurations are used and the dimensions of these parameters are not same and also the units of these parameters is mixed one.

For transistor circuits in place of h_{11}, h_{12}, h_{21} and h_{22} it is convenient to write h_i, h_r, h_f and h_o respectively and we shall use these notations in our discussions. The subscripts I, r, f and o have physical meaning these represent input, reverse, forward and output respectively. In Fig. 5.1 a four terminal two-port network is shown. It can be given an analytical form with the help of h-parameters. According to relation (1) the voltage V_1 is equal to the sum of fall of voltage $h_i I_1$ due to the current I_1 and a generator voltage $h_r V_2$. According to relation (2) the output current I_2 is equal to the sum of current produced by voltage V_2 in the admittance h_0 and a current $h_f I_1$ produced by a current generator. Thus the equations (1) and (2) can be represented by Fig. (5.1(b)). Hence Fig. (5.1 (b)) is the equivalent circuit of the two port network given in Fig. 5.1 (a).

5.2 CB, CE and CC Configurations of the Transistor and their Equivalent Circuits

Transistor is a three terminal (base, emitter and collector) device. If its one terminal is common to both input and output circuits. It acts as a four terminal network. On the basis of common terminal three configurations are possible for a transistor.

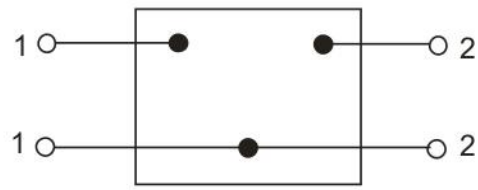


Figure: 5.2

(i) Common Base Configuration: In this configuration the base terminal is common to input and output and it is kept grounded. So it is also called grounded base configuration. The emitter and base terminals act as input terminals and collector and base terminals act as output terminals. Fig. (5.3). in this configuration

$V_1 = V_{EB}$, $V_2 = V_{CB}$

$I_1 = I_E$ and $I_2 = I_C$

$h_{11} = h_{ib}$, $h_{12} = h_{rb}$

$h_{21} = h_{fb}$, $h_{22} = h_{ob}$

The subscript 'b' represents common base.

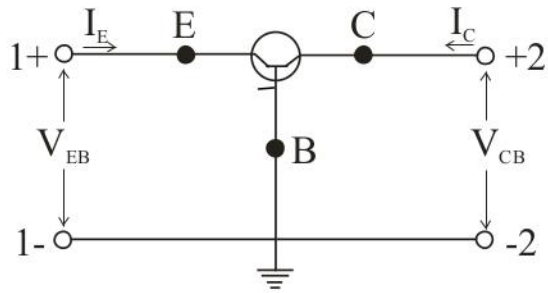


Figure: (5.3)

According to equations (1) and (2) of section (5.1) for the CB configuration

$$V_{EB} - h_{ib} I_E + h_{rb} V_{CB}$$

$$I_C = h_{fb} I_E + h_{ob} V_{CB}$$

For small amplitude signals, the h-parameters can be considered constant, so that

$$\Delta V_{EB} = h_{ib} \Delta I_E + h_{rb} \Delta V_{CB}$$

$$\Delta I_C = h_{fb} \Delta I_E + h_{ob} \Delta V_{CB}$$

The variations in voltage and currents can be written as small letters. Therefore we can write.

$$v_{eb} = h_{ib} i_e + h_{rb} v_{cb}$$

$$i_c = h_{fb} i_e + h_{ob} v_{cb}$$

On the basis of the above relations for small amplitude alternating signals, the equivalent circuit will be as shown in Fig. (5.3)

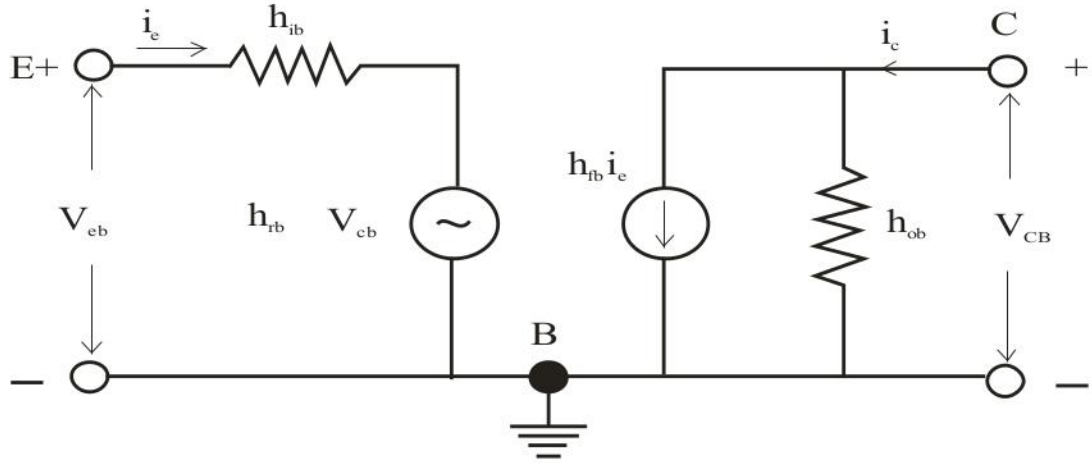


Figure: 5.3

(ii) Common Emitter Configuration: In this configuration the grounded emitter is common to input and output. This configuration is shown in fig 5.4. For small amplitude signals the following relations will true in terms of h-parameters.

$$v_{be} = h_{ie}\, i_b + h_{re}\, v_{ce}$$

$$i_c = h_{fe}\, i_b + h_{oe}\, v_{ce}$$

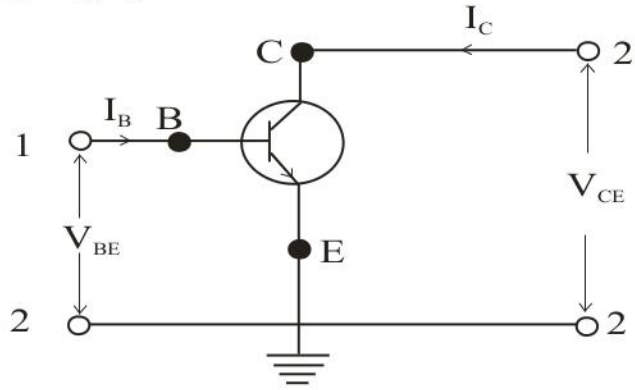


Figure : 5.4

According to this relation the equivalent circuit is given in fig. 5.5. In place of hoe an output resistance r_{oe} can also be shown but it should be remembered that h_{oe} is conductance (or in general admittance) in mho and r_{oe} is resistance in ohms. For the forward current gain h_{fe} the symbol β is also used.

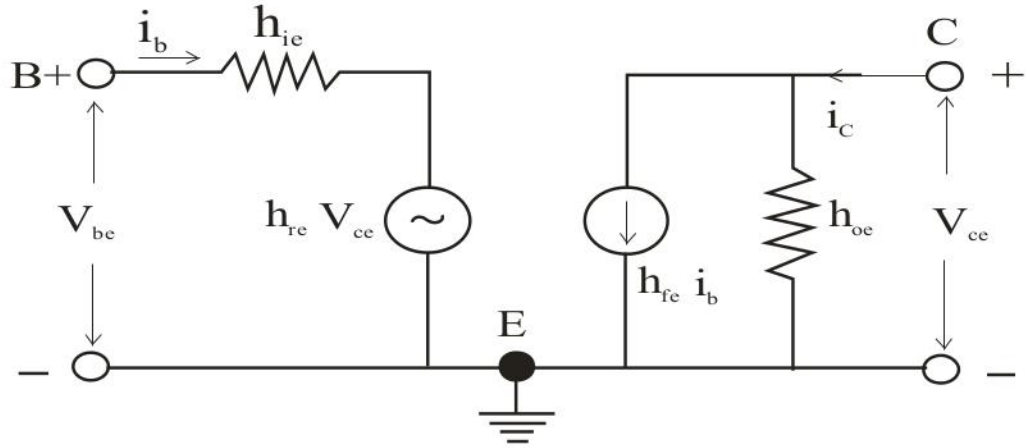


Figure (5.5)

(iii) Common Collector Configuration: In this configuration the collector is grounded and it is common in input and output fig. 5.6

For small amplitude alternating signals in common collector configuration

$$V_{bc} = h_{ic}\, i_b + h_{rc}\, V_{ec}$$

$$I_e = h_{fc}\, i_b + h_{oc}\, V_{ec}$$

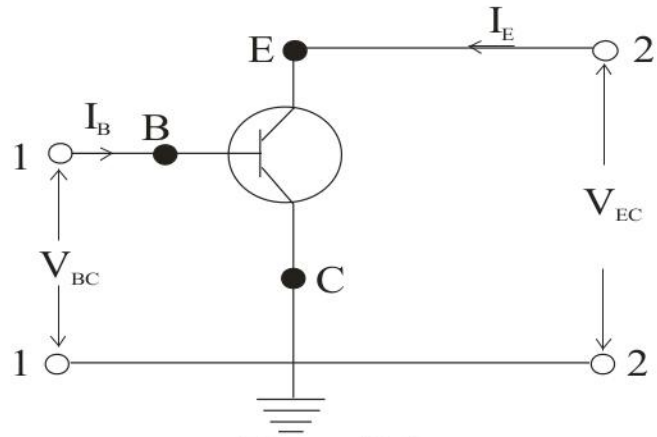


Figure: (5.6)

The equivalent circuit is shown in fig. 5.7

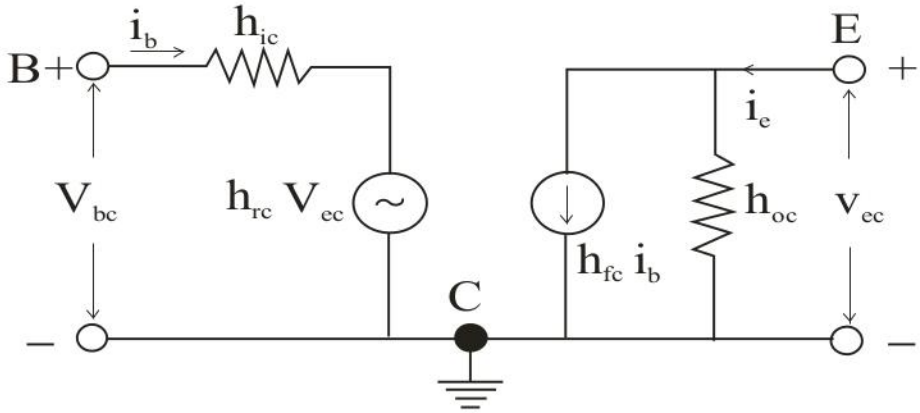


Figure 5.7

5.3 Analysis of A Transistor Amplifier Using h-Parameters

Transistor is a current operated device and it works as a four terminal network. To form a transistor amplifier the transistor is suitably biased and a signal voltage source V_s having a resistance R_s is connected in the input circuit and a load Z_L in the output circuit. A basic amplifier circuit is shown in fig. 5.8. This four terminal network c in BE for any configuration (CE, CB or CC) of the transistor and the load given here can be a pure resistance or impedance or any circuit.

In the analysis it is assumed that the h-parameters remain constant in the operating range. The equivalent circuits of Fig. 5.8 are given in Fig. 5.9 and 5.10. In Fig.(5.9) the input is in the form of a voltage source and in (5.10) it is in the form of a current source.

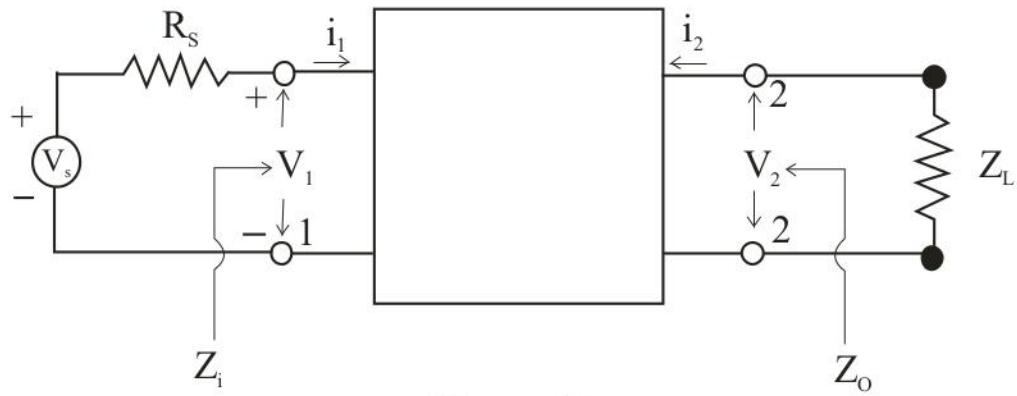


Figure: 5.8

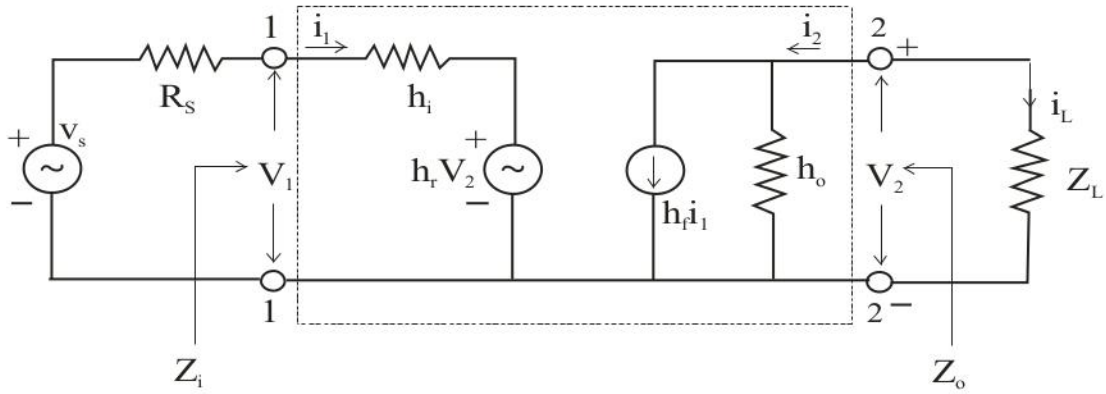


Figure: (5.9)

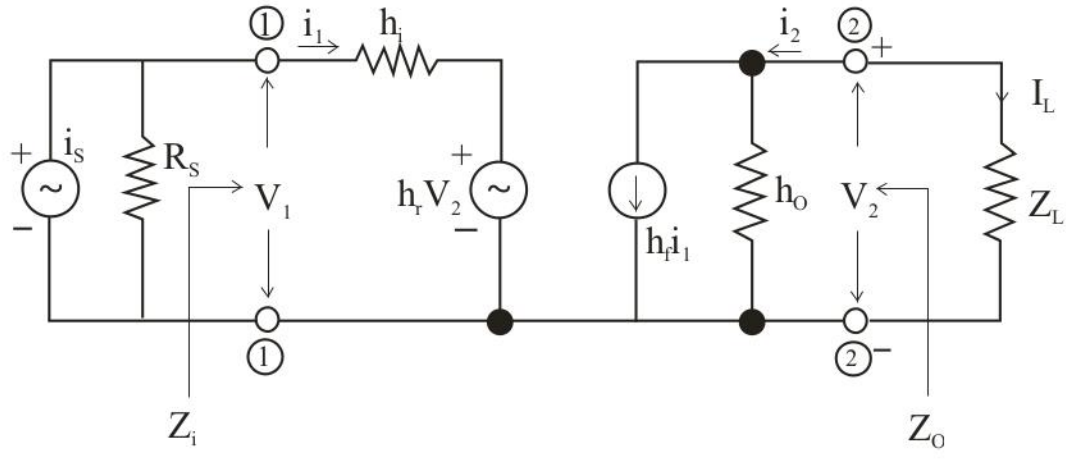


Figure: (5.10)

(i) Current gain or current amplification (A_i): For a transistor amplifier the ratio of output current and input current is called current gain. It is represented by

A_i.

Hence $A_i = \frac{i_2}{i_1}$

Using Kirchhoff's current law in the circuit of Fig. 5.9/5.10

$$i_2 = h_f\, i_1 + h_0\, V_2$$

But $V_2 = -\, i_2\, Z_L$

$\therefore$ $i_2 = h_f i_1 + h_o\,(-\, i_2 Z_L)$

So that $A_i = \frac{i_2}{i_1} = \frac{h_f}{(1 + h_o Z_L)}$ (1)

$$\therefore A_{iL} = \frac{I_L}{I_1} = -\frac{h_f}{(1 + h_o Z_L)}$$

(ii) Input Impedance Z_i : The input impedance of a transistor amplifier is the impedance observed at the input terminals of the transistor. Its value is equal to the ratio of voltage V1 at the input terminals to the current i1.

Hence $Z_i = \frac{V_1}{i_1}$

In the input circuit of Fig (5.9)/(5.10)

$$V_1 = h_i\, i_1 + h_r\, v_2$$

$\therefore$ $Z_i = \frac{h_i i_1 + h_r v_2}{i_1} = h_i + h_r \frac{v_2}{i_1}$

But $V_2 = -\, i_2 Z_L$ and $A_i = \frac{i_2}{i_1}$

$\therefore$ $z_i = h_i + h_r \left(\frac{-i_2 Z_L}{i_1} \right)$

$$= h_i - h_r\, A_i\, Z_L$$

Substituting the value of A_i from (1)

$\therefore$ $Z_i = h_i - \frac{h_r h_f Z_L}{1 + h_o Z_L}$ .... (2)

The value of $(h_r\, h_f)$ is very small and so the second term can be neglected in comparison to h_i i.e. $Z_i \approx h_i$

(iii) Voltage amplification or voltage gain : The ratio of output voltage V2 and input voltage V1 is called voltage amplification or voltage gain of the amplifier.

Hence voltage gain $A_v = \dfrac{\text{Output Voltage}}{\text{Input Voltage}} = \dfrac{V_2}{V_1}$

But $V_2 = - i_2 Z_L = A_i i_1 Z_L$

$$\therefore A_v = -\frac{A_i i_1 Z_L}{V_1} = -\frac{A_i Z_L}{Z_i}$$

$$\text{or } A_{VL} = A_{iL} \frac{Z_L}{Z_i}$$

Substituting the values of A_i and Z_i

$$A_v = -\left(\frac{h_f}{1+h_o Z_L}\right)\left(\frac{Z_L}{h_i - \dfrac{h_r h_f Z_L}{1+h_o Z_L}}\right)$$

$$= -\frac{h_f}{(1+h_o Z_L)} \times \frac{Z_L (1+h_o Z_L)}{h_i (1+h_o Z_L) - h_r h_f Z_L}$$

$$A_v = -\frac{h_f Z_L}{h_i + \Delta_h Z_L} \qquad (3)$$

where $\Delta_h = (h_i h_o - h_r h_f)$

The value of $\Delta_h Z_L$ is usually negligible in comparison to h_i hence

$$A_v = -\frac{h_f Z_L}{h_i} \qquad (4)$$

(iv) Output impedance Z_o : To obtain the value of output impedance the source voltage v_s in the circuit shown in Fig. (5.9) is made zero and the load impedance Z_L is replaced by a voltage sources V_2. In this condition if i_2 is current taken from the source V_2 then the ratio of V_2 and i_2 is equal to the output impedance, i.e.,

$$Z_o = \frac{V_2}{i_2}$$

$$i_2 = h_f i_1 + h_o v_2$$

In Fig (5.9) making $v_s = 0$ in the input circuit we have

$$R_s i_1 + h_i i_1 + h_r V_2 = 0$$

So that $i_1 = -\left(\frac{h_r v_2}{R_s + h_i}\right)$

Using the value of i_1 $i_2 = h_f -\left(\frac{h_r v_2}{R_s + h_i}\right) + h_o v_2$

$$= V_2\left(h_o - \frac{h_f h_r}{R_s + h_i}\right)$$

Hence $Z_o = \frac{v_2}{v_2\left(h_o - \frac{h_f h_r}{R_s + h_i}\right)}$

$$= \frac{(R_s + h_i)}{(h_o R_s + h_o h_i - h_f h_r)} \qquad (5)$$

Thus the output impedance is a function of resistance of the signal source in the input.

(v) Power gain A_p : For the transistor amplifier the ratio of the output ac power to the input ac power is called power gain. Hence

Power gain $A_p = \frac{\text{Output alternating power}}{\text{Input alternating power}}$

$$A_p = \frac{V_2 i_2}{V_1 i_1} = |A_v||A_i| \qquad (6)$$

i.e. the value of power gain is equal to the product of voltage gain Av and current gain A_i. Using the value of A_v.

$$A_p = \left(\frac{A_i Z_L}{Z_i}\right) A_i$$

$$= A_i^2 \left(\frac{Z_L}{Z_i}\right)$$

Substituting the value of A_i and Z_i

$$A_p = \left(\frac{h_f}{1 + h_o Z_L}\right)^2 \left(\frac{Z_L}{h_i - \frac{h_r h_f Z_L}{1 + h_o Z_L}}\right)$$

$$A_p = \frac{h_f^2 Z_L}{(1 + h_o Z_L)(h_i + Z_L \Delta_h)} \qquad (7)$$

In the above analysis the resistance R_s of the signal source has not been considered. In this state the quantities, current gain A_i and voltage gain A_v are respectively called internal transistor current gain and internal voltage gain. If the current gain and voltage gain are determined considering the resistance R_S of signal source also then these are respectively called overall current gain and overall voltage gain.

(a) Over all current gain : If the source at the input is a current generator is then its internal resistance R_s is taken in parallel to it as is shown in fig. 5.10. The overall current gain A is then

$$A_i = \frac{i_2}{i_s} = \frac{i_2}{i_1} = \frac{i_1}{i_s} = A_i \frac{i_1}{i_s} \tag{8}$$

From the input circuit of Fig. (5.10)

$$i_1 Z_i = (i_s - i_1) R_s$$

So that $\frac{i_1}{i_s} \times \frac{R_s}{Z_i + R_s}$

Hence $A_i = A_i \times \frac{R_s}{(Z_i + R_s)} = \frac{A_i R_s}{(Z_i + R_s)}$ (9)

If $R_s \to \infty$ then $A'_i = A_i$

Hence A_i is the current gain for an ideal current source.

(b) Overall voltage gain (A_v): If source voltage is V_s then voltage gain

$$A'_v = \frac{v_2}{v_s} = \left(\frac{v_2}{v_1}\right)\left(\frac{v_1}{v_s}\right)$$

$$A'_v = A_v \left(\frac{v_1}{v_s}\right) \tag{10}$$

In the input circuit of Fig. (5.9)

$$(R_s + Z_i)\, i_1 = v_s$$

But $i_1 = \frac{v_1}{Z_i}$

$$\therefore (R_s + Z_i) \frac{v_1}{Z_i} = v_s$$

or $\frac{v_1}{v_s} = \frac{Z_i}{(R_s + Z_i)}$ (11)

$$\therefore \quad A_v' = A_v\left(\frac{Z_i}{(R_s + Z_i)}\right) \tag{12}$$

But $A_v = -\frac{A_i Z_L}{Z_i}$

$$\therefore \quad A_v' = -\left(\frac{A_i Z_L}{Z_i}\right)\left(\frac{Z_i}{R_s + Z_i}\right)$$

$$= -A_i \frac{Z_L}{(R_s + Z_i)} \tag{13}$$

The equation (12). shows that if the resistance of source $R_s = 0$ then $A'_v = A_v$

i.e. with an ideal voltage source or with a source of zero internal resistance the voltage gain in equal to over all voltage gain.

5.4 Transistor Amplifier Circuits in Different Configurations and Their Analysis

A junction transistor can be used as an amplifier in three configurations:

(i) Common emitter transistor amplifier

(ii) Common base transistor amplifier

(iii) Common collector transistor amplifier

Now we will analyze these amplifiers using the method.

(i) Common emitter amplifier: In Fig. (5.11) the basic circuit of a common emitter amplifier using an NPN transistor is shown. The equivalent circuit for small amplitude alternating signals is given in fig. (5.12). the analysis of the equivalent circuit yields:

Current gain $A_{ie} = \frac{i_c}{i_b}$

But $i_c = h_{fe}\, i_b + h_{oe}\, v_o$

$= h_{fe}\, i_b - h_{oe}\, i_c\, R_L$ as $v_o = -\, i_c R_L$

or $i_c\,(1 + h_{oe} R_L) = h_{fe} i_b$

$$\therefore \quad A_{ie} = \frac{h_{fe}}{1 + h_{oe} R_L} \tag{1}$$

$$\approx h_{fe} \text{ as } h_{oe}\, R_L \ll 1$$

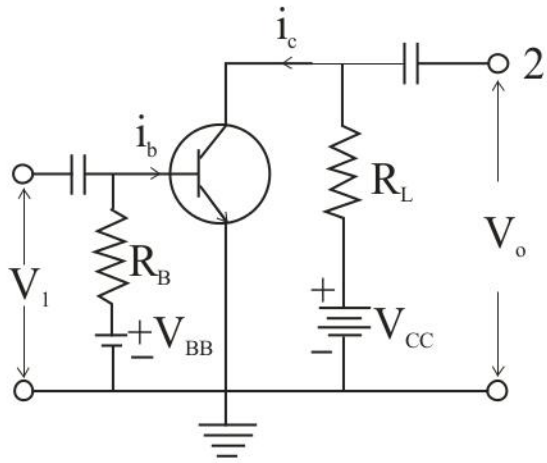


Figure: (5.11)

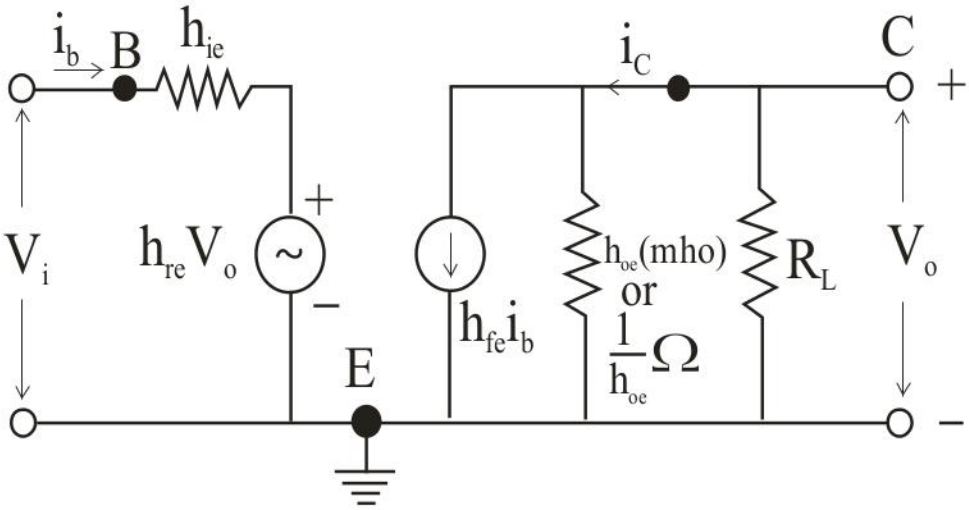


Figure: (5.12)

voltage gain $A_{ve} = \dfrac{V_o}{v_i}$

But $v_o = i_c R_L$

$v_i = i_b h_{ie} + h_{re} v_o = i_b h_{ie} - h_{re} i_c R_L$

$\therefore \quad A_{ve} = \dfrac{-i_c R_L}{i_b h_{ie} - h_{re} i_c R_L}$

$$= \frac{-h_{fe} R_L}{h_{ie} + \left(h_{ie} h_{oe} - h_{fe} h_{re}\right) R_L} = \frac{h_{fe} R_L}{h_{ie} + \Delta_h R_L} \qquad (2)$$

where $\Delta_h = \begin{vmatrix} h_{ie} & h_{re} \\ h_{fe} & h_{oe} \end{vmatrix} = \left(h_{ie} h_{oe} - h_{fe} h_{re}\right)$

Input resistance - by definition

$$R_{ie} = \frac{v_i}{i_b}$$

$$= \frac{i_b h_{ie} + h_{re} v_o}{i_b} = \frac{i_b h_{ie} + h_{re}(-i_c R_L)}{i_b}$$

$$= h_{ie} - A_{ie} h_{re} R_L = h_{ie} - \frac{h_{re} h_{fe} R_L}{1 + h_{oe} R_L} \qquad \ldots (3)$$

Output resistance $R_{oe} = \frac{v_o}{i_c}$

But $i_c = h_{fe} i_b + h_{oe} v_o$

and $i_b h_{ie} + h_{re} v_0 = 0$ or $i_b = -\frac{h_{re}}{h_{ie}} v_o$

$\therefore$ $$i_c = -\frac{h_{fe} h_{re}}{h_{ie}} v_o + h_{oe} v_o$$

$$= v_o \left(\frac{h_{oe} h_{ie} - h_{fe} h_{re}}{h_{ie}} \right)$$

$$= \frac{v_o \Delta h}{h_{ie}}$$

$\therefore$ $$R_{oe} = \frac{v_o h_{ie}}{v_o \Delta_h} = \frac{h_{ie}}{\Delta_h} \qquad \ldots (4)$$

If the resistance of signal source is not negligible then in place of h_{ie} we have to take $(h_{ie} + R_s)$ so that

$$R_{oe} = \frac{h_{ie} + R_s}{h_{oe}(h_{ie} + R_s) - h_{fe} h_{re}} = \frac{h_{ie} + R_s}{\Delta_h + h_{oe} R_s} \qquad \ldots\ldots (5)$$

(ii) Common base amplifier: The basic circuit of a common base amplifier and the equivalent circuit for small alternating signals are given in Fig. 5.13 and 5.14 respectively.

Current gain $A_{ib} = \frac{i_c}{i_c}$

$$i_c = h_{fb} i_e + h_{ob} v_o$$

$$= h_{fb}\, i_e + h_{ob}\,(-i_c R_L) \qquad [\because v_o = -\, i_c R_L]$$

$$\therefore \qquad i_c = \frac{h_{fb} i_e}{(1 + h_{ob} R_L)}$$

Hence $$A_{ib} = \frac{h_{fb}}{(1 + h_{ob} R_L)} \tag{6}$$

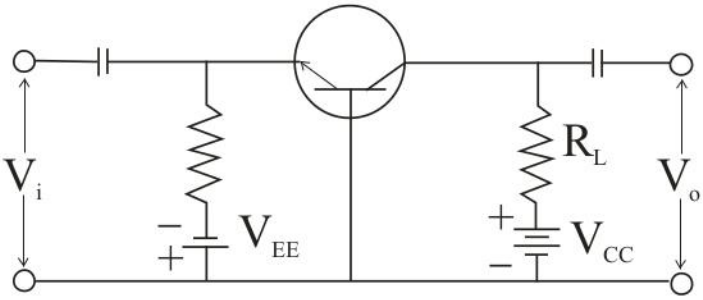


Figure: (5.13)

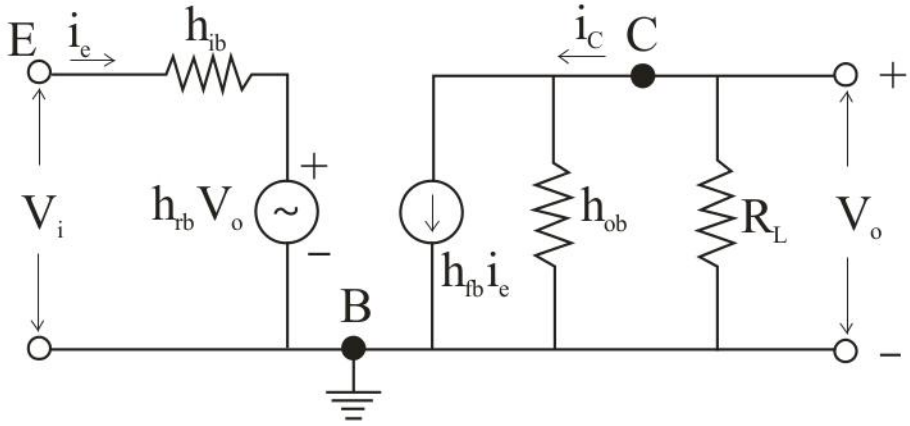


Figure: (5.14)

Voltage gain $$A_{vb} = \frac{v_o}{v_i}$$

$$= \frac{-i_c R_L}{h_{ib} i_e + h_{rb} v_o} = \frac{-i_c R_L}{h_{ib} i_e + h_{rb}\,(-i_c R_L)}$$

$$= -\frac{h_{fb} R_L}{h_{ib} + \Delta_h R_L} \tag{7}$$

Here $$\Delta_h = (h_{ib} h_{ob} - h_{fb} h_{rb})$$

Input resistance $R_{ib} = \frac{v_i}{i_e}$

$$R_{ib} = \frac{h_{ib} i_e + h_{rb} v_o}{i_e}$$

$$= h_{ib} - A_{ib} h_{rb} R_L \qquad [\because v_0 = -i_c R_L]$$

$$= h_{ib} - \frac{h_{fb} h_{rb} + R_L}{(1 + h_{ob} R_L)} \tag{8}$$

Output resistance $R_{ob} = \frac{v_o}{i_c}$

$$i_c = h_{fb} i_e + h_{ob} v_o$$

and $h_{ib} i_e + h_{rb} v_o = 0$

$$i_e = \frac{h_{rb} v_o}{h_{ib}}$$

$$\therefore \quad i_c = v_o \frac{(h_{ob} h_{ib} - h_{fb} h_{rb})}{h_{ib}}$$

$$= \frac{v_o \Delta_h}{h_{ib}}$$

$$\therefore \quad R_{ob} = \frac{h_{ib}}{\Delta h} \tag{9}$$

5.5 Self Learning Exercise-I

Section A: Very Short Answer Type Questions

Q.1 Write the h-parameter of a transistor in common base configuration.

Q.2 Why transistors act as a four terminal network device?

Q.3 Define current gain.

Section B: Short Answer Type Questions

Q.4 The transistor used in an amplifier in common base configuration has the parameters h_{ob} = 1μmho, h_{fb} = -0.95. If the load resistance in the amplifier is 200kΩ. Determine the current gain.

Q.5 The h parameter of a transistor used in common emitter configuration are:

h_{ie} = 1100 , h_{fe} =50, h_{oe} = 25 mho and $h_{re} = 2.5 \times 10^{-4}$.

If the source resistance in 1kΩ then determine output resistance.

Q.6 Write the formula of current gain, voltage gain input resistance and output resistance of the h-parameter of a transistor used in common emitter configuration.

5.6 Emitter Follower (Common Collector Amplifier)

In this circuit the load is connected between the emitter and ground terminal. The collector is grounded for the alternating signals. Between the base B and emitter E the resistance is h_{ie}. The input current is i_b and output current is i_e. In the collector circuit between the collector and emitter there is a current generator $h_{fe}i_b$ or $g_m v_{be}$.

$$g_m = \frac{I_c}{V_{be}}$$

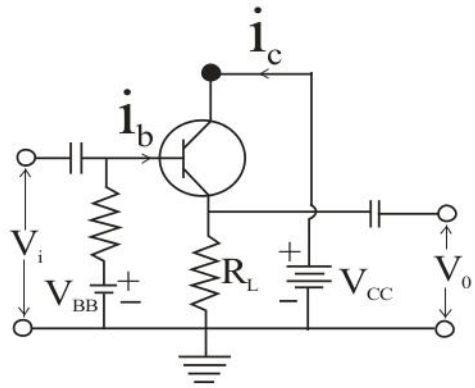


Figure : (5.15)

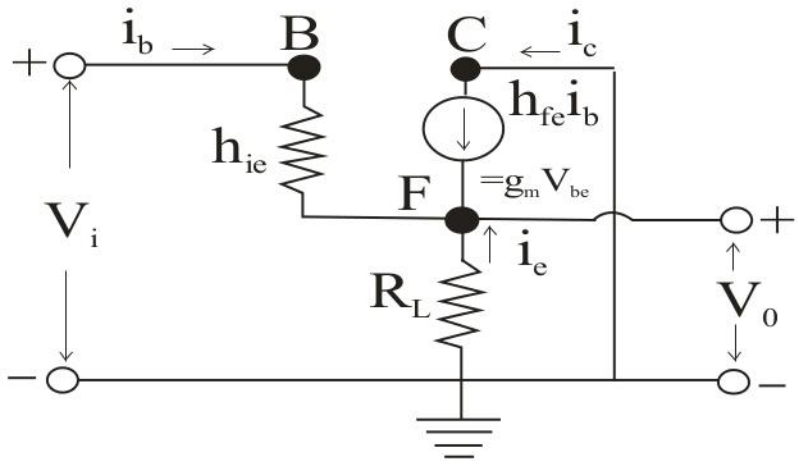


Figure: (5.16)

Current Gain : $A_{ic} = \frac{i_e}{i_b}$

$$= \left(\frac{-(i_b + i_c)}{i_b} \right) \qquad \text{(because } i_c = h_{fe} i_b\text{)}$$

$$= -\left(1 + \frac{h_{fe} i_b}{i_b}\right) = -(1+h_{fe}) \qquad (10)$$

But **voltage gain** $A_{vc} = \frac{v_o}{v_i}$

$$v_o = i_e R_L = (i_b + i_c) R_L$$

$$= i_b (1+h_{fe}) R_L$$

$$V_i = i_b h_{ie} + (i_b + i_c) R_L$$

$$= i_b h_{ie} + i_b (1+h_{fe}) R_L$$

$$= i_b [h_{ie} + (1+h_{fe}) R_L]$$

$$A_{ve} = \frac{v_o}{v_i} = \frac{i_b (1+h_{fe}) R_L}{i_b \left[h_{ie} + (1+h_{fe}) R_L \right]}$$

$$= \frac{1}{1 + \frac{h_{ie}}{(1+h_{fe}) R_L}} = \frac{g_m R_L}{1 + g_m R_L} \qquad (11)$$

But $\frac{h_{ie}}{(1+h_{fe}) R_L} \ll 1, h_{fe} \gg 1$

$\therefore \quad A_{vc} = 1$ i.e., $v_o \approx v_i$

Input resistance $R_{ic} = \frac{v_i}{i_b}$

$$= h_{ie} + (1+h_{fe}) R_L$$

$$\approx h_{ie} + h_{fe} R_L$$

$$= h_{ie} (1 + g_m R_L) \qquad (12)$$

Output resistance: To determine output resistance R_{oc}, take v_i equal to zero, the circuit will be of the following from Fig. (5.21)

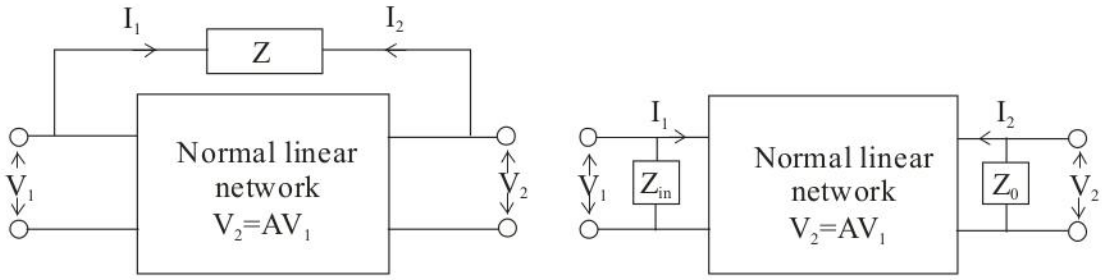


Figure: (5.17)

$$\therefore \qquad R_{oc} = \frac{h_{ie}R_L}{h_{ie}+R_L} \qquad (13)$$

Thus a CC amplifier or an emitter follower acts as an impedance transformer with unity voltage gain. Its input impedance is very high and output impedance low. So it is used for matching a source of high internal resistance with a load of low resistance. Its action is opposite to the action of CB circuit.

5.7 Miller's Theorem and its Dual

In a four terminal network whose voltage amplification factor is A, if the impedance between input and output parts is Z then this impedance is equivalent of an input impedance $\frac{Z}{(1-A)}$ and an output impedance $\frac{AZ}{(A-1)}$. This effect is called Miller effect.

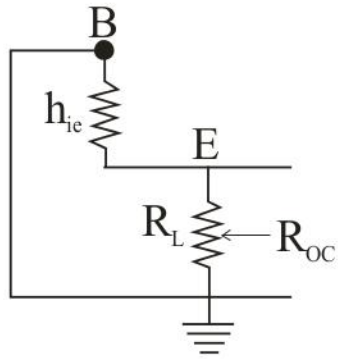


Figure : 5.21

Proof : For the normal linear network

$$V_2 = AV_1$$

The current I_1 flowing from the voltage source towards the impedance Z is

$$I_1 = \frac{V_1 - V_2}{Z} = \frac{V_1 - AV_1}{Z} = \frac{(1-A)V_1}{Z}$$

$$= \frac{V_1}{\frac{Z}{(1-A)}} = \frac{V_1}{Z_{in}}$$

where $Z_{in} = \frac{Z}{(1-A)}$

The current I_2 flowing from the voltage source V_2 towards the impedance Z is

$$I_2 = \frac{V_2 - V_1}{Z} = \frac{V_2 - V_2|A}{Z} = \frac{(A-1)V_2}{AZ}$$

$$= \frac{V_2}{AZ|(A-1)} = \frac{V_2}{Z_0}$$

where $Z_0 = \frac{AZ}{(A-1)}$

5.8 Self Learning Exercise- II

Section A: Very Short Answer Type Questions

Q.1 State Miller's theorem.

Q.2 Write the formula for Emitter follower for output resistance using h parameter.

Q.3 Write the formula for common emitter transistor amplifier using h parameter of voltage gain.

Section B: Very Short Answer Type Questions

Q.4 In an emitter follower, the parameters of the load transistor used are h_{ic} = 900Ω, h_{fc} = -25, h_{rc} = 1 and h_{oc} =22 × 10^{-6} mho. If source resistance is 10kΩ, Calculate output resistance.

Q.5 Derive an expression for current gain in an emitter follower.

Q.6 Compute the voltage gain of a single-stage common base transistor amplifier. If R_L = 10 KΩ, h_{ib} = 30 ohm, h_{fb} = - ^{6} 0.95, h_{ob} = 10^{-6} and h_{rb} = 1.25× 10^{-3}.

5.9 Summary

The h- parameters of a two port network

$$h_{11} = \left.\frac{V_1}{I_1}\right|_{V_2=O} \qquad h_{21} = \left.\frac{I_2}{I_1}\right|_{V_2=O}$$

$$h_{12} = \left.\frac{V_1}{V_2}\right|_{I_1=O} \qquad h_{22} = \left.\frac{I_2}{V_2}\right|_{I_1=O}$$

The h- parameters of a CE transistor

$$h_{ie} = \left.\frac{V_b}{I_b}\right|_{V_C=O} \qquad h_{fe} = \left.\frac{I_C}{I_b}\right|_{V_C=O}$$

$$h_{re} = \left.\frac{V_b}{V_c}\right|_{I_b=O} \qquad h_{oe} = \left.\frac{I_C}{V_C}\right|_{I_b=0}$$

Variation of CE h- parameters results due to change in collector Junction temperature and collector current. The CE amplifier has high current gain A_i, high voltage gain A_v, medium input resistance R_i and medium output resistance. The CB amplifier has low current gain A_i, high voltage gain A_v, low input resistance R_i and high output resistance. The CC amplifier has high current gain, low voltage gain, high input resistance and low output resistance.

If Z is the impedance connected betweens nodes 1 and 2 of an N-node network, then as per miller's theorem, current I_1 drawn from node 1 through the impedance Z may be obtained by disconnecting node 1 from Z and by bringing an impedance Z/(1-K) from node 1 to ground, where k is the voltage ratio V_2/V_1. This nothing but a common collector amplifier. The name emitter follower is given to it since the emitter voltage follows the input voltage.

5.10 Glossary

Amplifier: A device which takes in a weak electric signal and sends out a stronger one.

Transistor: Transistors are tiny electrical devices that can be found in everything from radios to robots.

h-Parameters: four-pole, hybrid characteristics of transistors.

h_{FE}: Current gain of the raw transistor in a common-emitter circuit.

h_{fe} : The small-signal current gain of the raw transistor in a common-emitter circuit.

h_{ie} : Short circuit input impedance.

$\mathbf{h_{oe}}$: Open circuit output admittance.

$\mathbf{h_{re}}$ **:** Open circuit reverse voltage transfer ratio.

5.11 Answers to Self Learning Exercises

Answer to Self Learning Exercise-I

Ans.1: h_{ib}, h_{rb}, h_{fb}, h_{ob}.

Ans.2: Transistor is a three terminal (base, emitter and collector) device. If its one terminal is common to both input and output circuits, it acts as a four terminal network. On the basis of common terminal three configurations are possible for a transistor.

Ans.3: For a transistor amplifier the ratio of output current and input current is called current gain. It is represented by A_i. Hence $A_i = \frac{i_2}{i_1}$

Ans.4: 0.79

Ans.5: 52.5 KΩ

Ans.6: (a) $A_{ie} = -\frac{h_{fe}}{1 + h_{oe} R_L}$ (b) $R_{ie} = h_{ie} + h_{re} A_{ie} R_L$

(c) $A_{ve} = +\frac{A_{ie} R_L}{R_{ie}}$ (d) $R_{oe} = \frac{h_{ie} + R_S}{h_{oe} h_{ie} + h_{oe} R_s - h_{fe} h_{re}}$

Answer to Self Learning Exercise-II

Ans.1: In a four terminal network whose voltage amplification factor is A, if the impedance between input and output parts is Z then this impedance is equivalent of an input impedance $\frac{Z}{(1-A)}$ and an output impedance $\frac{AZ}{(A-1)}$.

Ans.2: $\frac{R_s + h_{ie}}{h_{fe}}$

Ans.3: $A_{ve} = A_{ie} \frac{R_L}{R_{ie}}$

Ans.4: 432Ω.

Ans.5: - $(1+h_{fe})$

Ans.6: 225.

5.12 Exercise

Section A: Very Short Answer Type Questions

Q.1 Define h- parameters.

Q.2 Why is a common emitter circuit better than a common base circuit, explain.

Q.3 Define the h-parameters of a transistor in common emitter configuration.

Q.4 Define the h-parameter for a transistor in common base configuration.

Section B: Short Answer Type Questions

Q.5 Describe the method to determine the h-parameters of a transistor in common emitter configuration.

Q.6 The parameters of a transistor in common emitter (CE) configuration are h_{fe} = 330, h_{oe} = 20μ mho, h_{ie} = 3300Ω and $h_{re} = 2\times10^{-4}$, If in the amplifier the load resistance is 5kΩ and source resistance is 5kΩ, calculate current gain, voltage gain, input resistance, output resistance and power gain.

Q.7 State Miller's theorem and prove it.

Q.8 Analyzing a transistor amplifier obtain expression for the current gain, voltage gain, input and output resistance in terms of h-parameter.

Section C : Long Answer Type Questions

Q.9 Draw the circuit and equivalent circuit of a emitter follower amplifier. Obtain expression for current gain, voltage gain, input and output resistance in terms of h-parameters.

Q.10 For small signal voltage draw the h-parameter equivalent circuit of a basic transistor amplifier. Derive expressions for current gain, voltage gain and input impedance for this amplifier.

Q.11 Draw the circuit and equivalent circuit of a common base transistor amplifier. Obtain expression for current gain, voltage gain, input and output resistance in terms of h-parameters.

Q.12 Draw the circuit diagram of a transistor amplifier in common collector configuration in which the load is RL. Derive expression for expressions for input impedance, output impedance and voltage gain for this amplifier.

References and Suggested Readings

1. N.N. Bhargava, D.C. Kulshreshtha, S. C. Gupta, Basic electronics and linear circuits, Tata McGraw-Hill Publishing Company Limited, 1997.
2. Allen Mottershead, Electronic devices and circuits, Prentice Hall of India publisher, 1998.
3. Sedra and Smith, Microelectronic circuits, Oxford University Press, 2004.
4. Jacob Millman and Christos C. Halkias, Integrated Electronics: Analog and Digital Circuits and Systems, Tata McGraw-Hill Publication, 1991.
5. V. K. Mehta and Rohit Mehta, Principles of electronics by Tenth Edition, S Chand Publication, 2006.

UNIT-6
Transistor DC Biasing

Structure of the Unit

6.0 Objectives

6.1 Introduction of Transistor

6.2 Fixed bias circuit

6.3 Voltage divider bias

6.4 Self-Learning Exercise-I

6.5 Emitter feedback bias

6.6 Collector feedback

6.7 Collector and emitter feedback bias

6.8 Basic idea of photo transistor

6.9 Self-Learning Exercise-II

6.10 Summary

6.11 Glossary

6.12 Answers to self learning Exercises

6.13 Exercise

References ans Suggested Readings

6.0 Objectives

This unit introduces the concepts of transistor DC biasing, fixed bias circuit and voltage divider bias. After going through this unit you should be able to understand the collector and emitter feedback bias and basic idea of photo transistor and their applications.

6.1 Introduction of Transistor Biasing

For the normal operation of a transistor amplifier circuit, it is essential that the emitter-base junction should be forward biased while collector-base Junction

should be reversed biased. To achieve this, bias batteries may be used or associated circuit with the transistor may be employed. The latter method is more efficient and is frequently used. The circuit providing the desired biasing is known as biasing circuit. The transistor biasing may be defined as the proper flow of zero signal collectors current and the maintenance of collector- emitter voltage during the passage of the signal. When a transistor is not properly biased, it works inefficiently and produces distortion in the output signal. Special efforts are made for this purpose.

Following are the methods used for providing the biasing arrangement in a transistor.

(1) Base bias or fixed current bias

(2) Base bias with emitter feed back

(3) Base bias with collector feedback

(4) Voltage divider bias

We shall discuss these biasing arrangements one by one. The basic principle employed in each arrangement is that the desired base current (and hence collector current) is obtained from V_{CC}. In zero signal conditions and collector load R_C is selected in such a way that V_{CE} should not fall below 0.5 Volt for germanium transistor and 1 volt for silicon transistors.

6.2 Fixed Bias Circuit

In base bias, a high resistance Rb of several hundred **Ω** is connected between base and positive terminal of supply for NPN transistor (Fig. 6.1) and between base and negative terminal of supply for PNP transistor. The required zero signal base current is provided by V_{CC} which passes through R_b. By the proper selection of R_b, the required base current I_b (and hence $I_C = \beta I_b$) can made to flow. Now our aim is to calculate the value of R_b so that required base current (and hence collector current) flows under zero signal condition. Let I_C be the required zero signal collector current. Applying the Kirchhoff's voltage law to the closed circuit ABEDA, we get

$$V_{CC} = I_b R_b + V_{be}$$

$$\therefore \qquad R_b = \frac{(V_{CC} - V_{be})}{I_b} \qquad (1)$$

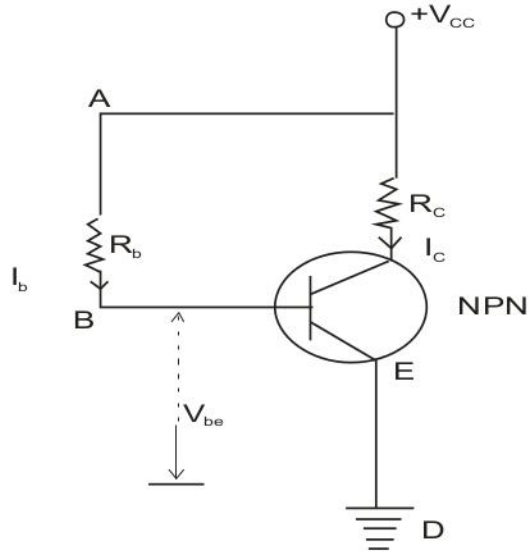


Fig. 6.1: Fixed Bias Circuit

Here V_{CC} and I_b are known and V_{be} can be seen from transistor manual, hence value of R_b can be found. Since V_{be} is generally quite small, hence

$$R_b = \frac{V_{CC}}{I_b} \qquad (2)$$

Applying Kirchhoff's voltage law to output section, we have

$$V_{CC} = I_C R_C + V_{Ce} \qquad (3)$$

where V_{Ce} = Voltage between collector and emitter terminals of the transistor.

It is clear from equation (3) that the voltage drop $I_C R_C$ can never be more that supply voltage V_{CC}, thus

$$I_C \leq \left(\frac{V_{CC}}{R_C}\right)$$

When the transistor is in saturation, V_{CE} is almost Zero and the saturation collector current is given by

$$(I_C)_{Sat} = \frac{V_{CC}}{R_C}$$

6.3 Voltage Divider Bias Method

This method is commonly used for providing biasing and stabilization. In this

method, two resistances R_1 and R_2 are connected across the supply voltage V_{cc} is shown in Fig. 6.2 (a). The emitter resistance R_e provides stabilization. The name voltage divider is derived from the fact that resistors R_1 and R_2 form a potential divider across V_{cc}. The voltage drop across R_2 Forward biases the base emitter junction while V_{cc} supply reverse biases the base collection Junction. The forward bias on base emitter Junction causes the base current and hence collector current flows in zero signal conditions.

In order to thevenin's the base circuit, we first break the base lead as shown in fig. 6.2 (b). The open circuit voltage is $V_{th} = \frac{R_2 V_{CC}}{R_1 + R_2}$ (V_{th} is generally denoted by V_b).

Again we reduce Vcc to zero as shown in Fig. 6.2 (c). Looking into the voltage divider, the venin's resistance is

$$R_{th} = \frac{R_1 R_2}{R_1 + R_2}$$

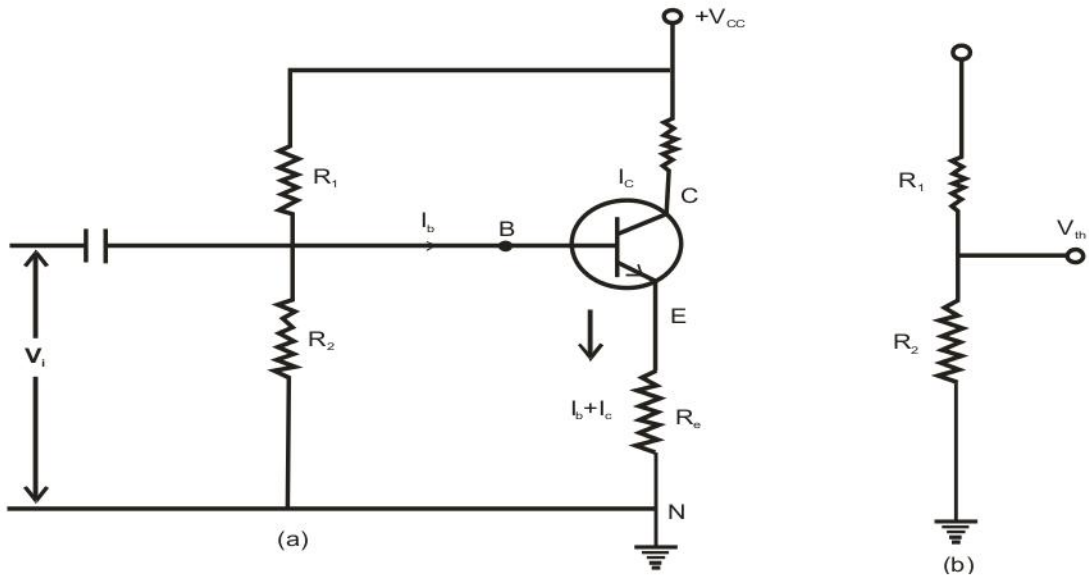


Fig. 6.2(a),(b): Self Biased Circuit and its equivalent circuit

This is generally denoted by R_b. The equivalent circuit is shown in Fig. 6.2 (d).

The collector current I_C can be calculated by applying Kirchhoff's law to the base circuit.

Let I_1 be the current through R_1 and R_2, then

$$I_1 = \frac{V_{CC}}{(R_1 + R_2)}$$

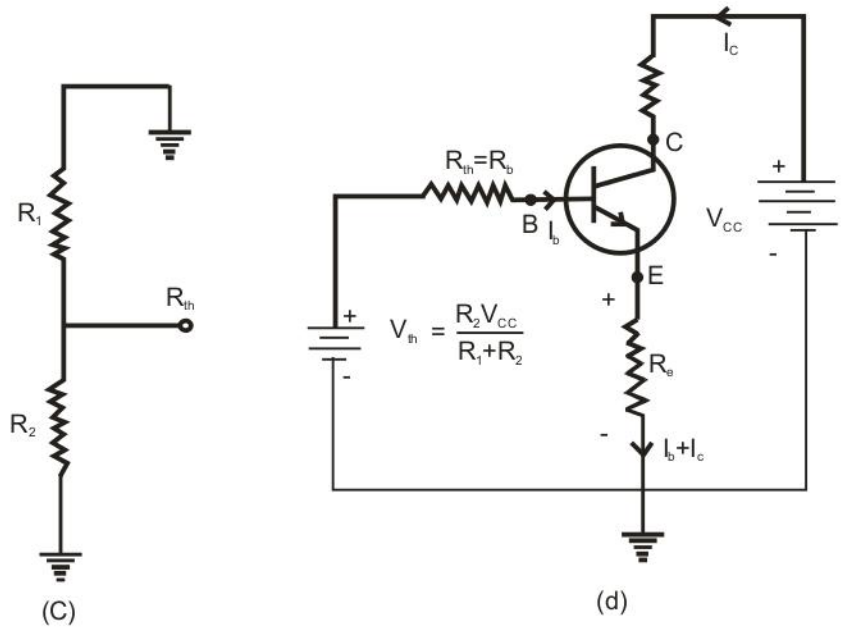


Fig. 6.2(c),(d): Self Biased Circuit and its equivalent circuit

The voltage across R_2 is given by

$$V_2 = \left(\frac{V_{CC}}{R_1 + R_2}\right) \times R_2$$

If Voltage across R_e be V_e, then

$$V_2 = V_{be} + V_e \text{ or } V_2 = V_{be} + I_e R_e$$

$$\therefore \quad I_e = \frac{V_2 - V_{be}}{R_e}$$

$$\text{or} \quad I_C = \frac{V_2 - V_{be}}{R_e} \qquad \left(\because I_e \approx I_C\right)$$

The collector - emitter voltage V_{cc} can be calculated by applying Kirchhoff's voltage law to the collector side.

Here

$$V_{CC} = I_C R_C + V_{Ce} + I_e R_e = I_C R_C + V_{Ce} + I_C R_e$$

$$V_{CC} = I_C (R_C + R_e) + V_{Ce}$$

$$\text{or} \quad V_{Ce} = V_{CC} - I_C (R_C + R_e)$$

The reason that how the operating point stability is improved can be understood as follows.

Let us consider that the collector current I_C tend to increase due to the increase in I_{CO} (leakage current) because of temperature rise. Now this increased current I_C causes an increased d. c. voltage drop across R_e. This reduces the net emitter to base forward bias. As a result the base current I_b is reduced.

The reduction in base current causes a reduction in collector current. Thus the self biasing resistor R_e reduces the increases in I_C and improves the operating point stability. Using Thevenin's theorem, the analysis is as follows:

$$V_{th} = I_b R_{th} + V_{be} + R_e I_e$$

$$I_e = (\beta+1) I_b$$

$$V_{th} = I_b R_{th} + V_{be} + R_e (\beta+1) I_b$$

$$V_{th} - V_{be} = I_b [R_{th} + (\beta+1)R_e]$$

$$I_b = \frac{V_{th} - V_{be}}{R_{th} + (\beta+1)R_e} \approx \frac{V_{th}}{R_{th} + \beta R_e}$$

Now $I_C = \beta I_b = \dfrac{\beta V_{th}}{R_{th} + \beta R_e}$

The collector to base, emitter voltage is given by

$$V_{Ce} = V_{CC} - (R_C + R_e) I_C$$

$$V_{Ce} = V_{CC} - \frac{(R_C + R_e)\beta V_{th}}{R_{th} + \beta R_e}$$

6.4 Self Learning Exercise-I

Section A : Very Short Answer Type Questions

Q.1 Fixed bias is applied in a transistor amplifier. The parameters of the operating point Q of the transistor are :

$V_{CC} = 10V$, $I_B = 40\mu A$, $V_{BE} = 0.14V$.

Determine the value of resistance R_B used for fixed bais.

Q.2 A fixed bias is used in an amplifier. At the operating point Q of the transistor $V_{CC} = 30V$, $I_B = 60\mu A$, $V_{BE} = 0.6V$. Calculate the bias resistance for the fixed bias.

Q.3 Calculate the Co-ordinates of the operating point with the given data

V_{CC} = 10V, R_C = 1KΩ, R_b = 100KΩ and β=60 for a common emitter configuration.

Section B : Short Answer Type Questions

Q.4 In a common emitter amplifier self bias is applied on the transistor. The parameters for the circuit are V_{CC} = 12V, R_L = 4kΩ, R_E = 2KΩ, R_1 = 105KΩ, R_2 = 50KΩ, and for the transistor β= 50,V_{BE} = 0.6V. Determine the dc operating point of the transistor.

Q.5 In a common emitter amplifier self bias is used for the NPN transistor. Its parameters are V_{CC} = 24V, R_L = 5000Ω, R_E=1000Ω, R_1 = 81000Ω, R_2 = 9000Ω, and for the transistor β = 60 and V_{BE} = 0.6V. Determine the dc operating point of the transistor.

Q.6 Find the value of IC for potential divider method if V_{CC} = 9V, R_e = 1KΩ, R_1 = 39 KΩ, R2 = 10KΩ, R_c = 2.7 KΩ V_{be} = 0.15V and β =90

6.5 Emitter Feedback Bias

The circuit arrangement of emitter feedback bias is shown in Fig. 6.3. In this circuit only an emitter resistor is connected to the base bias circuit. The difference voltages and current can be calculated as follows:

1. Base current can be calculated by writing Kirchhoff's voltage equation for base emitter loop.

   Here

$$V_{CC} = R_b I_b + V_{be} + R_e I_e$$

$$V_{CC} = I_b [R_b + (\beta+1)R_e] + V_{be}$$

$$V_{CC} - V_{be} = I_b [R_b + (\beta+1) R_e]$$

$$\therefore \quad I_b = \frac{V_{CC} - V_{be}}{R_b + (\beta+1)R_e} \quad - \quad (1)$$

2. Saturated collector current $(I_C)_{Sat}$. At saturation, V_{CC} is esentially zero

$$\therefore \quad (I_C)_{Sat} = \frac{V_{CC}}{R_C + R_e} \quad - \quad (2)$$

3. The value of I_C can be calculated as follows: applying Kirchhoff's voltage law to supply, base, emitter and ground route, we have

$V_{CC} = I_b R_b + V_{be} + I_e R_e$

$I_b = I_C/\beta$ and $I_e \cong I_C$

$$\therefore \quad V_{CC} \cong \frac{I_C R_b}{\beta} + V_{be} + I_C R_e$$

$$I_C = \frac{V_{CC}}{[R_e + R_b/\beta]} \qquad \because \text{ Neglecting } V_{be} \tag{3}$$

$$R_e \cong \frac{V_{CC} - V_{be} - (I_C R_b/\beta)}{I_C} \tag{4}$$

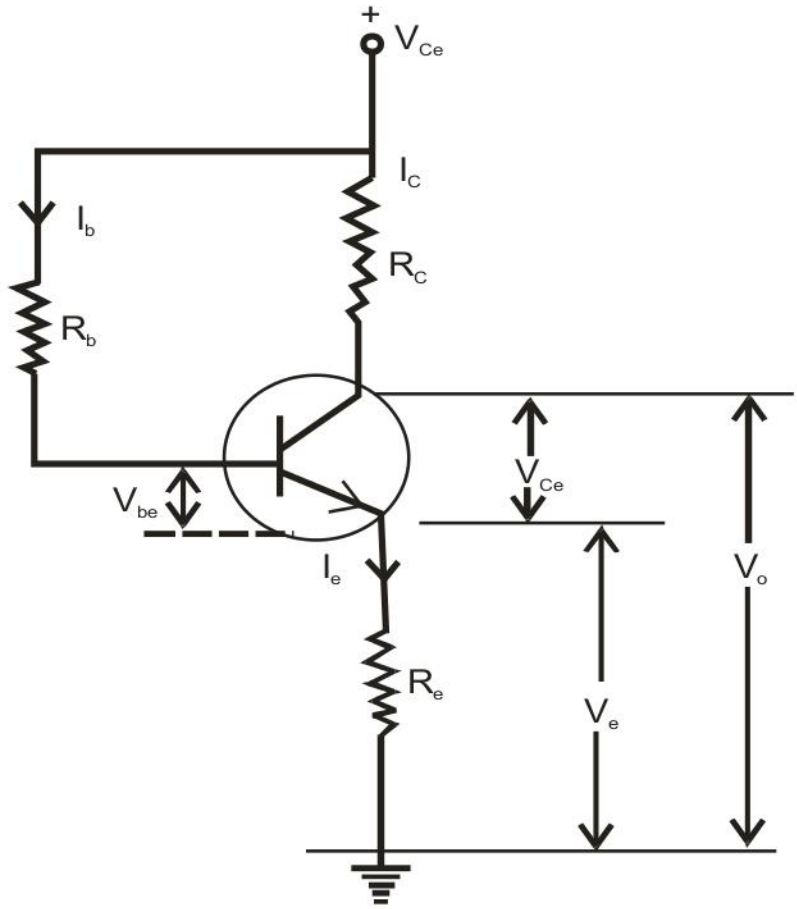


Fig. 6.3: Emitter Feed Back bias

4. The value of V_{Ce} can be calculated by writing Kirchhoff's voltage equation for collector and emitter loop here

$V_{Ce} = V_{CC} - I_C (R_C + R_e)$

5. Collector to ground Voltage $V_C = V_{CC} - I_C R_C$ (5)

6. Emitter to ground Voltage $V_e = I_e R_e = I_C R_C$ (6)

When the temperature rises, the leakage current increases as a result, the collector and emitter current increase. So, the voltage drop across R_e i.e., V_e also increases. As

$$V_{CC} = R_b I_b + V_{bC} + I_e R_e$$

or $$I_b \cong \frac{V_{CC} - I_e R_e}{R_b} \approx \frac{V_{CC} - V_e}{R_b}$$

Hence, the current I_b reduces. Now the collector current also reduces. Thus we see that the collector current is not allowed to increase to the extent it would have been in the absence of the resistor R_e.

6.6 Collector Feedback Bias

The circuit arrangement of collector feedback is shown in fig. 6.4 In this method, one end of R_b is connected to the base while the other end is connected to collector. It should be remembered that the required zero signal base current is determined by collector base voltage V_{cb} and not by V_{CC}.

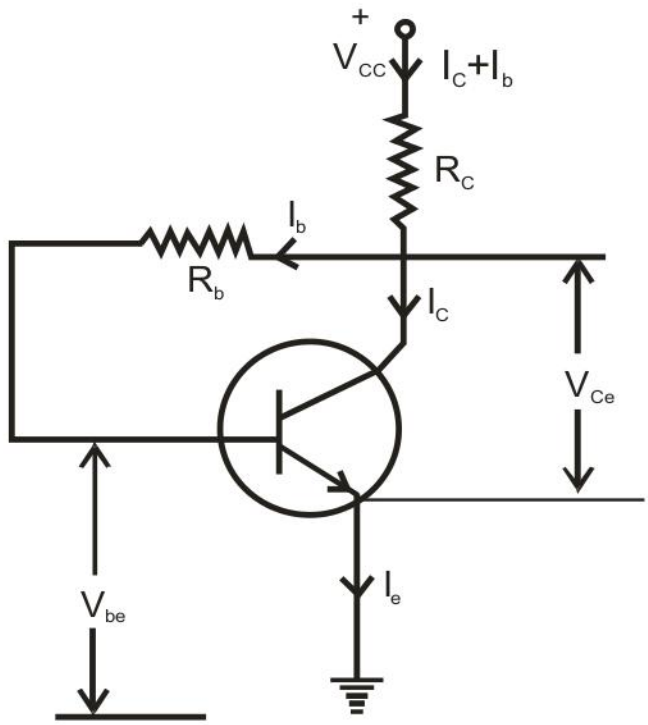


Fig. 6.4: Collector Feedback Bias

Here V_{Cb} forward biases the base emitter junction and base current I_b flows through R_b. This causes the zero signal collector current to flow in the circuit. It derives its name from the fact that since voltage for R_b is derived from collector, there exists a negative feedback which tends to stabilize I_C against changes in β.

The circuit parameters can be calculated as follows:-

(i) For the calculation of R_b, needed to give zero signal current I_c, we have

$$V_{CC} = I_C R_C + I_b R_b + V_{be}$$

$$\beta = \frac{I_C}{I_b}$$

$$I_C = \beta I_b$$

$$V_{CC} = \beta I_b R_C + I_b R_b + V_{be}$$

$$\therefore R_b = \frac{V_{CC} - V_{be} - \beta I_b R_c}{I_b} \qquad (1)$$

(ii) $$(I_C)_{Sat} = \frac{V_{CC}}{R_C} \qquad \because V_{Ce} = O \qquad (2)$$

(iii) $V_C = V_{CC} - I_C R_C$ and $V_C = I_b R_b + V_{be}$

Equating these two expressions, we have

$$I_b R_b + V_{be} = V_{CC} - I_C R_C$$

$$\frac{I_C R_b}{\beta} + V_{be} = V_{CC} - I_C R_C$$

$$\therefore \quad I_C = \frac{V_{CC} - V_{be}}{R_C + (R_b / \beta)}$$

$$I_C \cong \frac{V_{CC} - V_{be}}{R_C + (R_b / \beta)} \qquad (3)$$

6.7 Collector and Emitter Feedback Bias

A better stabilization of the operating point may be achieved by base bias with collector and emitter feed backs. The circuit arrangement is shown in Fig. 6.5. As the temperature rises, the leakage current increases. As a result, the collector current as well as emitter current increases. Now the voltage drop across Re also increases. So, the base current I_b also reduces. This results in the reduction of

collector current. In this way, the collector current is not allowed to increases to the extent it would have been in the absence of the resistance R_e.

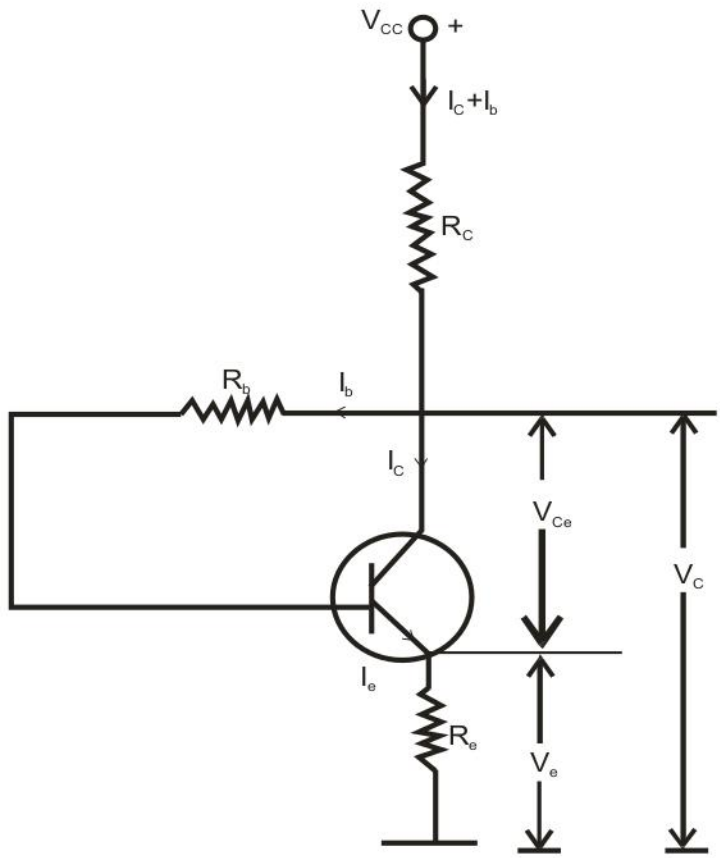


Fig. 6.5: Collector and Emitter Feedback bias

Here $(I_C)_{Sat} = \dfrac{V_{CC}}{R_C + R_e}$

$$I_C \cong \frac{V_{CC} - V_{be}}{R_C + R_e + (R_b / \beta)}$$

$$V_C = V_{CC} - (I_C + I_b) R_c \cong V_{CC} - I_C R_C$$

$$V_e = I_e R_e \cong I_C R_e$$

$$V_C = V_C - V_e$$

6.8 Basic Idea of Photo Transistor

A photo transistor also known as dipolar photo-transistor is a three level device having its base region exposed to illumination. It differs from a conventional bipolar transistor by having a large base-collector junction as the light collecting

element. It has been depicted in fig. 6.7, fig. 6.8

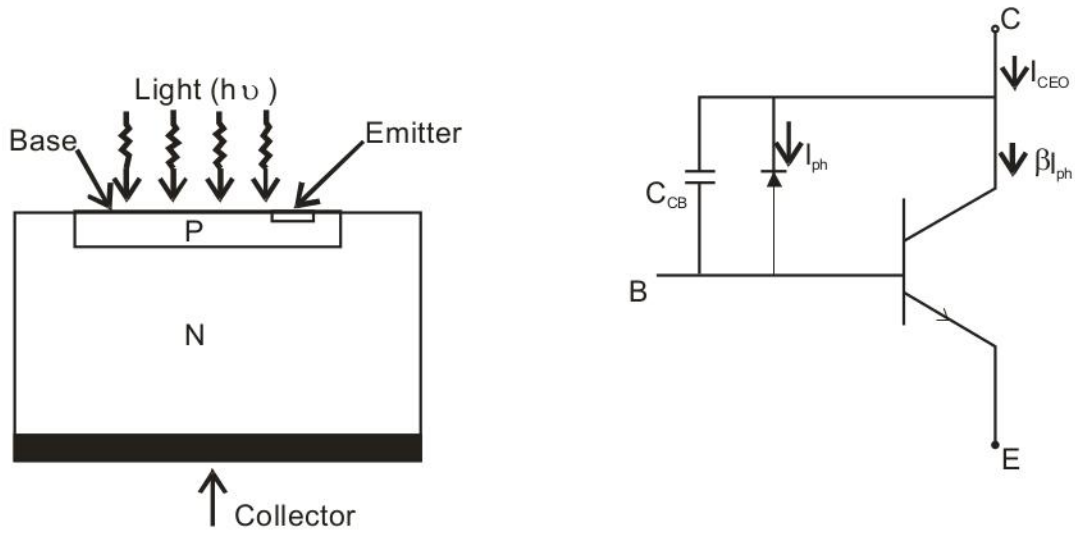


Fig.6.7(a)Photo Transistor (b)Equivalent Circuit Diagram

Show the circuit model of a photo transistor in which its base-collector junction is represented by a parallel combination of a diode and a capacitor.

Both P-N-P and N-P-N type photo-transistors are manufactured. They are usually connected in common emitter configuration. It is sometimes possible to forward bias the photo-transistor. They may have either the base connections open circuited or connected through a resistance to the common line as shown in fig. 6.8.

Fig. 6.8 depicts the basic circuit incorporating N-P-N photo-transistor.

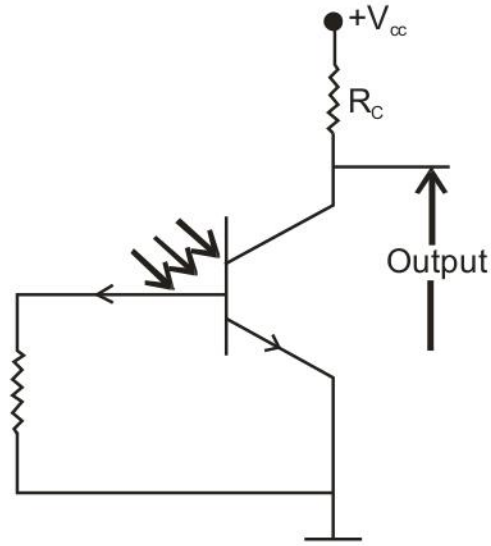


Fig. 6.8: N-P-N photo Transistor

N-P-N Photo Transistor

When the base lead is floating, the photo generated carries contribute a photo current I_{ph} in the collector. In addition to it, the holes generated in the base and the holes swept into the base from the collector, lower the base emitter potential consequently across the base. Therefore the total collector current.

$$I_{CEO} = I_{ph} + h_{FE}I_{ph} = (1+h_{FE}I_{ph})$$

where I_{ph} is the photo current components.

Characteristics

Typical collector characteristics curves of a photo transistor are shown in fig. 6.9. These curves depict the variation of collector current with collector emitter voltage for various values of illumination which is expressed in units of lumen/meter^2 or LUX.

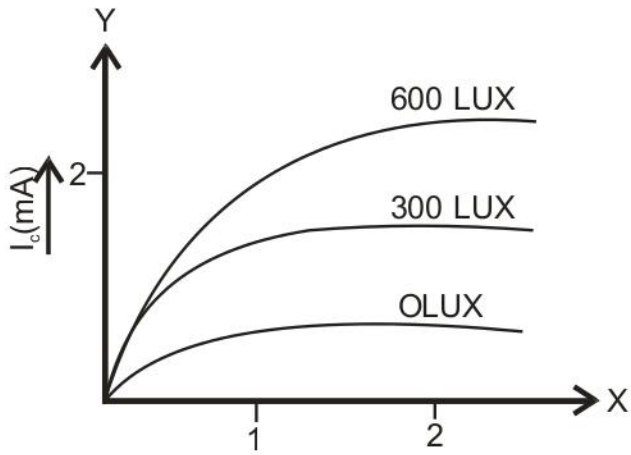


Fig. 6.9: V-I Characteristics of Photo Transistor

6.9 Self Learning Exercise-II

Section A: Very Short Answer Type Questions

Q.1 Calculate the collector voltage in voltage divider method with the following data:

$V_{CC} = 20V$, $R_1 = 10K\Omega$, $R_2 = 10K\Omega$, $R_c = 1K\Omega$ and $R_e = 5K\Omega$

Q.2 An NPN transistor circuit has $\alpha = 0.985$, $R_2 = 20k\ \Omega$, $R_e = 2K\ \Omega$ and $V_{be} = 0.3$ Volt. If $V_{CC} = 20$ Volt, Calculate R_1 and R_c to place Q point $I_C = 2mA$, $V_{Ce} = 5$ Volt.

Q.3 A BJT Circuit has β = 50 and V_{CC} = 20V and uses a potential divider bias circuit with R_C = 2K Ω, R_e = 0.1K Ω, R_1 = 100K Ω and R_2 =5K Ω, Find the Q point, Assume V_{be} = 0.2 V

Section B: Short Answer Type Questions

Q.4 In a single stage C.E. amplifier using a single battery supply V_{cc} = 20V and the transistor has a minimum β value of 20 and $I_C o$ = 10μA. Determine the value R_1, R_2 and R_e for a quiescent point at V_{ce} = 6 V and I_C = 2mA if R_c = 4k Ω.

Q.5 A Transistor with β = 45, is used with collector to base resistor Rb biasing with a quiescent value of 5 Volt for V_{Ce}. If V_{CC} = 24V, R_C= 10K Ω and R_e = 270 Ohm. Complete the value of R_b.

Q.6 It is required to set the operating point by biasing with feedback resistor at I_C = 1mA, V_{Ce} = 10 Volt. If β =100, V_{CC} = 15 Volt, V_{be} = 0.3 Volt how will you do it?

6.10 Summary

Main utility of transistor is its ability to amplify weak signals by using it in the amplifier circuits. Transistor achieves its amplifying action by virtue of its ability to transfer electrical signal from lower resistance port (emitter-base) to higher resistance port (collector-base). Best choice of the configuration for a voltage amplifier is common emitter (CE) configuration. This is due to the fact that in CE configuration input dynamic resistance is high (~1kW) and output dynamic resistance is low as compared to CB configuration. CC configuration input dynamic resistance is highest and output dynamic resistance is lowest as compared to other two configurations but it is not the suited one for voltage amplifiers since it has voltage gain less than unity. DC load line is used to represent DC load conditions. When it is superimposed over I-V characteristics of non-linear electronic device, we can get all possible operating points. One has to design an electronic circuit by taking care of quiescent conditions. This is accomplished by fixing the Q-point. Biasing of an electronic circuit is required to make its Q-point. These are defined as rate of change of collector current with respect to leakage current (I_{CO}), DC current gain (b) of transistor and base-emitter voltage (V_{BE}) respectively. Lower the values of these stability factors more stable the circuit.

Requirements of biasing circuits are

1. To establish Q-point at the center of active region where transistor works most linearly hence allow the faithful amplification.
2. To fix the Q-point at the designed point so that it cannot move into saturation or cut-off regions and make distortions in the output waveform.
3. To ensure I_C does not change against the temperature variations.
4. To make Q-point to be independent of transistor parameters so that when transistor get replaced Q-point does not shift.

In fixed biased circuit a high valued base resistance R_B is connected in the input circuit to provide single battery use in biasing. However it suffers from high instabilities. Remedy to fixed biased biasing circuitry is provided by Connecting base resistance to Collector terminal directly (Collector-to-base biasing), or Placing an emitter resistance between emitter terminal and ground. This causes negative feedback to achieve the stability of Q-point. In voltage divider and emitter biasing circuits, the direction of current flow through emitter resistance R_E *i.e.* I_E is such that it causes a voltage drop across R_E which reverse biases the emitter-base junction. However this fact is utilized to achieve the stability in voltage divider and emitter biasing. If collector current I_C increased either due to temperature increase or transistor replacement, then this increase causes emitter current I_E to increase. The increases in I_E increases the voltage drop across R_E, which in turn tend to reverse bias emitter junction. This decreases the base current value and hence the collector current. This sequence of events ensures the stability of the circuit.

6.11 Glossary

Amplifier: Transistor circuits used to increase the voltage or power levels of an electrical signal.

Biasing circuit: Circuit made by DC batteries and resistances, which ensure the Q-point to, desired position on DC load line.

DC load line: DC conditions present due to linear elements and power supply is represented on the I-V characteristics of non-linear element, by a straight-line with slope equal to reciprocal of load resistance.

Oscillator: Transistor circuits used to generate either sine or square wave electrical signals.

Q-point: Quiescent point: Point on the DC load line, which represents the quiescent conditions.

6.12 Answers to Self Learning Exercises

Answers to self learning Exercise-I

Ans.1: $R_B = 246K\Omega$

Ans.2: $R_B = 490\ K\Omega$

Ans.3: $I_C = 10\ \mu A$, $I_c = 6mA$, $(I_C)_{sat} = 10\ mA$, $V_{ce} = 4V$

Ans.4: $I_B = 28\ \mu A$, $I_C = 1.4\ mA$, $V_{CE} = 3.544V$

Ans.5: $V_{BB} = 2.4V$, $R_B = 8.1K\Omega$, $K_B = 26.1\mu A$

$I_C = 1.57\ mA$, $V_{CE} = 14.55\ V$

Ans.6: 1.5 mA

Answers to Self learning Exercise-II

Ans.1: $V_C = 18V$

Ans.2: $R_1 = 73.2\ K\Omega$, $R_C = 2.5\ K\Omega$

Ans.3: Q Point is $I_C = 3.8mA$ and $V_{Ce} = 12V$

Ans.4: $R_1 = 12.7\ K\Omega$, $R_2 = 6\ K\Omega$ and $R_e = 2.86\ K\Omega$

Ans.5: $R_b = 110\ K\Omega$

Ans.6: $R_b = 970\ K\Omega$

6.13 Exercise

Section-A (Very Short Answer type question)

Q.1 What is biasing in a transistor amplifier?

Q.2 Explain the necessity of biasing in a transistor amplifier.

Q.3 Draw the diagram of a fixed bias circuit.

Q.4 What do you understand by transistor biasing?

Q.5 What are its needs for transistor biasing?

Section-B (Short Answer type question)

Q.6 What do you mean by stabilization of operating point?

Q.7 Why transistor biasing circuit is not used commonly explain.

Q.8 Explain the working of transistor biasing.

Q.9 Calculate

(i) $(I_C)_{Sat}$ (ii) I_C (iii) V_C (iv) V_e and V_{Ce},

with the given values :

$V_{CC} = 40$ Volt, $R_C = 3K\ \Omega$, $R_e = 1K\ \Omega$, $R_b = 400\ K\ \Omega$ and $\beta = 100$.

Q.10 Determine the operating point of the voltage divider bias method with the following data:

$V_{CC} = 20$ Volt, $V_{be} = 0.7$ Volt, $R_1 = 15\Omega$, $R_2 = 5\Omega$, $R_c = 1K\Omega$ and $R_e = 3K\Omega$

Section C (Long Answer Type Question)

Q.11 Derive an expression of Fixed Bias Circuit.

Q.12 Describe the various methods used for transistor biasing state their advantages and disadvantages.

Q.13 Describe the voltage divider method in detail. How stabilization operating point is achieved by this method?

Q.14 Derive an expression for the stability factors in a

(i) Fixed bias and (ii) Collector to base bias.

Determine the bias resistor R_b for

(i) Fixed bias and (ii) Collector to base bias for the two circuits.

Given $V_{CC} = 12$ Volt, $R_C = 330\Omega$, $\beta = 100$, $I_b = 0.3$ MA and $V_{CC} = 6$ Volt.

Q.15 Describe the photo transistor.

References and Suggested Readings

1. N. N. Bhargava, D. C. Kulshreshtha, S. C. Gupta, Basic electronics and linear circuits, Tata McGraw-Hill Publishing Company Limited,1997.

2. Allen Mottershead, Electronic devices and circuits, Prentice Hall of India publisher, 1998.
3. Sedra And Smith, Microelectronic circuits, Oxford university press, 2004.
4. Jacob Millman and Christos C. Halkias, Integrated Electronics: Analog and Digital Circuits and Systems, Tata McGraw-Hill Publication, 1991.
5. V. K. Mehta and Rohit Mehta, Principles of electronics by Tenth Edition, S Chand publication, 2006.

UNIT-7
AC Model of Transistor &Multistage Amplifiers

Structure of the Unit

7.0 Objectives

7.1 Introduction

7.2 A C model of transistor

7.3 Voltage Amplifier

7.4 Self Learning Exercise-I

7.5 Two stage R - C Coupled Transistor Amplifier

7.6 Darlington Connection

7.7 Self learning Exercise-II

7.8 Summary

7.9 Glossary

7.10 Answers to Self Learning Exercises

7.11 Exercise

References and Suggested Readings

7.0 Objectives

This unit introduces you to the AC model of a transistor of a large signal amplifier.

After going through this unit we then study the voltage feedback seven input circuit as voltage amplifier and then two stage R-C coupled transistor amplifier and Darlington connection for emitter follower.

7.1 Introduction

A single stage of amplifier (using BJT or FET) can provide small voltage gain or current, usually insufficient for most of the applications. For obtaining high voltage gain or current gain, we connect several amplifier stages in cascade i.e. connected in such a way that the output of one stage becomes the input of the next stage.

Thus a multistage cascade amplifier may provide any desired high voltage or current gain, subject to the practical limitations. An important consideration is the choice of elements coupling one stage to the next. Different coupling methods give to different types of cascade amplifiers such as R.C. coupled amplifier, transformer coupled amplifier, direct coupled (d.c.) amplifiers etc. These are described in this unit.

7.2 AC Model of Transistor

Bipolar junction transistor (BJT) is basically used as an amplifier. Transistor when used as an amplifier may use either small signal operation or large signal operation. A large signal amplifier (or power amplifier) handles large amount of power and, therefore, it aims at obtaining high conversion efficiency i.e. high ratio of useful a.c. output power to the dc input power from the collector supply VCC. In this attempt to obtain high conversion efficiency, the transistor is usually operated over some nonlinear region of characteristics resulting in increased distortion. Thus high output power and high conversion efficiency are obtained at the cost of distortion. Large signal operation may be best studied graphically. On the other hand, small signal operation aims at linear amplification. Analysis of small signal amplifier may, no doubt, be done graphically but it is more convenient to use small signal equivalent circuit (or model) for transistor operating over the active region. The transistor may then be replaced by its model and the usual method of network analysis may be used to obtain expressions for such quantities as the voltage gain, current gain, power gain, input impedance, output impedance etc of the amplifier.

7.3 Voltage Feedback Series Input Circuits (Voltage Amplifier)

The circuit shown in Fig. (7.1) is an emitter follower in which the output voltage V_0 is obtained across the emitter resistance Re and this is fed with the signal voltage at the base of transistor by voltage series feedback.

Using the feedback theory its voltage gain, input resistance and output resistance can be determined. The equivalent circuit of this common-emitter amplifier using the h-parameter models shown in Fig. (7.2)

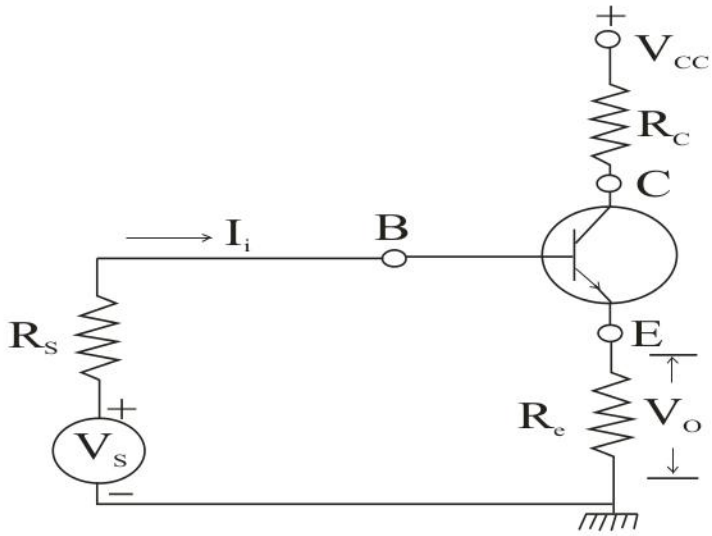


Fig. 7.1

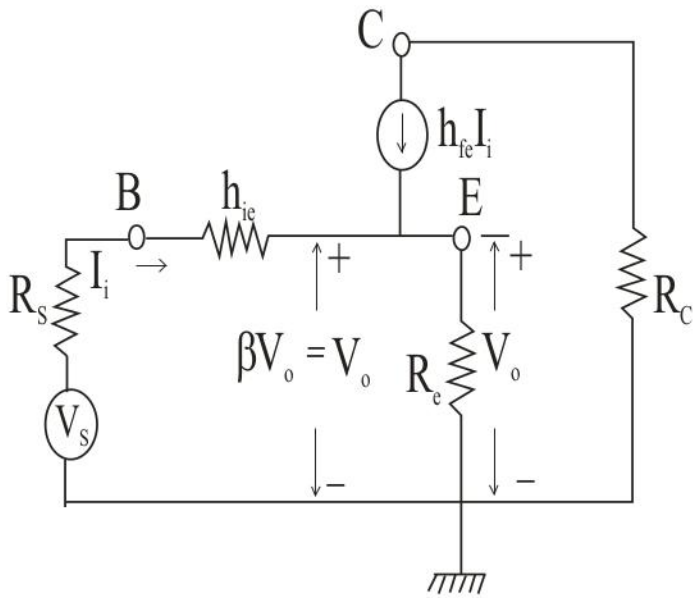


Fig. (7.2)

To determine the voltage gain of the amplifier without feed back the emitter E of the transistor is assumed to be grounded in the input circuit, so that

Single Voltage $V_S = (R_S + h_{ie})\, I_i$ (1)

But output voltage $V_0 = h_{fe}\, I_i R_e$ (2)

Substituting the value of I_i from equation (2) in equation (1)

$$V_0 = \frac{h_{fe} R_e V_s}{(R_s + h_{ie})} \quad (3)$$

∴ Voltage gain of the amplifier without feed back

$$A = \frac{V_o}{V_s} = \frac{h_{fe} R_e}{(R_s + h_{ie})} \quad (4)$$

As in this circuit the whole output voltage is fed back

∴ Feed back voltage $V_f = \beta V_0 = -V_0$

i.e. $\beta = -1$

(a) Voltage Gain

The voltage gain of the amplifier with feed back is

$$A_f = \frac{A}{(1 - A\beta)}$$

Using the values of A and β from equations (4) and (5)

$$A_f = \frac{h_{fe} R_e}{(R_s + h_{ie})\left(1 + \frac{h_{fe} R_e}{R_s + h_{ie}}\right)}$$

$$= \frac{h_{fe} R_e}{(R_s + h_{ie} + h_{fe} + R_e)} \quad (6)$$

Which is less than 1. The value of $(R_s + h_{ie})$ is very small compared to $h_{fe} R_e$, hence $A_f \approx 1$

(b) Input Resistance

The input resistance of the amplifier without feed back

$$R_i = \frac{V_s}{I_i} = (R_s + h_{ie}) \quad (7)$$

Output Voltage $V_o = h_{fe} I_i R_e$

Voltage gain $A = \frac{h_{fe} R_e}{(R_s + h_{ie})}$

With feed back the input resistance of the amplifier

$$R'_i = R_i (1 - A\beta)$$

Using equations (5), (7), and (8)

$$R'_i = (R_s + h_{ie}) \left(1 + \frac{h_{fe} R_e}{(R_s + h_{ie})}\right)$$

$$= (R_s + h_{ie} + h_{fe} R_e) \quad \textbf{(9)}$$

(c) Output Resistance

The resistance signal voltage zero, is the output resistance. In this circuit R_e acts as load resistance. In determining the output resistance the load resistance is assumed to be infinite i.e. the output terminals are considered to be open. In the given circuit if the resistance R_e is assumed to be infinite then the output resistance will become $\frac{\infty}{\infty}$ (indeterminate). Hence to determine it, R_e is assumed to being its position and in the final relation the limiting value with $R_e \to \infty$ is determined. Without feedback the output resistance of the circuit will be R_e and

$$A = \frac{h_{fe} R_e}{(R_s + h_{ie})} \quad (10)$$

$$\beta = -1$$

With feedback the output resistance of the circuit will be

$$R'_o = \frac{R_O}{1 - A\beta} = \underset{R_e \to \infty}{Lt} \frac{R_e}{\left(1 + \frac{h_{fe} R_e}{R_s + h_{ie}}\right)}$$

$$\underset{R_e \to \infty}{Lt} \frac{R_s + h_{ie}}{h_f e + \frac{R_s + h_{ie}}{R_e}} = \frac{R_s + h_{ie}}{h_{fe}} \quad (11)$$

Here the relation of voltage gain, input and output resistance have been derived from the feedback theory assuming that there is no forward transmission through the feed back network.

7.4 Self Learning Exercise- I

Section A : Very Short Answer Type Questions

Q.1 Calculate the voltage gain of voltage series feedback amplifier having gain A = -300, and β =0.05.

Q.2 Calculate the input impedance of voltage series feedback amplifier having

gain A = -300, R_{in} = 1.5kΩ and β = 0.5.

Q.3 Define amplifier.

Section B : Short Answer Type Questions

Q.4 Derive the expression for voltage gain of voltage series feedback amplifier.

Q.5 What is the input Resistance of voltage series feedback amplifier?

Q.6 What is the output Resistance of voltage series feedback amplifier?

7.5 Two Stage R-C Coupled Transistor Amplifier

In multistage or cascaded amplifier one stage is coupled to the other with the help of a coupling load or coupling network. The transistor amplifier in which the coupling load or coupling network is a resistance- capacitance network or R-C network is called R-C coupled transistor amplifier. This coupling network is used commonly as it is cheaper as well as it provides a high fidelity for a wide range of frequency. It is general used for voltage amplification.

(a) Circuit for a R-C coupled transistor amplifier: In fig (7.3) the circuit diagram of a two stage R-C coupled common emitter amplifier is shown.

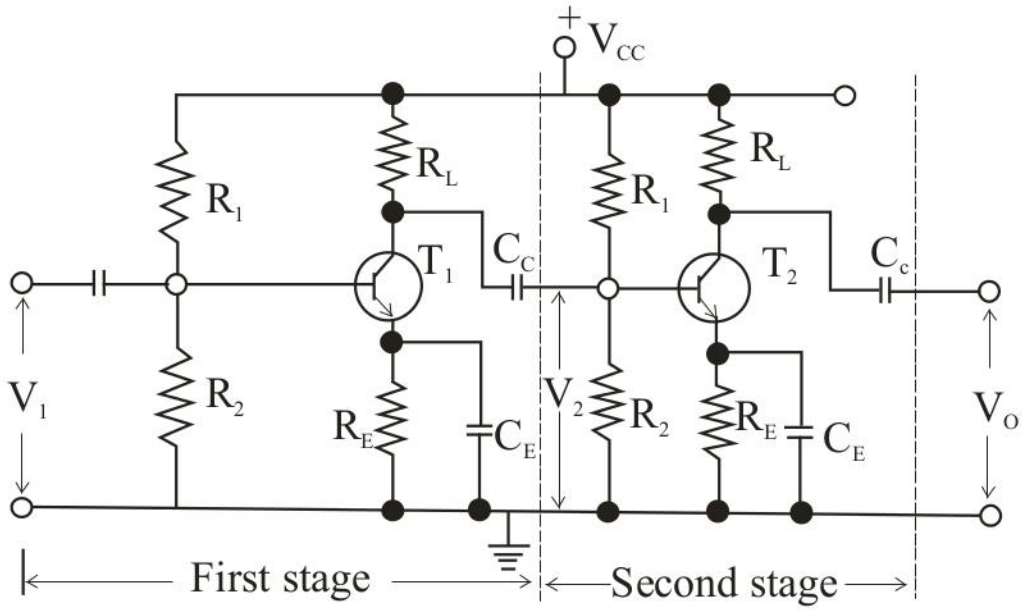


Fig. (7.3)

In the circuit shown T_1 and T_2 are respectively junction transistors used in first and second stage. R_L is load resistance and C_C is coupling capacitor. With the help

of these two the output of first stage is coupled to the second stage. The capacitor C_C allows the alternating part of the output of first stage to be passed on to second stage and stops the application of the dc part on the base of second transistor. So it is also called blocking condenser. Resistance R_1, R_2 and R_E along with R_L provide proper bias to the transistor. The capacitor C_E shunting the resistance R_E is a bypass condenser. Due to the presence of C_E the negative feedback of signal due to R_E becomes negligible. To reduce the cost of amplifier generally the condenser C_E which has to be of high capacity, is not used. It results in a loss of voltage gain but it is compensated by increasing the number of stages in the multistage amplifier.

(b) Frequency Response of Single Stage R-C Coupled Transistor Amplifier:- The analysis of frequency response of a single stage amplifier can be divided in three frequency ranges:

(i) Low Frequency range

(ii) Mid frequency range

(iii) High frequency range

(i) Mid frequency range:- The mid frequency range is in between 300 to 30000 HZ. In this range the coupling condenser CC and bypass condenser CE can be treated ineffective (short circuited) as their impedance $X_C = \frac{1}{\omega C}$ is negligibly small.

The stray capacitances and other capacitances which shunt the input and output have very high impedance in this range so these can be treated as open circuited. If the first stage of the transistor amplifiers represented by the h-parameter equivalent circuit then in this circuit values of h_{re} and h_{oe} are very small i.e., negligible, so these are not shown in the circuit.

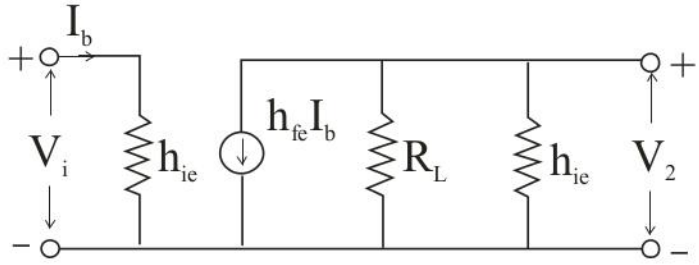


Fig. (7.4)

The h-parameter equivalent circuit for the mid frequency range of a single stage

transistor amplifier is shown in Fig. (7.4). The load R_1 is shown shunted by the input resistance h_{ie} of the second stage.

The effective load resistance of the first stage of amplifier is given by

$$\frac{1}{Z_L} = \frac{1}{R_L} + \frac{1}{h_{ie}}$$

$$= \frac{R_{ie} + R_L}{h_{ie} R_L}$$

or $$Z_L = \frac{h_{ie} R_L}{h_{ie} + R_L} \quad (1)$$

The output voltage $V_2 = - (h_{fe} i_b) Z_L$ (2)

Input voltage $V_1 = h_{ie} i_b$ (3)

Voltage gain in mid frequency range

$$A_{vm} = \frac{V_2}{V_1} = \frac{h_{fe} i_b Z_1}{h_{ie} i_b}$$

$$= -\frac{h_{fe}}{h_{ie}} Z_L \quad (4)$$

But $\frac{h_{fe}}{h_{ie}} = g_m$ (transconductance)

$\therefore$ $$A_{vm} = - g_m Z_L \quad (5)$$

Substituting the value of Z_L

$$A_{VM} = -\frac{h_{fe}}{h_{ie}} \left(\frac{h_{ie} R_L}{h_{ie} + R_L} \right) = -\frac{h_{fe} R_L}{(h_{ie} + R_L)} \quad (6)$$

$$= \frac{g_m R_L}{(1 + R_L / h_{ie})} \quad (7)$$

The minus sign in the above expressions shows that there is a phase difference of π - radian ($=180^0$) between output and input signal voltages.

(ii) Low frequency range:

In the low frequency region the impedance of the coupling condenser is significant and it can not be neglected but the impedance of stray capacitance which shunts the load is very high (almost infinite) so these can be considered as open circuit.

Under these conditions the h-parameter equivalent circuit of R-C coupled amplifier will be as shown in fig. (7.5)

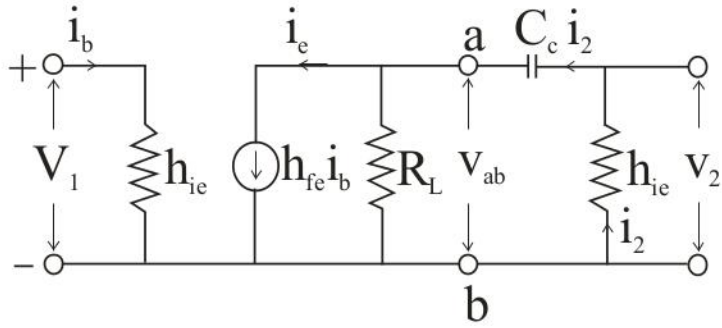


Fig. (7.5)

The input voltage $v_1 = i_b h_{ie}$

The output impedance of the first stage will be due to a parallel combination of R_L and $\left(h_{ie} \frac{1}{j\omega C_C} \right)$. Hence

$$\frac{1}{Z_L} = \frac{1}{R_L} + \left(\frac{1}{\left(h_{ie} - \frac{j}{\omega C_C} \right)} \right) \tag{8}$$

or $$Z_L = \frac{R_L \left(h_{ie} - j/\omega C_C \right)}{h_{ie} + R_L - j/\omega C_C} \tag{9}$$

The potential difference between a and b will be

$$V_{ab} = h_{fe}\, i_b\, Z_L \tag{10}$$

If the current flowing through h_{ie} and C_C is i_2 then

$$i_2 = \frac{V_{ab}}{\left(h_{ie} - j/\omega C_c \right)} = \frac{h_{fe} i_b Z_L}{h_{ie} - j/\omega C_C} \tag{11}$$

The output voltage for first stage will be the potential difference across h_{ie} i.e.

$$V_2 = - h_{ie}\, i_e$$

$$= -h_{ie} \frac{h_{fe} i_b Z_L}{\left(h_{ie} - j/\omega C_c \right)} \tag{12}$$

Hence voltage gain for low frequency region is

$$A_{ve} = \frac{V_2}{V_1}$$

$$= -\frac{h_{fe} Z_L}{\left(h_{ie} - j/\omega C_c\right)} \tag{13}$$

Substituting the value of Z_L from equation (9) in equation (13)

$$A_{ve} = -\frac{h_{fe}}{\left(h_{ie} - j/\omega C_c\right)} \times \frac{R_L\left(h_{ie} - j/\omega C_c\right)}{\left(h_{ie} + R_L - j/\omega C_c\right)} \tag{14}$$

Equation (14) shows that the voltage gain avl is a complex quantity whose magnitude and phase angle will be

$$|A_{vl}| = \frac{h_{fe} R_L}{\sqrt{\left(h_{ie} + R_L\right)^2 + \frac{1}{\omega^2 C_C^2}}} = \frac{A_{vm}}{\left[1 + \frac{j}{\omega^2 C_C^2 \left(h_{ie} + R_L\right)^2}\right]^{1/2}}$$

and phase angle $\phi_1 = \pi + \theta_1$ (15)

where $\theta_l = \tan^{-1}\left[\frac{1}{\omega C_C\left(h_{ie} + R_L\right)}\right]$ (16)

(iii) Lower half power frequency:

The frequency in the low frequency region at which the voltage gain is $\frac{1}{\sqrt{2}}$ or 0.707 times the voltage gain in mid frequency region is called lower half power frequency or lower cut off frequency. Hence at lower half power frequency

$$|A_{vl}| = \frac{|A_{vm}|}{\sqrt{2}}$$

If f_2 is lower half power frequency then from equation (15)

$$|A_{vl}| = \frac{|A_{vm}|}{\sqrt{1 + \left[\frac{1}{2\pi f_l C_c\left(h_{ie} + R_L\right)}\right]^2}} = \frac{|A_{vm}|}{\sqrt{2}} \tag{17}$$

So that $f_e = \frac{1}{2\pi C_C\left(h_{ie} + R_L\right)}$ (18)

Using the value of f_e from equation (18) in equation (14) the voltage gain in lower

frequency region will be obtained in terms of lower half power frequency.

$$A_{ve} = |A_{vl}| = \frac{|A_{vm}|}{1 - \frac{jf_1}{f}} \tag{19}$$

Hence when the frequency f decreases then A_{vl} decreases and the value of ϕ_e increases.

When $f = f_e$, $|A_{vl}| = \frac{|A_{vm}|}{\sqrt{2}}$ *and* $\theta_l = 45^0$

$$\phi_l = 180^0 + 45^0 = 225^0 \tag{20}$$

i.e. the phase difference between input and output signal is 225^0.

When $f \to 0, A_{vl} \to 0$ and $\phi_e \frac{3\pi}{2} = 270^0$

If the bypass capacitor C_E is also present in the circuit then due to this also the frequency response is affected like that due to C_C.

High frequency range:

In the high frequency range the effect of capacitor C_C is negligible and its impedance can be assumed zero. The effect of shunt capacitance is quite significant when the frequency is increases from mid frequency range. The equivalent circuit of single stage R-C coupled transistor amplifier in the high frequency range is shown in fig. (7.6). the capacitance C_d appears in the input of second stage due to Miller effect and this along with resistance h_{ie} appears in parallel with the load R_L.

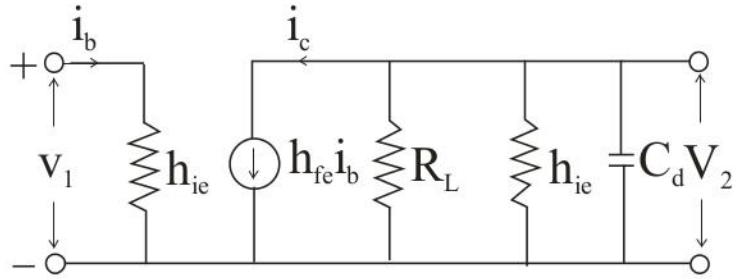


Fig. : (7.6)

The value of capacitance C_d is

$$C_d = C_{be} + (1 + g_m R_L)\, C_{bc} \tag{21}$$

If the effective load impedance is Z_L then

$$\frac{1}{Z_L} = \frac{1}{R_L} + \frac{1}{h_{ie}} + j\omega C_d$$

or
$$Z_L = \frac{h_{ie} R_L}{h_{ie} + R_L + j\omega C_d h_{ie} R_L} \tag{22}$$

Input signal voltage

$$V_1 = i_b h_{ie} \tag{23}$$

output signal voltage

$$V_2 = h_{fe}\, i_b\, Z_L \tag{24}$$

Hence the voltage gain in the high frequency region is

$$A_{vh} = \frac{V_2}{V_1} = -\frac{h_{fe} i_b Z_L}{h_{ie} i_b}$$

$$= \frac{h_{fe} Z_L}{h_{ie}} \tag{25}$$

Substituting the value of Z_L

$$A_{vh} = \frac{h_{fe} R_L}{h_{ie} + R_L + j\omega C_d h_{ie} R_L}$$

$$= \frac{A_{vm}}{1 + j\omega C_d \dfrac{h_{ie} R_L}{(h_{ie} + R_L)}} \tag{26}$$

The magnitude of voltage gain

$$|A_{vh}| = \frac{A_{vm}}{\left[1 + \left\{\dfrac{\omega C_d h_{ie} R_L}{(h_{ie} + R_L)}\right\}^2\right]^{1/2}} \tag{27}$$

Phase difference

$$\phi = 180^0 - \theta_h$$

where $\theta_h = \tan^{-1}\left(\dfrac{\omega C_d h_{ie} R_L}{h_{ie} + R_L}\right)$ (28)

Upper half power frequency: The frequency in the high frequency region at which the magnitude of gain is $\frac{1}{\sqrt{2}}$ or 0.707 times the voltage gain in mid frequency region is called upper half power frequency. If upper half power frequency is f_h then at $f = f_h$

$$|A_{vh}| = \frac{|A_{vm}|}{\sqrt{2}}$$

Using equation (27)

$$\frac{|A_{vm}|}{\sqrt{2}} = \frac{|A_{vm}|}{\sqrt{\left[1+\left\{\frac{2\pi f_h C_d h_{ie} R_L}{(h_{ie}+R_L)}\right\}^2\right]^{1/2}}}$$

So that $f_h = \frac{1}{2\pi C_d}\left(\frac{h_{ie}+R_L}{h_{ie}R_L}\right)$ (30)

Using equations (28) and (31) the voltage gain in high frequency region can be written in terms of upper half power frequency as

$$A_{vh} = \frac{A_{vm}}{1+j(f/f_h)} \quad (31)$$

or $$|A_{vh}| = \frac{|A_{vm}|}{\sqrt{1+(f/f_h)^2}} \quad (32)$$

Phase shift $$\theta_h = \tan^{-1}\left(\frac{f}{f_h}\right)$$

Total phase difference $\phi_h = 180^0 - \theta_h = 180^0 - \tan^{-1}\left(\frac{f}{f_h}\right)$ (33)

When frequency increases the voltage gain and total phase difference decrease in the high frequency region.

When $f = f_h$, $|A_{vh}| = \frac{|A_{vm}|}{\sqrt{2}}$

and $\phi_h = 180^0 - 45^0 = 135^0$. When $f \to \infty$, $A_{vh} \to 0$, $\phi_h \to \frac{\pi}{2} = 90^0$.

The dependence of voltage gain and phase difference between input and output signals on frequency are shown in fig. (7.7) respectively.

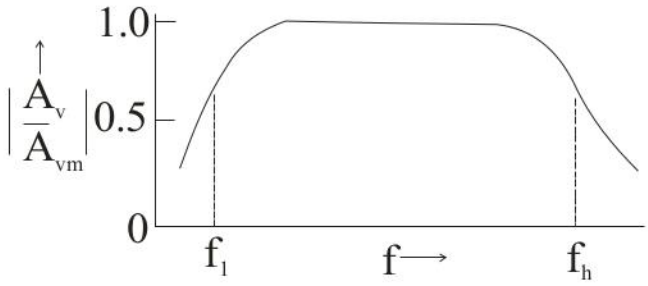


Fig. 7.7

7.6 Darlington Pair (Compound Transistor Configurations)

Composite Transistor Amplifiers are special purpose multistage Amplifiers considered as a single unit with special purposes. These configurations are popularly available in discrete Amplifiers as well as in Amplifiers using I_{Cs}. The various types of composite Transistor configurations are

1. (CC+CC) pair (Darlington configuration),
2. (CC+CB) pair and
3. (CC+CE) pair

7.6.1 Darlington (CC+CC) Pair

The Darlington pair configuration shown in Fig. 7.8 is named after its inventor Sidney Darlington, an engineer of Bell Telephone laboratories, USA.

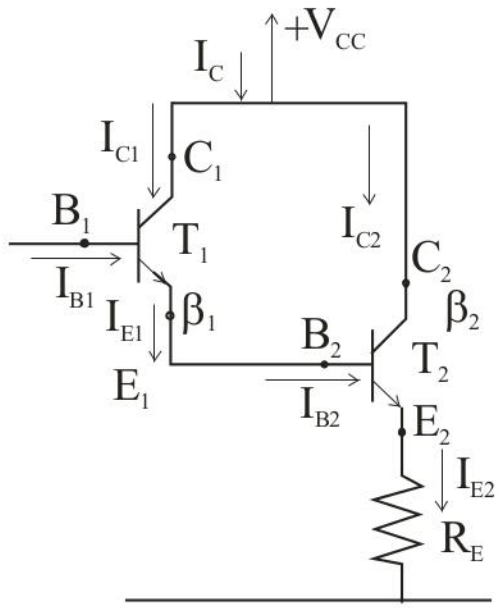


Fig. 7.8

It is a compound device using two Transistors in a single unit, wherein the Collector terminals are tied together; the Emitter current of one Transistor is connected to the Base of the second Transistor. Only three external connections base, Emitter and Collector are made available for application use. The configuration used is (CC + CC) Amplifier configuration or two Emitter follower circuits in one module.

Analysis :

$$I_{C1} = \beta_1 . I_{B1}$$

$$I_{B2} = I_{E1} = I_{B1} + I_{C1} = I_{B1} + \beta_1 . I_{B1} = [1+\beta_1]. I_{B1}$$

$$I_{C2} = \beta_2 . I_{B2}$$

But $I_C = [I_{C1}+I_{C2}] = \beta_1 . I_{B1} + \beta_2 . I_{B2}$

$$= \beta_1 . I_{B1} + \beta_2 . [1+\beta_1]. I_{B1} \quad (1)$$

$$\therefore \quad I_C = I_{B1} [\beta_1 + \beta_2 + \beta_1 . \beta_2] \quad (2)$$

Overall composite gain $[\beta_1 + \beta_2 + \beta_1 . \beta_2] \approx \beta_1 . \beta_2$. (3)

If β of each Transistor = 100, the overall current gain = 10, 200 from eq. (3), which is approximately equal to 10,000.

The two transistors need not necessarily be of NPN type. They can be PNP type Transistors also, which are used popularly in discrete circuit designs. But good quality PNP Transistors are rarely available. An alternative compound configuration in which one Transistor is PNP and another Transistor is NPN known as feedback pair is more popular in integrated circuits. Feedback a pair is shown in Fig. 7.9

The overall current gain of the feedback pair = $\beta_N \beta_P$ where β_N is the beta (β) parameter of NPN Transistor and β_P is the beta (β) parameter of PNP Transistor.

Advantages

- As the overall current gain is very large, they are used for large current applications to enhance the current supplying capacity of voltage regulators.
- High input impedance.

- They can be used as voltage follower with high current gain, high input resistance and overall voltage gain very close to unity.

Disadvantages

The leakage current if any in the first transistor in the Amplifier is amplified by the second Transistor results in high overall leakage current. This limits the usage of the Darlington pair for three or more stages.

The input impedance reduces when biasing arrangement is made in a Darlington pair. This disadvantage is overcome by adopting bootstrapping technique, wherein a capacitor is connected between the Collector of the first Transistor and the Emitter of the second Transistor and introducing a resistor in between the Base and the Collector circuit of the Transistor.

Darlington Emitter follower

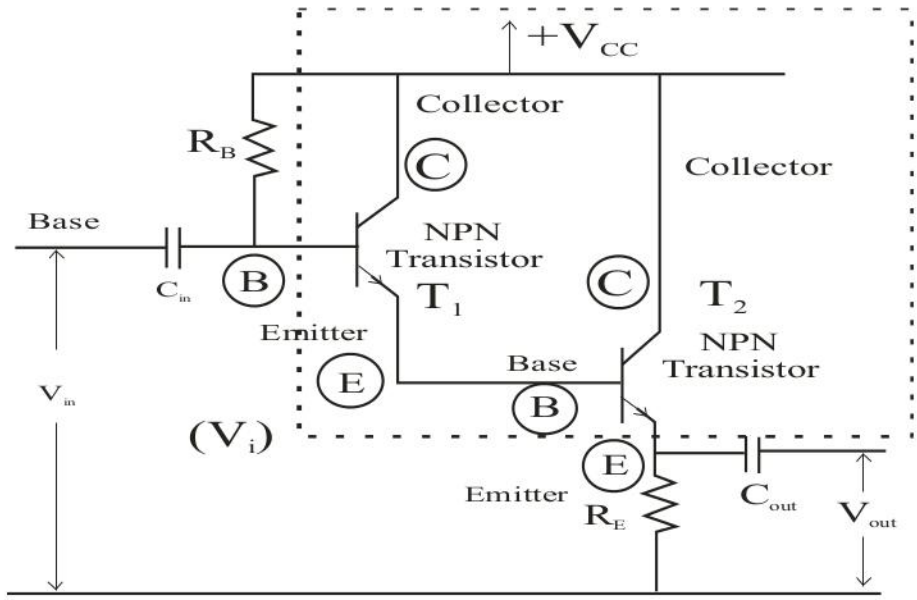


Fig. 7.9

Analysis of Darlington Emitter Follower

Figure 7.9 Shows Darlington Emitter follower circuit and Fig. 7.10 shows the small signal equivalent circuit of Darlington Emitter Follower.

Derivation for current gain A_1 of Darlington Emitter Follower

Current flowing through he input resistance R_{in} if the Transistor T_1 (SeeFigure 7.9)

$$I_B = \frac{[V_{in} - V_o]}{R_{in}} \tag{4}$$

Current of $I_B + \beta_D$. I_B flows through the Emitter resistor R_E where is β_D the overall β of the Darlington pair:

$$\therefore \quad [I_B + \beta_D . I_B].R_E = V_o. \tag{5}$$

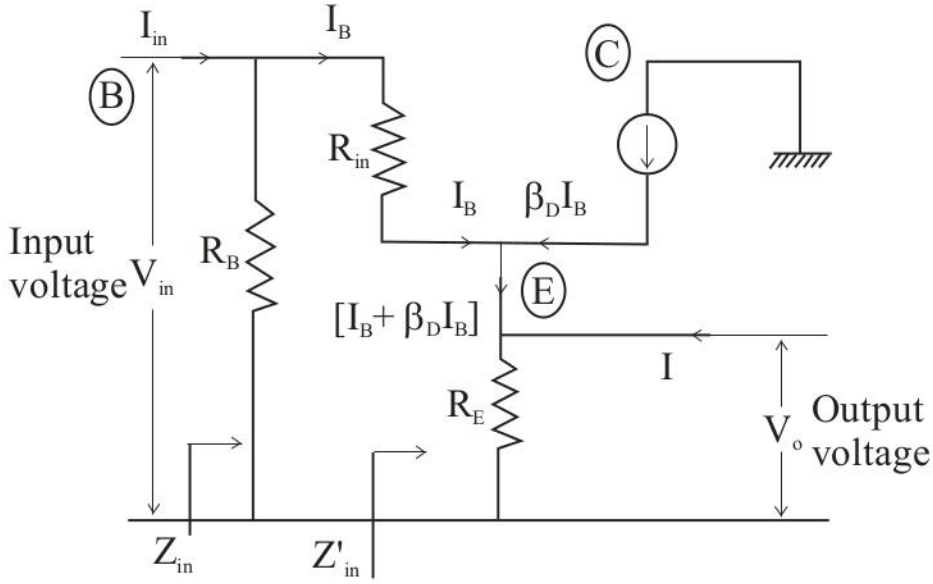


Fig. 7.10

Equation (4) can be written as

$$V_{in} = V_o + I_B . R_{in.}$$

Substituting Eq. (5) in the above equation,

$$V_{in} = [I_B + \beta_D . I_B].R_E + I_B . R_{in} = I_B . [(1 + \beta_D).R_E + R_{in}]$$

But $\beta_D \gg 1$. Therefore, $\beta_D . R_E \gg R_{in}$

$$\therefore \quad V_{in} = I_B \ [\beta_D . R_E]$$

Therefore, The impedance looking into the base of the Transistor $T_1 = Z'_{in}$:

$$Z'_{in} = \frac{V_{in}}{I_B} = \beta_D . R_E$$

Therefore, Input impedance

$$Z_{in} = \quad R_B \| Z'_{in} = R_B \| \beta_D . R_E \tag{6}$$

is very high. Hence, Darlington pair is a High input resistance circuit.

AC current gain

Current through R_E is I_o

$$I_o = I_B + \beta_D.I_B \approx \beta_D.I_B$$

$$\therefore \quad \frac{I_o}{I_B} = \beta_D. \tag{7}$$

Base current $I_B = \dfrac{R_B}{\left[\beta_D.R_E + R_B\right]} \times I_{in}$

$$\therefore \quad \frac{I_B}{I_{in}} = \frac{R_B}{\left[\beta_D.R_E + R_B\right]}. \tag{8}$$

Using Eqs. (7) and (8),

$$\text{Current gain} \quad A_I = \frac{I_o}{I_{in}} = \frac{I_o}{I_B} \times \frac{I_B}{I_{in}} = \frac{\beta_D.R_B}{\left[\beta_D.R_E + R_B\right]}. \tag{9}$$

AC output impedance Z_o

Simplified equivalent circuit to estimate Z_o is shown in Fig. 7.10

Make Source voltage zero in the equivalent circuit of Fig. 7.14 and apply a test voltage V at the output terminals to find I. Then the output impedance is the ratio of V to I. The method of finding the output impedance is shown in Fig. 7.11.

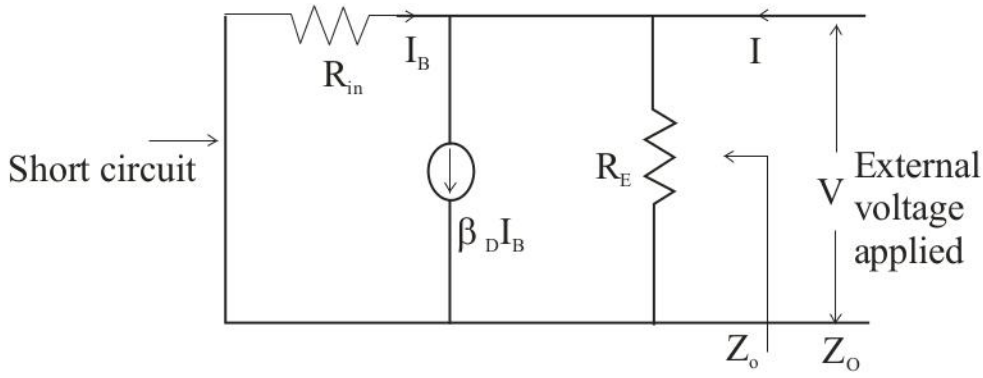


Fig. 7.11

$$I = \frac{V}{R_E} + \frac{V}{R_{in}} - \beta_D.I_B.$$

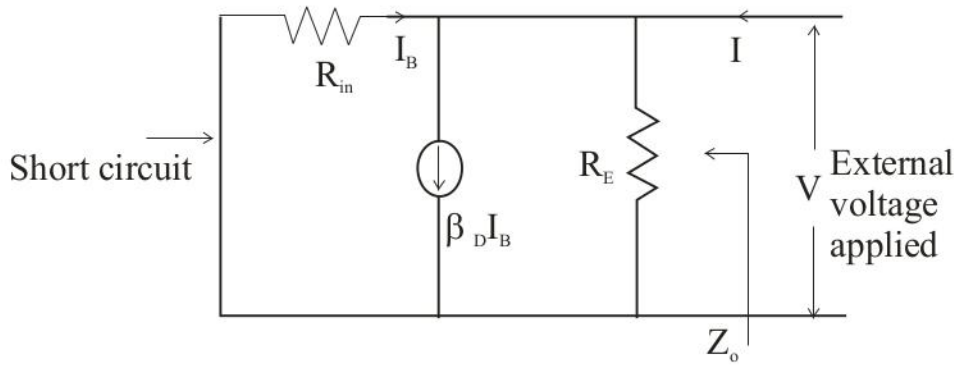


Fig. 7.12

Voltage gain A_V

Simplified equivalent circuit is shown in Fig. 7.13 to calculate voltage A_v.

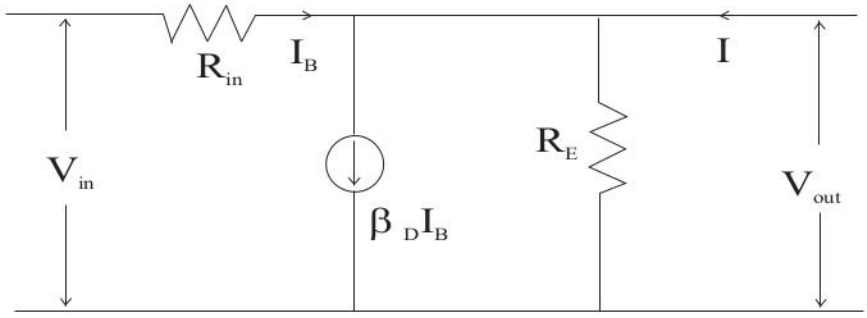


Fig. 7.16

$$V_{in} = I_B . R_{in} + I_B . R_E + I_B [\beta_D . R_E]$$

$$V_o = I_B . R_E + I_B . \beta_D . R_E$$

$$\therefore \text{ Voltage gain } \quad A_V = \frac{V_o}{V_{in}} = \frac{R_E [1+\beta_D]}{R_{in} + R_E [1+\beta_D]} \approx 1. \qquad (10)$$

From the above calculations, voltage gain of a Darlington pair is approximately 1.

7.7 Self learning Exercise-II

Section A: Very Short Answer Type Questions

Q.1 For a R-C coupled common emitter transistor amplifier the mid frequency voltage gain is $A_{vm} = 100$. The lower and upper cut off frequencies are f_e = 20 HZ and F_h = 20 KHz respectively. Determine the frequencies at which the voltage gain of the amplifier is 80.

Q.2 An RC couples amplifier uses transistor having $h_{ie} = 1k\Omega$ and

$h_{oe} = 25\times10^{-6}$ mho. The self bias resistors are R_1=18KΩ between the collector supply and base and R_2=2kΩ between base and ground. The load resistor R_C=1kΩ. What should be the minimum value of coupling capacitor in order to have lower 3dB frequency F_1 not exceeding 6HZ.

Q.3 What is the order of input resistance for Darlington emitter followers?

Section B: Short Answer Type Questions

Q.4 What is the phase shift introduced at high 3-DB frequency in a R-C coupled amplifier?

Q.5 What is the typical value of coupling capacitor in a R C coupled transistor amplifier?

Q.6 Give one advantage of RC coupled amplifier?

7.8 Summary

This technique consists in connecting several amplifier stages such that the output of one stage forms the input of the next stage. Cascading permits realization of any desired voltage or current gain subject, of course, to the practical limitations. R-C coupled amplifier is a cascade amplifier using a resistor as load impedance and a capacitor as the coupling element. It uses a resistor R_c as the load impedance and a high value capacitor C_b as the coupling capacitor. Darlington pair consists of two C_E transistors Q_1 and Q_2 in which emitter E_1 of Q_1 is connected to the base B_2 of Q_2, both collectors are directly connected to the V_{CC} supply and the emitter E_2 of Q_2 is grounded. It is an emitter follower using the darlington pair. It provides high input impedance R_i of the order of 2MΩ.

7.9 Glossary

Amplifier : Transistor circuits used to increase the voltage or power levels of an electrical signal.

Transistor : A semi conductor device with three or more elements , Emitter, base, Collector. A semiconductor device capable of providing power amplification.

Emitter : A region from which charge carriers that are minority (Carriers in the base are injected into the base.

Collector : A region through which a primary flow of charge carriers leaves the

base.

Darlington : Darlington Amplifier, a double emitter follower. Two transistor connector so they operate as if they are a single transistor. A Darlington pair can be either produced using in individual transistors or purchased as a single device.

7.10 Answers to Self Learning Exercises

Answers to Self Learning Exercise-I

Ans.1: 18.75

Ans.2: 24KΩ

Ans.3: Amplifier is defined as the transistor circuits used to increase the voltage or power levels of an electrical signal.

Ans.4: $A_F = \dfrac{h_{fe} R_e}{(R_s + h_{ie} + h_{fe} R_e)}$

Ans.5: $R_i = (R_s + h_{ie} + h_{fe} R_e)$

Ans.6: $R_o = \dfrac{R_s h_{ie}}{h_{fe}}$

Answers to Self Learning Exercise-II

Ans.1: In low frequency F= 26.66 HZ

In High frequency F = 15 KHZ

Ans.2: $C_b \geq$, 16.2μ farad.

Ans.3: 2 MΩ

Ans.4: 135^0

Ans.5: 10μ F

Ans.6: Economy

7.11 Exercise

Section-A (Very Short Answer type Questions)

Q.1 What is band width?

Q.2 What is -3dB frequency?

Q.3 What is the effect of coupling capacitor on voltage gain of an amplifier at different frequencies?

Q.4 Why are amplifiers connected in cascade?

Q.5 Which transistor configuration is most commonly used for multistage amplifier and why?

Section-B (Short Answer type Questions)

Q.6 Why is coupling capacitor in a common emitter R C coupled amplifier called blocking capacitor?

Q.7 Why is the gain of R C coupled amplifier constant is mid-frequency range?

Q.8 Define AC model of a transistor?

Q.9 What is a voltage amplifier?

Q.10 What is the difference between RC coupled and direct coupled amplifier?

Section-C (Long Answer Type Questions)

Q.11 Draw the circuit diagram of a two stage RC- coupled transistor amplifier. Explain why RC coupled amplifier circuits cannot be used to amplify slowly varying dc signals?

Q.12 Draw the circuit diagram of a R-C coupled transistor amplifier obtain the expression for voltage gain in lower, upper and mid-frequency range.

Q.13 Draw the voltage series feedback amplifier circuit. Derive the expression for voltage gain input and output resistance of the circuit.

Q.14 Draw the circuit of Darlington emitter follower. Derive expression for input resistance, over all voltage gain, over all current gain and output resistance.

References and Suggested Readings

1. G. K. Mithal, Electronics Devices and Circuits, Khanna Publishers, 2006.
2. B. Visvesvara Rao and K. Raja Rajeswari, Electronic Circuit analysis, Pearson, 2012.
3. Gupta and Kumar, Hand Book of Electronics, Pragati Prakashan, Meerut, 2000.
4. V. K. Mehta and Rohit Mehta, Principles of Electronics by Tenth Edition, S Chand and Company, 2006.

Unit 8
Power Amplifiers

Structure of the Unit

8.0 Objective

8.1 Introduction

8.2 Power amplifier classification

8.3 Some important terms

8.4 Self Learning Exercise-I

8.5 Illustrative examples

8.6 Class A amplifier

8.7 Class B amplifier

8.8 Class C amplifier

8.9 Self Learning Exercise-II

8.10 Illustrative examples

8.11 Summary

8.12 Answers to self learning exercises

8.13 Exercise

8.14 Answers to Exercise

References and Suggested Readings

8.0 Objectives

The aim of the unit is to familiarize the readers with the fundamentals of power amplifiers. The Class A, Class B and Class C type of power amplifier have been discussed in this unit with an object to acquaint the readers with their main features and application areas.

8.1 Introduction

An amplifier system usually receives a signal from some input source or some transducer, which is then amplified to a large signal suitable to drive an output

device. The amplifier system consists of several stages of amplifiers in cascade and at the final stage generally the power amplifiers are used to provide sufficient power in the form of large signal to the output load (speaker or transmitter). Thus power amplifiers are large signal amplifiers. Small signal amplifiers usually deal with small voltage and currents and therefore their power handling capacity is less. The main concern for small signal amplifier is for amplification linearity and the magnitude of gain instead of power efficiency. On the other hand the power amplifiers are usually used to handle large voltage signals with sufficiently large current levels, hence the main considerations with power amplifier is their power efficiency, power handling capacity and impedance matching.

8.2 Power Amplifiers Classification

The Power amplifiers are classified in terms of following classes.

Class A: The power amplifiers in which the output signal varies for a complete 360° cycle i.e. the collector current flows during the complete cycle is known as Class A amplifiers (Fig. 1).

Class B: In Class B amplifiers the variation of output signal is for one half the input signal cycle i.e. 180° of the signal. Two class B amplifiers are connected together to provide output of 360°. These are known as push-pull amplifiers.

Class C: The biasing of class C amplifiers is done so that the conduction takes place for much less than 180°. The power output from class C amplifiers is much more than class A, B or class AB but waveform is highly distorted. These are thus used for tuned circuits as tuned amplifiers at radio frequencies.

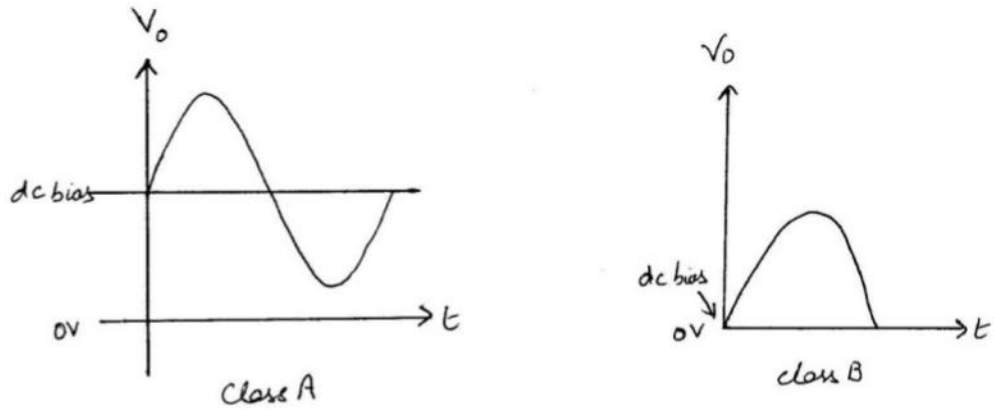


Fig. 1: Variation of output signal with time for Class A and Class B amplifier.

Class AB: In class AB amplifiers the output signal varies over more than one half the input signal cycles. The variation is between 180° to 360°.

Class D: Class D amplifiers are used to operate with digital or pulse signals. The power efficiency of class D amplifiers is very high but the input signal should be converted to pulse or digital waveform and after amplification the digital signal is again converted to sinusoidal signal.

8.3 Some Important Terms

(i) Power Gain– The ratio of the power delivered to the load to the power input or power delivered to the amplifiers. Thus

$$A_P = \frac{P_L}{P_i}$$

here A_p is the power gain, P_L in the power delivered by the amplifier to the load and P_i is the power delivered to the amplifier

As power can be expressed as

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

$$A_p = \frac{V_L^{\,2}(R_i)}{V_i^2(R_L)}$$

$$or\ A_p = A_v^{\,2}(\frac{R_i}{R_L})$$

$$where\ A_v = \frac{V_L}{V_i} = voltage\ gain$$

(ii) Efficiency – More important criterion than the power gain of a power amplifier is the conversion efficiency. The conversion efficiency or collector circuit efficiency is the ability of the power amplifier to convert the d.c. power from the supply into a.c. (signal) power output delivered to the load. The conversion efficiency is defined as

$$\eta = \frac{\text{a.c. output power delivered to the load}}{\text{d.c. power supplied by the battery}} \times 100\%$$

(iii) Distortion –

A pure sinusoidal signal varies over a cycle of 360° and has a single frequency.

The voltage variation is also by same amount in positive and negative parts. Ideally the output from an amplifier should be exactly of same shape as input, should vary 360° and has a single frequency but generally it does not happen. Whenever, there in deviation in output wave shape of an amplifier from the input wave shape this is known as distortion.

There can be amplitude distortion and frequency distortion. Amplitude distortion occurs due to nonlinearity of device characteristic. Frequency distortion occurs because the response of circuit elements to the input signal is different at different frequencies. According to Fourier analysis distorted periodic waveforms are described in terms of its fundamental frequency component and integer multiple of frequency components known as harmonics. Harmonic distortion in a signal implies there are harmonic frequency components apart from fundamental frequency. Harmonic distortion can be written as

$$\%HD_n = \frac{|A_n|}{|A_1|} \times 100\%$$

Where HD_n is n^{th} harmonic distortion, A_1 in amplitude of fundamental frequency and A_n in amplitude of nth frequency components.

Total harmonic distortion is expressed as

$$\%THD = \sqrt{HD_2^{\,2} + HD_3^{\,2} + HD_4^{\,2} + ...} \times 100\%$$

For small signals, transistor behaves as a linear device about the operating point and hence distortion is not problem for voltage amplifiers. On the other hand power amplifiers handle large signal and therefore the problem of distortion is more pronounced. An amplifier is considered better which has less distortion.

(iv) Heat sinking -

Power transistors handle large currents and get heated up during the operation which results in increase in temperature. The maximum power handled by the transistor and temperature of the collector junction of the device are related. Temperature acts as the limiting factor in power handling of the transistor. So it is importance to allow transistor to dissipate heat. Heat sinking plays an important role in maintaining the temperatures at lower values during operation. Generally power transistors are placed in large metal cases with increased surface area to allow the heat to escape. The metal case is connected to the collector terminal of

the transistor. An ideal infinite heat sink can hold the case temperature at ambient air temperature but practically no such heat sink is available. Thus power dissipation is allowed only up to a certain temperature. With use of a good heat sink the power dissipation capability of a power transistor is increased but above a certain temperature, the power dissipation capacity is reduced or derated. The magnitude of maximum power allowed for a particular transistor is reduced or derated with increase in case temperature. Mathematically,

$P_D\ (t_1) = P_D(t_0)-(t_1-t_0)$ (derating factor)

Here t_o is the temperature at which derating should start, t_1 is the temperature at which we have to find power dissipation, P_D represent the maximum power dissipation at a particular temperature. The value of derating factor can be obtained from the specification sheet of the device. Its unit is W/^{o}C.

The permissible power dissipation for a power transistor can be expressed as

$$P_T = \frac{T_{J\max} - T_a}{\theta}$$ where T_{jmax} = maximum junction temperature

T_a = ambient temperature

θ = Thermal resistance

Unit of θ is ^{o}C/Watt

8.4 Self Learning Exercise-I

Very Short Answer Questions

Q.1 Class A amplifier operates for how much of the input cycle?

Q.2 Why does amplifier distortion occurs?

Q.3 Which terminal of transistor is connected to the metal case?

Short Answer Type Question

Q.4 What is power gain for a power amplifier?

Q.5 Define conversion efficiency.

8.5 Illustrative Examples

Example 1 Determine the output power of the amplifiers given that input voltage is 700 mV peak to peak, $R_i = 3k\Omega$ and $A_p = 45000$.

Sol. The peak to peak input voltage is 700 mV

Thus input has value of voltage is $\frac{350mV}{\sqrt{2}} = 247.5mV$

Input power is $p_{in} = \frac{V_i^2}{R_i} = \frac{(247.5mV)^2}{3k\Omega}$

$$p_{in} = 20.4\mu W$$

The output power is given by

$$P_{out} = P_{in} A_p = (20.4\mu W)\ (45000)$$
$$= 0.918W.$$

Example 2 Find the harmonic distortion components for a signal given that fundamental frequency is 3V, second harmonic is 0.25V, third harmonic is 0.1 V and fourth harmonic is 0.03V.

Sol. The harmonic distortion components are

$$\%HD_2 = \frac{0.25V}{3V} \times 100 = 8.33\%$$
$$\%HD_3 = \frac{0.1V}{3V} \times 100 = 3.33\%$$
$$\%HD_4 = \frac{0.03V}{3V} \times 100 = 1\%$$

Example 3 Find the magnitude of maximum dissipation at 80°C and 100°C for a 100W silicon transistor rated at 25°C if derating is required above 25°C. The derating factor is given as 0.5W/ °C .

Sol. $P_D(80^oC) = P_D(25^oC) - (80^oC - 25^oC)\ (0.5W/^oC)$

$= 100W - (55^oC)\ (0.5W/^oC)$

$= 100W - 27.5W$

$= 72.5W$

$P_D(100^oC) = P_D(25^oC) - (100^oC - 25^oC)\ (0.5W/^oC)$

$= 100W - (75^oC)\ (0.5W/^oC)$

$= 100W - 37.5W = 62.5W$

8.6 Class A Amplifier

Class A amplifier operates for a full 360° in the linear region. For this the Q-point is biased such that the signal may swing up and down (Fig. 1) .The d.c. bias is at one-half the supply voltage level, this results in poor efficiency. The maximum efficiency of class *A* amplifier for series-fed connection is 25% and for transformer coupled class A amplifier it is 50%.

8.6.1 Series Fed Class A Amplifier

The fixed bias circuit is used in series-fed class A amplifier with power transistor as shown in Fig. 2.

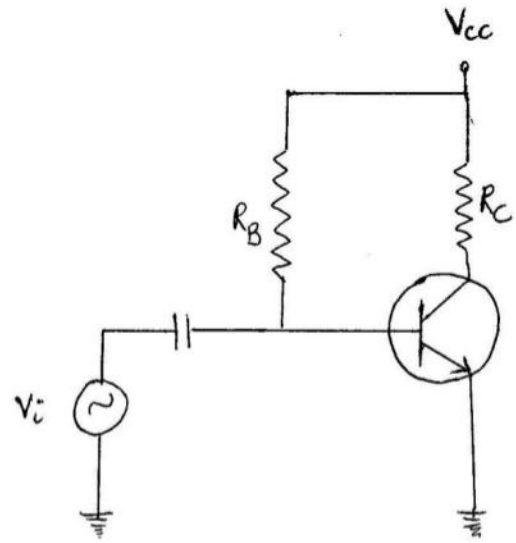


Fig. 2: Series fed Class A amplifier.

The Q-point is situated at the centre of the d.c. load line in order to achieve full swing. The efficiency of this amplifier is quite less and so it is rarely used.

The dc base-bias current is given by

$$I_B = \frac{V_{cc} - 0.7V}{R_B}$$

$I_C = \beta I_B$ and $V_{CE} = V_{CC} - I_C R_C$ for dc bias operation. For estimating the efficiency of the amplifier we require output ac power and input dc power.

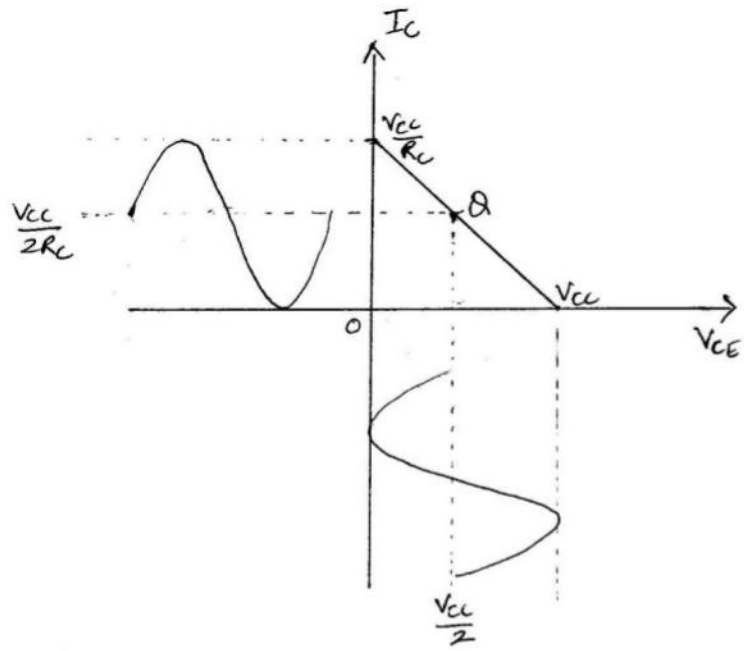


Fig.3: Q-point for series –fed Class A amplifier.

The input dc power from the supply is given as

$$P_{i_{dc}} = V_{CC} I_{CQ}$$

here I_{CQ} is the collector current at Q-point (Fig .3) which is $\frac{V_{CC}}{2R_C}$

Therefore $$P_{i_{dc}} = \frac{V_{CC}^2}{2R_C}$$

The ac output power given by

$$P_{o_{ac}} = \frac{v_{ce(p-p)}}{2\sqrt{2}} \times \frac{i_{c(p-p)}}{2\sqrt{2}}$$

where factor $2\sqrt{2}$ is used to convert peak to peak value in r.m.s. value

Thus $$P_{o_{ac}} = \frac{v_{ce(p-p)} i_{c(p-p)}}{8}$$

As $v_{ce(p-p)} = V_{CC}$, $i_{c(p-p)} = \frac{V_{CC}}{R_C}$

$$P_{o_{ac}} = \frac{V_{CC}^2}{8R_C}$$

Maximum efficiency is

$$\eta = \frac{P_{oac}}{P_{i_{dc}}} \times 100 = \frac{V_{CC}^{\ 2} / 8R_C}{V_{CC}^{\ 2} / 2R_C} \times 100\% = 25\%$$

Thus the maximum efficiency of series fed class A amplifier is only 25%, practically the value is even less than this hence it is not preferred.

8.6.2 Transformer Coupled Class A Amplifier

In this amplifier a transformer is used to couple the output signal to the load (Fig.4). Transformer coupling allows for better impedance matching and in reducing dc power loss thus maximum efficiency of this amplifier is 50%. The drawbacks of transformer coupling is size and cost of transformer and distortion problem.

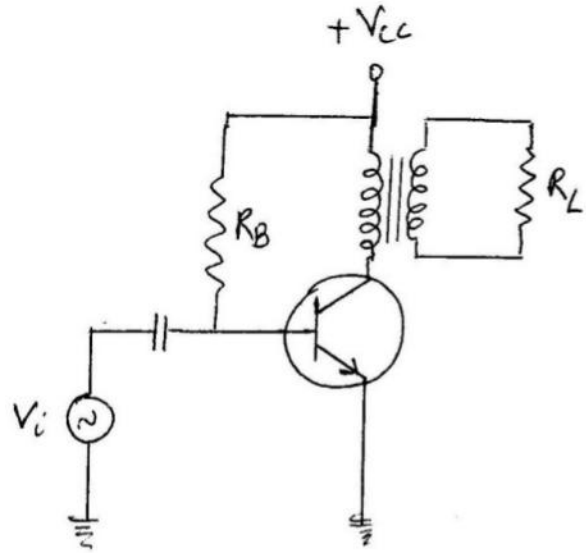


Fig.4: Transformer coupled Class A amplifier.

Impedance Transformation:

As per the simple transformer relations

$$\frac{V_1}{V_2} = \frac{N_1}{N_2} \quad \text{and} \quad \frac{I_1}{I_2} = \frac{N_2}{N_1}$$

where V_1 and V_2 are primary and secondary voltages respectively, I_1 and I_2 are primary and secondary currents respectively and N_1 and N_2 are the numbers of primary and secondary turns.

The turn ratio $n = \frac{N_2}{N_1}$

determines the stepping up or down of the voltage. If $N_2 < N_1$ voltage is stepped down and current is stepped up. A relation for resistance can be found as follows

$$\frac{V_1}{I_1} = \frac{1}{n^2}\frac{V_2}{I_2} \quad So \; R_{Leff} = \frac{R_L}{n^2}$$

here R_{Leff} is the effective input resistance $\frac{V_1}{I_1}$ and R_L is the output resistance $\frac{V_2}{I_2}$

Maximum efficiency :

The load line of the amplifier is shown in Fig. 5. The dc load line is straight vertical line as dc resistance is very small ideally taken as 0Ω.

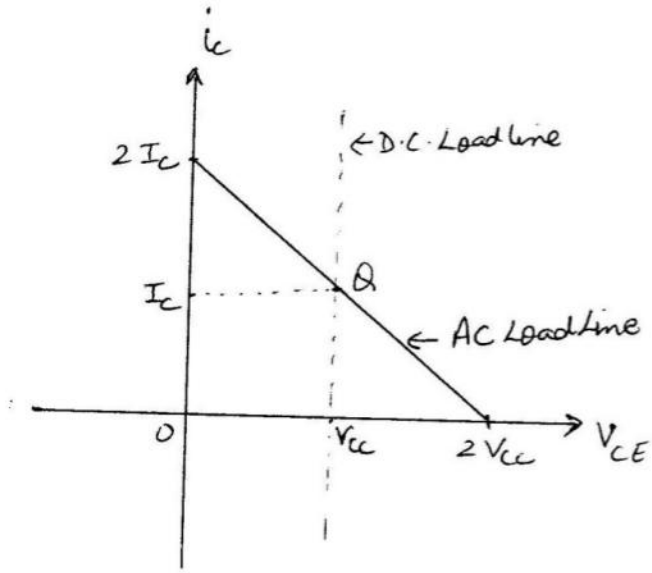


Fig.5: Load lines for transformer coupled Class A amplifier.

Due to presence of transformer the peak to peak signal swings are.

$$i_{C(p-p)} = 2I_C$$
$$V_{ce(p-p)} = 2V_{CC}$$

Input dc power $P_{dc} = V_{cc} I_c$

$$= I_c^2 R_{Leff} \; (\because V_{cc} = I_c R_{Leff})$$

Maximum ac output power $P_{ac\max} = \frac{v_{ce(p-p)} i_{c(p-p)}}{8}$

$$= \frac{1}{2} V_{cc} I_c$$

$$= \frac{1}{2} I_c^{\,2} R_{Leff}$$

The maximum collector efficiency

$$= \frac{P_{ac_{\max}}}{P_{dc}} \times 100$$

$$= \frac{\frac{1}{2}(I_C^{\,2} R_{Leff})}{I_C^{\,2} R_{Leff}} \times 100$$

$$= 50\%$$

Efficiency And Power Dissipation:

The ac power delivered to the load can be expressed as

$$P_{ac} = \frac{(v_{ce_{\max}} - v_{ce_{\min}})(i_{c_{\max}} - i_{c_{\min}})}{8}$$

DC power drawn from the supply

$$P_{dc} = V_{cc} I_{CQ}$$

$$\text{efficiency} = \frac{P_{ac}}{P_{dc}} \times 100$$

$$\text{Power dissipation} = P_{dc} - P_{ac}$$

Maximum theoretical efficiency based on signals is given by

$$\%\eta = 50\left(\frac{v_{ce_{\max}} - v_{ce_{\min}}}{v_{ce_{\max}} + v_{ce\min}}\right)^2 \%$$

8.7 Class B Amplifiers –Push Pull Circuit

Class B amplifiers operate in the linear region for 180° and are in cutoff for 180°.The transistor conducts for only one half of the cycle. In order to amplify the full cycle, two class B amplifiers are used which conduct on opposite half cycles. The combination is known as push-pull circuit. The class B amplifiers are more efficient than class A amplifiers. The block diagram of push- pull operation of class B amplifiers is shown in Fig. 6.

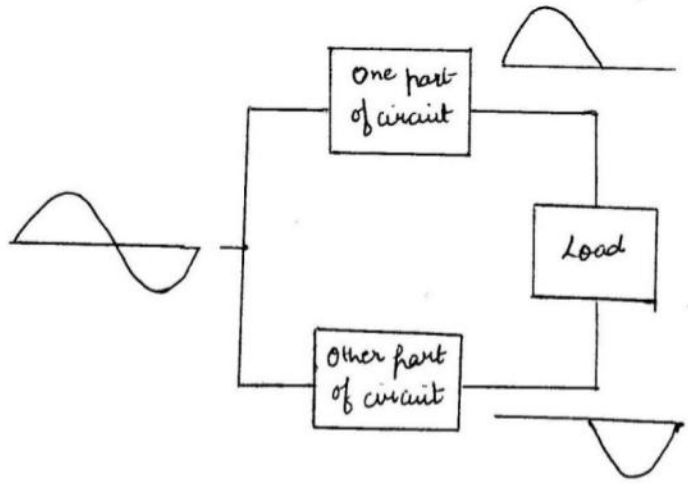


Fig. 6. Block diagram of push-pull operation of Class B amplifier.

The input dc power in expressed as

$$P_{i_{dc}} = V_{cc} I_{dc}$$

Where I_{dc} is the average dc current obtained from the power supplies. The average current in terms of peak value of the output current waveform I_p is given as

$$I_{dc} = \frac{2}{\pi} I_p$$

Thus $P_{i_{dc}} = \frac{2V_{cc} I_p}{\pi}$

The output ac power in terms of peak to peak output voltage $V_{L(p\text{-}p)}$ is given by

$$P_{o_{ac}} = \frac{V_{L\,(p-p)}^{\,2}}{8R_L} = \frac{V_{L\,(p)}^{\,2}}{2R_L}$$

where $V_{L(p)}$ is peak value of output voltage and is $V_{L(p)} = \frac{V_{L(p-p)}}{2}$

Efficiency:

The efficiency of class B amplifier is given as

$$\eta = \frac{P_{o_{ac}}}{P_{i_{dc}}} \times 100$$

$$\eta = \frac{V_{L_{(p)}}{}^{2} / 2R_L}{2V_{CC} I_P / \pi} \times 100 = \frac{\pi}{4} \frac{V_{L(p)}}{V_{cc}} \times 100$$

$$[\because I_p = \frac{V_{L(P)}}{R_L}]$$

This shows that efficiency increases with increase in the peak voltage. The maximum efficiency is when $V_{L(p)}=V_{cc}$

$$\eta_{max} = \frac{\pi}{4} \times 100\% = 78.5\%$$

Power dissipation:

The power dissipated by the two transistors in push-pull circuit is expressed as

$$P_{2Q} = P_{i_{dc}} - P_{o_{ac}}$$

The power dissipated by each transistor is

$$P_Q = \frac{P_{2Q}}{2}$$

Circuit Of Transformer Coupled Push-Pull Amplifier:

The push pull circuit consists of a phase splitting input transformer, push-pull circuit connection, output transformer and load on show in Fig.7. The phase splitting input transformer is a centre tapped transformer which produces opposite polarity signals which act an inputs for the two transistors.

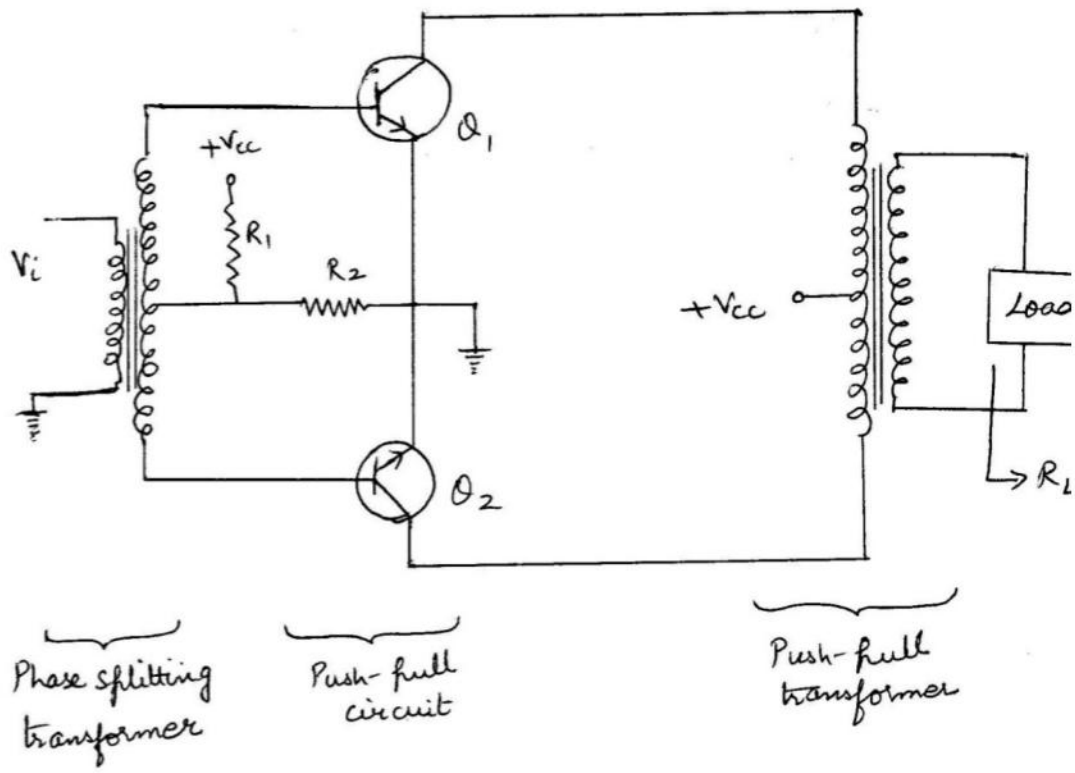


Fig. 7. Transformer coupled push-pull amplifier circuit.

The output transformer drives the load in push- pull mode and produces signals across load over the full cycle. The transistor Q_1 is driven into conduction during the first half-cycle of operation while transistor Q_2 is in off state. The current I_1during first half cycle produces signal to the load. In the next half cycle Q_2 conducts and Q_1 is off which results in current I_2. The current I_2 through the transformer produces second half cycle of output across load. Thus overall a signal is developed across the load for the full cycle of operation.

Complementary-Symmetry Circuit:

Full cycle output may also be obtained by using complimentary transistors (npn and pnp) an shown in fig 8. Here in this circuit one disadvantage is use of two separate supplies and another is crossover distortion. Crossover distortion is presence of some nonlinearity in the output signal when the signal crosses over from positive to negative. This is due to the fact that exact switching over of transistors does not take place.

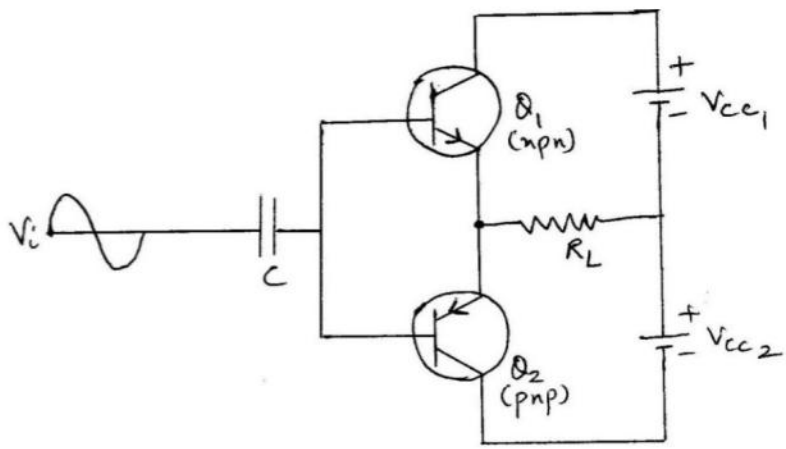


Fig. 8. Complementary-symmetry push-pull amplifier circuit.

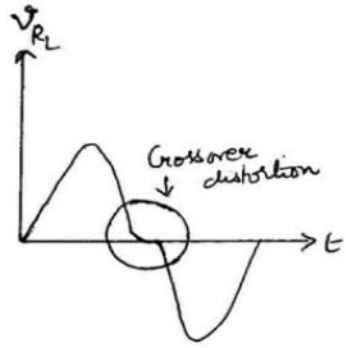


Fig. 9 Crossover distortion in complementary-symmetry push-pull amplifier

8.8 Class C Amplifiers

Class C amplifiers conduct for less than 180° of the cycle. These are more efficient than class A or Class B amplifiers but there is more distortion in the output waveform of Class C amplifiers as compared to Class A or Class B. amplifiers.

The basic resistively loaded Class C amplifier has been shown in Fig. 10, but this circuit is not used as output is not for full cycle and is highly distorted.

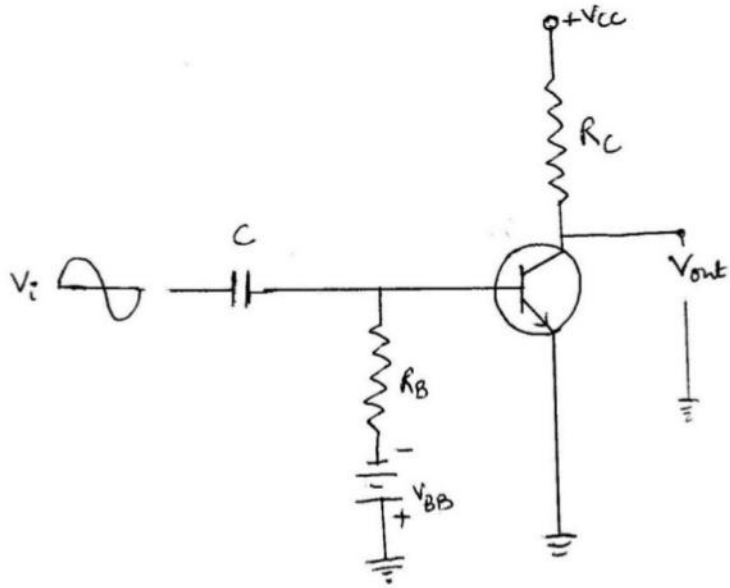


Fig.10. Resistively loaded Class C amplifier circuit.

Instead tuned Class C amplifier circuit is used which provides full cycle of output signal. The tuned Class C amplifier finds application in radio and communication. The basic circuit of tuned Class C amplifier has been shown in Fig 11. Here tank circuit is used instead of resistance.

The current pulses from the collector are responsible for initiation and sustenance of tank circuit oscillations. The charging-discharging of the capacitor and inductor produces a full cycle of output signal. Initially the current pulse charges the capacitor to V_{cc}. After the pulse ends the capacitor discharges and in turn charges the inductor. When capacitor completely discharges there is collapse in inductor's magnetic field which recharges capacitor in opposite direction. Like this the process goes on and one complete cycle is produced with peak to peak output

voltage of value around $2V_{cc}$. The regular current pulses from the collector sustain the oscillations of the tank circuit to constant amplitude.

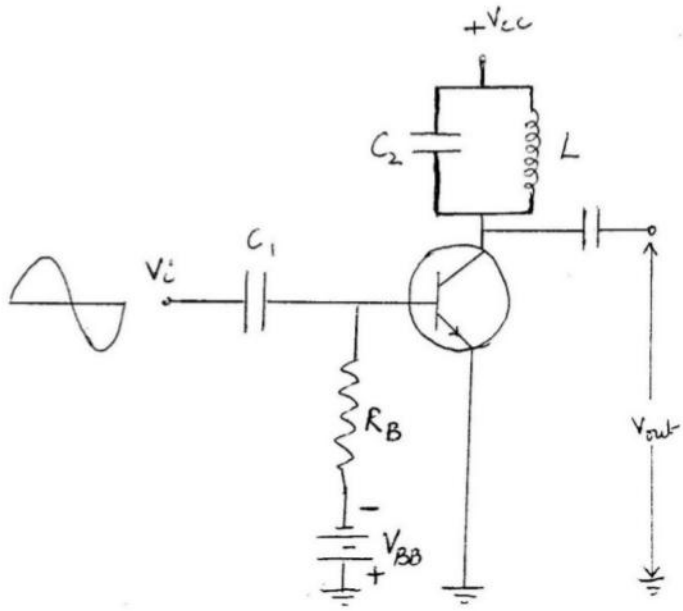


Fig. 11. Tuned Class C amplifier circuit.

When the tank circuit is tuned to the frequency of the input signal the output waveform has oscillations at fundamental frequency. Thus this circuit can be used at one fixed frequency, as used in communication circuit.

Power Dissipation :

The power dissipation of Class C amplifier is low as it conducts for only a small part of the input cycle. Using ideal pulse approximation the power dissipated during the time when transistor conducts is given by

$$P_{Don} = I_{Csat}\, V_{CEsat}$$

Where I_{Csat} is the maximum current amplitude and V_{CEsat} is minimum voltage amplitude.

The average power dissipation for the entire cycle is given by

$$P_{D_{avg}} = (\frac{t_{on}}{T}) P_{D_{on}}$$

where t_{on} is the time for which the transistor is on.

Efficiency:

The peak is peak output voltage value across the tank circuit is approximately $2V_{cc}$ and V_{rms} is $\frac{V_{cc}}{\sqrt{2}}$. Thus maximum output power can be expressed as $P_o = \frac{V_{rms}^2}{R_c} = \frac{V_{cc}^2}{2R_C}$

Here R_c is the equivalent of the collector tank circuit.

The efficiency of the amplifier is given by

$$\eta = \frac{P_o}{P_o + P_{D_{avg}}}$$

as P_{Davg} is very less the efficiency of class C amplifiers approaches unity i.e. around 100%.

8.9 Self-Learning Exercise -II

Very Short Answer Type Questions

Q.1 Where is the Q-point for Class B amplifiers?

Q.2 Which amplifier has minimum efficiency – Class A, B or C?

Q.3 Which power amplifiers operate in linear region at all time?

Short Answer Type Questions

Q.4 What are the benefits of transformer coupling in Class A amplifiers?

Q.5 Why is tank circuit used in class C amplifiers?

8.10 Illustrative Examples

Example 1 Find the maximum a.c. power output for a transformer coupled Class A amplifier given the turn ratio is 5, secondary load is 120 Ω and zero signal collector current is 100 mA.

Sol. n=5, I_C=100mA, R_L=120 Ω

The maximum a.c. power output $P_{ac_{\max}} = \frac{1}{2} I_C^2 R_{Leff}$

As $R_{Leff} = n^2 R_L$

$$P_{ac_{\max}} = \frac{1}{2} I_C{}^2 (n^2 R_l$$

$$= \frac{1}{2}(100mA)^2 (5)^2 120\Omega$$

$$= 15W$$

Example 2 Find the efficiency of a transformer-coupled Class A amplifier for outputs of 10V and 6V when the supply voltage is 10 V

Sol. The maximum theoretical efficiency is given by –

$V_{CEmax} = V_{CEQ} + V_0$

Where $V_{CEQ} = V_{cc}$ and V_0 is output voltage

For V_0=10V

$V_{CEmax} = 10V + 10V = 20V$

$V_{CEmin} = 10V - 10V = 0V$

$$\%\eta = 50(\frac{20V - 0V}{20V - 0V})^2 \% = 50\%$$

For V_0=6V

$V_{CEmax} = 10V + 6V = 16V$

$V_{CEmax} = 10V - 6V = 4V$

$$\%\eta = 50(\frac{16V - 4V}{16V - 4V})^2 \% = 30\%$$

Example 3 Find the input power, output power and efficiency of Class B amplifier which gives a 18V peak signal to a 15 Ω load. The power supply is 25V.

The peak load current $I_P = \frac{V_{L(p)}}{R_L} = \frac{18V}{15\Omega} = 1.2A$

The average dc current from the supply is

$$I_{dc} = \frac{2}{\pi} I_P = \frac{2}{\pi}(1.2A) = 0.764A$$

Input power $P_{i_{dc}} = V_{CC} I_{dc} = (25V)(0.764A)$

$= 19.1$ W

Output power $P_{o_{ac}} = \frac{V_{L(p)}^{\ 2}}{2R_L} = \frac{(18V)^2}{2(15\Omega)} = 10.8W$

Efficiency $\%\eta = \frac{P_{oac}}{P_{i_{dc}}} \times 100\% = \frac{10.8}{19.1} \times 100\%$

$= 56.5\%$

Example 4 Find average power dissipation and efficiency of Class C amplifier given that the transistor is on for $1\mu s$ and time power is $5\mu s$, $I_{C_{sat}} = 100mA, V_{ce(sat)} = 0.25V, V_{cc} = 24V$ and R_c is 120Ω.

Sol. Average power dissipation

$$= \left(\frac{t_{on}}{T}\right) I_{c(sat)} V_{ce(sat)}$$

$$= \left(\frac{1\mu s}{5\mu s}\right)(100mA)(0.25V) = 5mW$$

$$Po = \frac{V_{cc}^2}{2R_c} = \frac{(24)^2}{2x120} = 2.4\ W$$

Efficiency = $\frac{P_o}{P_0 + P_{Davg}} = \frac{2.4W}{2.4W + 5mW}$

$= 0.998$

$or\ \eta\% = 99.8\%$

8.11 Summary

The unit described the fundamentals of power amplifiers. Power amplifiers are used for power amplification and are large signal amplifiers. The power amplifiers are classified as Class A, Class B, Class C and Class D. The basic terms such as efficiency, distortion, power gain and power dissipation have been explained. The circuit and operation of Class A, Class B and Class C amplifiers has been described.

8.12 Answer To Self Learning Exercises

Answer To Self Learning Exercise-I

Ans.1: Full

Ans.2: Nonlinearity of device characteristic

Ans.3: Collector

Ans.4: Ratio of power delivered to the load to the input power

Ans.5: Ability of an amplifier to convert the d.c power from supply to a.c output signal

Answer To Self Learning Exercise-II

Ans.1: Cut off

Ans.2: Class A

Ans.3: Class A

Ans.4: Belter impedance matching and increase in efficiency

Ans.5: To provide full cycle output signal

8.13 Exercise

Very Short Answer Type Questions

Q.1 What is the maximum efficiency of Class B amplifier?

Q.2 Class C amplifiers are biased at which point?

Q.3 What is the unit of derating factor?

Q.4 Class D amplifiers operator with which type of signal?

Q.5 Where should be the Q point for class A amplifiers?

Short Answer Type Questions

Q.6 What are the drawbacks of transformer coupling in class A amplifiers?

Q.7 Where do class C amplifiers find application?

Q.8 What is crossover distortion?

Q.9 What is the importance of heat sink in power amplifiers?

Q.10 Why two class B amplifiers are used in push pull circuit?

Long Answer Type Questions

Q.11 Differentiable between Class A, Class B and Class C amplifiers.

Q.12 Describe Class B amplifiers push-pull circuit and determine the maximum efficiency for Class B amplifiers.

Q.13 Describes the operation of tuned class C amplifiers and determine efficiency for it.

Q.14 Find the average power dissipation for maximum output of Class C amplifiers which is on for 10 percent of the input cycle,

$V_{ce(sat)} = 0.18V \;\; I_{c_{sat}} = 25mA.$

Q.15 For a series fed Class A amplifier V_{ce}=10V, I_c=0.5A

$V_{ce} = 24V, i_b = 10mA, B = 25, R_c = 20\Omega$

Find output power, input power and collector efficiency of the amplifier.

8.14 Answers To Exercise

Ans.1: 78.5%

Ans.2: Below out off

Ans.3: W/^{0}C

Ans.4: Pulse or digital

Ans.5: Centered on the load line

Ans.14: 450μW

Ans.15: Output power P_{oac}=0.625W, P_{dc}=12W, $\eta = 5.2\%$

References and Suggested Readings

1. Electronic Devices and Circuit Theory, Robert L. Boylestad and Louis Nashelsky, Pearson Publication, 2008
2. Electronic Devices, Floyd, Pearson Publication, 2012.
3. Integrated Electronics, J. Millman and C.C. Halkias, Tata McGraw Hill,1991.
4. Principles of Electronics, V.K. Metha ,S. Chand & Co.,2008.
5. Electronic Device and Circuits, G.K. Mittal, Khanna Publishers, 2005.

UNIT-9
Field Effect Transistors

Structure of the Unit

9.0 Objectives

9.1 Introduction

9.2 JFET – construction

9.3 JFET – operation

9.4 JFET – characteristics

9.5 JFET – parameters

9.6 Self Learning Exercise -I

9.7 Illustrative examples

9.8 FET small signal model

9.9 MOSFET

9.10 Depletion – type MOSFET

9.11 Enhancement – type MOSFET

9.12 Self Learning Exercise – II

9.13 Illustrative Examples

9.14 Summary

9.15 Answers to self learning exercise

9.16 Exercise

References and Suggested Readings

9.0 Objectives

The aim of this unit is to familiarize the readers to the principle, operation and characteristics of the field effect transistors. The fundamentals of the junction field effect transistors and metal oxide semiconductor field effect transistors are described here. The FET parameters and the small signal model have also been discussed.

9.1 Introduction

Field effect transistors are used to a large extent in various applications where bipolar junction transistors (BJT) are used. But there are some basic differences between the two types of transistors. The conduction in bipolar junction transistor (BJT) depends on two types of charge carriers namely electrons and holes, whereas the conduction in field effect transistor (FET) is dependent on any one type of charge carrier and hence it is also known as unipolar device. Another major difference between BJT and FET is the controlling parameter. BJT is a current controlled device whereas FET is a voltage controlled device. In BJT the output current I_C is controlled by input current I_B whereas in FET the output current I_D is controlled by voltage V_{GS}. Further FET have high input impedance, and more stable with changes in temperature .It is usually smaller than BJT. The ac voltage gains are more for BJT amplifier as compared to FET whereas the noise levels are lower for FET when compared to BJT.

9.2 Junction Field Effect Transistor (JFET) –Construction

JFET is a three terminal voltage controlled device. As per construction JFET can be either n-channel or p-channel JFET. In an n-channel JFET, the major portion is of n type material which forms a channel between p-type materials (Fig.1). The p-type materials on both the side of the n-channel are joined together and the terminal is named as Gate (G). The other two terminals are Source (S) and Drain (D). Source terminal is at the bottom of the n-channel and Drain terminal is at the top of the n-channel.

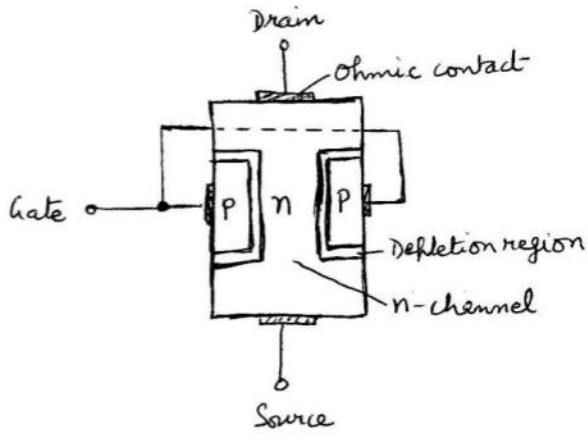


Fig.1. n-channel JFET

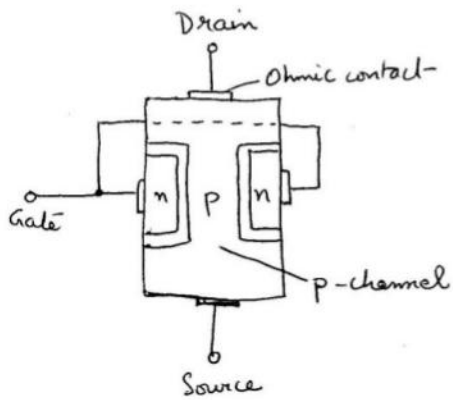
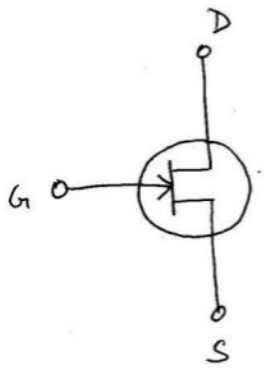


Fig.2. p-channel JFET



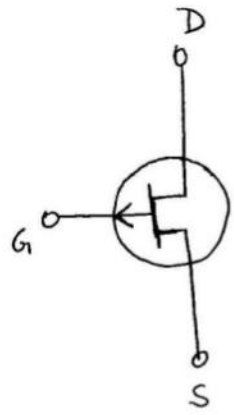


Fig.3. Symbol n-channel JFET **Fig.4.** Symbol p-channel JFET

Thus it can be seen from the figure that one pn junction is formed in the JFET under no bias condition. The current conduction through the n-channel is controlled by the voltage V_{GS} at the input circuit.

The construction of p-channel JFET is similar, the difference is that it has p-type material at the channel instead of n-type and there are two layers of n-type which are connected to form gate terminal with Ohmic contact (Fig.2). The symbols are shown in Fig.3.

9.3 JFET- Operation

The current through the channel i.e. flow of free electrons or holes is controlled by the gate-source voltage V_{GS}. JFET is always operated with the reverse-biased gate source junction. With the change in the reverse biased voltage of gate-source, the resistance of the channel and the width of the channel changes. This is due to the fact that the depletion region of the reverse biased pn junction depends on the magnitude of the reverse voltage. With the increase in reverse voltage V_{GS} the width of depletion region increases, which results in increase of the resistance and the decrease in source to drain current I_D. Thus V_{GS} controls the magnitude of the drain current I_D. We are considering here n-channel JFET.

For the operation of the JFET the gate-source junction is reverse biased and a positive voltage V_{DS} is applied between drain and source. To understand the working of JFET let us consider the condition V_{GS} = 0V and V_{DS} = positive value initially. Under this condition gate and source are at the same potential and the drain is at higher potential, the flow of conventional current I_D is from drain to source as electrons flow from source to drain. The reverse bias and the width of

depletion region will be more towards top. As the voltage V_{DS} increases the width of depletion region increases and at a point these two depletion regions appear to touch each other. This condition is known as pinch off and the voltage V_{DS} at this point is known as pinch off voltage V_p. At this point the current I_D reaches saturation value known as I_{DSS} (drain to source current when gate to source is shorted), which is maximum value of drain current for JFET.

When $V_{GS} < 0V$ i.e. the gate-source junction is reverse biased the depletion regions width increase resulting in narrowing of n-channel. As V_{GS} becomes more negative, the drain current decreases. Thus the source to drain current is controlled by the application of voltage (or electric field) at gate-source junction, which is the reason for the name junction field effect transistor. The value of V_{GS} for which the I_D becomes zero is defined as $V_{GS(off)}$ and practically $V_{GS(off)} = -V_p$.

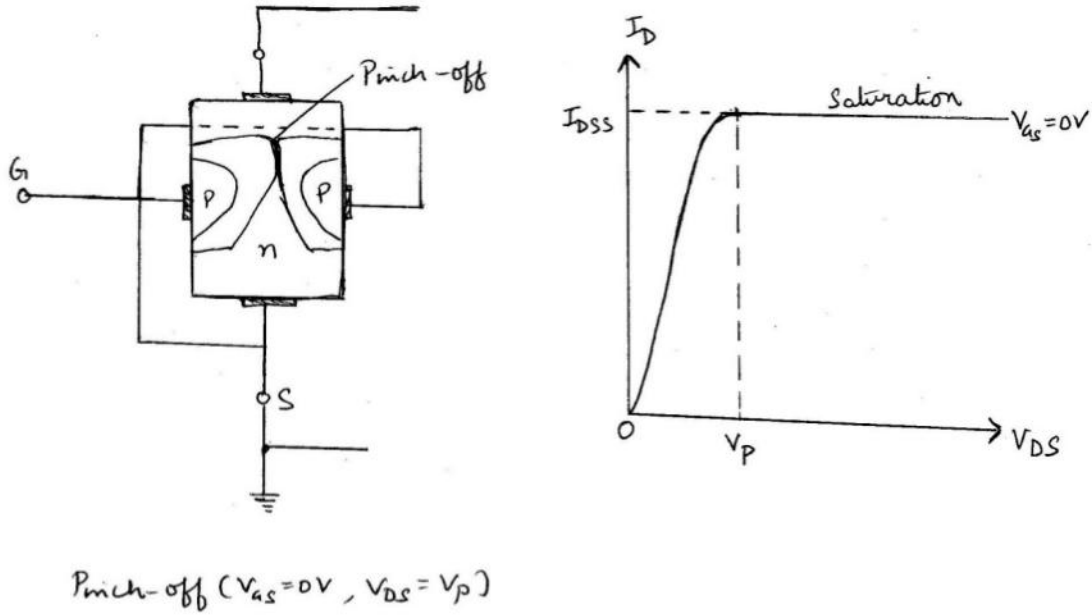


Fig.5. Pinch-off voltage and saturation of drain current.

9.4 JFET Charecteristics

The V-I characteristic curves of the JFET are shown in Fig.6 as drain characteristic curve and transfer characteristic curve. For drain characteristic curve the values of I_D and V_{DS} are plotted for various values of V_{GS}. From Fig.6 it can be seen that initially I_D increases proportionally with increase in V_{DS} and hence the region is known as Ohmic region as in this region I_D and V_{DS} follow Ohmic law.

After that with increase in V_{DS}, the value of I_D remains almost constant, this region is known as constant current area and the saturation value of current as described earlier is I_{DSS}. Beyond this area if V_{DS} is further increased then drain current increases rapidly and it can result in breakdown of the device.

For n-channel JFET as $V_{GS} < 0V$ and is made more and more negative, the value of I_D and I_{DSS} decrease as can be seen from the Fig.6.

The transfer characteristic curve of the JFET is plotted between I_D versus V_{GS} where I_D is output quantity and V_{GS} is input controlling quantity. The name transfer characteristic is given as the curve shows a transfer from input to output variables. It can be seen that there is a nonlinear relationship between I_D and V_{GS}. When $V_{GS} = 0V$ the drain current is maximum given by I_{DSS} and when $V_{GS} = V_{GS(off)}$ the drain current becomes zero.

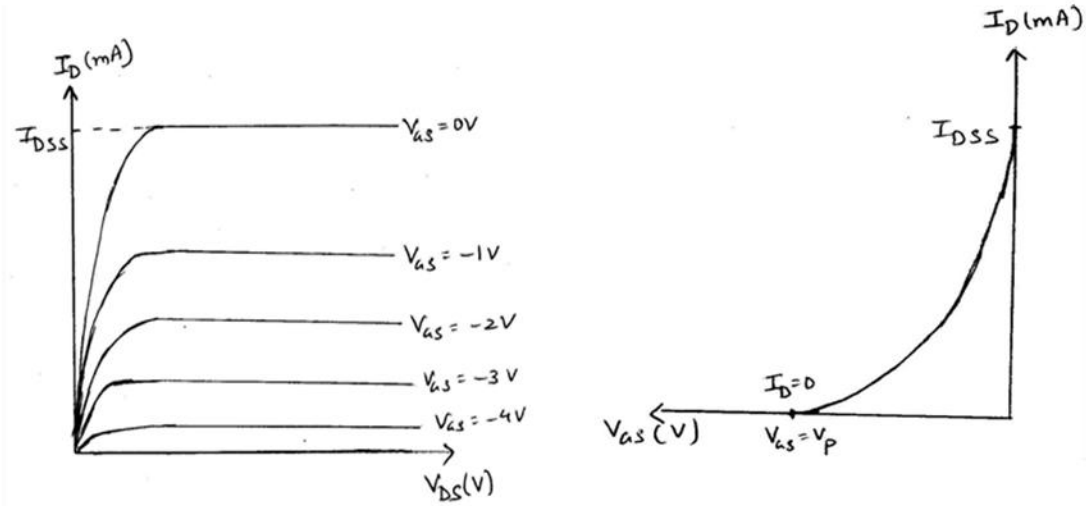


Fig. 6. Drain and transfer characteristics of JFET (n-channel with negative V_{GS}).

The transfer characteristics of the JFET can be expressed mathematically by Shockley's equation.

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

The magnitude of $V_{GS(off)}$ is equal to the value of Vp i.e. $\left|V_{GS(off)}\right| = V_p$ but these are opposite in sign. Here it is important to note that Vp is measured at $V_{GS} = 0V$ and it is value of V_{DS} at which the drain current becomes constant. When V_{GS} is negative

for n-channel JFET the pinch off occurs at voltage values of V_{DS} less than Vp. $V_{GS(off)}$ is that value of V_{GS} for which the drain current becomes zero.

9.5 JFET Parameters

(i) Forward Trans conductance (g_m):- It is the ratio of the change in drain current (ΔI_D) to the change in gate-to-source voltage (ΔV_{GS}) with the drain-to-source voltage (V_{DS}) being constant. Thus

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

The unit of g_m is Siemens (S) or is expressed in mA/volt or micro mho. It measures the control of gate voltage on the drain current. The value of g_m is greater near the top of the transfer curve than the value near the bottom of the curve due to non-linearity of transfer curve. Generally the value of g_m at $V_{GS} = 0V$ is quoted in the specification sheet and is termed as $g_{m(O)}$.

If $g_{m(O)}$ is known, the approximate value of g_m can be calculated through

$$g_m = g_{mo}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)$$

(ii) Input resistance:-

The input resistance of JFET is quite high as it operates with its gate-source junction in reverse bias. Input resistance is determined by the ratio of certain gate-to-source voltage V_{GS} and gate reverse current I_{GSS}

$$R_{in} = \left|\frac{V_{as}}{I_{GSS}}\right|$$

(iii) Drain-to-source resistance (r_d):

It is defined as the ratio of the change in drain-source voltage (ΔV_{DS}) to the change in drain current (ΔI_D) when gate-source voltage V_{GS} is kept constant

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

(iv) Amplification factor (μ)

The ratio of change in drain to source voltage and change in gate to source voltage when drain current is constant is known as the amplification factor i.e.

$$\mu = -\frac{\Delta V_{DS}}{\Delta V_{GS}}$$

The relation between parameters of JFET is given by

$$\mu = r_d \times g_m$$

Since $r_d = \frac{\Delta V_{DS}}{\Delta I_D}$, $g_m = \frac{\Delta I_D}{\Delta V_{GS}}$ *and* $\mu = \frac{\Delta V_{DS}}{\Delta V_{GS}}$

9.6 Self Learning Exercise-I

Q.1 Name the three terminals of FET.

Q.2 What is the bias between gate - to -source for a JFET?

Q.3 What is the main difference between BJT and JFET?

Q.4 Why are FETs known as unipolar device?

Q.5 Why is FET voltage controlled device?

9.7 Illustrative Examples

Example 1 Find the value of drain current if a JFET has following parameters

$$V_{GS} = -4V, \quad V_{GS(off)} = -8V, I_{DSS} = 28\,mA$$

Sol. We Know $I_D = I_{DSS}\left[1 - \frac{V_{GS}}{V_{GS(off)}}\right]^2$

$$= 28\left[1 - \frac{(-4)}{(-8)}\right]^2 mA$$

$$= 28\left[\frac{1}{2}\right]^2 mA \quad = \frac{28}{4} = 7mA$$

Example 2 Find the value of drain current of the JFET for $V_{GS} = 0V, -2V, -4V, -8V$ and $-10V$ when it is given that $I_{DSS} = 12\,mA$ and $V_{GS(off)} = -8V$

Sol. For $V_{GS} = 0V$ $\qquad I_D = I_{DSS} = 12mA$

For $V_{GS} = -2V$

$$I_D = I_{DSS}\left[I - \frac{V_{GS}}{V_{GS(off)}}\right]^2$$

$$= 12\left[1 - \frac{(-2)}{(-8)}\right]^2 = 6.75mA$$

For $V_{GS} = -4V$

$$I_D = 12\left[1 - \frac{(-4)}{(-8)}\right]^2$$

$$= \frac{12}{4} = 3mA$$

For $V_{GS} = -8V$

$$I_D = 0mA$$

For $V_{GS} = -10V$

$I_D = 0mA$ as JFET is cut off as $V_{GS} = -8V$ and for any negative voltage greater in magnitude the JFET remain cut off.

Example 3 The JFET shown in the fig.7 has $V_{GS(off)} = -5V$ and $I_{DSS} = 10mA$. Find the minimum value to V_{DD} in order to operate the device in constant current region.

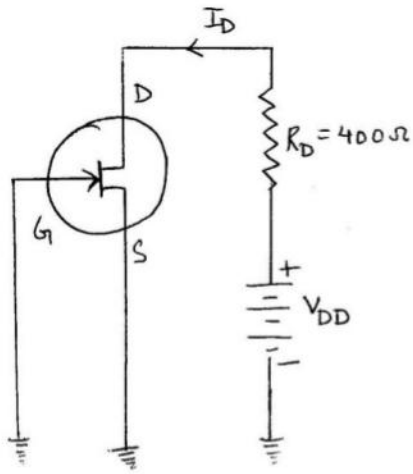


Fig. 7

Sol. As given $V_{GS(off)} = -5V$ it implies $V_p = 5V$ and thus the minimum value of $V_{DS} = 5V$ for constant current operation of JFET.

In constant current region $I_D = I_{DSS} = 10mA$ when $V_{GS} = 0V$

Applying Kirchhoff's voltage law

$$V_{DD} = I_D R_D + V_{DS}$$
$$= (10mA)(400\Omega) + 5V$$
$$= 4V + 5V = 9V$$

Thus the minimum value of V_{DD} required to operate the JFET in constant current region is 9V.

Example 4 For a JFET the drain – to – source voltage is 5V. What is V_p if gate-to-source voltage is zero?

Sol. Given $V_{DS} = 5V$ and $V_{GS} = 0V$

By definition V_p is the value of V_{DS} at which the drain current becomes constant and is measured at $V_{GS} = 0V$, thus in this use $V_p = 5V$.

9.8 The FET Small Signal Model

A small signal is an AC signal superimposed on a large signal. Any nonlinear device can be described using an expression with linearization about a bias point by taking partial derivatives of the expression with respect to all governing variables. These partial derivatives are linked to physical quantities such as resistance, conductance, capacitance etc. and circuit diagram is drawn relating these. During small signal modeling nonlinear components are modeled as linear components. In order to describe the small signal model for an FET the notations used are same as that for BJT for voltage and currents. The drain current for an FET may be expressed as a function of the gate-source voltage and drain – source voltage i.e.

$$i_D = f(V_{as}, V_{Ds}) \tag{1}$$

Considering the gate voltage and drain voltage as variable, the change in drain current Δi_D may be approximately expressed through Taylor series approximation

Taking first two terms the expression is

$$\Delta i_D = \left.\frac{\partial i_d}{\partial v_{GS}}\right|_{v_{DS}} \Delta v_{GS} + \left.\frac{\partial i_D}{\partial v_{DS}}\right|_{v_{GS}} \Delta v_{DS} \tag{2}$$

Using the definition of FET parameters

Transconductance $g_m = \left.\frac{\partial i_D}{\partial V_{GS}}\right|_{V_{DS}}$ (3)

and

Drain resistance $r_d = \left.\frac{\partial v_{DS}}{\partial i_D}\right|_{V_{GS}}$ (4)

We have $\Delta i_D = g_m \Delta v_{GS} + \frac{1}{r_d}\Delta v_{DS}$ (5)

If the variations in gate and drain voltages is such that the drain current is not changed i.e. Δi_D =0 then

$$g_m r_d = -\left.\frac{\Delta v_{DS}}{\Delta v_{GS}}\right|_{\Delta i_D = 0} \tag{6}$$

We know amplification factor $\mu = -\frac{\Delta v_{DS}}{\Delta v_{GS}}$ (7)

Thus $g_m r_d = \mu$ (8)

Equation (2) may be expressed in conventional small signal notations where $\Delta i_D, \Delta v_{GS}$ and Δv_{ds} are replaced by i_d, v_{gs}, v_{ds}. Thus equation (2) may be rewritten as

$$i_d = g_m v_{gs} + \frac{1}{r_d} v_{ds} \tag{9}$$

This equation is the basic equation which gives the change in drain current in terms of g_m, r_d, v_{gs} and v_{ds}. On basis of this equation the small signal model of FET for low frequency is developed as shown in Fig. 8(a).

The model has a Norton's output circuit with a dependent current source $g_m v_{gs}$. The output resistance is drain resistance. The input resistance is infinite as gate to source is reverse biased and gate current is considered zero.

The FET small signal high frequency model as shown in Fig. 8(b) may be obtained by adding three capacitances to the low frequency model. The three capacitances are (i) C_{gs} barrier capacitance between gate and source. (ii) C_{ds} drain to source channel capacitance. (iii) C_{gd} barrier capacitance between gate and drain. The effect of these capacitances is seen in the feedback from output to input and drop in the voltage gain with increase in frequency.

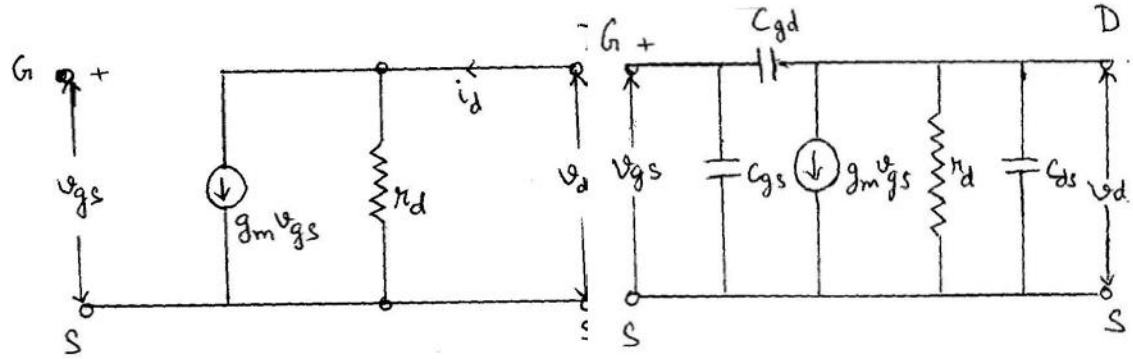


Fig. 8 (a) FET small signal low frequency model

Fig. 8 (b) FET small signal high frequency model

9.9 Metal Oxide Semiconductor Field Effect Transistor (MOSFET)

MOSFETs are sometimes also referred to as IGFETs (Insulated Gate FET) since the gate of the MOSFET is insulated from the channel by a silicon dioxide (SiO_2) layer. There are basically two types of MOSFETs (i) Depletion type MOSFET (D-MOSFET) and (ii) Enhancement type MOSFET (E-MOSFET). The main difference between D-MOSFET and E-MOSFET is that there is a channel present in D-MOSFET whereas a channel is induced in E-MOSFET.

9.10 Depletion –Type MOSFET (D-MOSFET)

Construction

Depletion type MOSFET can be n channel or p-channel. Here construction of n-channel D-MOSFET is described. For n-Channel D-MOSFET a p-type silicon base is taken as substrate on which the device is constructed. The source and drain terminals are connected to n- doped regions joined by an n-channel. The gate

terminal is connected to a metal contact but insulated from the n-channel by a dielectric layer of SiO_2.The insulating layer of SiO_2 is responsible for the high input impedance of the device.

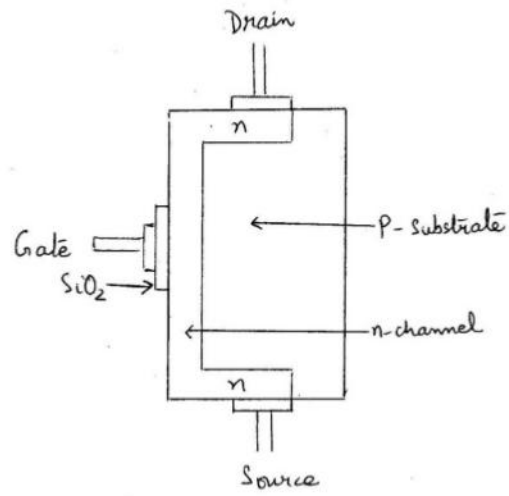


Fig. 9(a) n-channel D-MOSFET

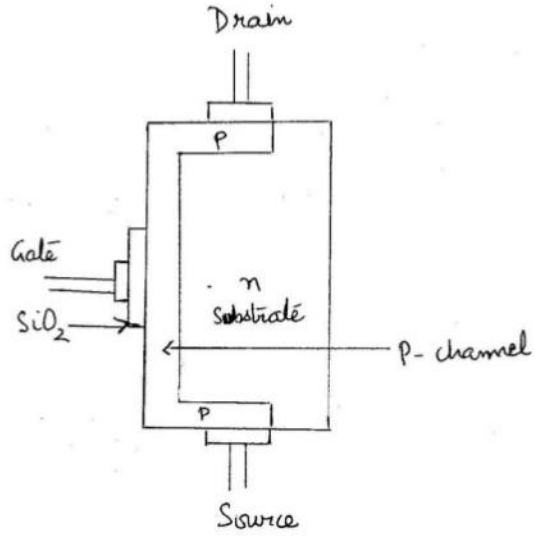


Fig. 9(b) p-channel D-MOSFET

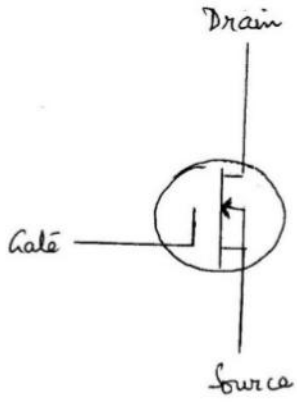


Fig. 10(a) Symbol n-channel D-MOSFET

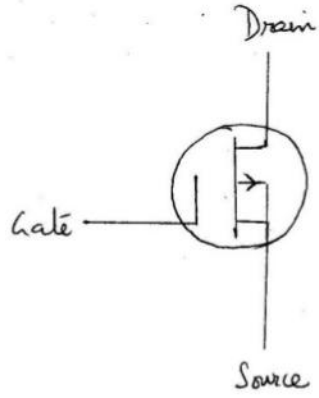


Fig. 10(b) Symbol p-channel D-MOSFET

Operation and characteristics :

The operation of D- MOSFET can be in two modes i) depletion mode (ii) enhancement mode as either a positive or negative gate voltage can be applied. When a negative gate to source voltage is applied the MOSFET operates in the depletion mode and when a positive gate- source voltage is applied the MOSFET operates in enhancement mode.

The operation of the MOSFET can be understood by considering gate as one plate of a parallel plate capacitor and the channel as another plate of the capacitor separated by SiO_2 layer as dielectric.

In depletion mode due to the negative gate-to-source voltage the negative charges at the gate repel the electrons from the n- channel and attract holes from p-substrate, and thus the channel is depleted of the free electrons available for conduction. As the negative on the gate increases, the drain current decreases and becomes zero at a particular negative gate to source voltage $V_{GS(off)}$.

In enhancement mode due to the positive gate-to-source voltage the positive potential at gate attracts more electrons into the channel and hence the channel conductivity is enhanced.

The characteristic curves for the D-MOSFET are shown in fig. 11. The region on the characteristic curve with positive V_{GS} is known as enhancement region and the region between cutoff and saturation level of I_{DSS} is known as the depletion region. The value of current where $V_{GS} = 0$ is $I_{DSS,}$ and where $I_D = 0$ the point corresponds to $V_{GS(off)}$. Similar to JFET for D- MOSFET $V_{GS(off)} = -V_P$

Shockley equation or square- law expression is valid for D-MOSFET also for both the depletion and enhancement regions but signs are to be carefully included in the expression.

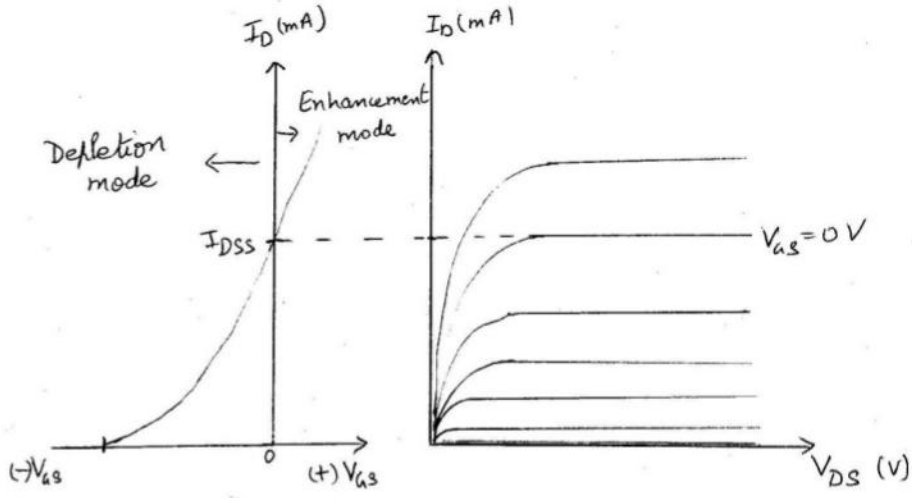
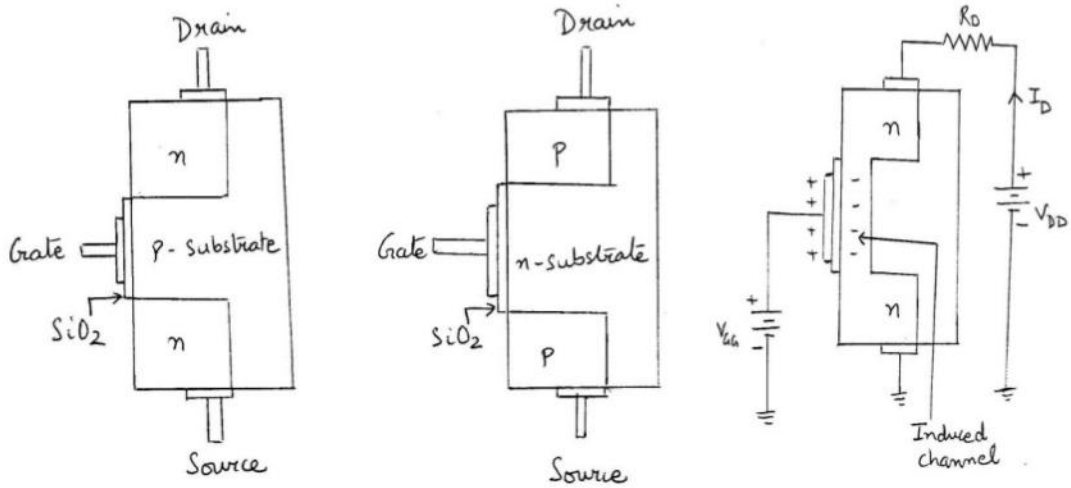


Fig. 11. Characteristic curves for D-MOSFET.

9.11 Enhancement type MOSFET (E-MOSFET)

Construction – E- MOSFET can be n- channel or p-channel. The basic construction of n- channel E – MOSFET is shown in the fig.12. It consists of a p-type substrate which acts as the base. The source and drain terminals are connected to n- doped regions through metal contacts, but there is no channel between the two n – doped regions. The absence of the channel between the two n- doped regions makes E-MOSFET differ in construction from D – MOSFET. The gate terminal is connected to metal contact and separated from the p- type substrate by an insulating SiO_2 layer.



(a) n-channel (b) p-channel (c) induced n-channel

Fig. 12. Enhancement type MOSFET.

Operation and characteristics

E- MOSFET operates in enhancement – mode only and for this positive gate – to-source voltage is applied for n- channel E- MOSFET. When $V_{GS} = 0V$ and a voltage is applied between drain and source, there is effectively no current due to absence of the channel. When both V_{GS} and V_{DS} are set to positive voltages, the positive potential at gate attracts the electrons from p- substrate. These electrons accumulate near the surface of SiO_2 layer which separates gate from the substrate, thus an induced n – region begins to form between the two n- doped regions. With the increase in the value of V_{GS} the induced n- region increases and supports conduction of electrons between drain and source. This induced n-region thus acts

as n-channel and as the current increases with increasing positive voltage of V_{GS}, the MOSFET is known as enhancement type MOSFET or E-MOSFET.

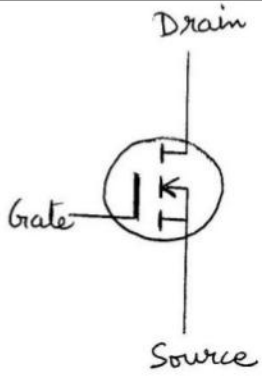


Fig.13(a) Symbol n-channel E-MOSFET

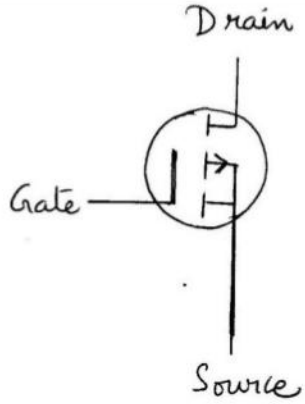


Fig.13(b) Symbol p-channel E-MOSFET

The value of V_{GS} at which the drain current in induced is referred to as threshold value $V_{GS(Th)}$. Below the threshold gate voltage there is no channel and hence no current.

The characteristic curves are shown in Fig. 14. Above the threshold voltage $V_{GS(Th)}$ the current begins to increase. If V_{GS} is kept constant and V_{DS} is increased, the drain current will reach saturation due to pinch-off as shown in the drain characteristics curve.

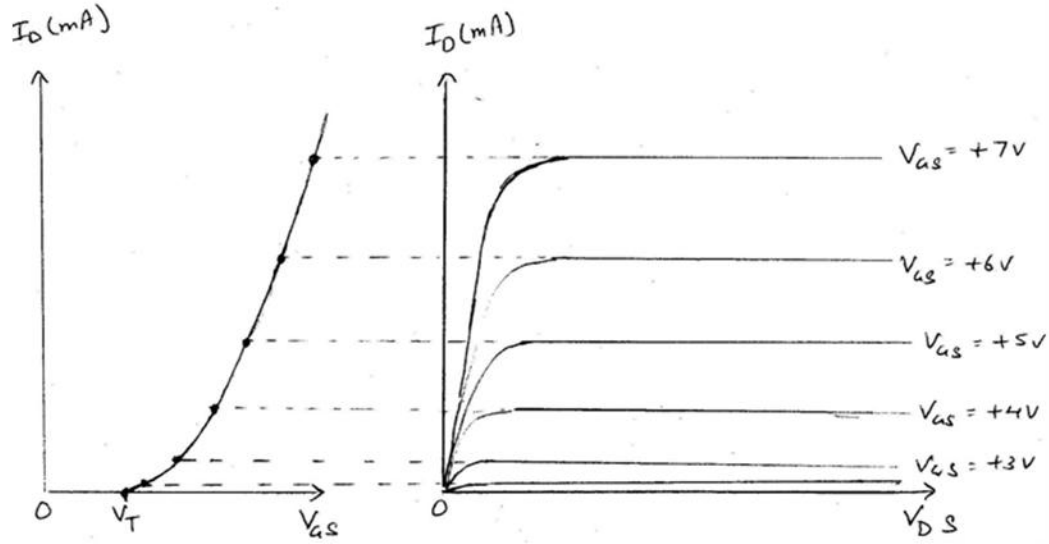


Fig. 14. Characteristic curves for E-MOSFET.

The relation between the drain current and V_{GS} is given by

$$I_D = K\left(V_{GS} - V_{GS(Th)}\right)^2$$

Here K is a constant which depends on the device construction. The value of K can be obtained if $I_{D(on)}$ is taken from the specification sheet, thus

$$K = \frac{I_{D(on)}}{\left(V_{GS} - V_{GS(Th)}\right)^2}$$

9.12 Self Learning Exercise–II

Q.1 Give the names of two types of MOSFETs.

Q.2 D-MOSFET operates in which two modes?

Q.3 What is the difference in construction of D-MOSFET and E-MOSFET?

Q.4 What is threshold voltage $V_{GS(Th)}$?

Q.5 What is the difference between MOSFET and JFET?

9.13 Illustrative Examples

Example 1 For a D-MOSFET it is given $I_D = 5mA$, $V_{GS} = -4V$ and $V_{GS(off)} = -8V$. Find I_{DSS}

Sol. We know $I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$

Substituting the values

$$5mA = I_{DSS}\left(1 - \frac{-4}{-8}\right)^2$$

$$5mA = I_{DSS}\left(1 - \frac{1}{2}\right)^2$$

$$5mA = \frac{I_{DSS}}{4}$$

$\Rightarrow$ $I_{DSS} = 5 \times 4mA = 20mA$

Example 2 For an E-MOSFET it is given that $V_{GS(Th)} = 1V$ and $I_D = 450mA$ at V_{GS}=11V. Find drain current for V_{GS}=5V.

Sol. Let us first find K from the expression

$$K = \frac{I_{D(on)}}{\left(V_{GS} - V_{GS(Th)}\right)^2} = \frac{450mA}{\left(11V - 1V\right)^2} = \frac{450mA}{100V^2} = 4.5mA/V^2$$

Now using equation

$$I_D = K\left(V_{GS} - V_{GS(Th)}\right)^2$$
$$= \left(4.5mA/V^2\right)\left(5V - 1V\right)^2$$
$$= 4.5 \times 16mA = 72mA$$

9.14 Summary

The unit discussed about the basic construction, operation and characteristics of field-effect transistor. Field – effect transistors are unipolar devices with three terminals source, drain and gate. Here junction field effect transistors and metal oxide semiconductor field effect transistors have been described. Junction Field-effect transistor (JFET) works with a reverse bias pn junction. Metal oxide semiconductor field-effect transistor (MOSFET) does not have a pn junction structure and the gate is insulated from channel by a dielectric layer of SiO_2. MOSFETs are of two types D-MOSFET and E-MOSFET. D-MOSFET can operate in deflection and enhancement modes whereas E-MOSFET operates in enhancement mode only. The operation and characteristics of JFET, D-MOSFET and E-MOSFET has been discussed in the unit.

9.15 Answers to Self-Learning Exercises

Answers to Self-Learning Exercise-I

Ans.1: Gate, source and drain

Ans.2: Reverse bias

Ans.3: JFET is a voltage controlled unipolar device whereas BJT is a current controlled bipolar device.

Ans.4: The conduction in FETs is due to one type of charge carrier hence these are known as unipolar devices.

Ans.5: FET is voltage controlled device as the output current I_D is controlled by voltage V_{GS}.

Answers to Self-Learning Exercise-II

Ans.1: D-MOSFET and E-MOSFET

Ans.2: Deflection mode and enhancement mode.

Ans.3: E-MOSFET does not have channel in between the two doped regions as in D-MOSFET.

Ans.4: The value of voltage V_{GS} for E-MOSFET at which the current becomes significant (more than zero)

Ans.5: The difference between MOSFET and JFET is that MOSFET does not have pn junction structure as JFET.

9.16 Exercise

Section-A (Very Short Answer Type Questions)

Q.1 The channel of JFET is between which two terminals?

Q.2 An n-channel D-MOSFET with negative V_{GS} operates in which mode?

Q.3 An n-channel D-MOSFET with positive V_{GS} operates in which mode?

Q.4 For a D-MOSFET the $V_{GS(off)} = -6V$. Is it n-channel or p-channel D-MOSFET?

Q.5 The drain current will increase or decrease if V_{as} in n-channel E-MOSFET is made more positive?

Section-B (Short Answer Type Questions)

Q.6 Define I_{DSS}.

Q.7 Define pinch off voltage.

Q.8 What do you understand by MOSFET?

Q.9 Draw $V-I$ characteristic curve for p- channel JEFT .

Q.10 Draw transfer characteristic curve for P- channel E- MOSFET.

Section-C (Long Answer Type Questions)

Q.11 Explain the construction and operation of a n- channel JFET.

Q.12 What are the differences between JEFT and BJT ?

Q.13 Define the FET parameters and establish the relation between them. $\mu = g_m r_d$

Q.14 Explain the operation of D- MOSFET and E- MOSFET.

Q.15 For the given values of $I_{DSS} = 10mA$ and $V_{GS(off)}$ = - 7 V for an n- channel D- MOSFET, Calculate the value of I_D *at* $V_{GS} = -3.5$ and $V_{GS} = +3.5V$.

9.17 Answers to Exercise

Ans.1: Source and drain

Ans.2: Depletion mode

Ans.3: Enhancement mode

Ans.4: n- channel

Ans.5: Increase

Ans.15: Hint use $I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$

For $V_{GS} = -3.5\,\text{V}$, $I_D = 2.5mA$

For $V_{GS} = +3.5V$, $I_D = 22.5mA$.

References and Suggested Readings

1. Electronic Devices and Circuit Theory, Robert L. Boylestad and Louis Nashelsky, Pearson Publication, 2008
2. Electronic Devices, Floyd, Pearson Publication, 2012.
3. Integrated Electronics, J. Millman and C.C. Halkias, Tata McGraw Hill,1991.
4. Principles of Electronics, V.K. Mehta ,S. Chand & Co.,2008.
5. Electronic Device and Circuits, G.K. Mittal, Khanna Publishers, 2005.

UNIT -10
Operational Amplifier

Structure of the Unit

10.0 Objectives

10.1 Introduction of operational Amplifier

10.2 Differential and common mode operation

10.3 Op-AMP Basics

10.4 Self - Learning Exercise-I

10.5 Inverting Amplifier

10.6 Non-Inverting Amplifier

10.7 Input and Output Impedances of Op-AMP

10.8 Comparision between inverting and Non-inverting Amplifier.

10.9 Unity Follower

10.10 Self-Learning Exercise-II

10.11 Summary

10.12 Glossary

10.13 Answer to self Learning Exercises

10.14 Exercise

References and Suggested Readings

10.0 Objectives

In this unit we study the configuration of OP-AMP. First we derive the formulae for the voltage gain of both configurations of OP-AMP and then we distinguish with between inverting and Non-inverting configuration.

10.1 Introduction Of Operational Amplifier

Integrated circuits have a uital role in modern electronics. Semi conductor devices (diode, transistors etc) have many advantages but the fabrication of circuits with discrete units is quite complicated and expensive. According to the needs of

space technology and computer technology it is essential to keep the size of the instruments as small as possible. Now it is possible to fabricate a complete circuit on a small silicon chip by monolithic integrated fabrication helps in achieving reduction of size, control of thermal effects, more reliability and less cost.

Operational amplifiers (in short OP AMP) is a direct coupled amplifier whose gain is very high and its operation is controlled by negative feedback. This amplifier can be used for performing many mathematical linear and non linear operations, that is why it is called operational amplifier. Operational amplifier is an important building block of integrated circuits. In this chapter we shall study the principle and applications of operational amplifiers. In addition we shall also study in detail differential amplifier which is the main gain element of integrated circuits.

10.2 Differential Amplifier And Common Mode Operation

It is the basic unit of an operational amplifier with differential input. In general the differential amplifier the difference between two input signals.

If the input signals are v_1 and v_2 then the ouptput of ideal differential amplifier will be.

$$v_o = A_d(V_1 - V_2) \qquad (1)$$

Where A_d is the gain of the differential amplifier. In an ideal state according to equation (1) the value of v_o does not depend on the mean level of input signals.

For example in a certain state if $V_1 = 25\mu$ V and $V_2 = -25\ \mu$v and in another state $v_1 = 525$ v μ and $v_2 = 475\ \mu$ v then in both the states the output signal must be according to $(v_1-v_2) = 50\ \mu$ v even when in these two states the mean levels (0 and 500 μ v) are different. In actual case the output signal depends on their mean level i.e. common mode signal $V_c = \frac{1}{2}(v_1 + v_2)$ also along with the difference signal $V_d = (v_1 - v_2)$.

If in the absence of the second signal the gain for the first signal is A_1 and in the absence of first signal the gain for second signal is A_2 then by principle of superposition in the presence of both the signals the output signal will be

$$v_0 = A_1v_1 + A_2v_2 \qquad (2)$$

But $v_d = (v_1 - v_2)$

and $v_c = \frac{1}{2}\left(v_1 + v_2\right)$ or $2v_c = (v_1 + v_2)$

$\therefore \quad v_1 = \frac{v_d + 2v_c}{2} = v_c + \frac{1}{2}v_d$

and $v_2 = \frac{2v_c - v_d}{2} = v_c - \frac{1}{2}v_d$

So that $v_o = A_1\left(v_c + \frac{1}{2}v_d\right) + A_2(v_c - \frac{1}{2}v_d)$ (3)

$= \frac{1}{2}\left(A_1 - A_2\right)v_d + \left(A_1 + A_2\right)v_c$

$= A_d V_d + A_c V_c$ (4)

Hence the gain for difference signal is

$$A_d = \frac{1}{2}\left(A_1 - A_2\right) \quad (5)$$

and gain for Common - mode signal is

$$A_C = (A_1 + A_2) \quad (6)$$

For an ideal differential amplifier A_c must be zero. In fact the value of A_d is expected to be as large as possible as compared to A_c.

The efficiency of a differential amplifier is measured by the magnitude of ratio of A_d and A_c and this ratio is called Common - mode rejection ratio or in brief CMRR.

$$\text{CMRR} = \rho = \frac{\text{Gain For difference Signal}}{\text{Gain For common mode signal}} = \left|\frac{A_d}{A_c}\right| \quad (7)$$

Indecible CMRR = $20 \log_{10} \left|\frac{A_d}{A_c}\right|$

By equation (4) $v_o = A_d v_d \left(1 + \frac{A_c V_c}{A_d V_d}\right)$

Or $V_o = A_d v_d \left(1 + \frac{1}{\rho}\frac{v_c}{v_d}\right)$ (8)

According to relation (8) a differential amplifier must be designed to have the value of ρ as large as possible compared to v_c/v_d. In such a state.

$V_o = A_d v_d$

as happens for an ideal differential amplifier.

To determine CMRR let we first take the input voltages as

$v_1 = 0.5v$, $v_2 = -0.5v$

So that $v_d = (v_1 - v_2) = 1V$

And $v_c = \frac{1}{2}(v_1 + v_2) = 0$

In this state the output voltage

$(v_o)_1 = A_d v_d + A_c V_c = A_d$

Now We take input voltages as

$v_1 = 1v$, $v_2 = 1v$

So that $v_d = (v_1 - v_2) = 0$

And $v_c = \frac{1}{2}(v_1+v_2) = 1V$

In this state the output voltage

$(v_o)_2 = A_d v_d + A_c v_c = A_c$

That is, in this case output voltage will provide directly the value of Ac. Thus we can determine both A_d and A_c and find CMRR. For better performance of the differential amplifier Ad should be large and Ac should be zero, i.e. CMRR should approach infinity. In other words, larger the value of CMRR, better is the differential amplifier. Specification for commercial devices do not quote values for A_c but give values for CMRR. The ratio of $(v_0)_1$ and $(v_0)_2$ is then equal to (A_d/A_c) i.e ρ.

Emitter coupled differential amplifier:-

The circuit of an emitter coupled differential amplifier is shown in Fig (10.1). Two identical transistors Q_1 and Q_2 are used in this circuit. Both these transistors are located near each other in the integrated circuit fabricated on a chip so that their temperature remains same and changes in their parameters are equal. Due to this there is no effect of these changes in the output signal in the differential mode. The collectors of Q_1 and Q_2 and connected to the power supply V_{cc} through equal load resistance R_L. The emitters of Q_1 and Q_2 are connected to each other and then connected to the emitter power supply V_{EE} through a high resistance R_E. The input signals V_1 and V_2 are applied to the bases of Q_1 and Q_2 respectively. The output signal is obtained between the collectors of Q_1 and Q_2 i.e. between A and B as shown in the fig. This circuit is symmetrical about the line a, a' as shown.

In common-mode the input signals are equal in magnitude as well as in same phase i.e. V_1 and V_2 In actual operation the common-mode signals are not applied but are produced due to asymmetry of the circuit or any other reason. It is desired that the signals in common mode be at the lowest level. The input in common mode is shown in Fig. (10.3)

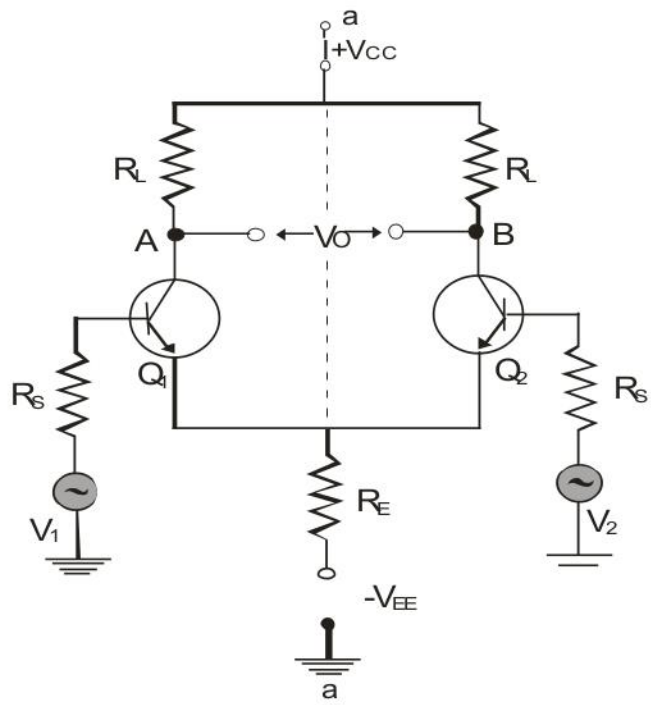


Fig. (10.1)

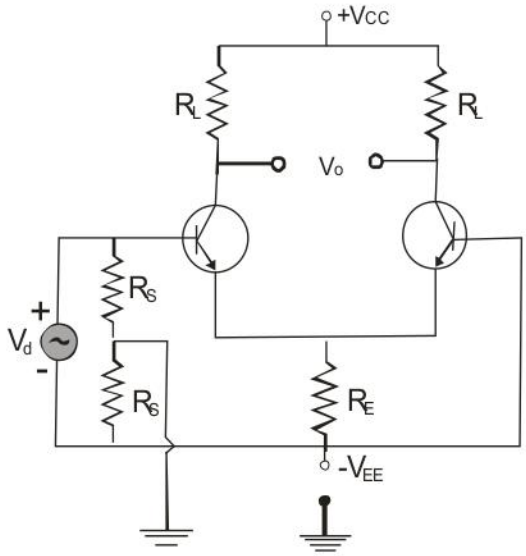


Fig. (10.2)

(Input in differential mode)

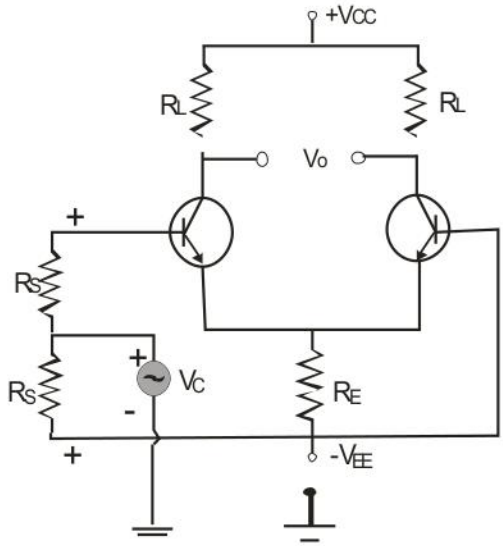


Fig. : (10.3)

(Input in Common Mode)

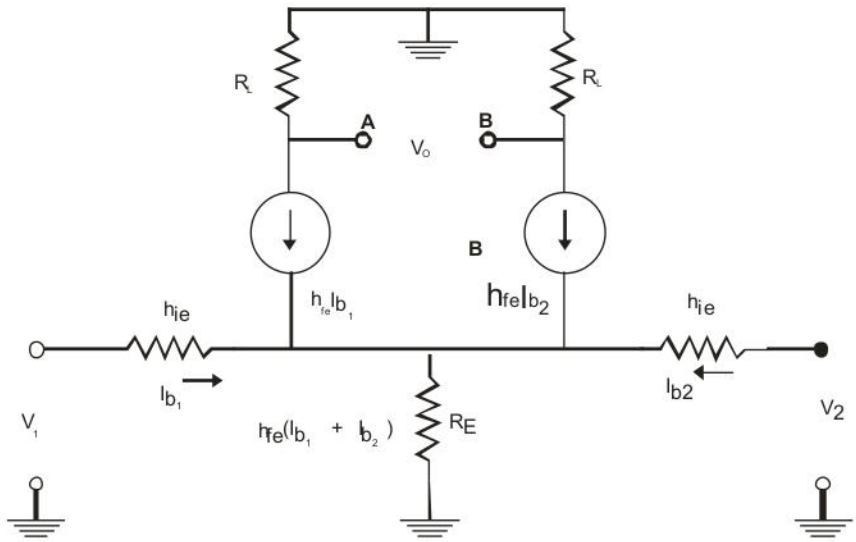


Fig.: (10.4)

Assuming the transistors Q_1 and Q_2 to be identical the equivalent circuit of the differential amplifier is shown in Fig. (10.4)

In the differential mode as the input signals V_1 and V_2 are equal in magnitude but in opposite phase hence there will be increase in collector current of one transistor and decrease in collector current of the second transistor. For small signal amplitudes the mutual characteristics of the transistor can be assumed to be linear

so that by applying signals in differential mode the increase in collector current of one transistor will be equal to decrease in collector current of the other. The current in the emitter resistance R_E will be equal o the sum of these currents will remain constant in this mode. Thus the output signal V_0 will be obtained between the collector terminals but no differential signal will appear across resistance R_E.

According to Fig. 10.4 - in the differential mode.

$$v_1 = h_{ie}\, Ib_1 + h_{fe}\, (Ib_1 + Ib_2) R_E$$

and $$v_2 = h_{ie}\, Ib_2 + h_{fe}\, (Ib_1 + Ib_2) R_E$$

so that $$v_d = v_1 - v_2 = h_{ie}\left(Ib_1 - Ib_2 \right) \qquad (9)$$

and output signal $v_o = (-h_{fe}\, I_{b1}\, R_L) - (-h_{fe}\, I_{b2}\, R_L)$

$$= -h_{fe}\, R_L\, (I_{b1} - I_{b2}) \qquad (10)$$

Hence gain in differential mode

$$A_d = \frac{v_o}{v_d} = -\frac{h_{fe}\, R_L}{h_{ie}} = -\, gm\, R_L \qquad (11)$$

If signals are present in common mode then due to these equal changes will occur in the collector currents of the two transistors so that the output signal will be zero. In this state the current in emitter resistance will increase and due to negative feedback the gain of common-mode will decrease and CMRR will increase. By increasing resistance R_E the negative feedback can be increased but with a larger value of R_E higher emitter supply voltage V_{EE} will be required. To solve this problem constant current transistor circuits are used in place of R_E.

10.3 Operational Amplifier (Op- Amp Basis)

Operational Amplifier is an amplifier having a very high gain and whose input impedance is very large but the output impedance is very low. With negative feedback the resultant gain of this amplifier depends only on the feed-back factor $A_f = \frac{A}{1+A\beta} = \frac{1}{\beta}$. In the beginning such amplifiers were used for mathematical

operations as summation, subtraction, integration, differentiation etc. due to which these were called operational amplifiers and in short OP AMP.

Now their use has become very wide and these are used as comparator, voltage regulator, wave generator etc.

This is a differential amplifier which amplifies the difference between two input signals. The specialty of this amplifier is that it rejects the unwanted voltage signals.

(i) Characteristics Properties of OP AMP-

Ideal OP AMP has the following properties:-

1. Infinite voltage gain.
2. Infinite band width
3. Infinite Input impedance
4. Zero output impedance
5. Perfectly balanced circuit i.e. when both the input signals are identical the output voltage is zero.
6. Perfechy stable i.e., due to changes in temperature voltages etc. the characteristics properties do not change.

The actual (Practical) OP AMP is almost an ideal OP AMP.

(ii) Circuit Symbol of OP AMP

The circuit symbol of OP AMP is shown in Fig. (10.5). It is has two input terminals, one output terminal and a ground terminal which is coupled to all the three other terminals. In Fig. (10.5). 1 and 2 are input terminal and o is output terminal. The ground terminal is not shown. The minus Sign at terminal 1 means that for an input signal at this terminal the polarity of output signal is reversed i.e., For this input the amplifier produces a phase change of 180^0. So this terminal is called inverting input terminal. The positive sign at terminal 2 means that for an input signal at this terminal there is no change in the polarity of output signal. In other words the phases of input voltage at 2 and output voltage at o remain same. This terminal 2 is called non-inverting input terminal.

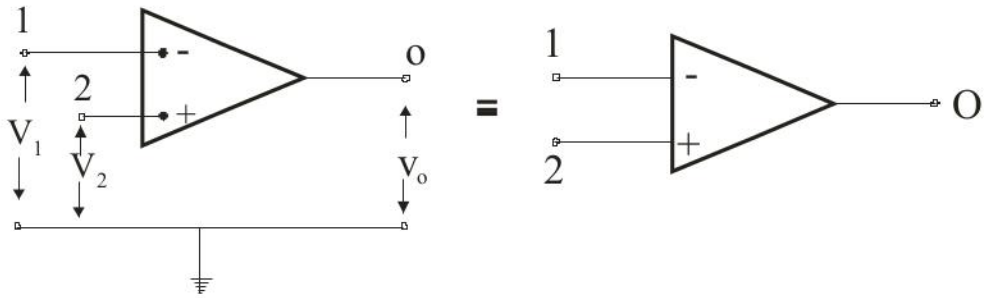


Fig.: (10.5)

The output voltage at terminal o is proportional to the difference of input voltages applied at terminals 1 and 2 i.e.

$$v_o \ \alpha \ (V_2 - V_1) \quad or \ V_o = A\ (v_2 \ - v_1)$$

Here the proportionality constant A is called voltage gain of operational amplifier (OPAMP).

For a particular 8 - Pin linear OP AMP the Pin- connections are shown in Fig. (10.6). The Pins 1 and 8 are not used

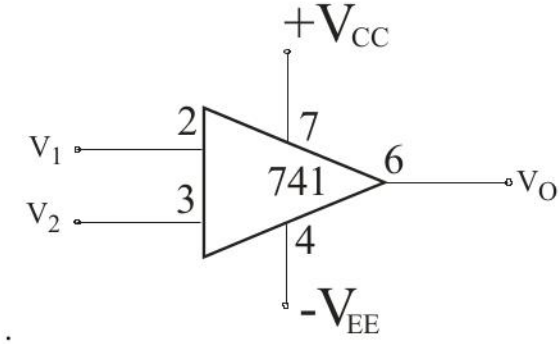


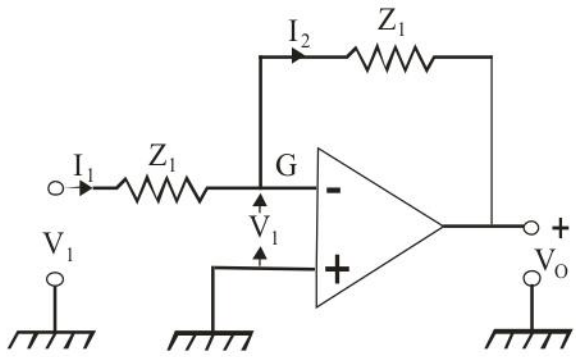
Fig.: (10.6)

The position of Pin 1 is shown by a special sign like dot, notch etc.

(iii) Working of OP AMP:-

An impedance Z_1 is connected to the inverting terminal of OP AMP and for negative feed-back a feed-back impedance Z_F is connected between the inverting terminal and output terminal is shown in Fig. (10.7) and (10.8). An input voltage V_i can be applied in the circuit of OP AMP in two ways. When the input voltage is applied between inverting terminal and ground terminal as shown in Fig. (10.7) then the positive terminal (non-inverting terminal) is connected to the ground and the circuit is called inverting amplifier. When inverting terminal in grounded

through Z_1 and signal voltage is applied between non-inverting terminal and ground then the circuit is called non-inverting amplifier, Fig. (10.8).



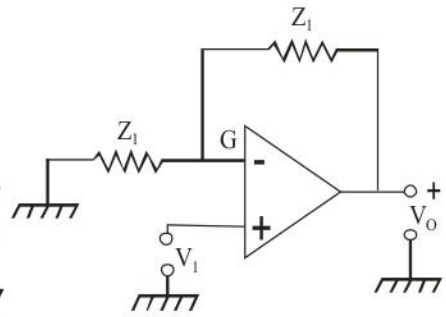


Fig. (10.7) **Fig. (10.8)**

10.4 Self Learning Exercise-I

Very Short Answer Type Questions

Q.1 What is the input impedance of an operational amplifier?

Q.2 What is the voltage gain of an operational amplifier?

Q.3 What are main features of operational amplifer circuit?

Short Answer Type Questions

Q.4 What is an operational Amplifier?

Q.5 Draw the basic circuit and symbol of an OP-AMP.

Q.6 Why this circuit is known as operational amplifier?

10.5 Inverting Amplifier

Circuit diagram of basic inverting OP-AMP is shown in Fig. (10.7) In this mode of operation, the positive input terminal of the amplifier is grounded and the input signal is applied to the negative input terminal through impedance Z_1. The feedback applied through the impedance Zf from the output to input terminal is negative. Feedback impedance Z_F and input impedance Z_L determine the inverting operation of the amplifier as we shall infer from forthcoming equation (5).

As OP - AMP is considered as ideal, it will have infinite voltage gain A. But

$$\left|A_F\right| = \frac{v_o}{v_i}$$

Therefore with finite value of V_o, the output voltage, gain A can be infinite only if input voltage, v_i is zero. The negative feedback from output to input through Z_F serves to keep the voltage v_i at zero. Thus terminal 1, will be at the potential of terminal 2. As terminal 2 is grounded, terminal 1 is thus virtually grounded, and no current flows from input terminal 1 to input terminal 2. Current I flowing through Z_1 will also flow through Z_f (since input impedance of an ideal OP AMP is infinite) Thus the current flowing through Z_1 is

$$I_1 = \frac{v_i - v_1}{z_1} \tag{1}$$

As the input impedance of OP AMP is very large it can be assumed that the current I_2 flowing through feedback impedance Z_F is equal to I_1 i.e. ($I_2 \simeq I_1$)

Now $I_2 = \frac{v_1 - v_o}{Z_f}$ (2)

$$\therefore \quad \frac{v_i - v_1}{z_1} = \frac{v_1 - v_o}{Z_f} \tag{3}$$

Or $$\frac{v_o}{Z_f} = \frac{v_1}{Z_f} + \frac{v_1}{z_1} - \frac{v_i}{z_i}$$

$$= v_1\left[\frac{1}{Z_f} + \frac{1}{Z_1}\right] - \frac{v_i}{z_1}$$

But since $A = -\frac{v_o}{v_1}$ we put $V_1 = -\frac{v_o}{A}$ in above equation so that

$$\frac{v_o}{Z_f} = -\frac{v_o}{A}\left(\frac{1}{Z_f} + \frac{1}{Z_1}\right) - \frac{v_i}{z_1}$$

$$\frac{v_o}{v_i} = -\frac{Z_f}{Z_1}\frac{1}{\left[1 + \frac{1}{A}\left(1 + \frac{Z_f}{Z_1}\right)\right]}$$

The ratio of output voltage V_o and input voltage v_i is resultant voltage gain A_F

$$\therefore \quad \text{Voltage gain } A_F = \frac{v_o}{v_i} = -\frac{Z_f}{Z_1}\left[1+\frac{1}{A}\left(1+\frac{Z_f}{Z_1}\right)\right]^{-1} \tag{5}$$

Where $\frac{v_o}{v_i} = A_f$ is called closed loop gain (with feedback) while A is called open loop gain (without feedback).

As the value of A is very large i.e. A >>1

$$A_F = -\frac{Z_f}{Z_1} \tag{6}$$

Thus the voltage gain in an inverting amplifier depends on the ratio of feed-back impedance Z_f and impedance Z_1 used at input. This does not depend on the internal gain A of the amplifier.

10.6 Non-Inverting Op - Amp

When we need an amplifier whose output is equal to and is phase with the input and in addition source and the load are in effect isolated (for which $Z_1 = \infty$ and Z_o=o) This is the case of Non-inverting OP-AMP. In non-inverting OP-AMP output is equal to and in phase with the input voltage. Like an emitter follower, here the source and the load are effectively isolated, i. e. $Z_i = \infty$ and Z_o = o. As input impedance $Z_1 = \infty$, no current flows into either input terminals of the OP-AMP. That is I_1 = o and If I_f= o. The circuit diagram is shown in Fig. (10.8). The input signal v_i is applied directly to the non-inverting terminal 2, so no phase inversion results at the output. Some of the output is feedback to the inverting input providing effectively some voltage vi at the inverting terminal of the same polarity as the non-inverting input with $Z_1 = \infty$, A $= -\infty$ and V_o = A (v_i-v_f) We conclude that for finite v_o (while A is infinity)

$v_i - v_f = o \quad or \; v_i = v_f$

Thus in this non- inverting amplifier the input voltage v_i is applied between the positive terminal and ground terminal of OP-AMP.

The feed-back occurs at the negative (inverting) terminal. Thus in this mode differential input (v_i-v_f) is applied in the amplifier where v_f is feed-back voltage.

Due to high input impedance of the amplifier the current flow from the amplifer at the input is negligible. Hence.

$$v_f = \frac{Z_1}{Z_1 + Z_f} v_o \tag{1}$$

$$\therefore \quad v_o = A (v_i - v_f) = A \left[v_i - \frac{Z_1}{Z_1 + Z_f} v_o \right] \tag{2}$$

So that the resultant voltage gain

$$A_f = \frac{v_o}{v_i} = \left[\frac{1}{A} + \frac{Z_1}{Z_1 + Z_f} \right]^{-1} \tag{3}$$

But A >> 1

$$A_f = \frac{Z_1 + Z_f}{Z_1} = \left(1 + \frac{Z_f}{Z_1} \right) \tag{4}$$

When $Z_f >> Z_1$ then $A_f = \frac{Z_f}{Z_1}$ as is obtained for inverting amplifier.

Thus we find that output is dependent on the ratio (Z_f/Z_1). IF $Z_1 = \infty$, then $v_o = v_i$ i. e. output voltage follows the input voltage, i.e., OP-AMP circuit acts as a voltage follower.

10.7 Input And Output Impedances of Op Amp

In an ideal OP AMP the input impedance is very high and output impedance is very low but due to feed-back by shunt impedance the input and output impedances can be changed. In this section we shall determine the input and output impedances for both the modes.

(i) Inverting Amplifier

In this amplifier voltage v_i is applied at the negative input terminal through an impedance z_1 with feed-back the input impedance is

$$z_i = \frac{v_i}{I_1} = \frac{v_i - v_1}{I_1} + \frac{v_1}{I_1} \bigcirc$$

But $\frac{v_i - v_1}{I_1} = Z_1$

and voltage gain without feedback $A = -\frac{v_o}{v_1}$

$$\therefore \quad Z_i' = Z_1 - \frac{1}{A}\frac{v_o}{I_1}$$

But $Z_f = \frac{v_1 - v_o}{I_2}$, $I_2 = I_1$ and at the point G, $v_1 \simeq 0$

$$Z_f = -\frac{v_o}{I_2} = -\frac{v_o}{I_1}$$

Hence the input impedance of inverting operational amplifier is

$$Z_i^1 = Z_1 + \frac{Z_f}{A} = Z_1 \qquad \text{As A} >> 1 \qquad (1)$$

Thus the input impedance z'_i of inverting amplifier depends on z1 and not on the voltage gain of OP AMP. Now we will determine the output impedance of inverting amplifier.

As $I_1 \simeq I_2$

$$\therefore \quad \frac{v_i - v_1}{Z_1} = \frac{v_1 - v_o}{Z_f}$$

or $v_i = v_1 + \frac{v_1 Z_1}{Z_f} - \frac{v_o Z_1}{Z_f}$ (2)

The internal gain of OP AMP is $A = -\frac{v_o}{v_1}$, so $v_1 - -\frac{v_o}{A}$. Using this value

$$v_i = -\frac{v_o}{A}\left[1 + (A + 1)\frac{Z_1}{Z_f}\right]$$

and $v_o = -Av_i\left[1 + (A + 1)\frac{Z_1}{Z_f}\right]^{-1}$ (3)

Dividing by output current I_o

$$\frac{v_o}{I_o} = -\left(\frac{Av_i}{I_o}\right)\left[1 + (a+1)\frac{Z_1}{Z_f}\right]^{-1} \tag{4}$$

If there is no feed-back then $Z_f = \infty$ i.e. the output impedance without feedback will be

$$Z_o = \frac{v_o}{I_o} = -\left(\frac{Av_i}{I_o}\right) \tag{5}$$

Using equation (4) and (5) the output impedance of inverting amplifier with feed-back will be

$$Z_o' = Z_o\left[1 + (A+1)\frac{Z_1}{Z_f}\right]^{-1} \tag{6}$$

$$Z'_o = \frac{Z_o}{1+\frac{AZ_1}{Z_f}}$$

As the voltage gain A of OP-AMP is very large the output impedance remains low.

(ii) Non- inverting Amplifier:-

To determine the input and output impedance of non-inverting amplifier we use equation (4) to (10) of last section (10.3)

$$A_f' = \frac{v_o}{v_i} = \left(1 + \frac{Z_f}{Z_1}\right)$$

$$v_o = v_i\left(1 + \frac{Z_f}{Z_1}\right) \tag{7}$$

Dividing by input current I_1

$$\frac{v_i}{I_1} = \frac{v_o}{I_1}\quad\frac{1}{\left(1+\frac{Z_f}{Z_1}\right)} \tag{8}$$

It the internal gain of OP-AMP is A then in non-inverting mode A = $\frac{v_o}{v_1}$ or

$v_o = Av_1$

$$\therefore \quad \frac{v_i}{I_1} = \frac{v_1}{I_1}\frac{A}{\left(1+\frac{Z_f}{Z_1}\right)} \tag{9}$$

The input impedance of OP AMP without feedback is $Z_i = \frac{v_1}{I_1}$ so the input impedance of non-inverting amplifier with feedback will be

$$Z'_i = \frac{v_i}{I_1} = \frac{AZ_i}{\left(1 + \frac{Z_f}{Z_1}\right)} \tag{10}$$

As A is very large , $Z'_i \gg Z_i$

For output impedance using equation (8) of last section (10.3)

$$\frac{v_o}{I_o} = \frac{A}{I_o}\left[v_i - \frac{Z_1}{Z_1 + Z_f}v_o\right] \tag{11}$$

Where I_o is output current.

In the absence of Feedback $Z_f = \infty$

$$\therefore \quad \text{Output impedance } Z_o = \frac{v_o}{I_o} = \frac{Av_i}{I_o} \tag{12}$$

Hence the output impedance with feedback is

$$Z'_o = \frac{AV_i}{I_o} - \frac{AZ_1}{Z_1+Z_f}\frac{V_o}{I_o} = Z_o - \frac{AZ_1}{Z_1+Z_f}Z_o^1$$

$$\text{So that } Z'_o = \frac{Z_o}{\left[1+\frac{AZ_1}{Z_1+Z_f}\right]} \tag{13}$$

As A is very large $Z'_o \ll Z_o$.

10.8 Comparision Between The Inverting And Non-Inverting Amplifier

Let us now compare the characteristics of the inverting and non-inverting type amplifiers. If the two configurations are identical except for the interchange of the input terminals and ground connections. However the expression for the voltage gain differs.

Inverting Amplifiers	**Non-Inverting Amplifiers**
1. Voltage gain $A_f = -\frac{Z_f}{Z_1}$	1. Voltage gain $A_f = 1 + \frac{Z_f}{Z_1}$
2. Z_1 and Z_f from a voltage division network for both input and the feedback signals.	2. Voltage division network takes place only in the use of the feedback signals.
3. Z_1 feeds a virtual ground at the negative point. This makes the input impedances equal to Z_i.	3. The feedback signal is effectively in series with the input signal and opposes it. This leads to high input impedance.
4. CMRR is not a problem with the positive terminal and the negative input at the virtual as ground, the input common mode voltage is zero	CMRR is a problem. Feedback causes the voltage at the inverting input to follow that at the non-inverting input. The input common mode voltage thus varies directly with the input voltage.

10.9 Unity Follower

A useful and simplest application of an op AMP is unity gain buffer circuit. This circuit is shown in Fig. (10.9). If in the non-inverting amplifier $Z_f = 0$ or $Z_1 = \infty$ then the gain of the amplifier becomes unity as shown in section (10.5) equation (4). As for op AMP the input resistance is very large and output resistance very small such circuit can be used for impedance matching between a high impedance source and low impedance load. Further to isolate a circuit from the load also this circuit is useful.

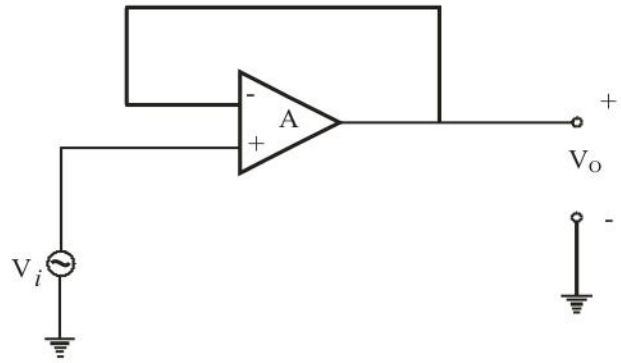


Fig. (10.9)

10.10 Self Learning Exercise-II

Very Short Answer Type Questions

Q.1 An inverting Op-AMP has R_1 = 100 kΩ and Rf = 500 kΩ. Find the output voltage for an input of 2V.

Q.2 What is the feedback gain of an inverting amplifier?

Q.3 What is the gain of an unity buffer amplifer?

Short Answer type Questions

Q.4 Calculate the output voltage of a non-Inverting OP-AMP having R_1 = 100 KΩ, R_f = 500 KΩ, when the input signal is of 2v.

Q.5 Draw the circuit diagram of an inverting OP-AMP.

Q.6 Draw the circuit diagram of an non-inverting OP-AMP?

10.11 Summary

All OP AMPS have infinite values of voltage gain, band width and input impedance. There values result in a zero output impedance. The common- mode rejection ratio of an OP-AMP is the ratio of differential to the common mode gain. Larger the CMRR, Smaller is the Common-Mode output Voltage.

The voltage series negative feedback configuration is commonly called a non-inverting amplifier while the voltage - shunt feedback configuration is called an inverting amplifier. The ideal voltage gain of these two amplifier depends only on the feedback network (Components) and is independent of the internal gain A of the OP-AMP. The use of negative feedback is a non-inverting amplifier increases the input impedance and band width and decrease the output resistance,

total offset output voltage and the effect of varying environmental conditions on the gain. The same results are obtained for the inverting amplifier using negative feedback except for input resistance. The input resistance of the inverting amplifier is relatively smaller because it depends on the resistance connected in series with the input signal source.

10.12 Glossary

Amplifier - A device that increases the strength of an electrical signal by drawing energy from a separate source to that of the signal.

Common Mode Rejection Ratio - Peak to peak change in equivalent input offset voltage over this range.

Input Offset Current - Difference between the currents into the two input terminals.

Differential - The result of Mathematical differentiation.

10.13 Answers to Self Learning Exercises

Answers to Self Learning Exercise-I

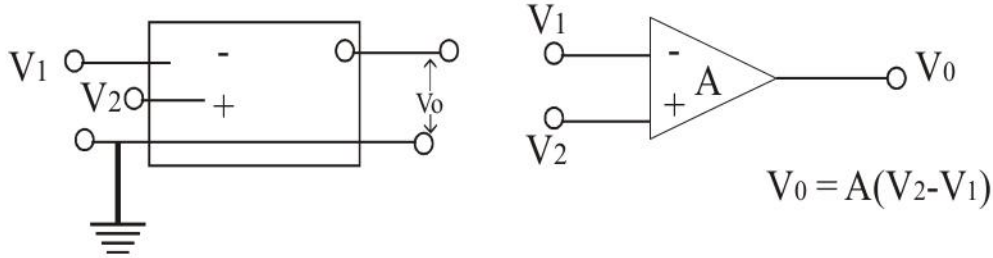
Ans.1: Infinite.

Ans.2: Infinite

Ans.3: A simple operational amplifier is a direct coupled amplifier whose characteristics are controlled by feedback.

Ans.4: An operational amplifier is basically a differential amplifier used to amplify the difference between two input signals. It is characterized by a high voltage gain, a very high input resistance and a low output resistance.

Ans.5:



Answers to Self Learning Exercise-II

Ans.1: -10V

Ans.2: $A_f = \frac{R_f}{R_1}$

Ans.3: One

Ans.4: +12v

10.14 Exercise

Section A (Very short answer type question)

Q.1 What do you understand by differential amplifier?

Q.2 Define common mode rejection ratio (CMRR).

Q.3 What are the voltage gain of inverting OP-AMP?

Q.4 What are the voltage gain of non-inverting OP-AMP?

Section-B (Short Answer type question)

Q.5 Is the closed loop gain of an OP-AMP Independent of its characteristics? If yes how?

Q.6 How do you distinguish between the inverting and non-inverting mode of operation of an OP-AMP?

Q.7 What are the main characteristics of OP-AMP when used as voltage amplifier?

Q.8 Can an OP_AMP be used as current amplifier? Yes What are its characteristics?

Q.9 What are the characteristics of an OP-AMP when used as voltage to current converter?

Section-C (Long Answer the type Question)

Q.10 What is meant by a differential amplifier? Draw the circuit diagram of emitter coupled differential amplifier and give the analysis for differential and common mode inputs.

Q.11 What do you understand by common-mode-rejection ratio (CMRR)? How

is CMRR of a differential amplifier determined? What is CMRR of an ideal differential amplifier?

Q.12 What is an operational amplifier (OP AMP)? Describe its specialities and give some of its applications.

Q.13 Give the characteristic properties of an ideal OP-AMP. Draw circuit diagrams of inverting and non-inverting OP AMP amplifiers and explain their working.

Q.14 Obtain expressions for input and output impedances of an OP-AMP Working as inverting and non- inverting amplifier.

Q.15 Using an OP AMP draw the circuit diagram of a unity gain buffer and explain its working. Explain the utility of this circuit.

References and Suggested Readings

1. OP - AMPS and linear Integrated circuits, Ramakanth Gayakwad, PHI learning private limited.
2. Hand Book of electronics, Gupta kumar Pragati Prakashan
3. A Text book of electronics, 2e S.L. Kakani and K.C. Bhandari, New Age International (P) Limited, Publishers.

UNIT-11
Applications of Operational Amplifier

Structure of the Unit

11.0 Objectives

11.1 Introduction

11.2 Summing Amplifier

11.3 Integrator

11.4 Differentiator

11.5 Self Learning Exercise-I

11.6 Input Off Set current

11.7 DC Offset parameters

11.8 Frequency Response and stability

11.9 Slew rate

11.10 Constant gain multiplier

11.11 Multiple Stage gains

11.12 Active Filters

11.13 Low pass filters

11.14 High pass filter

11.15 Band Pass filter

11.16 Wide Band- Pass filter

11.17 Narrow Band - Pass filter

11.18 Self Learning Exercise-II

11.19 Summary

11.20 Glossary

11.21 Answer to self learning Exercises

11.22 Exercise

References and Suggested Readings

11.0 Objectives

The use of OP-AMP in implementing mathematical operations. After going through this unit you will be able to know Offset parameters, slew rate etc. The gain decreases as the operating frequency increases. This variation in gain as a function of frequency and stability of it is discussed in this unit.

11.1 Introduction

Basic Construction, characteristic Parameters and various configurations are studied in the previous unit. Now our next logical step is to discuss linear and non linear applications of OP-AMP. Each category will include the most important application in that area. This unit presents a collection of circuits connected to linear applications using OP-AMP, which are basic to the development of more complex circuits. In linear circuits the output signal is same as nature as the input and varies in accordance with the input within the limit set by the saturation level and slew rate.

11.2 Summing Amplifier

(1) Adder:

The use of an OP-AMP as an adder is shown in Fig. (11.1). By an OP-AMP the sum of different input voltages with their weighting factors can be obtained. The different voltages V_A, V_B …….V_N are applied on the input terminal (Negative terminal in figure) through resistances R_A R_B … R_N respectively. The feed-back is accomplished by an appropriate resistance R_f.

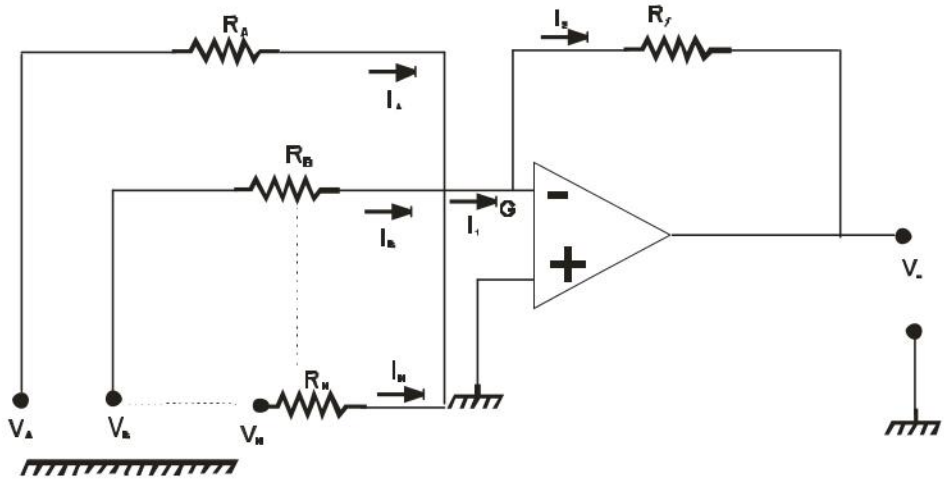


Fig. (11.1)

Due to very high input impedance of OP-AMP the voltage at G is almost Zero $V_1 = 0$ i.e. apparently the point G can be assumed to be grounded. Hence the sum of incoming currents a point G will be equal to the sum of outgoing currents i.e.

$$\therefore \quad I_A + I_B + \ldots\ldots + I_N = I_1 = I_2$$

But $I_A = \frac{V_A}{R_A}$, $I_B = \frac{V_B}{R_B}$ ..... $I_N = \frac{V_N}{R_N}$

$$\therefore \quad \frac{V_A}{R_A} + \frac{V_B}{R_B} + \frac{V_C}{R_C} + \ldots\ldots + \frac{V_N}{R_N} = -\frac{V_o}{R_F}$$

$$V_o = -R_f \left[\frac{V_A}{R_A} + \frac{V_B}{R_B} + \frac{V_C}{R_C} + \ldots.. + \frac{V_N}{R_N} \right]$$

$$= -\left[\frac{R_f}{R_A} V_A + \frac{R_f}{R_B} V_B + \ldots. + \frac{R_f}{R_N} V_N \right]$$

This with weighting factors $\frac{R_f}{R_A}$, $\frac{R_f}{R_B}$, ... etc. the sum of different voltages is equal to the output voltage V_o.

If $R_a = R_b = R_c = \ldots.. = R_n = R_f$ then

$$V_o = -(V_A + V_B + \ldots.. + V_N) \qquad (1)$$

or the output voltage numerically is equal to the sum of all input voltages. So this circuit is called an adder.

If the number of input voltages is n and $R_A = R_B = \ldots = R_N = nR_f$ then

$$V_o = -\frac{1}{n}\left(V_A + V_B + \ldots. + V_n \right)$$

i.e. the circuit will provide the average value of the input signal voltages.

(ii) Subtractor :-

In the circuit of OP AMP shown in Fig. (11.2) when two voltages are fed at the input the output voltage is proportional to the difference of the two inputs. Thus the circuit can be used as a subtractor.

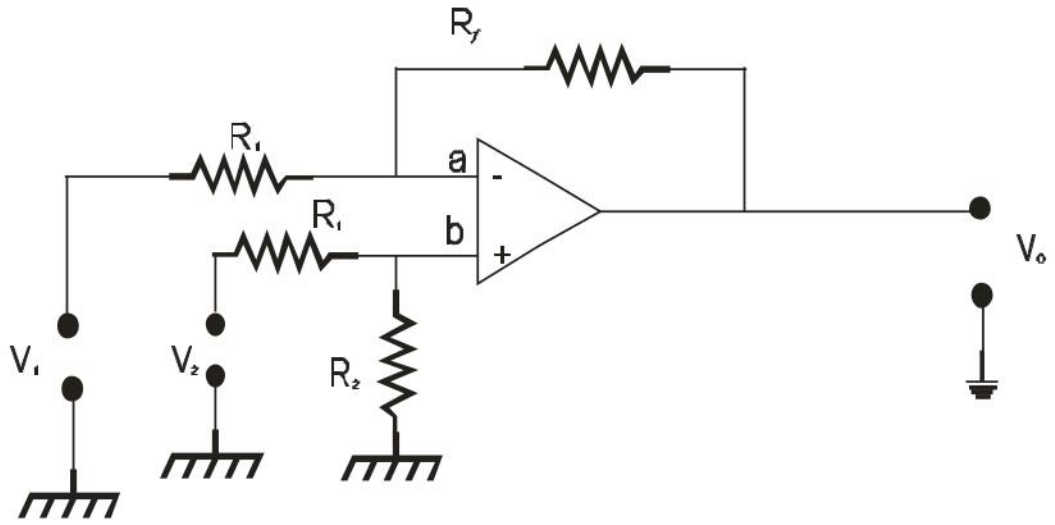


Fig. (11.2)

Suppose two input voltage V_1 and V_2 with respect to the ground are applied at the negative and positive terminals of an OP-AMP through equal resistances R_1 each. The impedance $Z_f = R_f$ in the Feed - back path is taken very large as compared to $Z_1 = R_1$ ($Z_f >> Z_1$)and ($R_2 >> R_1$) so that the gain in inverting and non-inverting modes are equal $\left[|A'| = \frac{Z_f}{Z_1} = \frac{R_f}{R_1}\right]$.

If due to the input V_1 at negative terminal the output voltage is V_{o1} and due to the input V_2 at positive terminal the output voltage is V_{o2} then the resultant output voltage is

$$V_o = V_{o1} + V_{o2} = -\frac{R_f}{R_1}V_1 + \frac{R_f}{R_2}V_2$$

$$= \frac{R_f}{R_1}(V_2 - V_1) \qquad (2)$$

11.3 Integrator

A circuit in which the output voltage wave form is the integration of the input voltage wave form (i.e. If the input is sine wave the output will be cosine wave; if the input is a square wave the output will be a triangular wave). Such an integrator circuit is obtained by using a basic inverting amplifier configuration if the feedback resistor R_f is replaced by a capacitor C (Fig. 11.3).

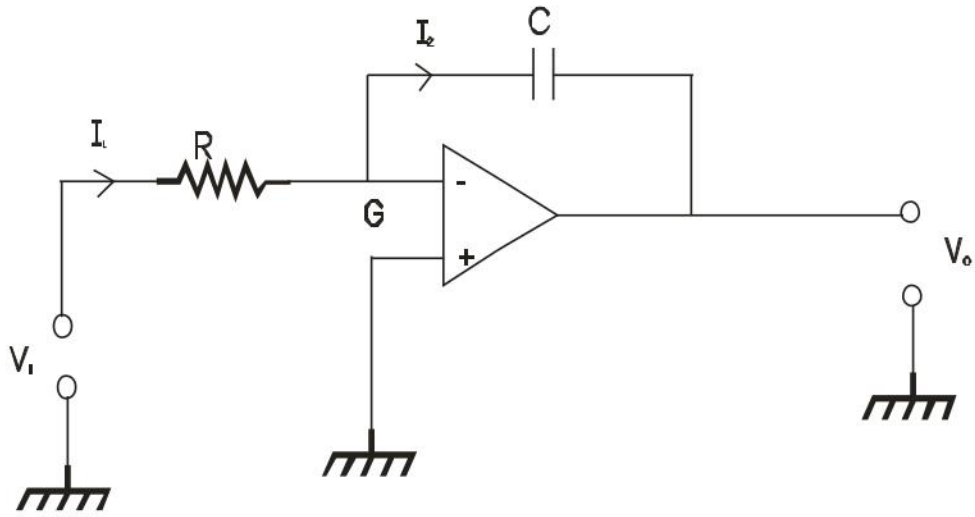


Fig. (11.3)

As we know due to the very high input impedance of OP-AMP the point G can be considered apparently as grounded so that the current through resistance R at a time t is

$$i_1 = \frac{V_i(t)}{R} \tag{1}$$

This current will be equal to the current i_2 flowing through the capacitor C.

$$i_2 \frac{dq}{dt} = \frac{d}{dt}\left[-Cv_o(t)\right]$$

$$= -C\frac{dv_o(t)}{dt} \tag{2}$$

Thus from equations (1) and (2)

$$-C\frac{dv_o(t)}{dt} = \frac{v_i(t)}{R}$$

or $$v_o(t) = -\frac{1}{RC}\int v_i(t)dt \tag{3}$$

If the time constant RC is taken equal to 1 (RC=1 as R = 1 MΩ and C = 1μF) then

$$V_o(t) = -\int v_i(t)dt \tag{4}$$

Hence in the integrator circuit the output voltage at any time t is numerically equal to the integral of input voltage with respect to time. The OP-AMP integrator may thus be used to convent a square wave into a triangular wave.

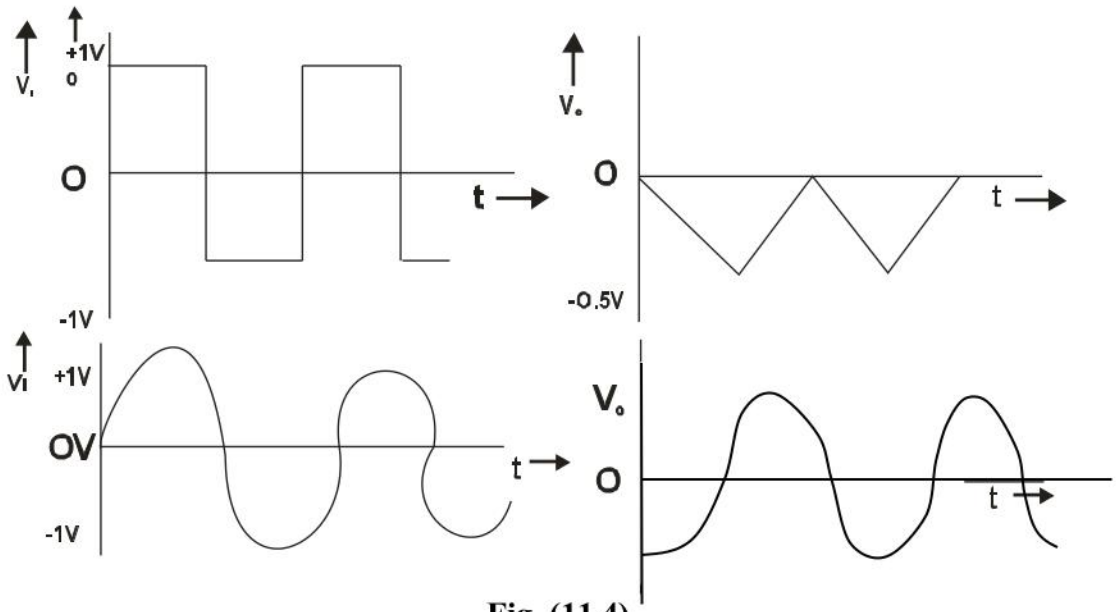


Fig. (11.4)

When $v_i = 0$ the integrator of Fig (11.3) works as an open loop integrator. This is because the capacitor (acts as an open circuit ($x_c = \infty$) to the input off set voltage V_o and the part of the input current charging the capacitor C produce the error voltage at the output of the integrator. Therefore in the practical integrator shown in Fig. (11.5), to reduce the error voltage at the output resistor R_F is connected across the feedback capacitor C. Thus R_F limits the low frequency gain and hence minimizes the variation in the output voltage.

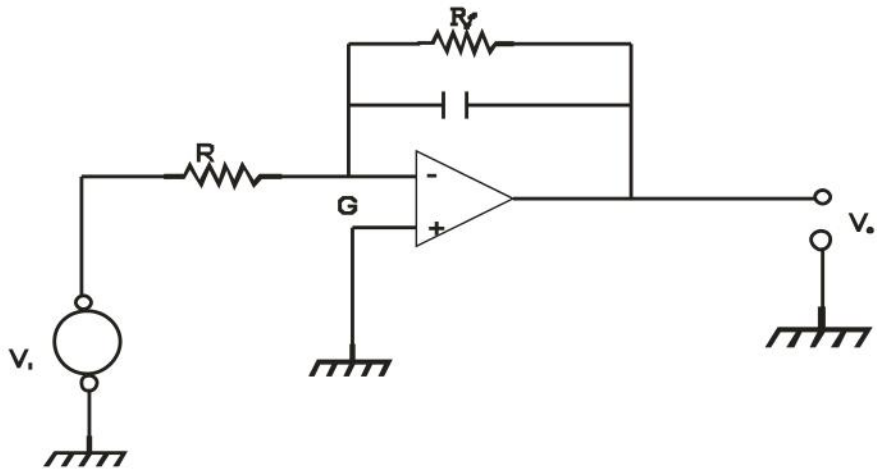


Fig. (11.5)

The integrator is most commonly used in analog computers and analog - to - digital (A/D) or (ADC) and signal wave shaping circuit.

11.4 Differentiator

The circuit diagram of a simple differentiator is shown in the fig. (11.6) as its name implies the circuit perform the mathematical operation of differentiation i.e. the output wave form is the derivative of the input wave form. Then such a differentiator circuit is obtained by using a basic inverting amplifier configuration if the input resistor R is replaced by a Capacitor C.

In this also assuming voltage V= 0 at the point G, the potential difference across the capacitor C at time t is

$$v_i(t) = -\frac{q(t)}{C} \tag{1}$$

Where q (t) is amount of charge on the condenser at time t.

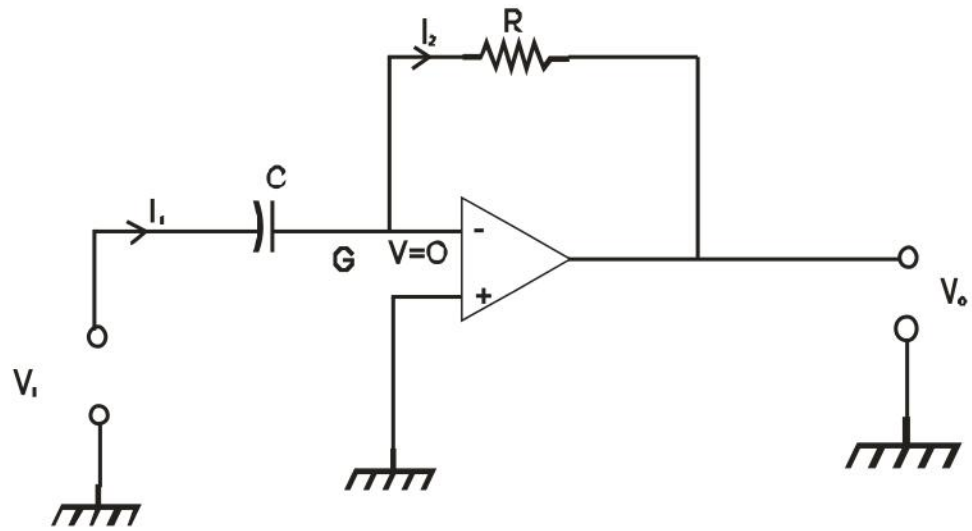


Fig. (11.6)

Therefore current $i_1(t) = \frac{dq(t)}{dt} = C\frac{dV_i(t)}{dt}$ (2)

The current in the feed-back resistance R will be

$$i_2(t) = -\frac{V_o(t)}{R} \tag{3}$$

$$\therefore \quad -\frac{V_o(t)}{R} = C\frac{dV_i(t)}{dt}$$

$$\text{or } V_o(t) = -RC\frac{dV_i(t)}{dt} \tag{4}$$

If the time constant RC= 1 then

$$V_o(t) = \frac{dV_i(t)}{dt}$$

Thus the output voltage $V_o(t)$ is equal to the RC times the negative instantaneous rate of change of input voltage V_i with time. Since the differentiator performs the reverse operation of the integrator, a cosine wave input will produce a sine wave output or triangular input wave will produce square wave output. However the differentiator of figure11.6, will not do this because it has some practical problem. The gain of the circuit R/x_c increases with increase in frequency at the rate of 20dB/decade. This makes the circuit unstable. Also the input impedance X_c decreases with increase in frequency, which makes the circuit may susceptible to high frequency noise when amplified this noise can completely over ride the differential output signal.

Both the stability and high frequency noise problem can be corrected by the addition of the two components R and $C_f$as shown in the fig (11.7)

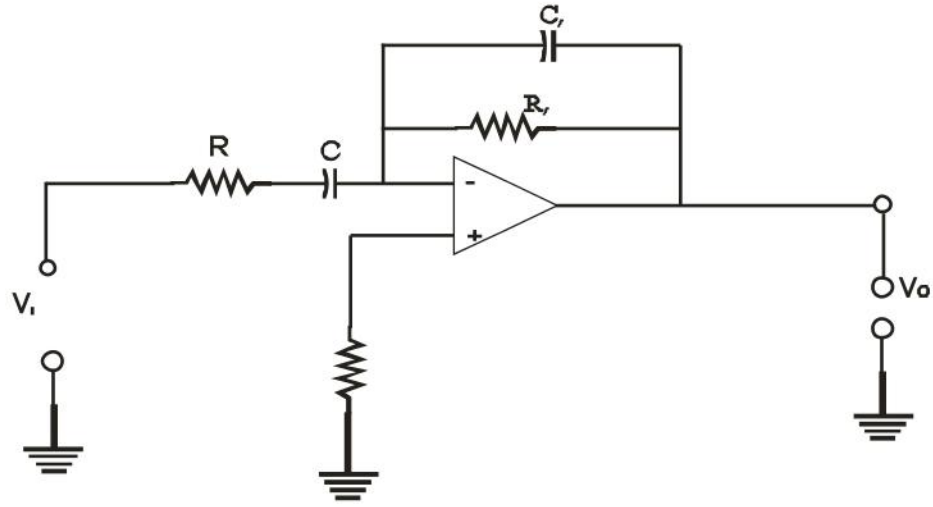


Fig. (11.7)

Fig. (11.8) Shows the sine wave and square wave inputs and results differentiated output for the practical differentiator.

The differentiator most commonly used in wave shaping circuits to detect high frequency component in an input signal and also as a rate of change detector

in F.M. Modulators.

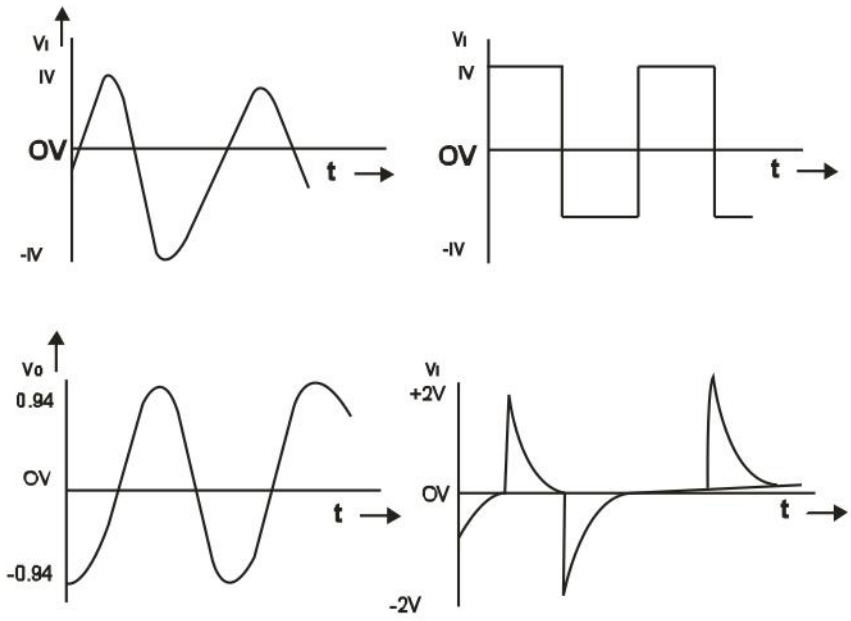


Fig. (11.8)

11.5 Self Learning Exercise-I

Very Short Answer Type Questions

Q.1 What is the output voltage of an Adder?

Q.2 What is the output voltage of an integrator Circuit?

Q.3 What is the output voltage of an differentiator?

Short Answer Type Questions

Q.4 Draw a circuit diagram of an Integrator Circuit.

Q.5 What is an output wave form of an Integrator circuit?

Q.6 What is an output wave form of a differentiator circuits?

11.6 Input Offset Current

If no voltage or equal voltage is applied on the input terminals of a differential amplifier of OP AMP then in ideal state the output voltage must be zero i.e. $V_o = 0$ if $V_1 = V_2$ or $V_1 = V_2 = 0$. But in practice it is found that due to mismatch of transistors or imbalance of input signal voltages or defects in fabrication, the output voltage is not zero when the input voltages are zero or equal. Due to certain reasons the direct current I_1 and I_2 at the two input terminals are not equal inspite of being very small i.e. for zero input the output voltage does

not become zero. This difference in biasing current is called input offset current.

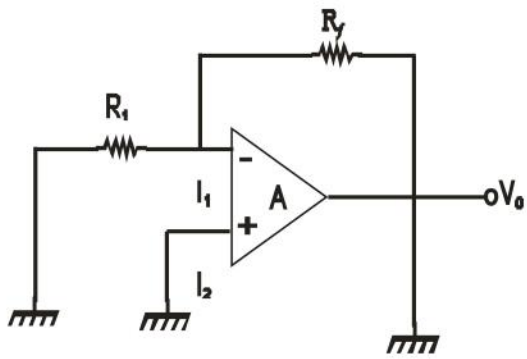


Fig.(11.9)

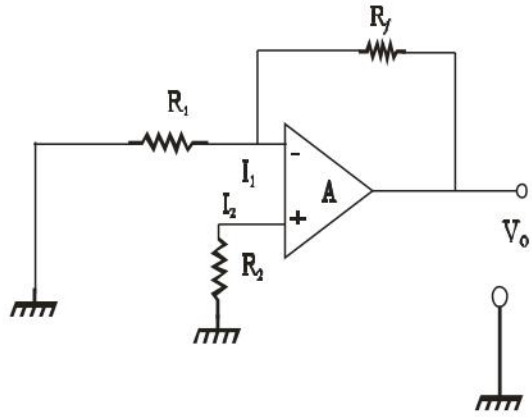


Fig. (11.10)

As the output impedance of an OP AMP is very small the output terminal of feed-back resistance R_f can be considered as grounded for direct current, as is shown in Fig. (11.9). the current by the parallel circuit of R_1 and R_f. Hence at this terminal the input voltage $V_1 = \dfrac{R_1 R_f}{R_1 + R_f} I_1$

This input voltage at the negative terminal off sets the amplifier as the voltage at positive terminal is zero. This can be reduced by connecting a resistance R_2 at the positive terminal as shown in Fig. (11.10). in this state the input voltage at positive terminal

$V_2 = I_2 R_2$

If the voltage gain of OP AMP is A then output voltage

$V_o = A (V_2 - V_1)$

$$V_O = A\left[I_2 R_2 - \frac{R_1 R_f}{R_1 + R_f} I_1 \right]$$

To minimize the off set out the output $R_2 = \dfrac{R_1 R_f}{R_1 + R_f} I_1$

This value of resistance R_2 minimizes the effect of input offset current $I_o = (I_2 - I_1)$ When I_2 and I_1 are not equal. By taking a proper value of R_2, V_o can be made zero but this equilibrium will be valid for definite values of I_1 and I_2. I_1 and I_2 are not constant and their values depend on the state of the circuit and temperature. Due to difference of biasing current $(I_2 - I_1)$ or off-set current the off-set voltage will be

$V_o = AR_2 (I_2 - I_1)$ which is equivalent to off-set voltage at input $V_{os} = \frac{V_o}{A} = R_2 (I_2 - I_1)$. As shown in Fig. (11.11) if an additional voltage of magnitude V_{os} is applied at the positive terminal then the in equilibrium can be removed. This additional voltage can be applied in the outer circuit or internally in the amplifier. Usually a potentiometer is used for this purpose where $V_{os} = \frac{VR_2}{(R + R_2)}$, Fig. (11.12).

Keeping both input grounded the ratio of off-set voltage V_o obtained at the output and gain of the amplifier A is called input off-set voltage.

$V_{os} = \frac{V_o}{A}$ (With zero input voltage)

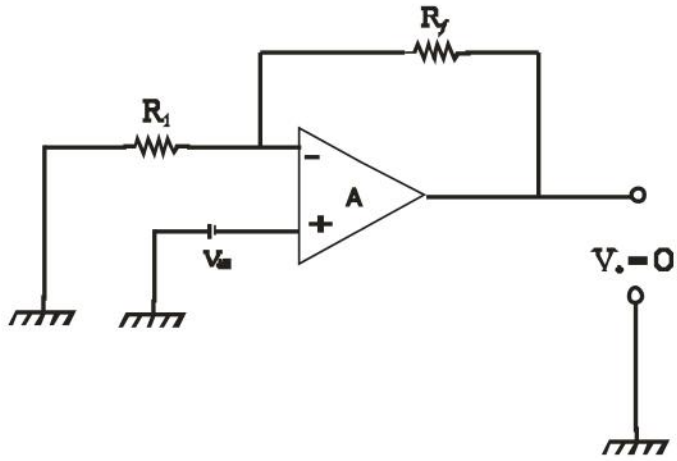


Fig. (11.11)

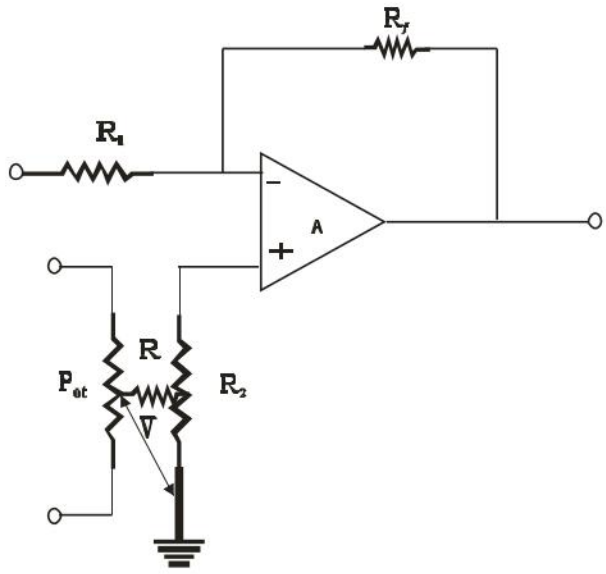


Fig. (11.12)

11.7 DC Offset Parameters

The OP-AMP is widely used in d. c. or a. c. amplifiers. In d. c. amplifier applications, certain electrical characteristics of the OP-AMP can cause large errors in the output voltage. The ideal output voltage should be equal to the product of the d. c. input signal and the amplifiers closed loop voltage gain. However, the output voltage may have an added error component. This error is due to differences between an ideal OP-AMP and a real OP-AMP. If the error component is comparable to or even larger than the ideal value of output voltage then steps are taken to minimize the error. OP-AMP characteristics that add error components to the d. c. output are: input bias current, input offset current, input offset voltage, drift.

When OP-AMP is used in an a.c. amplifier, coupling capacitor eliminate d. c. output- voltage errors mentioned above so that they are unimportant in a. c. applications. But there are other problems for a. c. amplifiers viz., frequency response, slew rate.

(1) Input Bias Current- In an ideal OP-AMP output is zero when the two inputs are identical (V_o =0 if V_1 = V_2). In practice, due to mismatch of input transistors, unequal bias current flows through the input terminals. Thus input bias current is the current flowing into each of the two input terminals when they are biased to the same voltage levels or the average of the currents into the two input terminals (fig. 11.13)

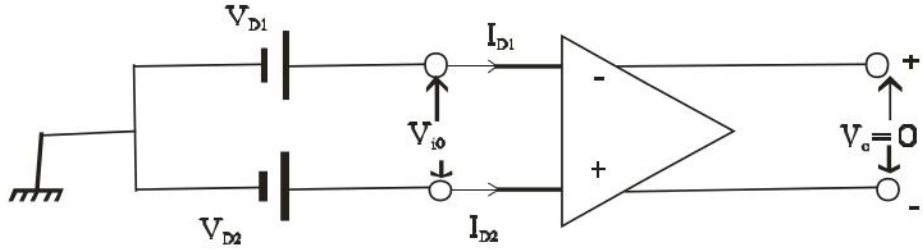


Fig.: (11.6.1)

With output at zero volts, is input bias current. That its

$$I_{bias} = \frac{I_{D1} + I_{D2}}{2} \qquad (1)$$

(2) Input Offset Current:- The input offset current I_{io}, is the difference of the currents into the two input terminals with the output at zero volts (V_o–0). Thus

with V_o=0, we have

$$I_{io} = I_{D1} - I_{D2} \tag{2}$$

Input offset current typically lies in the 20 to 60 MA. This current drifts with change in temperature. If ΔI_{io} is the change in input offset current due to the change ΔT in the temperature, then input offset current drift

$$= \frac{\Delta I_{io}}{\Delta T} \tag{3}$$

(3) Input Offset Voltage:- It is input voltage which must be applied across the input terminals to obtain zero output voltage.

In practical OP-AMP it is found that even if equal voltages are applied to the input terminals, output voltage is not zero. To set output voltage to zero, we required an input offset voltage, v_{io}, which typically lies in the range 1 to 4mV.

Input offset voltage drifts with change in temperature. If ΔV_{io} is the drift uninput offset voltage for a change ΔT in temperature, then input offset voltage

$$drift = \frac{\Delta I_{io}}{\Delta T} \tag{4}$$

(4) Input common mode range:- It is the maximum differential signal that can be applied safely to OP-AMP inputs.

(5) Output Offset Voltage :- It is the difference between the d.c. voltages present at the two output terminals when the two input terminals are grounded.

(6) Power supply voltage Rejection Ratio:- It is defined as input offset voltage change per volt of supply voltage change, i.e.

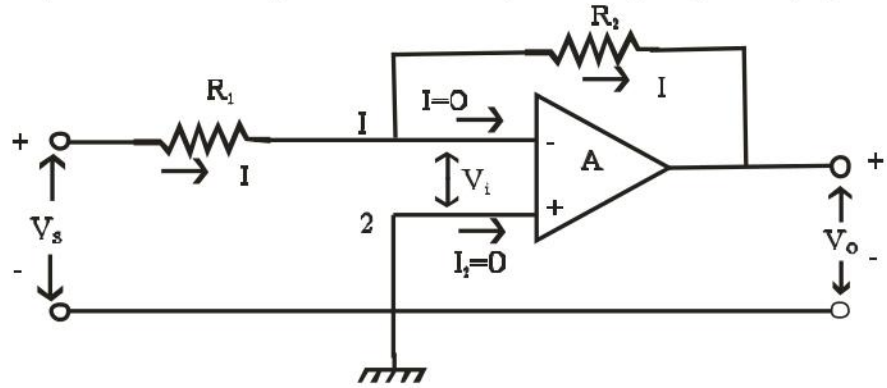
$$\text{PSRR} = \frac{\Delta V_O}{V_{cc}} \tag{5}$$

(7) Output Voltage swing (range):- It is the maximum peak to peak output voltage which can be obtained without waveform distortion. Output voltage swing is a function of the supply voltage and may range 50 to 80 percent of the supply voltage.

(8) Full Power Bandwidth:- It is the maximum frequency over which the full output voltage swing can be obtained.

11.8 Frequency Response And Stability

An OP-AMP is required to be unconditionally stable so that it will not break into oscillations for any amount of negative feedback. The excess phase shift associated with the gain stage leads to oscillations. Hence it is necessary to compensate frequency performance of the OP-AMP circuit in closed loop response. Let us diagram a closed loop inverting amplifier (Fig. 11.14).



(Fig. 11.14).

Since $R_i \longrightarrow \infty$, no current flows into the input terminals of the amplifier and the current I which flows through R_1 also flows through R_2 so that

$$I = \frac{V_s - V_i}{R_1} = \frac{V_i - V_o}{R_2}$$

$$\text{Or } \frac{V_s}{R_1} = \frac{V_i}{R_2} + \frac{V_i - V_o}{R_2} = V_o\left[\frac{V_i}{V_o R_1} + \frac{(V_i/V_0) - 1}{R_2}\right]$$

Defining $+A(\omega) = +\frac{V_o}{V_i} = -A(\omega)$ we have

$$\frac{V_s}{R_1} = \frac{V_o}{A(\omega)}\left[\frac{1}{R_1} + \frac{1 + A(\omega)}{R_2}\right]$$

$$\text{or } \frac{V_o}{V_s} = \frac{R_2}{(R_1 + R_2)}\left(\frac{\frac{-A(\omega)}{1 + R_1 A(\omega)}}{R_1 + R_2}\right)$$

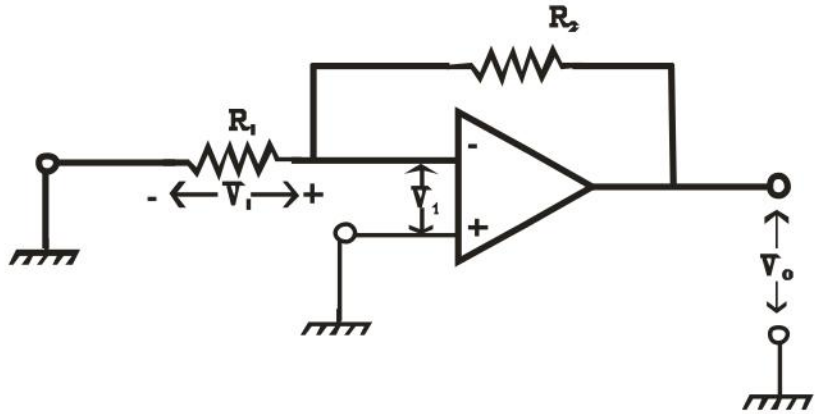
The denominator term is $\left(1 + \frac{R_1 A(\omega)}{R_1 + R_2}\right)$ is the term of interest regarding stability

and is defined as the loop gain. The gain A (ω) is complex, having magnitude as well as phase.

If $\frac{R_1|A(\omega)|}{(R_1+R_2)} = 1 < 180^0$ That is, it is unity in magnitude and becomes negative (or the phase angle of A (ω) becomes 180^0) then Vo/Vs→∞. It means we shall have an output signal with no input- a condition for oscillations and the circuit no more acts as amplifier. This oscillating condition is to be avoided. By sketching the Bode plot of an amplitude and phase of the loop gain function A (ω)/ (R_1+ R_2) versus frequency, it can be determined whether the magnitude is greater than zero decibels when the phase angle passes through 180^0 so that amplifier is unstable.

11.9 Slew Rate

It is the maximum rate of change of output voltage for a step input. The response of an OP-AMP to a large change in input signal is not as fast as might be expected. Suppose that a step function of large amplified is applied as V_1 (Fig. 11.15). The internal capacitances of the amplifier and feedback loop cannot change voltage rapidly so that a finite time is required for V_o to respond to the step input. Further as the feedback voltage, applied through R_2 to input, is proportional to V_o, the feedback is delayed in its returns to the input, and the amplifier input voltage v_i, is momentarily large. As a result, until the feedback voltage responds, input voltage v_i drives the amplifier into saturation and a distorted output wave form appear as a transient.



(Fig. 11.15)

Output voltage, V_o can rise only as fast as the capacitance can change or slew in

voltage. The maximum possible rate of change of output voltage is defined as the slew rate, S, where $S = \left(\frac{dv_o}{dt}\right)_{\max}$

For a sinusoidal voltage $V_o = |v_O| \sin\omega t$

So that $\frac{dv_O}{dt} = \omega |v_O| \cos\omega t$ or $\left(\frac{dv_o}{dt}\right)_{\max}$ Or $\geq \omega |v_O|$

So that maximum possible amplified at a frequency f is $|v_o| = \frac{\left(dv_o/dt\right)_{\max}}{\omega} = \frac{s}{\omega}$

This slew rate is determined by the capacitances in the amplifier further.

(i) As the input is a very fast step of amplitude which is much larger than the amplifier can handle linearly, in slew rate determination OP-AMP operates non-linearly. Thus slew rate permits us to determine the extent to which the linear range of operation of an OP-AMP is affected by the speed of the input signal.

(ii) As the large step input drives the output abruptly from one limit to other, the feedback cannot be effective and therefore rate is rather independent of the amount of feedback.

11.10 Constant Gain Multiplier

One of the most common OP-AMP circuits is the inverting constant- gain multiplier, which provides a precise gain or amplification Fig. (11.16) shows a standard circuit connection, with the resulting gain being given by $A = -\frac{R_f}{R_1}$

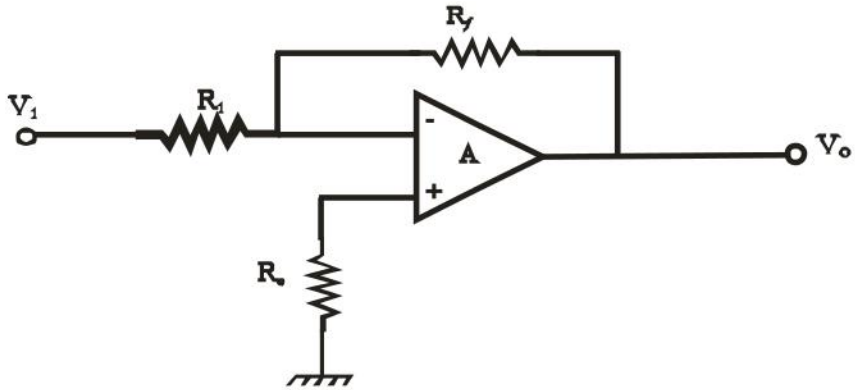


Fig.(11.16)

A non inverting constant gain multiplier is provided by the circuit of Fig. (11.17)

with the gain given by $A = 1 + \frac{R_f}{R_1}$

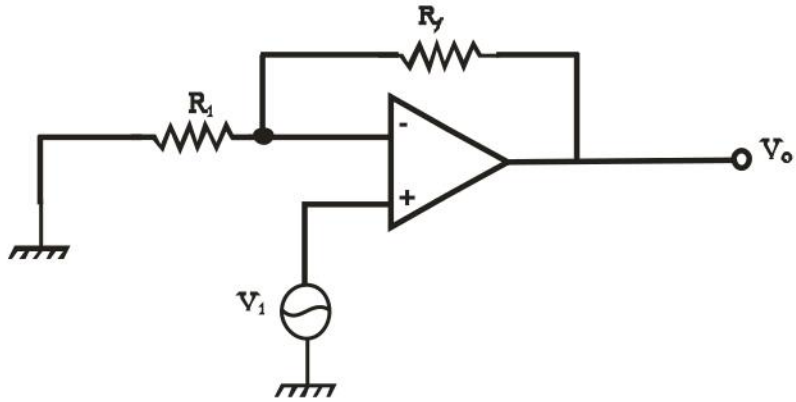


Fig. (11.17)

11.11 Multiple Stage Gains

When a number of stages are connected in series, the overall gain is the product of the individual stage gains. Fig.(11.18) shows a connection of three stages. the first stage is connected to provide non inverting gain is given by

$A = 1 + R_f / R.$

The next two stages provide an inverting gain given by

$A = -R_f / R.$

The overall circuit gain is then non inverting and is calculated by

$A = A_1 A_2 A_3$

Where, $A_1 = 1 + R_f / R_1$, $A_2 = -R_f / R_2$, $A_3 = R_f / R_3$

$$A = \left(1 + \frac{R_f}{R_1}\right)\left(-\frac{R_f}{R_1}\right)\left(-\frac{R_f}{R_3}\right)$$

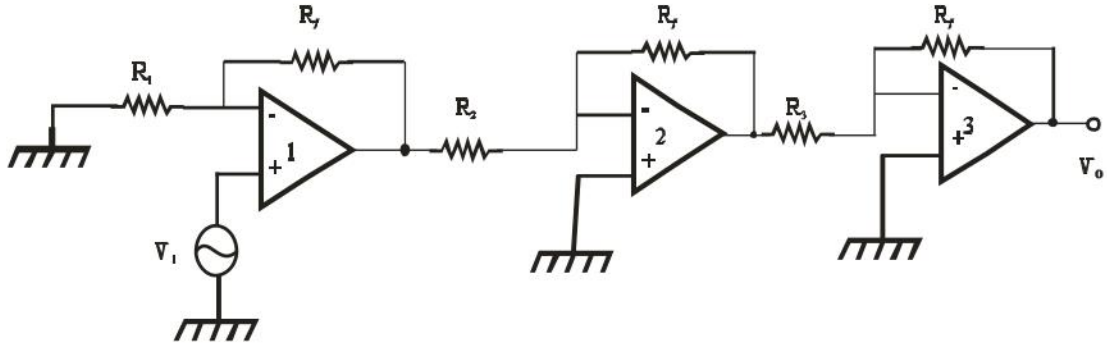


Fig. (11.18)

11.12 Active Filters

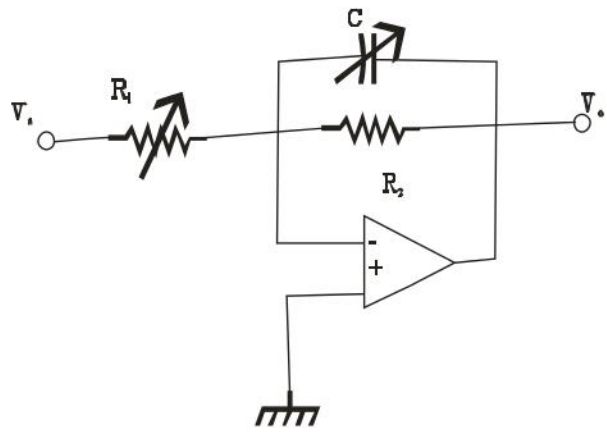
Active filters are filters that employ passive element, usually resistors and capacitors, in conjunction with active elements, such as operational amplifiers. They have the advantages.

(i) As active circuits contain no inductors, it is possible to integrate them.

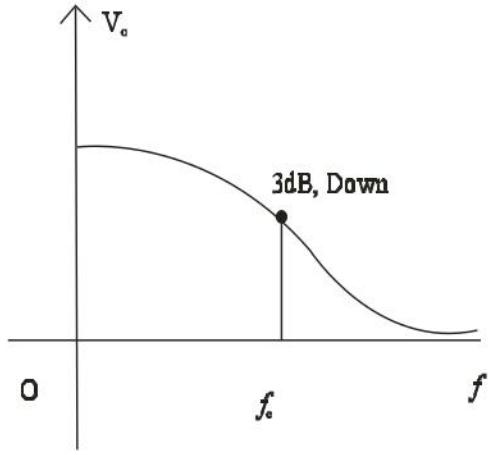
(ii) They can be made to work at very low frequencies where it is in practical to use inductors.

11.13 Low Pass Filter

In fig. (11.19) an active low pass filter with output response is shown. At low frequencies the capacitor appears open and the circuit acts as an inverting amplifier with a voltage gain of $-R_2/R_1$ as the frequency increases, capacitive reactance decreases with a consequent drop in voltage gain, till at infinite frequency the gain approaches zero because then capacitor appears shorted.



(Fig. 11.19)



(Fig. 11.20)

From figure $Z_1 = R_1$ and $\frac{1}{Z_2} = \frac{1}{R_2} + \frac{1}{1/j\omega c}$

$$\frac{1}{Z_2} = \frac{1}{R_2} + j\omega c$$

$$\frac{1}{Z_2} = \frac{1 + j\omega R_2 C}{R_2} \qquad \text{or } Z_2 = \frac{R_2}{1 + j\omega R_2 C}$$

The ratio of the impedances is

$$\frac{Z_2}{Z_1} = \frac{R_2}{R_1}\frac{1}{1+j\omega R_2 C} \qquad \text{or} \qquad \left|\frac{Z_2}{Z_1}\right| = \frac{R_2}{R_1}\frac{1}{\sqrt{[(1+(\omega R_2 C)^2]}} \tag{1}$$

The output response shows that output, V_o is maximum at low frequencies. When the frequency reaches the critical or cutoff frequency, the output is down 3db. The 3dB corner frequency occurs when

$$\omega R_2 C = 1 \quad \text{or} f_c = \frac{1}{2\pi R_2 C} \tag{2}$$

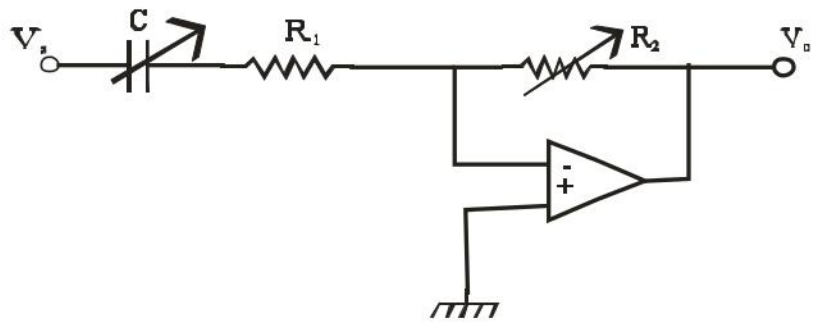
By varying C we vary the corner frequency and by varying R_1, we control the gain A because

$$A = \frac{V_o}{V_s} = -\frac{Z_2}{Z_1} = -\frac{R_2}{R_1}\frac{1}{\sqrt{\left[1+(\omega R_2 C)^2\right]}} \tag{3}$$

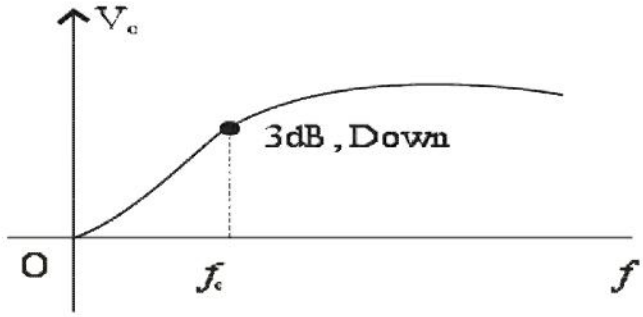
Where mines sign stands for phase inversion.

11.14 High Pass Filter

In fig.: 11.21 an active high pass filter with output response is shown. At low frequencies, the capacitor appears open and the voltage gain approaches zero. But at high frequencies, the capacitor appears as shorted and the circuit becomes an inverting amplifier with a voltage gain of - R_2/R_1.



(Fig.: 11.21)



(Fig.: 11.22)

From Figure

$$Z_1 = R_1 + \frac{1}{j\omega c} = \frac{1 + j\omega cR_1}{j\omega c} \text{ and } Z_2 = R_2$$

$$\text{So that } \frac{Z_2}{Z_1} = \frac{R_2 j\omega c}{1 + j\omega cR_1} \text{ or } \left|\frac{Z_2}{Z_1}\right| = \frac{\omega cR_2}{\sqrt{[1 + (\omega cR_1)^2]}} \quad (1)$$

For cut off frequency which corresponds to 3dB down, we put

$$\omega CR_1 = 1 \quad \text{or} \quad 2\pi fc \; CR1 = 1 \text{ or } fc = \frac{1}{2\pi R_1 c} \quad (2)$$

11.15 Band Pass Filters

A Band pass filter has a pass band between two cut off frequencies f_H and f_L such that $f_H > f_L$. Any input frequency outside this pass band is attenuated.

Basically, there are two types of band-pass filters: (i) wide band pass, and (2) narrow band pass. Unfortunately there is no set dividing line between the two. However, we will define a filter as wide band pass if its figure of merit or quality factor Q<10. On the other hand, if Q>10. We will call the filter a narrow band-pass filter. Thus Q is a measure of selectivity, meaning the higher the value of Q, the more selective is the filter or the narrower its band width (BW). The relationship between Q, the 3-dB band width and the centre frequency f_c is given by

$$Q = \frac{fc}{BW} = \frac{fc}{f_H - f_L} \quad (1)$$

For the wide band pass filter the center frequency f_c can be defined as

$$f_c = \sqrt{f_H f_L} \quad (2)$$

Where f_H = high cut off frequency (Hz)

f_L = Low cut off frequency of the wide band-pass filter (Hz)

In a narrow band -pass filter, the output voltage peaks at the center frequency.

11.16 Wide Band Pass Filter

A wide band-pass filter can be formed by simply cascading high pass and low pass sections and is generally the choice for simplicity of design and performance. To obtain a ± 20 dB band pass, high pass and low pass sections are cascaded. In other words, the order of the band pass filter depends on the high pass and low pass filter

section.

Fig. (11.23) Shows the ± 20 dB wide band pass filter which is composed of high-pass filter and low pass filters.

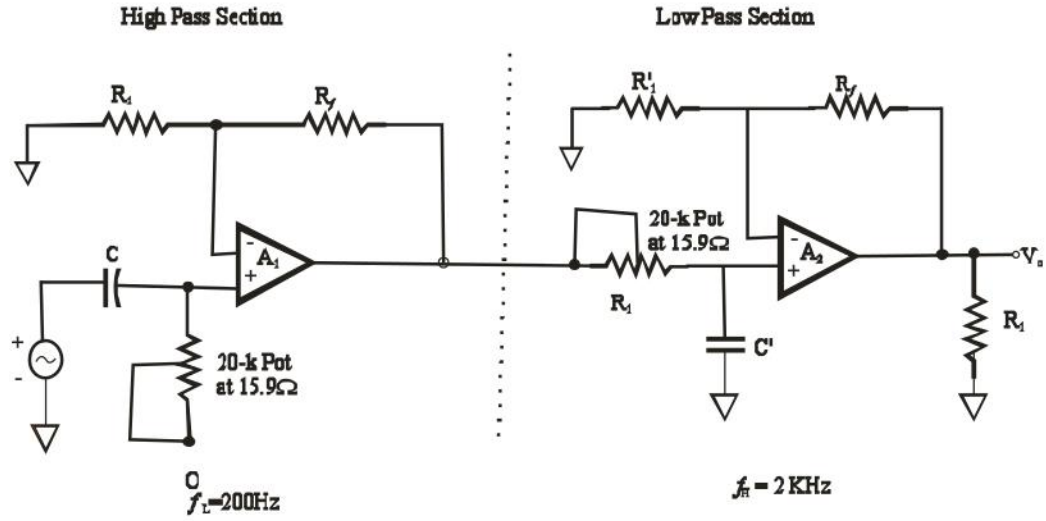


Fig. (11.23)

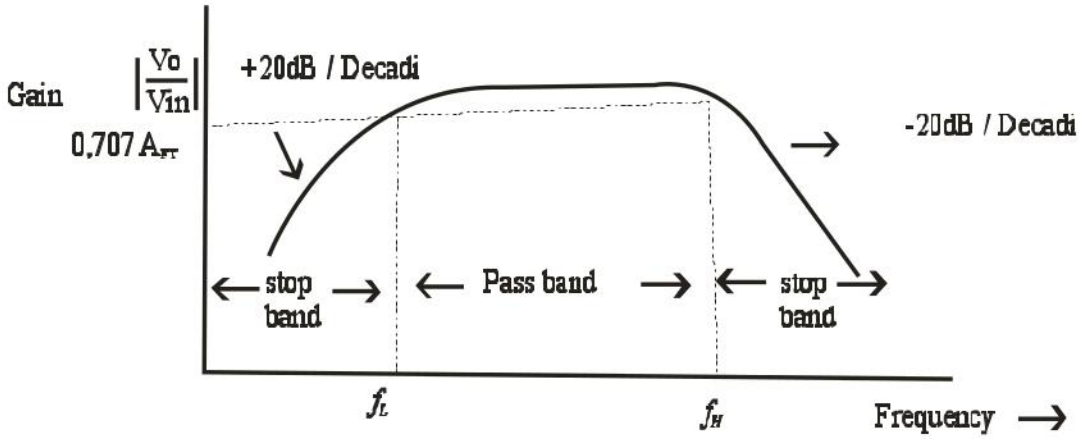


Fig. (11.24)

Generally, the narrow band pass filter is designed for specific values of centre frequency f_c and Q or f_c and band width. The circuit components are determined from the following relationships.

Let $C_1 = C_2 = C$

$$R_1 = \frac{Q}{2\pi f_c C A_F} \qquad (1)$$

$$R_2 = \frac{Q}{2\pi f_c C\left(2Q^2 - A_F\right)} \tag{2}$$

$$R_3 = \frac{Q}{2\pi f_c C} \tag{3}$$

Where A_F is the gain at f_c, given by

$$A_F = \frac{R_3}{2R_1} \tag{4}$$

The gain A_F, however must satisfy the condition

$$A_F < 2Q^2 \tag{5}$$

Another advantage of the multiple feedback filter of Fig. (11.24) is that its centre frequency f_c can be changed to a new frequency f'_c without changing the gain or band width. This is accomplished simply by changing R_2 to R_2' so that

$$R_2' = R_2\left(\frac{f_c}{f'_C}\right)^2 \tag{6}$$

11.17 Narrow Band-Pass Filter

The narrow band-pass filter using multiple feedback is shown in Fig. 11.25. As shown in this figure, the filter use only one OP-AMP compared to all the filers discussed so far, this filter is unique in the following respects:

(1) It has two feedback paths, hence the name multiple feedback filter.

(2) The OP-AMP is used in the inverting mode.

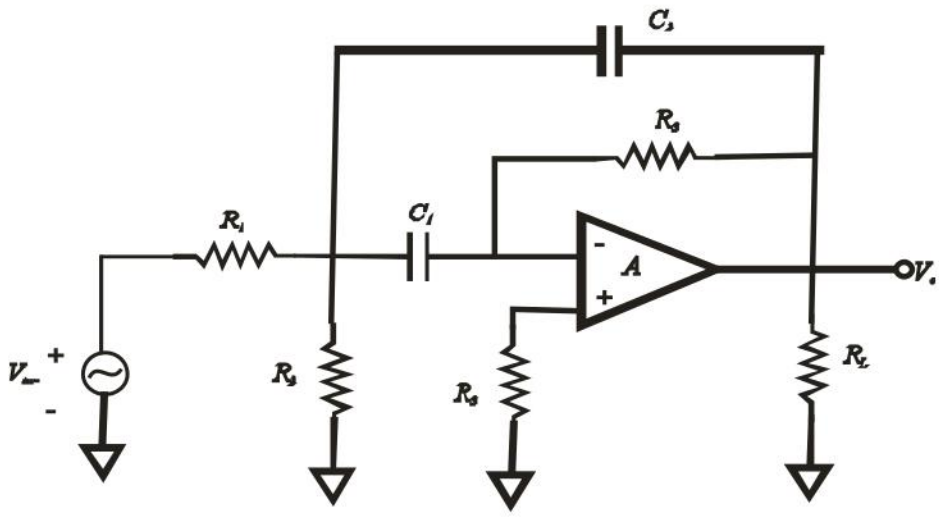


Fig. (11.25)

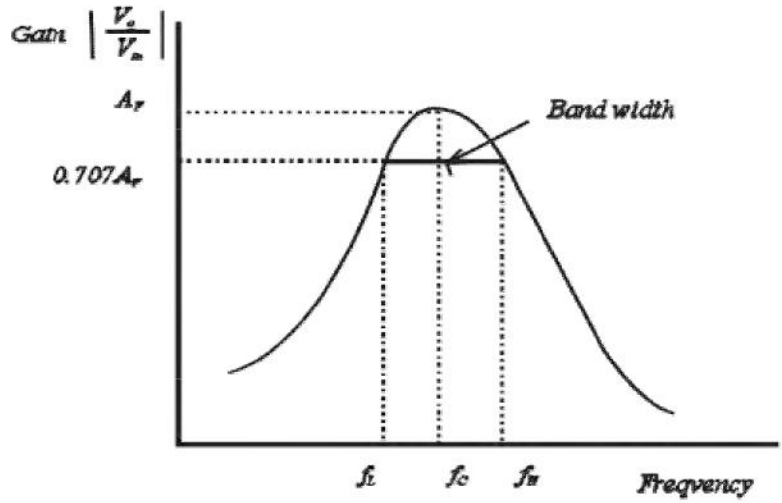


Fig. (11.26)

11.18 Self Learing Exercise-II

Very short Answer type question.

Q.1 Write the formula of Input offset current?

Q.2 Define power supply voltage rejection ratio?

Q.3 Write the relationship between Q and bandwidth?

Short Answer type question:-

Q.4 Draw the circuit diagram of a slew rate.

Q.5 Define Multiple stage gain.

Q.6 Draw the circuit diagram of a narrow Band-Pass filter.

11.19 Summary

Summing, subtracting, differentiating and integrating circuits can be constructed by using inverting configuration. The integrator and differentiators are most commonly used in signal wave shaping applications. In addition, the integrator is used in analog computers and differentiator is used as a rate of change detector in FM modulation. These values result in a dc output offset voltage. Offset voltages and current can be nullified by using compensating network. Frequency response is the ability of the amplifier to amplify signals at different frequencies. Generally for OP-AMPS, as the operating frequency increases two effects are observed: gain of the OP- AMP decreases and the phase shift between the output and the input signal

increases. The open loop gain of the OP-AMP is relatively constant for frequencies below the corner frequency but successively decreases at the rate of -20dB/decade. The useful frequency on open loop OP-AMP is very small. Where as in closed loop OP-AMPs the band width is much larger and advantage. For stability of OP-AMP, the phase angle at the paint of intersection of an ideal closed loop gain with the open loop response should be less than 180^0. Slew rate limiting occurs with larger amplitudes and high frequency signals. If the slew rate is exceeded, distortion in set in at the output stage.

A filter is often a frequency-selective circuit that passes a specified band of frequencies and alternates signals of frequencies outside this band. A low pass filter has a constant gain from OHZ to a certain frequency called the cut off frequency f_H, at which the gain is down by 3dB. Above f_H, the gain decreases with increases in frequency. On the other hand, a high-pass filter passes all the frequencies above a certain frequency called the low cut off frequency f_L. The upper cut off frequency of the high pass filter, however, depends on the band width of the OP-AMP. A band-pass filter has a pass band between two cut off frequencies f_H and f_L such that $f_H > f_L$ Any input frequency outside this pass band is attenuated.

11.20 Glossary

Amplifier : A device that increases the strength of an electrical signal by drawing energy from a separate source to that of the signal.

Input offset Current: difference between the currents into the two input terminals.

Slew rate: When the input is over driven, such as with a square wave, the output of the amplifier will slew at some maximum rate.

11.21 Answer To Self Learing Exercises

Answer To Self Learing Exercise-I

Ans.1: $V_o = -(V_A + V_B + \ldots\ldots + V_N)$

Ans.2: $V_o(t) = -\frac{1}{R_c}\int v_i(t)dt$

Ans.3: $V_O(t) = -R_C \frac{dv_i(t)}{dt}$

Answer To Self Learing Exercise-II

Ans.1: $I_{io} = I_{D1} = I_{D2}$

Ans.2: It is defined as input offset voltage change per volt of supply voltage change.

Ans.3: $\frac{f_C}{B_W} = \frac{f_C}{f_H - f_L}$

11.22 Exercise

Very Short Answer Type Questions

Q.1 Draw the circuit diagrams of differentiator OP-AMP.

Q.2 How does an operational amplifier work as an integrator?

Q.3 What is input Off-Set current?

Q.4 Define a filter.

Q.5 What is band pass filter.

Short Answer Type Questions

Q.6 Draw the circuit diagrams of differentiator and integrator? Explain its working.

Q.7 Explain the effect of negative feedback on the frequency response.

Q.8 What is the difference between compensated and non-compensated OP-AMPS?

Q.9 Explain the difference between the slew rate and transient response.

Q.10 Draw the frequency response curves and circuit diagram for the Low pass filters.

Long Answer Type Questions

Q.11 What is input Off-Set current? Describe a method to minimize it.

Q.12 How does an OP AMP act as an adder and subtractor? Explain by drawing circuits diagrams.

Q.13 Explain the use of OP-AMP as integrator and differentiator. Draw the necessary circuits.

Q.14 Define input Offset voltage and explain why it exists in all OP-AMPS

Q.15 Discuss the use of operation amplifier in active filters sketch the circuits of band pass, high and low pass active filters.

References and Suggested Readings

1. OP-AMP and Liner Integrated circuits, Ramkanth Gayakwad, PHI learning private Limited.cf
2. Hand Book of electronics, Gupta Kumar, Pragati Prakashan.
3. A text book of electronics, 2 e, S.L. Kakani and K.C. Bhandari, Newage International (P) Limited, Publishers.

UNIT-12

Feedback Amplifier

Structure of the Unit

12.0 Objectives

12.1 Introduction

12.2 Terminology for Feedback Amplifier Circuits

12.3 Feedback connection topologies

12.4 Feedback Gain formula

12.4.1 Voltage series feedback circuit

12.4.2 Voltage shunt feedback circuit

12.5 Types of feedback

12.6 Merits and Demerits of negative feedback

12.7 Self learning Exercise-I

12.8 Input Impedance and output Impedance with negative feedback

12.8.1 Effect of negative feedback on input impedance

12.8.2 Effect of negative feedback on output impedance

12.9 Stabilization of gain

12.10 Reduction in Noise and Distortion

12.11 Extension of bandwidth

12.12 Self learning Exercise –II

12.13 Summary

12.14 Glossary

12.15 Answers to Self Learning Exercises

12.16 Exercise

12.17 Answers to Exercise

References and Suggested Readings

12.0 Objectives

Amplifier is an electronic device which raises the power level of input signal. It is most common circuit in electronic devices. An ideal amplifier should possess the properties such as high bandwidth, low noise and distortion, high stability of gain, high input impedance and low output impedance. For any basic practical amplifier these parameters are more or less constant and cannot be controlled. Feedback allows playing with these parameters without any change in the basic amplifier circuit. Moreover, it is a technique to changes in the parameters to improve the performance an amplifier circuit.

This chapter includes the basic concept of feedback for amplifiers and different types of feedback connections. It includes derivation of mathematical formulas for negative feedback amplifier. It explains the various advantages of negative feedback for a amplifier circuit.

12.1 Introduction

In an amplifier or an oscillator circuit the input does not know what is going on at the output. If any reason the output changes or distorted the net input still remains unaffected and this distortion continue. Feedback term is not an electronic term, it is in uses day to day life e.g. if a teacher teaching in the class, the feedback provided by student helps to improve the teaching and in actual provide the information to the teacher about his level of approach to the students. Somehow, in an electronic circuit *feedback is the process of taking a part of output signal and providing back at input of the circuit.*

12.2 Terminology Used In Feedback Circuit

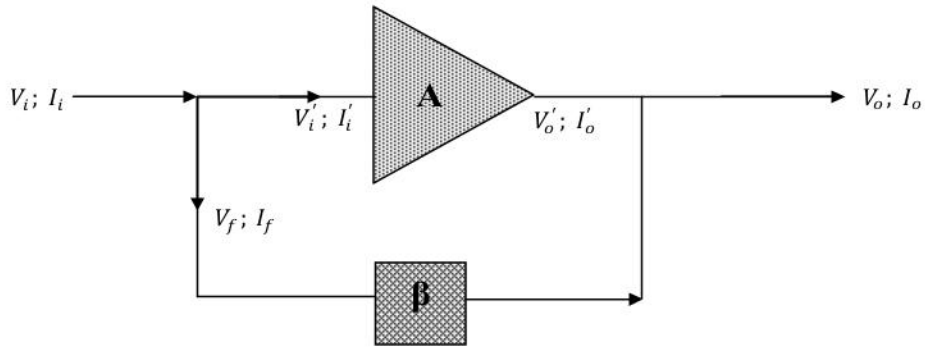


Fig.1: Feedback in amplifiers.

A_f	Gain with feedback; overall gain		
β	Feedback ratio	V_f	Feedback voltage
A	Gain without feedback	I_f	Feedback current
V_i	Input signal voltage to the circuit	V_i'	Actual input signal voltage to amplifier
V_o	Output signal voltage from the circuit	V_o'	Actual output signal voltage from amplifier(without feedback)
I_i	Input signal current to the circuit	I_i'	Actual input signal current to amplifier
I_o	Output signal current from the circuit	I_o'	Actual output signal current from amplifier

12.3 Feedback Connection Topologies

There are four basic ways of connection to provide the feedback. Current and voltage both can provide at input as feedback:

a) Voltage series feedback: A part of output voltage (V_o) is feedback at input in series in the form of feedback voltage (V_f).

b) Voltage shunt feedback: A part of output voltage (V_o) is feedback at input in parallel in form of feedback current (I_f).

c) Current series feedback: A part of output current (I_o) is feedback at input in series in form of feedback voltage (V_f).

d) Current shunt feedback: A part of output current (I_o) is feedback at input in parallel in the form of feedback current (I_f).

Block diagrams for such types of feedback connections are shown below:

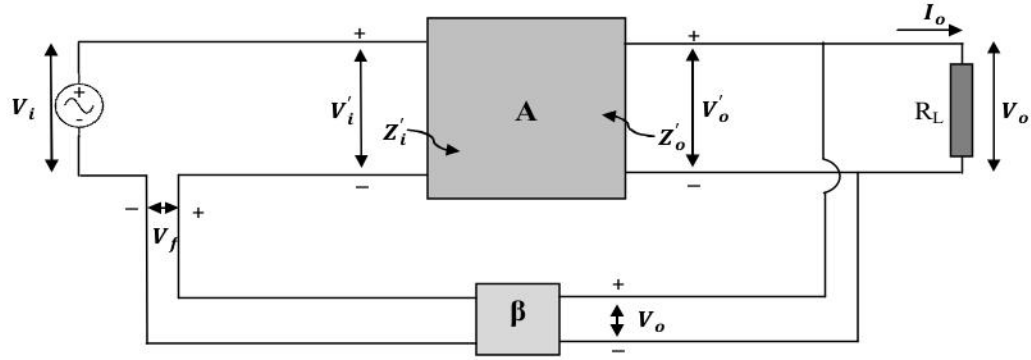


Fig.2: Voltage series feedback ($V_i' = V_i - V_f; V_o = V_o'$)

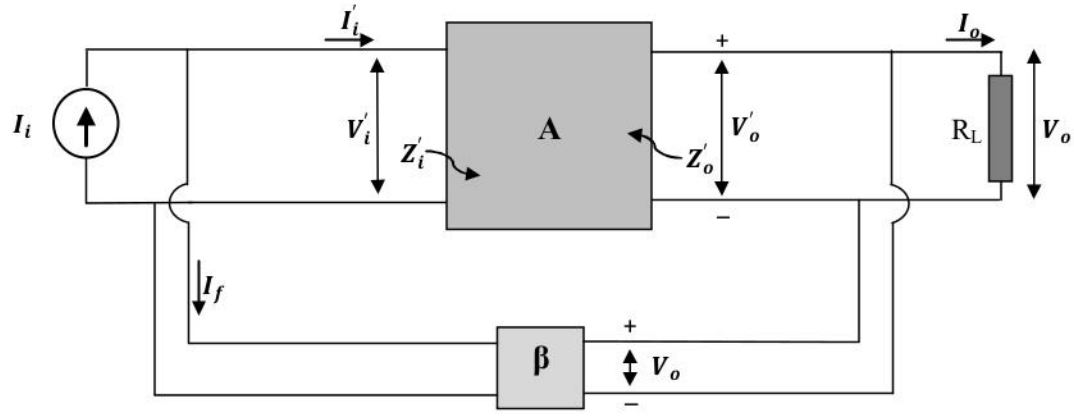


Fig.3: Voltage shunt feedback ($V_o' = V_o$; $I_i' = I_i - I_f$).

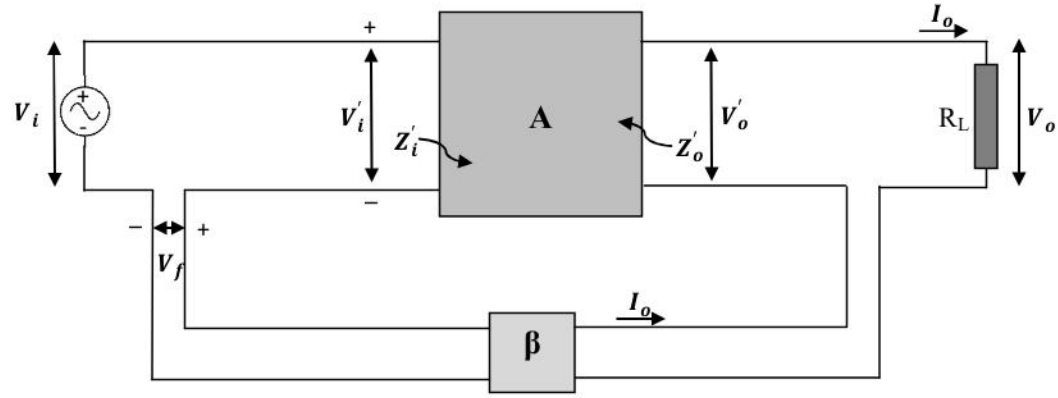


Fig.4: Current series feedback.

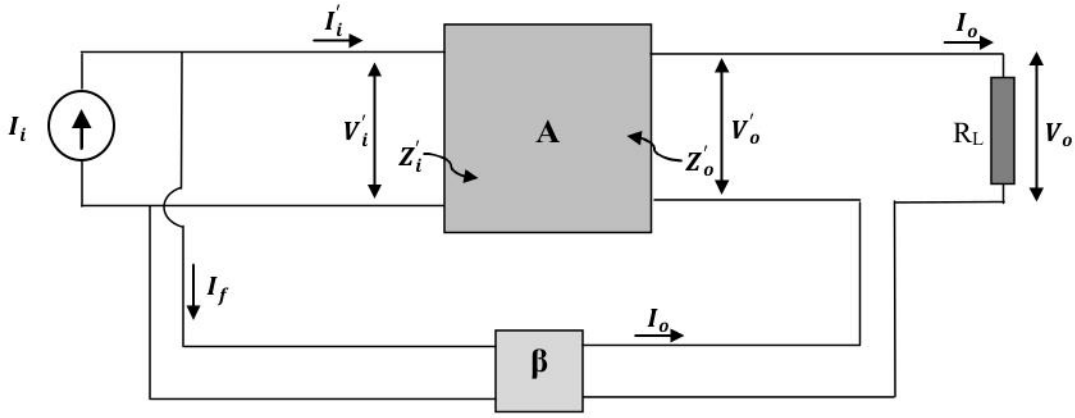


Fig.5: Current shunt feedback

Table 1: Various parameters for different feedback topologies

	Voltage-series	Voltage shunt	Current series	Current shunt
Gain with feedback (A_f)	$\frac{V_o}{V_i}$	$\frac{V_o}{I_i}$	$\frac{I_o}{V_i}$	$\frac{I_o}{I_i}$
Feedback (β)	$\frac{V_f}{V_o}$	$\frac{I_f}{V_o}$	$\frac{V_f}{I_o}$	$\frac{I_f}{I_o}$
Gain without feedback (A)	$\frac{V_o'}{V_i'}$	$\frac{V_o'}{I_i'}$	$\frac{I_o'}{V_i'}$	$\frac{I_o'}{I_i'}$
	$V_o' = V_o$ $V_i' \neq V_i$	$V_o' = V_o$ $I_i' \neq I_i$	$I_o' = I_o$ $V_i' \neq V_i$	$I_o' = I_o$ $I_i' \neq I_i$

12.4 Feedback Gain Formula

It is obvious that after employing the feedback the resultant feedback gain is changed.

12.4.1Voltage series feedback circuit:

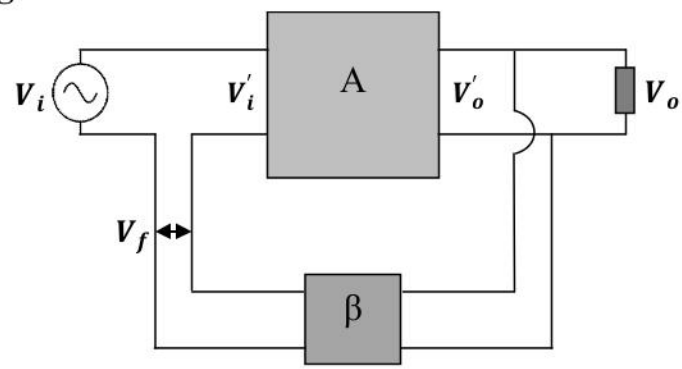


Fig. 6: Block diagram for voltage series feedback.

Figure shows the voltage series feedback circuit. Expressions for various voltages in this circuit can be written as:

$$V_i' = V_i \pm V_f$$

Here negative sign related to opposite polarity (out of phase) of feedback voltage to the input signal voltage while positive sign indicates the same phase of feedback voltage and input signal voltage. In results, the actual voltage to the amplifier (V_i') will be changed and it will be reduced for out of phase type feedback whereas, it

will be increased in the case of feedback with same polarity. Further,

$$V_i' = V_i \pm \beta V_o$$

$$\frac{V_o}{A} = V_i \pm \beta V_o$$

$$V_o = AV_i \pm A\beta V_o$$

$$\frac{V_o}{V_i} = \frac{A}{1 \mp A\beta}$$

Since, $\beta = \frac{V_f}{V_o}$

Also $A = \frac{V_o'}{V_i'} = \frac{V_o}{V_i'}$

So, overall gain of amplifier with feedback ($A_f = V_o/V_i$) can be given as below:

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

12.4.2 Voltage Shunt Feedback Circuit

Figure shows the block diagram of voltage shunt feedback circuit:

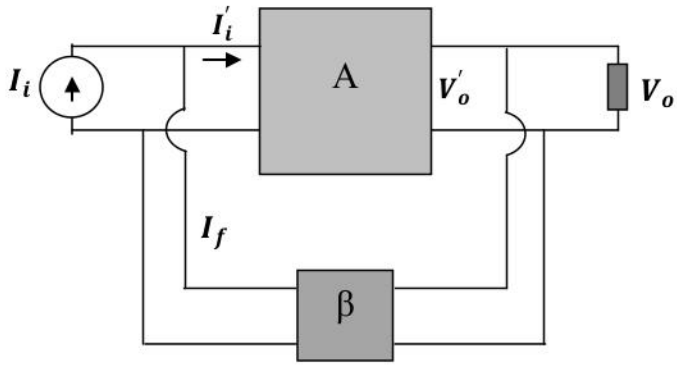


Fig. 7: Block diagram for voltage shunt feedback.

Here, total current to the amplifier can be given by as below:

$$I_i' = I_i \pm I_f$$

As like previous section here the negative sign related to opposite phase and plus sign is for the same phase of feedback at the input. The actual current to the amplifier (I_i') will be reduced in opposite phase type feedback whereas, it is reverse for the same polarity of feedback. Further,

$$I_i' = I_i \pm \beta V_o$$

$$\frac{V_o}{A} = I_i \pm \beta V_o$$

$$V_o = AI_i \pm A\beta V_o$$

$$\frac{V_o}{I_i} = \frac{A}{1 \mp A\beta}$$

Since for shunt feedback,

$$\beta = \frac{I_f}{V_o}$$

Also $A = \frac{V_o'}{I_i'} = \frac{V_o}{I_i'}$

So, overall gain of amplifier with feedback ($A_f = V_o/I_i$) can be given as below:

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

12.5 Types of feedback

In an electronic circuit the overall gain with the feedback given by the above expression. There are three cases are possible:

(i) *Negative feedback*: If the feedback signal is out of phase to the input signal is called as negative feedback. It most commonly used in amplifier circuit.

In this case $|1 + A\beta| > 0$ (negative) so $|A_f| < |A|$ and feedback gain is expressed for negative feedback as below

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

(ii) *Positive feedback:* If the feedback signal is in the same phase to the input is known as positive feedback.

In this case $|1 - A\beta| < 0$ so, $|A_f| > |A|$ and feedback gain is expressed for negative feedback as below

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

(iii) *Sustained Oscillations:* if the feedback signal is in such a way that $A\beta = 1$ so in this case, $|1 - A\beta| = 0$ and $|A_f| = \infty$. Amplifier provides sustained oscillations at output without any signal at input with such type of feedback. This condition is most commonly used in oscillators.

12.6 Merits And Demerits Of Negative Feedback

In a negative feedback current or voltage is provided at input out of phase to the input signal and it reduce the voltage or current level of actual input to the circuit hence, reduce the overall gain of circuit. Meanwhile, the voltage series negative feedback improves the performance of the amplifier circuit in various terms as below:

a) Increase input impedance

b) Reduce the out impedance

c) Improve the stability of gain

d) Reduce the noise and distortion

e) Increase the bandwidth of amplifier circuit.

f) More linear operation.

12.7 Self Learning Exercise-I

Q.1 What is the effect of negative feedback on the gain of amplifier?

Q.2 How many types of feedback possible for a amplfier?

Q.3 What is the formula of overall gain for a negative feedback amplifier?

Q.4 What are the advantages of negative feedback for amplifier circuit?

Q.5 What is the condition of feedback to achieve sustained oscillations?

12.8 Input Impedance And Output Impedance With Negative Feedback

Different feedback connection affects the input and output impedance of the amplifier circuit**.** Series feedback connections at input tend to increase the input impedance while shunt feedback connections tend to decrease the input impedance. Voltage feedback tends to decrease the output impedance whereas current feedback tends to increase the output impedance.

Typically the lower output impedance and higher input impedance are desired for cascading of amplifiers. Hence, the voltage series feedback connection is most useful feedback circuit for amplifiers. In the study of feedback, the voltage series feedback connection will be considered.

12.8.1 Effect Of Negative Feedback On Input Impedance

(i) *Series feedback:* As it mentioned above that a series feedback at input tend to increase the input resistance which improve the performance of an amplifier circuit. For a series negative feedback circuit $V_i' = V_i - V_f$ and $I_i = I_i'$

If the input impedance of amplifier circuit is Z_i' then the input current can be given by:

$$I_i = I_i' = \frac{V_i'}{Z_i'}$$

$$I_i = \frac{V_i - V_f}{Z_i'}$$

Since, for voltage series feedback $\beta = \frac{V_f}{V_o}$

$$I_i = \frac{V_i - \beta V_o}{Z_i'}$$

$$I_i Z_i' = V_i - \beta A V_i'$$

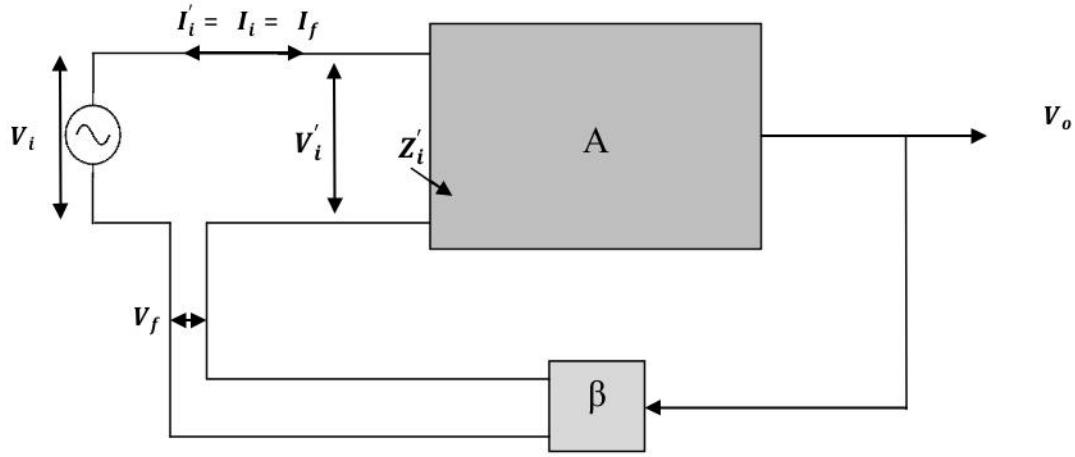


Fig. 8: Input Impedance calculation for series feedback circuit.

Input impedance with negative feedback can be given as

$$Z_{if} = \frac{V_i}{I_i} = Z_i' + \beta A Z_i' = Z_i'(1 + \beta A)$$

Since, for current series feedback $\beta = \frac{V_f}{I_o}$

$$I_i = \frac{V_i - \beta I_o}{Z_i'}$$

(As for current series feedback circuit the gain of amplifier $A = \frac{I_o}{V_i'}$)

$$I_i Z_i' = V_i - \beta A V_i'$$

$$Z_{if} = \frac{V_i}{I_i} = Z_i' + \beta A Z_i' = Z_i'(1 + \beta A)$$

So, from either of feedback connections at output it is voltage or current, if the feedback is given in series at input, the overall input impedance will be

$$\boxed{Z_{if} = Z_i'(1 + \beta A)}$$

As the $(1 + \beta A) > 1$

So, $Z_{if} > Z_i'$

So, the negative feedback in series at input will enhance the effective input

impedance of circuit which is a desirable characteristic of amplifier circuit.

(ii) *Shunt feedback: A* shunt feedback at input tend to decrease the input impedance which is not actually desired in an amplifier circuit. For a shunt negative feedback circuit $I_i^{'} = I_i - I_f$ and $V_i = V_i^{'}$. The block diagram of an amplifier circuit with voltage-shunt feedback is shown below:

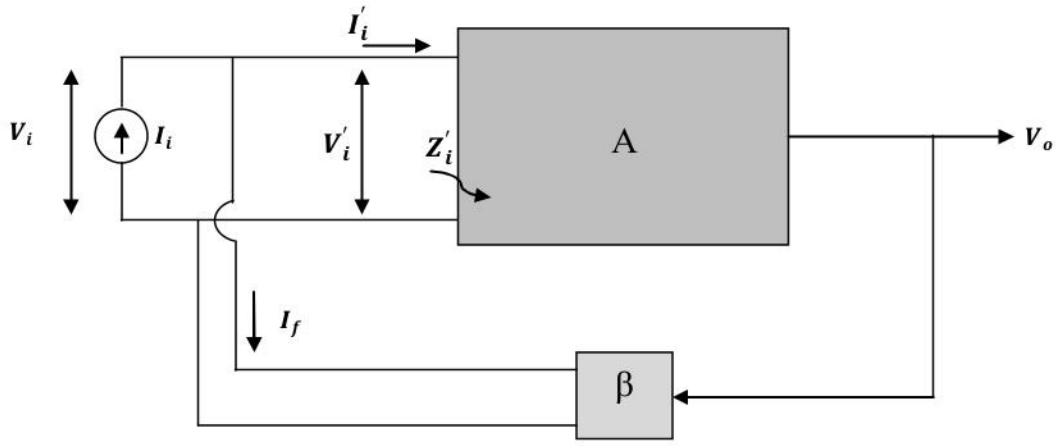


Fig. 9: Input Impedance calculation for shunt feedback circuit.

The overall input impedance (with feedback) of the above circuit with negative shunt feedback can be expressed as

$$Z_{if} = \frac{V_i}{I_i} \qquad \text{Since, } I_i' = I_i - I_f$$

$$Z_{if} = \frac{V_i^{'}}{I_i^{'} + I_f} \qquad \text{and } V_i = V_i'$$

(a) For a voltage shunt feed back	(b) For current shunt feed back
$\beta = \frac{I_f}{V_o}$ $Z_{if} = \frac{V_i'}{I_i' + \beta V_o}$ $Z_{if} = \frac{\frac{V_i'}{I_i'}}{\frac{I_i'}{I_i'} + \beta \frac{V_o}{I_i'}}$ As, $A = \frac{V_o}{I_i'}$	$\beta = \frac{I_f}{I_o}$ $Z_{if} = \frac{V_i'}{I_i' + \beta I_o}$ $Z_{if} = \frac{\frac{V_i'}{I_i'}}{\frac{I_i'}{I_i'} + \beta \frac{I_o'}{I_i'}}$ (Since, $I_o' = I_o$ for current feedback) and $A = \frac{I_o'}{I_i'}$

Hence, the input impedance with shunt negative feedback as below:

$$Z_{if} = \frac{Z_i'}{1 + \beta A}$$

As the $(1 + \beta A) > 1$

So, $Z_{if} < Z_i'$

12.8.2 Effect of negative feedback on output Impedance

(i) Series feedback: The impedance at output terminals in a negative feedback amplifier can be calculated using Thevenin equivalent circuits of amplifier with negative feedback. The equivalent circuit of an amplifier with negative feedback is shown in fig. 10. A signal voltage source (V_{of}) a*t* output can be considered which providing the current (I_{of}). The input terminals have been shorted so, $V_i' = -V_f$

a) For voltage series feedback: The output side has been replaced by an equivalent voltage source of (AV_i') with impedance Z_o' in the series.

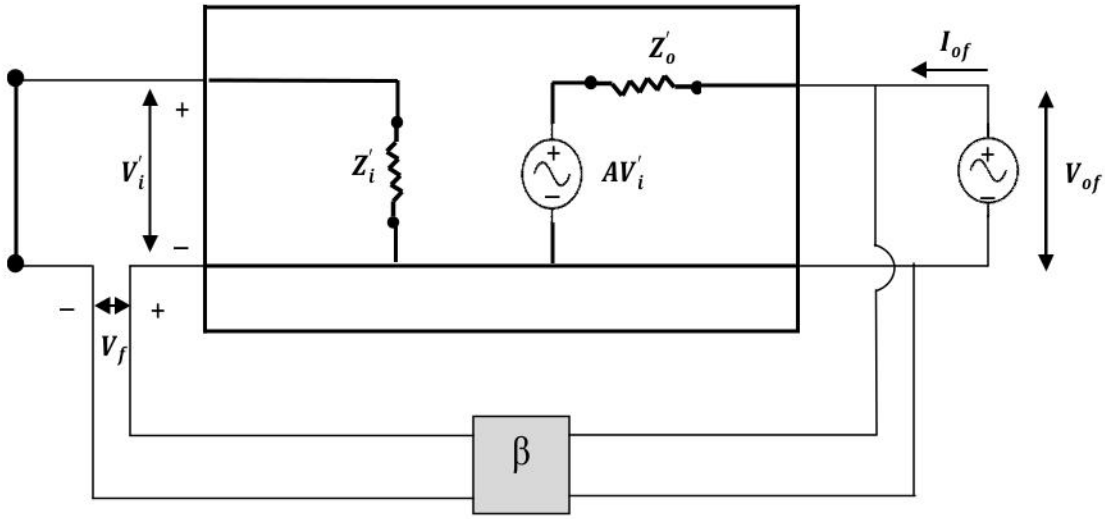


Fig.10: Output impedance calculation for voltage series feedback circuit.

If it can be assumed that the input impedance of the feedback network is very high it can be express,

$$V_{of} = I_{of}Z_o' + AV_i'$$
$$V_{of} + AV_f = I_{of}Z_o'$$
$$V_{of} + A\beta V_{of} = I_{of}Z_o'$$

$$\frac{V_{of}}{I_{of}} = \frac{Z_o'}{1 + A\beta}$$

$$Z_{of} = \frac{Z_o'}{1 + A\beta}$$

$Z_{of} < Z_o'$ Since, $(1 + A\beta) > 1$

So, for the voltage series feedback, the output impedance of an amplifier with negative feedback (Z_{of}) will be lower than the output impedance without negative feedback (Z_o') with a factor of $(1 + A\beta)$.

b) Output impedance for current series feedback: The Thevnin equivalent circuit *is* display *in fig.11.* The output side has been replaced by an equivalent current source of (AV_i') with parallel impedance Z_o'.

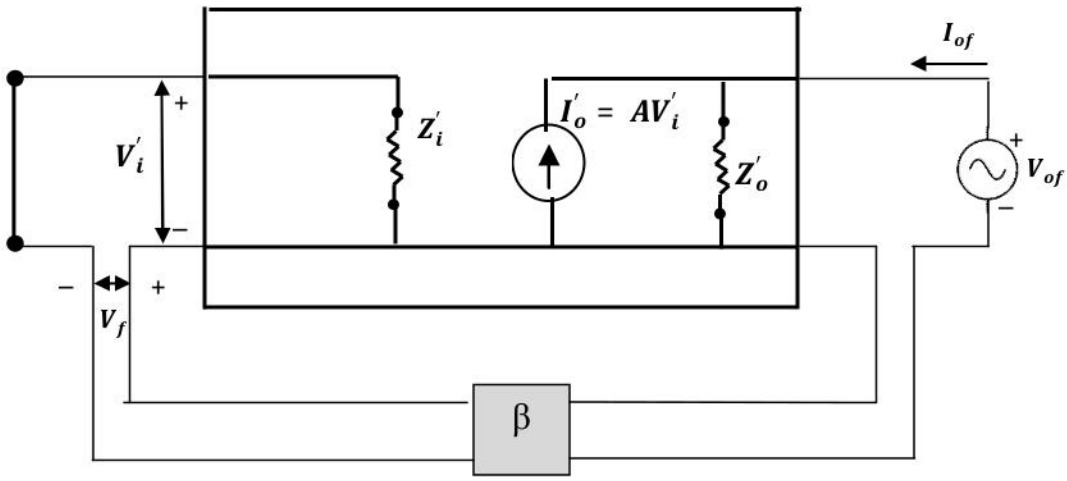


Fig.11: Output impedance calculation for current series feedback circuit.

$$I_{of} = \frac{V_{of}}{Z_o'} + AV_i'$$

$$I_{of} = \frac{V_{of}}{Z_o'} - AV_f$$

$$I_{of} = \frac{V_{of}}{Z_o'} - A\beta I_{of}$$

$$\frac{V_{of}}{I_{of}} = Z_o'(1 + A\beta)$$

$$Z_{of} = Z_o'(1 + A\beta)$$

$Z_{of} > Z_o'$ Since, $A\beta > 1$

The output impedance of an amplifier with negative feedback (Z_{of}) will be higher

than the output impedance without negative feedback (Z_o') with a factor of $(1 + A\beta)$.

(ii) Shunt feedback: The out impedance in a negative feedback amplifier can be calculated using Thevnin equivalent circuits. The equivalent circuit of an amplifier with negative feedback is shown in fig. 12. A signal voltage source (V_{of}) a*t* output can be considered which providing the current (I_{of}). The input terminals have been shorted and $I_i' = -I_f$

a) Voltage shunt feedback: The output side has been replaced by an equivalent voltage source of (AI_i') in the series with impedance Z_o'.

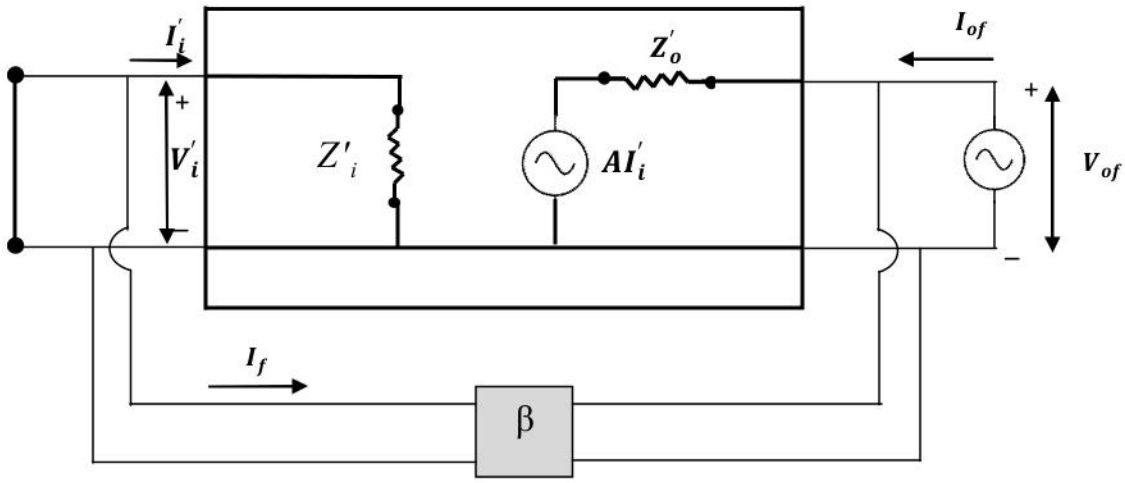


Fig.12: Output impedance calculation for voltage shunt feedback circuit.

$$V_{of} = I_{of}Z_o' + AI_i'$$

$$V_{of} = I_{of}Z_o' - AI_f$$

$$V_{of} = I_{of}Z_o' - A\beta V_{of}$$

$$V_{of}(1 + A\beta) = I_{of}Z_o'$$

$$Z_{of} = \frac{Z_o'}{1 + A\beta}$$

$Z_{of} < Z_o'$ Since, $A\beta > 1$

The output impedance of an amplifier with negative feedback (Z_{of}) will be lower than the output impedance without negative feedback (Z_o') with a factor of $(1 + A\beta)$.

b) Current shunt feedback: The output side has been replaced by an equivalent current source of (AI_i') with parallel impedance Z_o'

$$I_{of} = \frac{V_{of}}{Z_o'} + AI_i'$$

$$I_{of} = \frac{V_{of}}{Z_o'} - AI_f$$

$$I_{of} = \frac{V_{of}}{Z_o'} - A\beta I_{of}$$

$$\frac{V_{of}}{I_{of}} = Z_o'(1 + A\beta)$$

$$Z_{of} = Z_o'(1 + A\beta)$$

$$Z_{of} > Z_o' \text{ Since, } A\beta > 1$$

The output impedance of an amplifier with negative feedback (Z_{of}) will be higher than the output impedance without negative feedback (Z_o') with a factor of $(1 + A\beta)$.

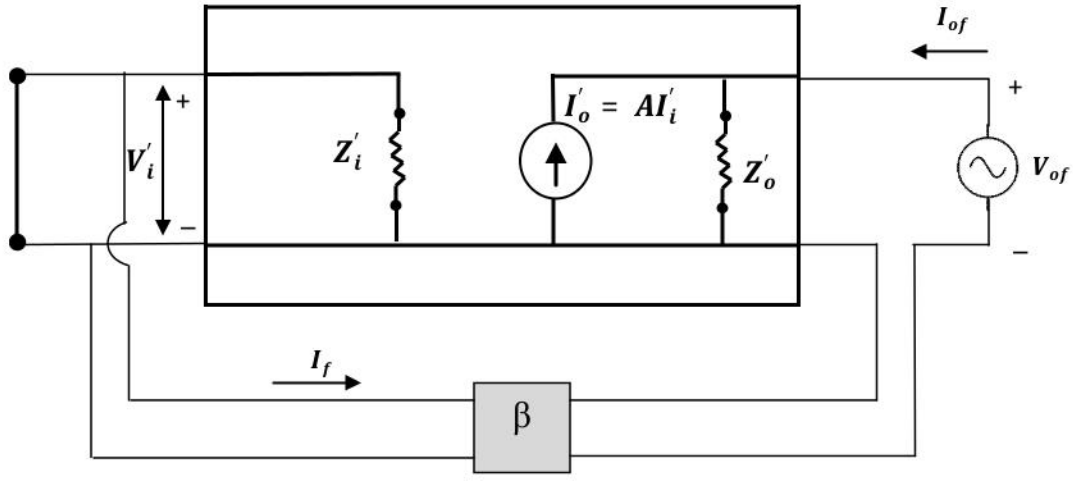


Fig.13: Output impedance calculation for current shunt feedback circuit

Table 2: Input and output impedance in various feedback connections

	Voltage-series	**Current series**	**Voltage shunt**	**Current shunt**
Output Impedance(Z_{of})	$\frac{Z_o'}{1 + A\beta}$	$Z_o'(1 + A\beta)$	$\frac{Z_o'}{1 + A\beta}$	$Z_o'(1 + A\beta)$
Input Impedance(Z_{if})	$Z_i'(1 + A\beta)$	$Z_i'(1 + A\beta)$	$\frac{Z_i'}{1 + A\beta}$	$\frac{Z_i'}{1 + A\beta}$

12.9 Stabilization of Gain By Negative Feedback

There are several reasons by which in an electronic circuit of amplifier the gain of amplifier (A) can be changed such as change in input voltage, fluctuations, and change in parameters of active device in amplifier circuit etc. Ideally, the gain of amplifier should be independent of these changes.

Negative feedback in amplifiers provides such an advantage. The gain of negative feedback amplifier can be express as below:

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

If the circuit is designed in the way that $A\beta \gg 1$ we can write $(1 + A\beta) \approx A\beta$

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

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

So the overall gain of amplifier with feedback will not depend on gain of amplifier (A) it will depend on feedback ratio only (β). Feedback ratio is decided by mostly combinations of passive components (R, L and C) which possess fairly constant values. So, the gain is stabilized. Moreover, if anyhow it is not possible that $A\beta \gg 1$ still it can be seen even there is change in gain of amplifier but the change in overall gain of will very less. The overall gain in negative feedback circuit is:

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

Differentiating it

$$dA_f = \frac{dA}{(1 + A\beta)^2}$$

Dividing this equation by the previous one

$$\frac{dA_f}{A_f} = \frac{1}{(1 + A\beta)} \frac{dA}{A}$$

$$\frac{dA_f}{A_f} \times 100 = \frac{1}{(1 + A\beta)} \frac{dA}{A} \times 100$$

$$Change\ in\ overall\ gain\ (\%) = \frac{1}{(1 + A\beta)}\ change\ in\ the\ gain\ (\%)$$

As $(1 + A\beta) > 1$ the change in overall gain (A_f) with feedback will be very less in compare to change in the actual gain of the amplifier (A).

12.10 Reduction In Distortion And Noise

The reduction in distortion is another advantageous characteristic of negative feedback for an amplifier. If an amplifier under use with cascading of stages to achieve desirable voltage level of signal but without feedback, the extra stages might introduce as much as noise back into the system which can enough distort the output signal. This problem can be resolved by using negative feedback. It reduces the overall gain but the factor of (1+Aβ) reduces both input noise and resulting nonlinear distortion. It can be understand by mathematical analysis also as below:

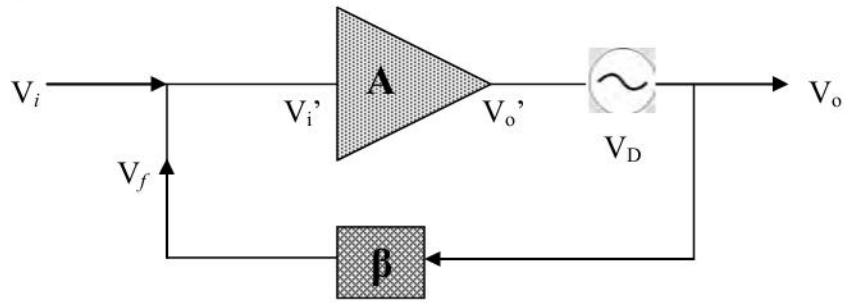


Fig. 14: A Reduction in noise and distortion by negative feedback.

If the distortion or noise represented by a alternator source of voltage V_D then the output voltage can be given as

$$V_o = V_o' + V_D$$
$$V_o = AV_i' + V_D$$

If no feedback then $V_i' = V_i$ and output voltage can be given by

$$(V_o)_{no\ feedback} = AV_i + V_D$$

But if there is negative feedback then $V_i' = V_i - V_f$ and output voltage will be:

$$V_o = A(V_i - V_f) + V_D$$
$$V_o = A(V_i - \beta V_o) + V_D$$
$$V_o(1 + A\beta) = AV_i + V_D$$
$$V_o = \frac{AV_i}{(1 + A\beta)} + \frac{V_D}{(1 + A\beta)}$$
$$V_o = A_f V_i + \frac{V_D}{(1 + A\beta)}$$

Comparing this output with feedback to the output without feedback signal output voltage will be less $[(V_o)_{no\ feedback} > V_o$ since, $A_f < A]$. However, by using

negative feedback the noise or distortion can be reduced with the factor of $(1 + A\beta)$.

12.11 Extension In Bandwidth

In a practical amplifier the open loop gain droops at high frequencies due to active device and stray capacitances. The gain at low frequencies also drops due to capacitive coupling of various stages. By using negative feedback although gain gets lower however, can be increased in band width. Negative feedback enhances the upper cut of frequency (f_h) and reduces the lower cut of frequency (f_l) with the factor of $(1 + A\beta)$. In a single stage RC coupled amplifier the voltage gain at frequency f can be express as below:

$$A_l = \frac{A_m}{(1-\frac{jf_l}{f})} \text{ (At lower frequency range)}$$

$$A_h = \frac{A_m}{(1+\frac{jf}{f_h})} \text{ (At higher frequency range)}$$

We know that for negative feedback amplifier overall gain is given by

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

(i) For lower frequency range:

$$A_{lf} = \frac{A_l}{1 + A_l\beta}$$

$$A_{lf} = \frac{\frac{A_m}{(1-\frac{jf_l}{f})}}{1 + \frac{A_m}{(1-\frac{jf_l}{f})}\beta}$$

$$A_{lf} = \frac{\frac{A_m}{(1 + A_m\beta)}}{1 + \frac{j}{f}\frac{f_l}{(1 + A_m\beta)}}$$

$$A_{lf} = \frac{A_{mf}}{1 + \frac{j}{f}f_{lf}}$$

Here, gain for mid frequency range is given by $A_{mf} = \frac{A_m}{1+A_m\beta}$ and lower half power frequency

$$f_{lf} = \frac{f_l}{(1 + A_m\beta)}$$

Since, as the $(1 + \beta A) > 1$ So, $f_{lf} < f_l$

Hence negative feedback reduces the value of lower cut off frequency.

(ii) For lower frequency range:

$$A_{hf} = \frac{A_h}{1 + A_h\beta}$$

$$A_{hf} = \frac{\frac{A_m}{(1 + \frac{jf}{f_h})}}{1 + \frac{A_m}{(1 + \frac{jf}{f_h})}\beta}$$

$$A_{hf} = \frac{\frac{A_m}{(1 + A_m\beta)}}{1 + \frac{j}{f} f_h(1 + A_m\beta)}$$

$$f_{hf} = f_h(1 + A_m\beta)$$

Since, as the $(1 + \beta A) > 1$ So, $f_{hf} > f_h$

Hence negative feedback enhances the value of higher cut off frequency.

The band width for a amplifier with negative feedback:

$$(f_{hf} - f_{lf}) > (f_h - f_l)$$

$B_f > B$

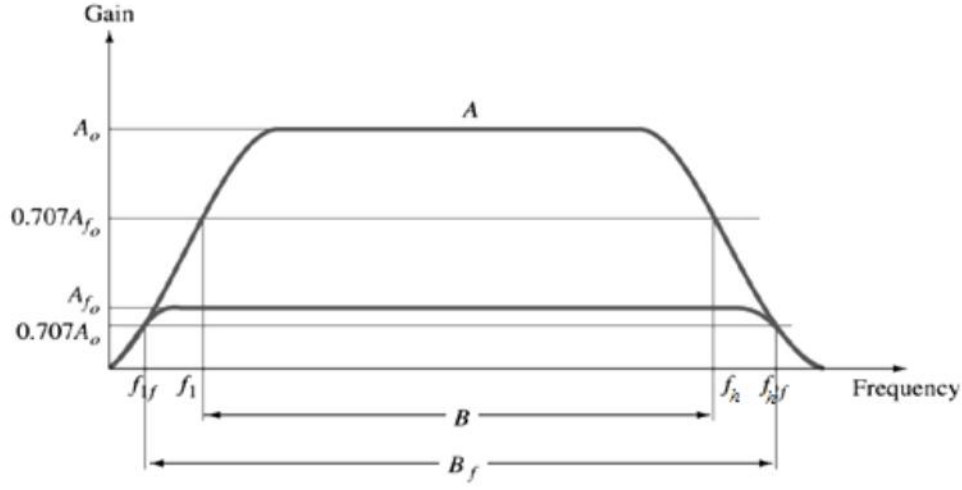


Fig. 14: Bandwidth extension by negative feedback.

12.12 Self learning Exercise –II

Q.1 What is the effect of voltage series negative feedback on input and output impedance of amplifier?

Q.2 What is the effect of voltage shunt negative feedback on input and output impedance of amplifier?

Q.3 What is the effect of current series negative feedback on input and output impedance of amplifier?

Q.4 What is the effect of current shunt negative feedback on input and output impedance of amplifier?

Q.5 What is the condition of feedback to achieve sustained oscillations?

12.13 Summary

In an electronic circuit, the feedback is the process of taking a part of output signal and providing back at input of the circuit. It can be of two types one is if the feedback signal is out of phase to the input signal is called as negative feedback and another is if the feedback signal is in the same phase to the input is known as positive feedback. Four types of negative feedback circuits are possible (i) voltage series; (ii) voltage shunt; (iii) current series and (iv) voltage series feedback circuit.Voltage series type of negative feedback is r*ef*err*ed* in amplifier circuit as it increse the input impedance and decrased the output impedance which is actually desired. Negative type feedback although decrese the overall gain but it improve the performance of aplifier in various manners like it Increase input impedance; Reduce the out impedance; Improve the stability of gain; Reduce the noise and distortion; Increase the bandwidth of amplifier circuit; More linear operation.

12.14 Glossary

Input impedance: The *input impedance of an amplifier* is the impedance at input terminals felt by the source driving the input of the amplifier. If it is too low, it can have poor loading effect on the previous stage in results affects the frequency response and output signal level of that stage.

Output impedance: The output impedance, source impedance, or internal impedance of an electronic device is the resistance exhibited by its output

terminals to an alternating current (AC) of a particular frequencyy. It is the Thevenin equivalent impedance looking back into the output terminals.

Cascading: A cascade amplifier is any two-port network constructed from a series of amplifiers, where each amplifier sends its output to the input of the next amplifier in a chain.

Noise: It can be defined as unwanted signals in the system which degrades the desired signal content.

Distortion: Distortion is the alteration of the original shape of something, such as an object, image, sound or waveform. Distortion is usually unwanted.

Bandwidth: The bandwidth represents the range of frequencies that the amplifier is most effective in amplifying. The bandwidth of an amplifier is the difference between the frequency limits of the amplifier.

12.15 Answers to Self Learning Exercises

Answers to Self Learning Exercise-I

Ans.1: It will reduce the gain of amplifier.

Ans.2: Two types negative and positive.

Ans.3: $A_f = \frac{A}{1+A\beta}$

Ans.4: Increase input impedance; Reduce the out impedance; Improve the stability of gain; Reduce the noise and distortion; Increase the bandwidth of amplifier circuit; More linear operation.

Ans.5: It reduces the overall gain but the factor of $(1+A\beta)$ reduces both input noise and resulting nonlinear distortion. by adjusting the gain of amplifier the noise can be minimize meanwhile no effect on output level of amplifier

Answers to Self Learning Exercise-II

Ans.1: Input impedance will be incresed and output impedance will be decresed by factor of $(1 + A\beta)$.

Ans.2: Input impedance will be decreased and output impedance will also be decresed by factor of $(1 + A\beta)$.

Ans.3: Input impedance will be incresed and output impedance will also be incresed by factor of $(1 + A\beta)$.

Ans.4: Input impedance will be decresed and output impedance will be incresed by factor of $(1 + A\beta)$.

Ans.5: $|1 - A\beta| = 0$

12.16 Exercise

Q.1 What do you understand by feedback in a amplifier?

Q.2 What do you understand by positive and negative feedback?

Q.3 Prove for the negative fedback the overall gain will $A_f = \frac{A}{1+A\beta}$.

Q.4 Draw the block diagram of the four types of negative feedback feedback circuits. which one of these is used to amplifier circuits to improve the performance?

Q.5 Expplan how the negative feedback in an amplifier helps in reducing the noise and distortion.

Q.6 Calculate the voltage gain, input impedance, output impedancein for an amplifier having voltage series negative feedback with parameters A= ¬100, Z_i= 10 KΩ; Z_o=20 KΩ; and feedback ratio β=¬0.1.

Q.7 If an amplifier with gain of -1000 and feedback ratio of -0.1 has a gain change 30% due to temperature. Calculate the change in gain of the feedback amplifier.

Q.8 An amplier has internal gain 200 and its output impedance is 1 KΩ. A negative feedback of 0.02 is employed then calculate the ouput impedance of the feedback amplifier.

Q.9 A amplifier with internal gain of -100 and providing 2 MHz bandwidth in frequency responce. if a negative feedback of 0.05 is employed then what will be the new bandwidth?

Q.10 A apmlifier with internal gain of -200 producing harmonic distortion around 8%. if a negative feedback of 0.075 is employed to this upto which level the

distortion will be reduced?

12.17 Answers to Exercise

Ans.1: In an electronic circuit feedback is the process of taking a part of output signal and providing back at input of the circuit.

Ans.2: If the feedback signal is out of phase to the input signal is called as negative feedback. If the feedback signal is in the same phase to the input is known as positive feedback

Ans.3: see section 12.4

Ans.4: see section 12.3. Voltage series negative feedback.

Ans.5: see section 12.10.

Ans.6: $A_f = -9.09$; $Z_{if} = 110\ K\Omega$; $Z_{of} = 1.82\ K\Omega$.

Ans.7: 0.3%

Ans.8: $Z_{of} = 200\ K\Omega$

Ans.9: 12 MHz

Ans.10: 0.5%

References and Suggested Readings

1. Electronic Devices and Circuit Theory by R.L. Boylestad and L. Nashelsky; published by Pearson Education.
2. Basic Electronics and Linear Circuits b7y NN Bhargava, D C Kulshreshtha, and SC Gupta; published by McGrawHill Education.

.3 Electronic Devices and Circuits by David A. Bellpublished by Oxford ; .University Press

UNIT-13
Oscillators

Structure of the Unit

13.0 Objectives

13.1 Introduction

13.2 Classification of Oscillators

13.3 Working Principle of a sinusoidal oscillator

13.4 Positive feedback in an oscillator

13.5 Condition for self-excitation

13.6 LC Oscillators (Tuned Oscillator)

13.6.1 Hartley Oscillator

13.6.2 Colpitts Oscillator

13.7 Self-learning Exercise-I

13.8 RC Oscillators

13.8.1 RC Phase Shift Oscillator

13.8.2 Wien Bridge Oscillator

13.9 Crystal Oscillator

13.10 Self-learning Exercise –II

13.13 Summary

13.14 Glossary

13.15 Answers to Self-Learning Exercises

13.16 Exercise

13.17 Answers to Exercise

References and Suggested Readings

13.0 Objectives

Now a day's oscillators are essential circuits in communication and broadcasting.

These circuits require to produce alternating current (ac) with high frequencies. This chapter includes the basic principle of working of various oscillator circuits.

This Chapter deals with

- Learning about various forms of feedback, which can originate the oscillations.
- Analysis of oscillator circuits that uses resonant circuits to determine their frequencies.
- Working of LC and RC oscillator and derivation for condition of self-excitation.
- Concept and working of crystal oscillator

13.1 Introduction

An electronic circuit, which generates the sinusoidal wave(ac) is known as oscillator. An oscillator circuit consists of a tuning circuit to decide the frequency of alternating current, amplifier and direct current (dc) power supply for biasing of active device. The circuit is not provided any kind of ac input; it converts, in actual, dc energy to ac energy.

There is difference between alternator (or commonly known as generator) and oscillator Alternator also generate ac current. Alternator consists of magnetic poles and an armature and the armature rotates inside the magnet to generate the ac power. This type of arrangement has some limitations and cannot produce ac current at high frequencies. To generate the high frequency by an alternator either number of poles or the rotation speed should be very high, which is not possible practically.

The high frequency current is essential in all communication systems such as television and radio broadcasting. More examples are Radar, cell phones, etc. Oscillators can provide the ac signal at wide range of frequencies. Moreover, a signal generator is essential to test various instruments in laboratories such as performance of stereo amplifier. Furthermore, high frequency oscillators are used in induction and dielectric heating, microwave ovens, etc.

13.2 Classification of Oscillators

Oscillators can be divided mainly into two parts:

(i) *Sinusoidal Oscillators*: Oscillators which generates the sinusoidal waveform of ac current are known as sinusoidal oscillators.

(ii) *Non sinusoidal Oscillators*: Oscillators which generates any waveform except sinusoidal such as square wave, pulse, saw tooth, triangular wave forms. Such type of oscillators circuits are useful in testing of various devices in a practical laboratory. Saw-tooth wave oscillator is essential part of cathode ray oscilloscope. These are more commonly known as waveform generators.

Sinusoidal oscillators for radio and audio frequencies can be further classified on the basis of tuning circuit or frequency generator used in the oscillator as:

(i) LC oscillators

(ii) RC oscillators

(iii) Crystal oscillators.

13.3 Working Principle of a Sinusoidal Oscillator

An oscillator circuit consists of frequency selection circuit (mostly a resonant circuit) which activated by an amplifier. Basic working of any sinusoidal oscillator can be understand in following points:

(a) *Generation of oscillations*: As the power is switched on to the circuit, a noise voltage is generated in the active device (BJT/FET) which may be due to random motion of electrons in the circuit. This noise voltage contains almost every sinusoidal frequency.

(b) *Selection of frequency*: Among these all frequencies a desired frequency is selected by resonant circuit.

(c) *Sustained Oscillations*: However, in start the amplitude of these oscillations is not on desired level. Further, the amplitude will also be decreased with the time due to loss of energy.

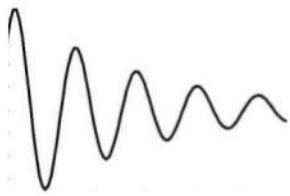


Fig.1: Damped Oscillations($A\beta < 1$).

The energy loss of oscillations may be due to any of the reasons such as resistance of the medium or circuit. Furthermore, if this ac energy is taken out at the output

by any load resistance, this also will cause the energy loss in the resonant circuit.
The oscillations can be maintained at constant level if we supply additional energy exactly in the same amount to the energy which is taken out at the output plus the energy that is lost in circuit. This additional energy is provided by amplifier attached with the resonant circuit.
The amplifier takes small part of signal from resonant circuit and provide the amplified signal. The current with desired frequency can be collect at the output of amplifier. This amplifier also maintains the amplitude of oscillations in the resonant circuit. Amplifier itself takes the energy from biasing dc batteries to amplify the signal. So, overall we can say that the dc energy, in actual, is getting converted in ac energy.

13.4 Positive feedback in Oscillators

An oscillator generates ac without any signal at input. A part of the output is feedback to input of the amplifier. The overall gain of the amplifier with feedback can be given as below

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

If the feedback signal is in such a way that the loop-gain $A\beta = 1$ so, $|1 - A\beta| = 0$ and $|A_f| = \infty$. In this case oscillator provides sustained oscillations at output without any signal at input with such type of feedback. This condition is most commonly known as *Barkhausen Criterion* for sustained oscillations.

Hence, to achieve the sustained oscillations with required amplitude following three essential conditions should be followed:

(i) The signal sent back to the amplifier should be in the same phase, hence the positive type of feedback.

(ii) Initially the loop gain should be greater than unity i.e. $A\beta > 1$. Since, the amplitude level of oscillations is build up to the desired level.

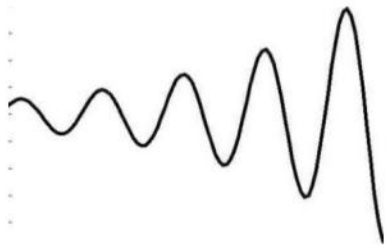


Fig. 2: Oscillations build up ($A\beta > 1$).

(iii) As the amplitude of oscillations is reached to the desired level then the signal should be amplified by amplifier exactly in the same amount as there is loss in the circuit and at the output i.e. $A\beta = 1$

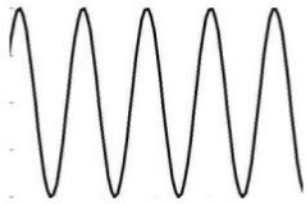


Fig. 3: Sustained Oscillations ($A\beta = 1$).

13.5 Condition for Self-Excitation

An oscillator circuit with Bipolar Junction Transistor in Common Emitter mode (CE-BJT) can be analyzed using hybrid parameters. A typical circuit of an oscillator with CE-BJT is shown in figure 4. Here we have considered three passive impedances X_1, X_2 and X_3 to make positive feedback (phase shift of $2n\pi$ in the feedback loop). The voltage drops across X_1 is feed back to the input of active device (so called feedback voltage V_f). The hybrid parameters for CE-BJT with standard notifications are taken as h_{ie}, h_{re}, h_{oe}, and h_{fe} are. The trans-conductance for CE-BJT is g_m while v_{be} is input voltage which is in actual voltage drop across base and emitter junction of CE-BJT. For a common emitter mode, the inverse voltage gain (h_{re}) and output admittance(h_{oe}) can be neglected i.e. $h_{re} = 0$ and $h_{oe} = 0$.

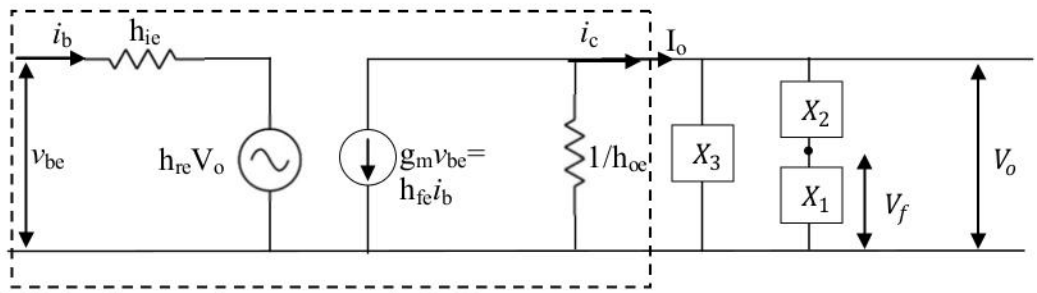


Fig 4. Hybrid equivalent circuit of a BJT resonant circuit.

We can write output voltage

$V_o = Output\ current \times Output\ impedance$

$V_o = I_o X_R$

$V_o = -g_m v_{be} X_R$

Here,

$$I_o = -g_m v_{be} = h_{fe} i_b$$

$$g_m = \frac{i_c}{v_{be}} = \frac{h_{fe}}{h_{ie}}$$

$$h_{fe} = \frac{i_c}{i_b}$$

$$X_R = \frac{X}{1 + Xh_{oe}} (\text{since } X \| \frac{1}{h_{oe}})$$

$$X = \frac{X_3(X_1 + X_2)}{X_1 + X_2 + X_3} \{\text{since } (X_1 + X_2) \| X_3\}$$

Voltage gain without feedback

$$A = \frac{V_o}{v_{be}} = -g_m X_R$$

And feedback ratio β is the part of output which is dropped on $X_1$and it can be expressed by

$$\beta = \frac{V_f}{V_o} = \frac{X_1}{X_1 + X_2}$$

So loop gain is

$$A\beta = -g_m X_R (\frac{X_1}{X_1 + X_2})$$

$$A\beta = -g_m \left[\frac{X_1 X_3}{(X_1 + X_2 + X_3) + X_3(X_1 + X_2)h_{oe}}\right]$$

In a practical oscillator circuits these passive elements are either capacitive or inductive. So,

$$X_1 = jx_1, X_2 = jx_2, X_3 = jx_3$$

$$A\beta = g_m \left[\frac{x_1 x_3}{j(x_1+x_2+x_3) - x_3(x_1+x_2)h_{oe}}\right] \quad (1)$$

For zero phase in the loop gain there should not be any imaginary part (in a complex number imaginary part related to phase)

$$x_1 + x_2 + x_3 = 0 \quad (2)$$

or $x_1 + x_3 = -x_2$

This equation implies that to make total phase zero at least one of the impedance should be of different nature from the other two i.e. if one is capacitive then other two passive elements will be inductive or vice versa.

By using this condition from equation 2 in equation 1, we can obtain

$$A\beta = g_m \left[\frac{x_1 x_3}{-x_3(x_1 + x_2)h_{oe}}\right]$$

$$A\beta = -\frac{g_m}{h_{oe}} \left[\frac{x_1}{x_1 + x_2}\right]$$

$$A\beta = \frac{g_m}{h_{oe}} \left[\frac{x_1}{x_3}\right]$$

According to Barkhausen Criterion loop gain $A\beta=1$

$$\frac{g_m}{h_{oe}} = +\frac{x_3}{x_1}$$

Indicates that x_1 and x_3 will be of same nature i.e. either inductive or capacitive.

3.6 LC Oscillators (Tuned Oscillator)

LC oscillators are widely used for radio frequencies as high as 500 MHz. These type of oscillators have the resonant circuit which is in actual combination of capacitor with a coil. In the resonant circuit a charged capacitor discharges over the coil and filled it with magnetic energy. This magnetic energy does not allow the change in current (not let reduce) so it forces the flow of current through capacitor. This charges the capacitor again while the coil become empty from magnetic energy and this process continues. This LC resonant circuit is also called as lumped or tank circuit. It gives the sinusoidal oscillations which can be taken out at load.

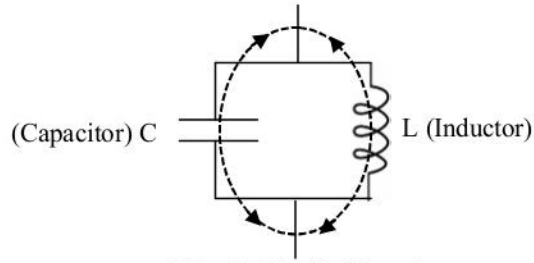


Fig.5: Tank Circuit.

Any oscillator circuit has the following essential parts to generate the oscillations

(i) Tuning circuit or oscillation generator
(ii) Feedback circuit
(iii) Amplifier unit

At the time of switching on of the power supply noise is created in active device(BJT or FET) which contains all frequencies. Frequency is selected by resonant circuit and further the level of oscillations is maintained by the amplifier. The frequency of LC resonant circuit is given by

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

The resonant circuit is attached to the collector of CE-BJT or drain of FET where the signal is already in opposite polarity to the input. Additional phase difference of 180 degrees is produced by the feedback network. Initially the loop gain should

be greater than unity to build up the oscillation to the desired level further as the desired level is achieved, it should be decreased to unity. This condition is achieved by proper biasing e.g. gate-leak biasing in a FET device [1].

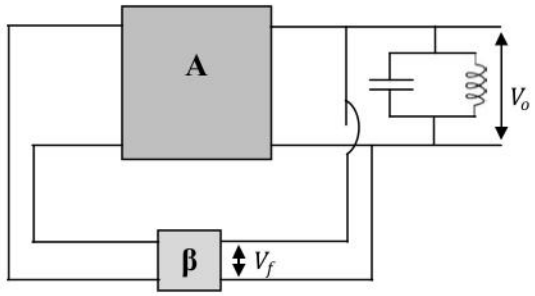


Fig.6: Block diagram of a typical LC oscillator

Generated oscillations are taken out at load either by capacitive or inductive coupling networks. If the load resistance is high, it may damp the oscillations in resonant circuit. To avoid such problem a buffer amplifier is used for impedance matching in between load and oscillator circuit.

There are two most common LC oscillators

(i)Hartley Oscillator (ii) Colpitts Oscillator

13.6.1 Hartley Oscillator

The circuit diagram is shown below with CE transistor amplifier.

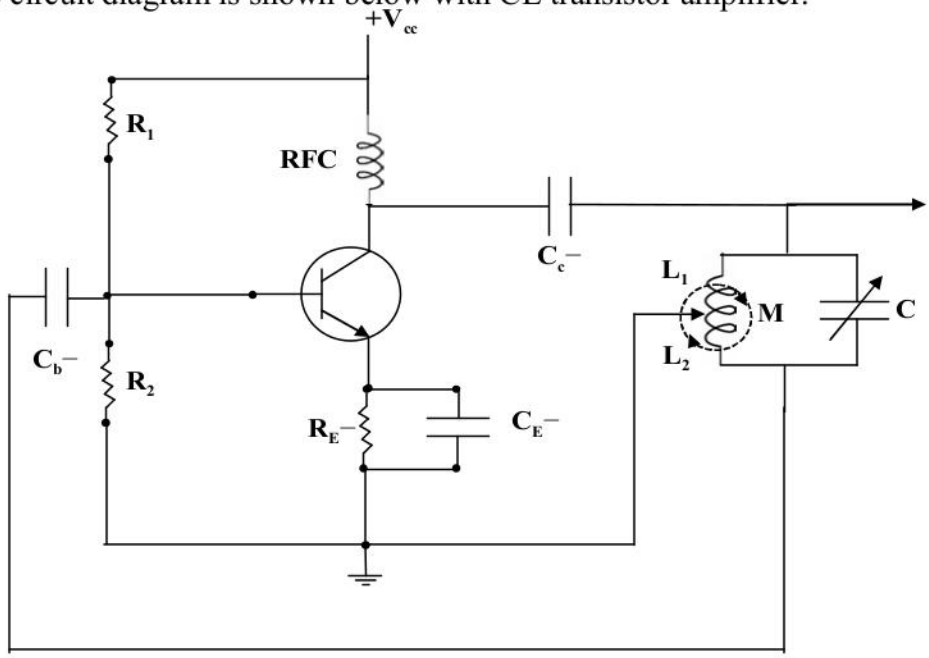


Fig.7: Circuit diagram of Hartley Oscillator with BJT

In the Hartley Oscillator a split tank inductor is used instead of single coil. The voltage generated across one part of the coil is used to give the feedback at input. Radio frequency choke (RFC) allows only dc and interrupt the ac to go back towards the battery. The ac voltage drop across the RFC is taken out at load. The presence of the coupling and blocking capacitor C_C and C_b respectively in Hartley oscillator is to isolate the dc current from the tank circuit. Here M is mutual inductance between the coils L_1 and L_2.

By using hybrid parameters, equivalent circuit of Hartley oscillator with CE-BJT can be drawn as shown in figure 8 (here the parameters h_{re} and h_{oe} are taken zero as for common emitter configurations these parameters are with negligible values). Here, the voltage drop across L_2 will be equal to voltage drop across h_{ie} and this voltage is working as feedback voltage (V_f).

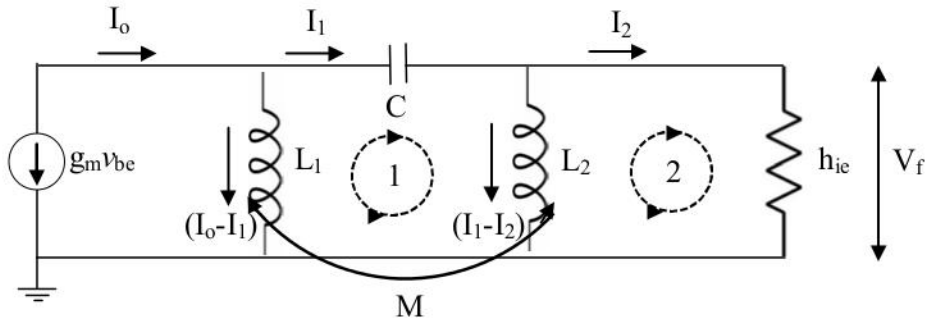


Fig.8: Equivalent circuit using hybrid parameters of Hartley Oscillator with CE-BJT.

Let the reactance applied by different reactive elements are:

$X_{L1} = j\omega L_1, X_{L2} = j\omega L_2, X_M = j\omega M$ and $X_C = \frac{1}{j\omega C}$

From loop 2

$$h_{ie}I_2 = (I_1 - I_2)X_{L2} - (I_O - I_1)X_M$$

$$I_2[h_{ie} + X_{L2}] + I_O X_M = I_1(X_{L2} + X_M) \quad (3)$$

From loop 1

$$X_{L1}(I_O - I_1) - X_M(I_1 - I_2) = X_C I_1 + X_{L2}(I_1 - I_2) - X_M(I_O - I_1)$$

$$I_O(X_{L1} + X_M) + I_2(X_{L2} + X_M) = I_1(X_C + X_{L1} + X_{L2} + 2X_M)$$

Let $$X = X_C + X_{L1} + X_{L2} + 2X_M \quad (4)$$

So, $$I_O(X_{L1} + X_M) + I_2(X_{L2} + X_M) = I_1 X \quad (5)$$

Now from equation 3 and 4 we can obtain

$$\frac{I_2}{I_O} = \frac{(X_{L1}+X_M)(X_{L2}+X_M)-X_M X}{X(X_{L2}+h_{ie})-(X_{L2}+X_M)^2} \quad (6)$$

Further gain of amplifier $A = \frac{V_o}{v_{be}}$ | $I_o = -g_m V_{be} = h_{fe} i_b$

Feedback ratio $\beta = \frac{V_f}{V_o}$

So, loop gain will be

$$A\beta = \frac{V_f}{v_{be}}$$

By using $V_f = h_{ie} I_2$ and $I = -g_m v_{be}$

$$A\beta = -g_m h_{ie} \frac{I_2}{I_o}$$

From the definition of Trans-conductance $h_{fe} = g_m h_{ie}$

$$A\beta = -h_{fe} \frac{I_2}{I_o}$$

$$A\beta = -h_{fe} \frac{(X_{L1} + X_M)(X_{L2}+ X_M) - X_M X}{X(X_{L2} + h_{ie}) - (X_{L2}+X_M)^2} \quad \text{(From equation 6)}$$

Putting the values of X_{L1}, X_{L2} and X_M.

We can consider $X = j\omega x$, here $x = (L_1 + L_2 + 2M - \frac{1}{\omega^2 C})$ (From equation 4)

$$A\beta = -h_{fe} \frac{(L_1+ M)(L_2+ M)-Mx}{\left(\frac{h_{ie}}{j\omega}\right)x+ L_2 X-(L_2+ M)^2} \quad (7)$$

There should be $2n\pi$ (n = 0, 1, 2, 3…..) phase in the loop for a oscillator so there should no imaginary part in the term of loop gain, hence $\frac{h_{ie}}{\omega} x = 0$ (since $h_{ie} \neq 0$ and $\omega \neq 0$),So, $x = 0$

Or $\left[(L_1 + L_2 + 2M) - \frac{1}{\omega^2 C}\right] = 0$ (From equation 4)

$$\omega = \frac{1}{\sqrt{(L_1 + L_2 + 2M)C}}$$

The resonant frequency of Hartley oscillator can be given by following

$$f = \frac{1}{2\pi\sqrt{(L_1 + L_2 + 2M)C}}$$

Also using $x = 0$ in equation 7

$$A\beta = h_{fe}\frac{(L_1 + M)}{(L_2 + M)}$$

Since for self-excitation and sustained oscillations

$$A\beta \geq 1$$

$$\boxed{h_{fe} \geq \frac{(L_2 + M)}{(L_1 + M)}}$$

It is condition on transistor and coils used in Hartley oscillator for self-excitation.

(Note: It can be noted for an oscillator which following the Barkhausen criterion i.e. $A\beta = 1$, the current passing through h_{ie} in loop 2 will be equal to i_b i.e. $I_2 = i_b$ if $A\beta = 1$).

13.6.2 Colpitts Oscillator

The Colpitts oscillator consists of split tank capacitor instead of split tank inductor, remaining circuit of Colpitts as similar as the Hartley. The voltage dropped across the capacitor C_2 is used for regenerative feedback for sustained oscillations. This type of oscillators are used to generate the sinusoidal frequency above 1 MHz.

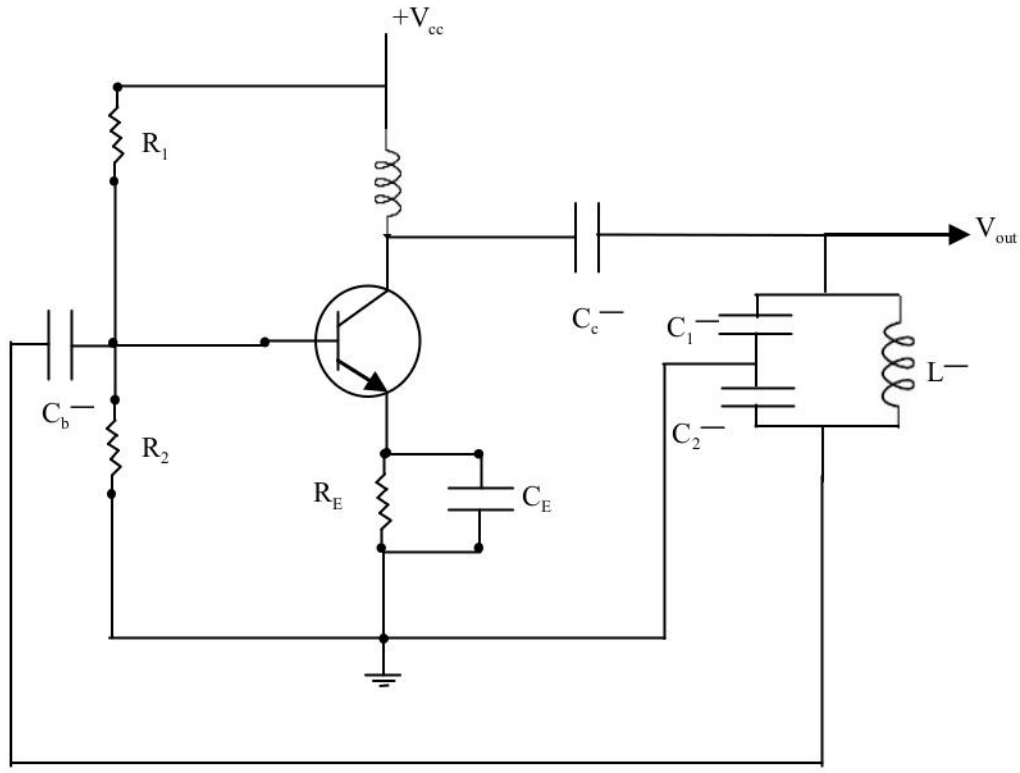


Fig. 9: Colpitts Oscillator.

By using hybrid parameters, equivalent circuit can be shown as below (similarly as in last section the parameters h_{re} and h_{oe} are taken zero as for common emitter configurations these parameters are with negligible values).Here, the voltage drop across C_2 will be equal to voltage drop across h_{ie}. This voltage is working as feedback voltage (V_f).

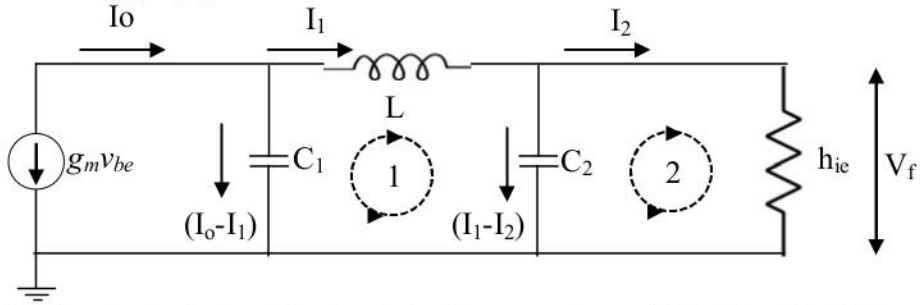


Fig. 10: Equivalent circuit using hybrid parameters of Hartley Oscillator with CE-BJT.

Let the reactance applied by different reactive elements are:

$$X_{C1} = \frac{1}{j\omega C_1}, X_{C2} = \frac{1}{j\omega C_2}, X_L = j\omega L$$

From loop 2 we can see that

$$\frac{1}{X} = \frac{1}{h_{ie}} + \frac{1}{X_{C2}}$$

$$X = \frac{h_{ie}X_{C2}}{X_{C2} + h_{ie}}$$

Further, from this loop

$$I_2 h_{ie} = (I_1 - I_2) X_{C2}$$

$$I_2(h_{ie} + X_{C2}) = I_1 X_{C2}$$

Or we can write $$I_1 = \frac{h_{ie}}{X} I_2 \quad (8)$$

Now from loop 1

$$(I_O - I_1)X_{C1} = I_1(X_L + X) \qquad (since\ X = h_{ie} \| X_{C2})$$

$$I_O = \frac{(X_L + X + X_{C1})}{X_{C1}} I_1 \quad (9)$$

$$I_O = \frac{(X_L + X + X_{C1})h_{ie}}{XX_{C1}} I_2$$

$$\frac{I_2}{I_O} = \frac{XX_{C1}}{(X_L + X + X_{C1})h_{ie}} \quad (10)$$

Further gain of amplifier $A = \frac{V_o}{v_{be}}$

Feedback ratio $\beta = \frac{V_f}{V_o}$

So loop gain $A\beta = \frac{V_f}{v_{be}}$

By using $V_f = h_{ie} I_2$ and $I_O = -g_m v_{be}$

$$A\beta = -g_m h_{ie} \frac{I_2}{I_O}$$

$$A\beta = -g_m \frac{XX_{C1}}{(X_L + X + X_{C1})}$$

By putting the values of $X_{C1}, X, X_{C2} and\ X_L$ we can achieve

$$A\beta = -g_m \frac{h_{ie}}{(1-\omega^2 LC_1)+j\omega(C_2+C_1-\omega^2 LC_1C_2)} \quad (11)$$

For positive feedback the phase difference should be zero

$$\omega(C_2 + C_1 - \omega^2 LC_1C_2) = 0$$

So we can write one of the following

$$\omega = \frac{1}{\sqrt{LC}}$$

The resonant frequency of Colpitts oscillator

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

Where $\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$

Further also using the value of resonant frequency in equation 11

$$A\beta = -g_m h_{ie} \frac{1}{(1 - \omega^2 LC_1)}$$

$$A\beta = g_m h_{ie} \frac{C_2}{C_1}$$

Since for self-excitation and sustained oscillations

$$A\beta \geq 1$$

$$g_m h_{ie} \geq \frac{C_1}{C_2}$$

So condition for self-excitation of Colpitts Oscillator

$$h_{fe} \geq \frac{C_1}{C_2}$$

13.7 Self Learning Exercise-I

Q.1 What are the oscillators?

Q.2 What is difference between an ac generator and a oscillator?

Q.3 What are the requiremnts to obtain sustained oscillatoins from an oscillator?

Q.4 Why the positive feedback is necessary for oscillators ?

Q.5 Why an LC tank circuit without excitation by amplifier does not produce sustain oscillation?

13.8 RC Oscillators

The RC oscillators are special class of oscillators which are used to produce low frequency normally in the range of audio frequencies. The LC circuits have several limitations in this range. RC oscillators are compatible with IC technology, since RC network is the only feasible on a chip. It is very difficult to make an inductance, that too of high value in an integrated circuit as required for generation of low frequencies. Therefore, RC oscillators are becoming increasingly popular.

There are two most common RC oscillators widely used

(i) RC phase shift oscillator
(ii) Wien bridge oscillator

13.8.1 RC Phase Shift Oscillator

The following parts are essential for this type of oscillator circuit:

(i) Feedback circuit
(ii) RC phase shift network
(iii) Amplifier unit

During switch on of the active device due to random motion of electrons every frequency is generated and a particular desired frequency is selected by phase shift circuit which transmitted by this network without attenuation. The feedback circuit took appropriate ratio of signal and feed in the Phase shift network which produce 180° extra phase while the signal had already phase of 180°due to active device (CE or FET). So, the signal is feed back to the input of active device in the same phase which was amplified by it, which fulfil the essential condition to produce sustain oscillations. The frequency of oscillations by RC phase shift

oscillator is given by

$$f = \frac{1}{2\pi RC\sqrt{10}}$$

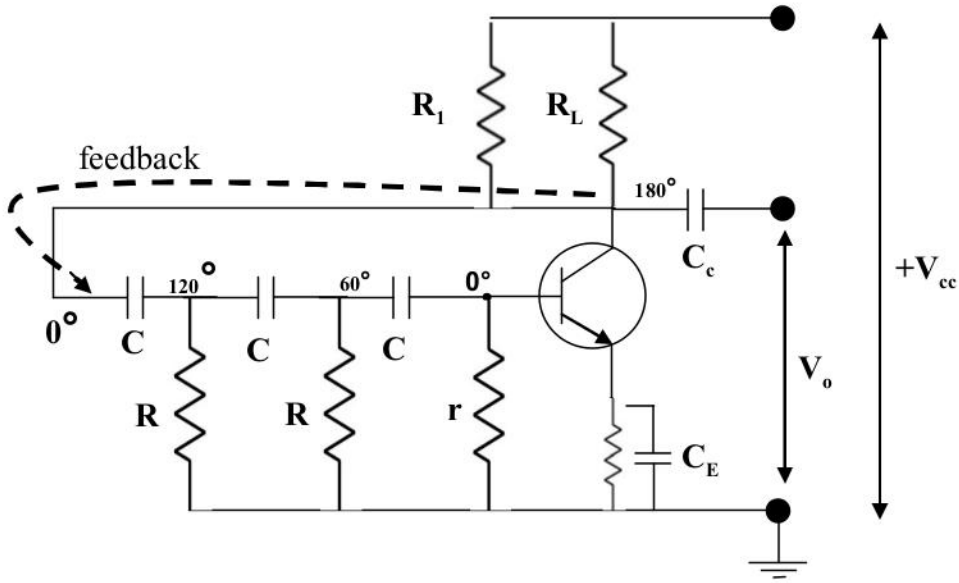


Fig. 11: A typical RC phase shift oscillator using CE- BJT [2].

The reactance by capacitor is given by

$$X_c = \frac{1}{\omega C}$$

And resultant impedance by one section of R and C

$$Z = \sqrt{R^2 + X_c^2}$$

And the phase angle produce by one section

$$\emptyset = \tan^{-1}\frac{X_c}{R}$$

The values of R and C have been selected in the way that at the required frequency the phase difference created between output voltage and the input voltage should be 60°. Then the phase angle between each successive RC section increases by another 60° giving a phase difference between the input and output of 180°.

Any single RC section can produce a phase change of up to 90°, so logically only two such sections are enough to produce required extra phase of 180°. However, using two filters with each producing a 90°phase shift would mean that, any drift in frequency would have effect on the 180° phase shift produced so frequency stability would be poor [3].

Alternatively, with three filters producing 60° each, or four filters providing 45° each, to make up the required 180°, frequency stability will be much better. However, adding more sections, it will increase the loading effect and hence will added complexity [3]. Here, loading effect refers to the adding of resistance from one RC section to the another RC section since, the resistors from consecutive sections will appear in parallel combination to each other. So, the four-filter 45° phase shift model is rarely used, even though using four filters does give slightly better frequency stability. Also the poor frequency stability of the two-filter model makes the three-filter oscillator the most usual choice.

RC oscillators with CE-BJT are easy to fabricate on single IC chip but they have following limitations:

a) Practically there is a wide difference (around 7%) between the calculated frequency value and the actual frequency produced.
b) The waveform amplitude is generally not well stabilised so a good wave shape cannot be ensured.
c) It is difficult to design RC oscillator with variable frequency form as this would involve clouding of either 3 variable capacitors or 3 large variable resistors, and such components are not easily available.

In the phase shift network, three sections with resistance R and capacitor C are selected. Each section provides the phase change of 60°. The input resistance of CE-BJT (h_{ie}) is come in the series of the resistance in the third section. So, the value of resistance in the third section (r) is selected in the way that the total resistance is equal to R (i.e. $R = r + h_{ie}$). The voltage drop on h_{ie} is feedback to the amplifier (CE-BJT). The current flow through h_{ie}(i.e. in the third section) can be taken equal to base current i_b if the oscillator is working with *Barkhausen* criterion (Aβ=1).

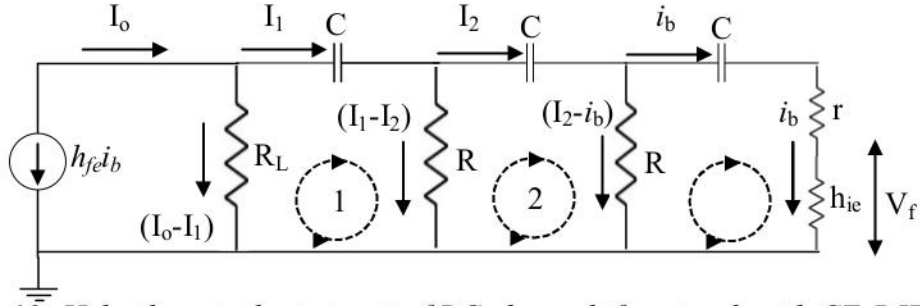


Fig. 12: Hybrid equivalent circuit of RC phase shift network with CE-BJT (neglecting h_{re} and h_{oe}).

From loop 1 $-R_L I_0 + R_L I_1 + X_C I_1 + RI_1 - RI_2 = 0$

$$\text{Or} \quad I_1(R_L + X_C + R) + I_2(-R) + i_b(R_L h_{fe}) = 0 \tag{12}$$

Here $X_C = \frac{1}{j\omega C}$ is the reactance applied by capacitor C on the ac current and

$$I_0 = -h_{fe} i_b .$$

Similarly, from loop 2

$$I_1(-R) + I_2(X_C + 2R) + i_b(-R) = 0 \tag{13}$$

And from loop 3

$$I_1\,(0) + I_2\,(-R) + i_b\,(r + h_{ie} + X_C + R) = 0$$

$$\text{Or} \quad I_1\,(0) + I_2\,(-R) + i_b\,(R + X_C + R) = 0$$

$$\text{Or} \quad I_1\,(0) + I_2\,(-R) + i_b\,(2R + X_C) = 0 \tag{14}$$

Rearranging equations 9, 10 and 11 in the matrix form

$$\begin{bmatrix} R_L + X_C + R & -R & R_L h_{fe} \\ -R & 2R + X_C & -R \\ 0 & -R & 2R + X_C \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ i_b \end{bmatrix} = 0$$

Since $I_1 \neq 0;\ I_2 \neq 0\ and\ i_b \neq 0$. So,

$$\begin{bmatrix} R_L + X_C + R & -R & R_L h_{fe} \\ -R & 2R + X_C & -R \\ 0 & -R & 2R + X_C \end{bmatrix} = 0$$

Solving this matrix, we can obtain

$$X_C^3 + X_C^2(R_L + 5R) + X_C(4RR_L + 6R^2) + (3 + h_{fe})R^2 R_L + R^3 = 0$$

$$\text{Or} \quad X_C^2(R_L + 5R) + X_C(4RR_L + 6R^2 + X_C^2) + (3 + h_{fe})R^2 R_L + R^3 = 0$$

Putting the value of $X_C = \frac{1}{j\omega C}$ in the above equation and rearranging it we can obtain

$$-\frac{1}{\omega^2 C^2}(R_L + 5R) + \frac{1}{j\omega C}\left(4RR_L + 6R^2 - \frac{1}{\omega^2 C^2}\right) + (3 + h_{fe})R^2 R_L + R^3 = 0$$

According to *Barkhausen* criterion the phase difference in the feedback loop should be $2n\pi$ (n = 0, 1, 2, 3…..). So, there should not be any imaginary term

$$4RR_L + 6R^2 - \frac{1}{\omega^2 C^2} = 0$$

$$\omega = \frac{1}{RC\sqrt{6 + 4\frac{R_L}{R}}}$$

In the RC phase shift oscillator, the value of R is selected mostly equal to the load resistance R_L=R. Hence,

$$\omega = \frac{1}{RC\sqrt{10}}$$

The resonant frequency of RC phase shift oscillator

$$\boxed{f = \frac{1}{2\pi RC\sqrt{10}}}$$

Using this value of resonant frequency and R_L=R in equation 15 we can obtain

$$h_{fe} = 56$$

So for the self-excitation in RC phase shift oscillator if R_L=R then the current gain of common emitter configuration should be

$$\boxed{h_{fe} \geq 56}$$

13.8.2 Wien-Bridge Oscillator

Wien bridge oscillator is used to produce the frequencies in the range of 10 Hz to 1 MHz. Wien bridge oscillator consists of two stage amplifier to produce 360° phase; this is called lead-lag circuit. RC network just provide appropriate feedback and select the desired frequency for sinusoidal(see figure13). It is used in all commercial audio generators. The frequency generated by Wien bridge oscillator can be given as below

$$f = \frac{1}{2\pi\sqrt{R_1C_1R_2C_2}}$$

There is ganging of series and parallel combination of R and C. The working of *Wien* bridge oscillator can be understood by combination of low pass and high pass RC filters.

(i) Series combination of R and C act as high pass filter and give zero signal for low frequencies. It creates high reactance for lower range of frequencies.

(ii) Parallel combination of R and C act as low pass filter and provide short path to the high frequencies.

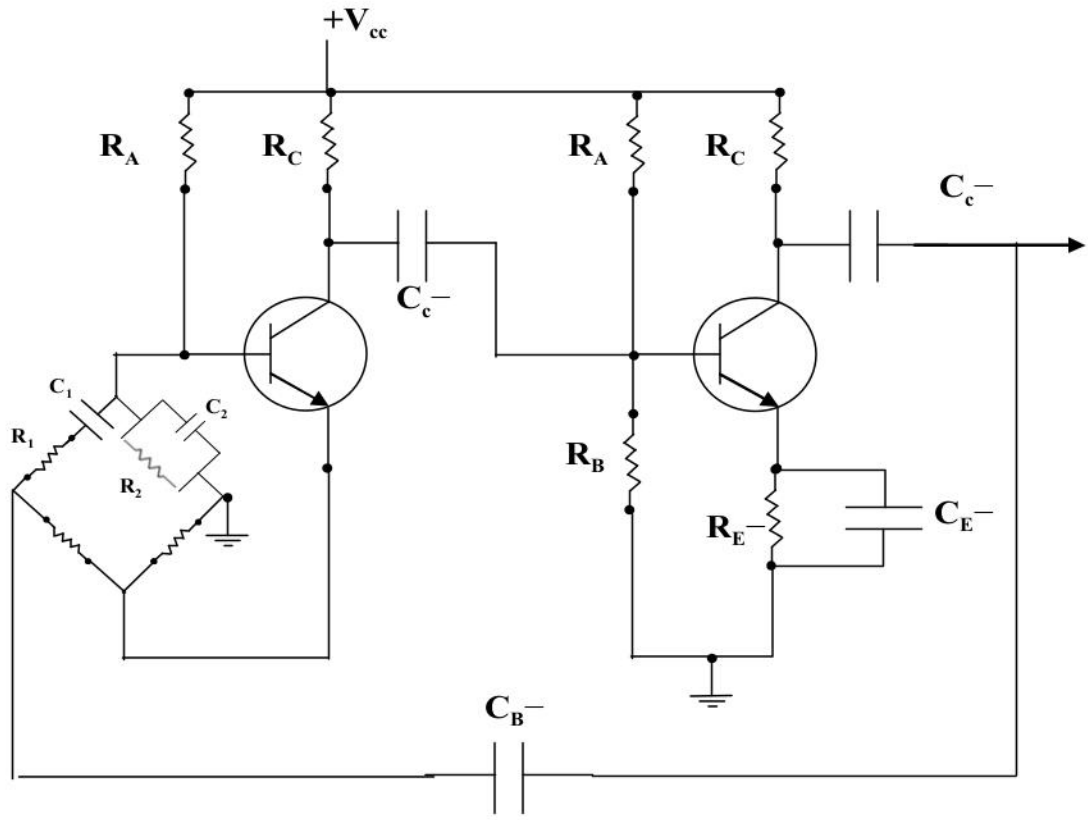


Fig. 13: Wien bridge oscillator.

Only a band of frequencies around resonant frequency can be pass through this combination of low and high pass filters. The resonant frequency can be defined here by saying a frequency for which the phase change by RC network is zero. The output voltage is maximum for resonant frequency and it is one third of the input voltage if the value of resistance $R_1 = R_2$ is taken as the real part of reactance by capacitor $C_1 = C_2$.

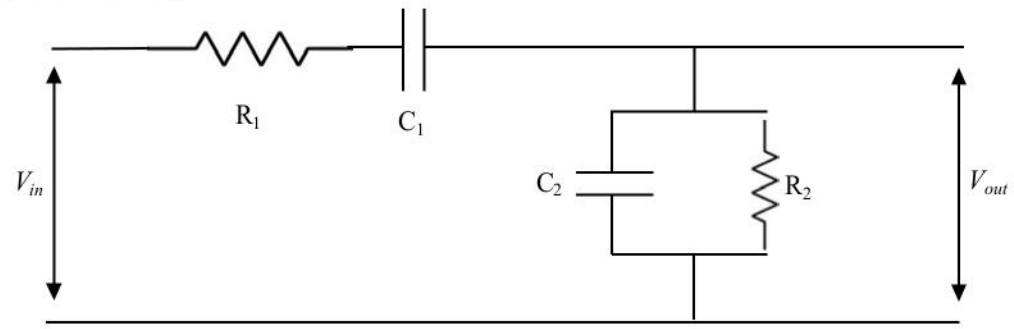


Fig. 14: RC network in Wien bridge oscillator.

For the *Wien* bridge RC network, if we select $R_1 = R_2$ and $C_1 = C_2$ we can establish

$$V_{out} = \frac{X}{(\mathrm{R} + \mathrm{X}_C + X)} V_{in}$$

Here $X = \frac{R\mathrm{X}_C}{(R+\mathrm{X}_C)}$ and $\mathrm{X}_C = \frac{1}{j\omega C}$

$$V_{out} = \frac{R\mathrm{X}_C}{R^2 + X_C^2 + 3R\mathrm{X}_C} V_{in}$$

$$V_{out} = \frac{1}{3 - j\left(\frac{1}{\omega CR} - \omega CR\right)} V_{in}$$

$$\left|\frac{V_{out}}{V_{in}}\right| = \frac{1}{\left[9 + \left(\frac{1}{\omega CR} - \omega CR\right)^2\right]^{\frac{1}{2}}}$$

And change in phase between input and output voltage by RC network

$$\phi = \tan^{-1} \frac{\left(\frac{1}{\omega CR} - \omega CR\right)}{3}$$

We can see that for the maximum value of $\frac{V_{out}}{V_{in}}$, $R = \frac{1}{\omega C}$

So, $V_{out} = \frac{1}{3} V_{in}$ and $\phi = 0$

and resonant frequency of *Wien* bridge oscillator

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

13.9 Crystal Oscillator

A crystal oscillator in principle is a tuned oscillator. A piezoelectric crystal is used as a resonant tank circuit. Piezoelectric crystals exhibit a unique property that these crystals mechanically expend or squeeze when a voltage is applied. Conversely if mechanical force is applied to expend or squeeze the crystal a voltage is generated. It is due to their internal crystal geometry. Paul Langevin first investigated quartz resonators for use in sonar during World War I [4]. The first crystal-controlled oscillator, using a crystal of Rochelle salt, was built in 1917 and patented in 1918 by Alexander M. Nicholson at Bell Telephone Laboratories [5]. Cady built the first quartz crystal oscillator in 1921 [6]. Quartz crystal oscillators were developed for high-stability frequency references.

Due to any change in physical parameters for active device the frequency generated by oscillators can diverse. In Communications a particular frequency is offered to a station for broadcasting but frequency diversion can create big trouble and noise. So, the frequency stability is very important in communication. These type of oscillators are used to provide high stability in frequency, especially in communications and digital clocks. The Q factor of crystal oscillator can be achieved as high as 10^6 in comparison to LC tank circuit which possess Q factor around 100.

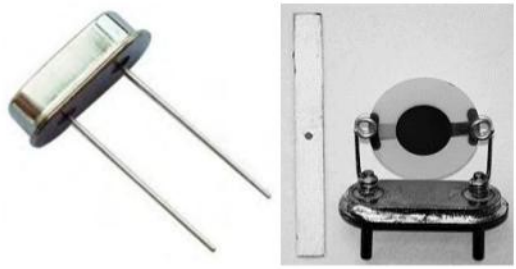


Fig. 15: A typical crystal oscillator [7].

A piezoelectric crystal is suitably cut and then mounted between two metal electrodes. It acts as capacitor (lets this capacitance as C_p) since the material of crystal act as a dielectric material between two electrodes. If a ac voltage is applied across these metal electrode plates, the crystal start vibration with the frequency of applied ac voltage. The vibration of crystal has its own natural frequency. If the frequency of ac voltage match with the natural frequency of crystal vibrations, then the amplitude of vibration reaches to its maximum value. This electro-mechanical resonance can be represented as equivalent electrical LCR series resonant circuit with capacitance Cs.

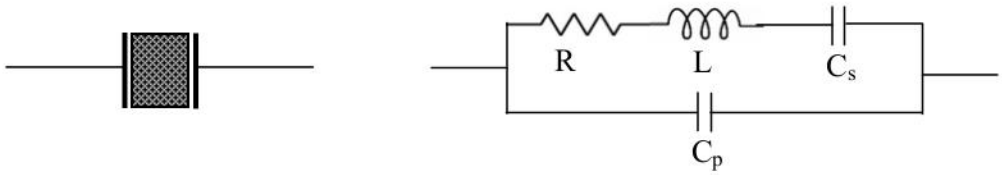


Fig. 16: Equivalent resonant circuit for a crystal oscillator.

The crystal possesses two frequencies, one is resonant frequency (f_s) due to series combination of L, R and Cs. The other one is f_p which is slightly higher than the *fs* is due to the resonance between series branch and parallel capacitance C_m. The

crystal as a whole behaves as an inductor only in the oscillator circuit for the frequencies lies between f_s and f_p.

Equivalent impedances of resonant circuit

$$\frac{1}{j\omega X} = \frac{1}{R + j\omega L + \frac{1}{j\omega C_s}} + \frac{1}{1/j\omega C_p}$$

Especially in quartz crystal damping is very low we can neglect so we can write

$$j\omega X = \frac{1 - \omega^2 LC_s}{j\omega C_s + j\omega C_p(1 - \omega^2 LC_s)}$$

$$X = \frac{1 - \omega^2 LC_s}{\omega^2\{C_s + C_p(1 - \omega^2 LC_s)\}}$$

We obtain two resonant frequencies

(i) Series Resonance: When the reactance is zero when there is resonance in LCR series combination i.e. $X_s = 0$

$$X_s = \frac{1 - \omega^2 LC_s}{\omega^2\{C_s + C_p(1 - \omega^2 LC_s)\}} = 0$$

$$\text{So, } 1 - \omega^2 LC_s = 0$$

$$f_s = \frac{1}{2\pi\sqrt{LC_s}}$$

(ii) Parallel resonance: When the total reactance is zero in parallel combination of L and C so $1/X_p = 0$

$$\frac{1}{X_p} = \frac{\omega^2\{C_s + C_p(1 - \omega^2 LC_s)\}}{(1 - \omega^2 LC_s)} = 0$$

$$\omega^2\{C_s + C_p(1 - \omega^2 LC_s)\} = 0$$

$$f_p = \frac{1}{2\pi}\sqrt{\frac{(C_s + C_p)}{LC_sC_p}}$$

In practice $C_p \gg C_s$

So, $f_p \approx f_s$

It gives very high frequency stability.

13.10 Self-Learning Exercise –II

Q.1 What are the advantages of RC oscillators over LC oscillators?

Q.2 Which range of frequencies are produced by RC oscillators?

Q.3 What is structural difference between RC phase shift and *Wien* bridge oscillator?

Q.4 Why only three sections are used in RC phase shift oscillators i.e. neither less nor more?

Q.5 What are the advantages of crystal oscillator over the RC and LC oscilattors?

13.11 Summary

Oscillators are circuits that produce an output waveform without an external signal source. The key to oscillator operation is positive feedback. A positive feedback network produces a feedback voltage that is in phase with the input signal.

Oscillators are characterized broadly in two types sinusoidal and non-sinusoidal. LC, RC and crystal oscillators are widely used to provide sinusoidal waves. LC oscillators produce essential phase of 360° in two parts; 180° by active device and other 180° by the reactance either capacitive or inductive. Hartley and Colpitts are most common LC oscillators to produce sinusoidal radio frequencies.

RC oscillators have the advantages that they can be easily fabricated on the chip and these are feasible for IC technology. These oscillators are mostly used to generate the frequencies lies in audio frequency range. Two most common RC oscillators in use are RC phase shift oscillator and *Wien* bridge oscillator. RC phase shift network oscillator produce essential 360° phase in two parts 180° by active device itself and next 180° by three section of R and C. The values of R and C are selected in such a way that each section produce 60° phase difference. *Wien* bridge consist lead-lag network with two active devices or operational amplifier working in non-inverting mode to produce 360° phase.

Crystal oscillators are the option to provide better frequency stability with high Q factor and high temperature stability. These oscillators work on the principle of piezoelectricity. These oscillators are used to produce radio frequencies.

13.12 Glossary

Frequency: The number of cycles per second.

Low pass filter: A low-pass filter is a filter that passes signals with a frequency lower than a certain cut-off frequency. It produces high reactance on signals with frequencies higher than the cut-off frequency.

High pass filter: A high-pass filter is a filter that passes signals with a frequency higher than a certain cut-off frequency. It produces short path for signals with frequencies higher than the cut-off frequency

Communication and Broadcasting:Broadcasting is the supply of audio and/or video content to a discrete audience via any electronic mass communications medium, but typically one using the electromagnetic spectrum (radio waves).

13.13 Answers to Self-Learning Exercises

Answers to Self-Learning Exercise-I

Ans.1: An electronic circuit which generates the alternating current (ac) known as oscillator.

Ans.2: There is difference between alternator or commonly known as generator which generate ac current. It consists magnetic poles and an armature which rotate inside the magnet and generate the ac power.

Ans.3: There are three condition to be followed to achieve sustain oscillations

The signal sent back to the amplifier should be in the same phase, hence the positive type of feedback. Initially the loop gain should be greater than unity i.e. $A\beta > 1$. Since, the amplitude level of oscillations is build up.

As the amplitude of oscillations is reached to the desired level then the signal should be amplified by amplifier exactly in the same amount as there is loss in the circuit and at the output i.e. $A\beta = 1$

Ans.4: Amplifier provides sustained oscillations at output without any signal at input with such type of feedback.

Ans.5: LC tank circuit loses the energy due to resistance of circuit

Answers to Self-Learning Exercise-II

Ans.1: RC oscillators are feasible with IC technology.

Ans.2: Frequencies in audio range.

Ans.3: RC phase shift oscilltor has three section and produce the 180 degree phase change while in the *Wien* bridge oscillator there is no phase change by RC network.

Ans.4: More than three section will give loading effect while two sections will give poor stability.

Ans.5: Crystal oscillators provide high frequency stability.

13.14 Exercise

Q.1 What do you understand by an electronic oscillator and what are its significances?

Q.2 How the oscillation starts in a oscillator?

Q.3 Why the oscillators cane not replaced by generators?

Q.4 What is Barkhausen criterion?

Q.5 What is role of amplifier in a oscillator circuit?

Q.6 What is piezoelectric effect?

Q.7 What are the limitations of LC oscillators?

Q.8 What are the limitations of RC oscillators?

Q.9 What are the advantages of RC oscillators?

Q.10 Why the crystal oscillator posseses two resonant frequencies?

Q.11 What are the adavantages of Crystal oscillator?

Q.12 What will the frequency of oscillations for *Wien* bridge oscillator if R=220 KΩ and C=250 pF?

Q.13 A Hartley Oscillator is designed using the components: L_1 = 250 mH, L_2 = 290 mH, M = -25 mH, C = 100 pF. What will the frequency of resonace of this oscillator?

Q.14 A Colpitts Oscillator with resonant frequency of 16 MHz have inductor of 10 µH. What will be the value of capacitors?

Q.15 Calculate the resonant frequency for RC phase shift Oscillator if R = 100 KΩ, C=100 pF

13.15 Answers to Exercise

Ans.1: An electronic circuit which generates the alternating current (ac) known as oscillator. These electronic circuits are important in broadcasting and communications, testing laboratory instruments etc.

Ans.2: The power is switched on to the circuit, a noise voltage is generated in the active device (BJT/FET) which may be due to random motion of electrons in the circuit. This noise voltage contains almost every sinusoidal frequency.

Ans.3: Generators can not generate oscillatiosn with high frequency.

Ans.4: The overall gain of the amplifier with feedback can be given as below

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

If the feedback signal is in such a way that the loop-gain $A\beta = 1$ so in this case, $|1 - A\beta| = 0$ and $|A_f| = \infty$. Amplifier provides sustained oscillations at output without any signal at input with such type of feedback. This condition is most commonly known as *Barkhausen Criterion* for sustained oscillations.

Ans.5: It keep maintaining the amplitude in the tank circuit.

Ans.6: Piezoelectric crystals exhibit a unique property that these crystal mechanically expend or squeeze when a voltage is applied. Conversely if mechanically forced is applied to expend or squeeze the crystal a voltage is generated. This property known as piezoelectricity.

Ans.7: LC oscillators can not be fabriacted on a chip and not feasible with IC technology. Moreover, to generate the curreny with low frequency the inductors and capacitiors are requitred with too thier too high values which are not readily availabel.

Ans.8: RC oscillators with CE-BJT are easy to fabricate on single IC chip still have following limitations:

a) Practically there a wide difference is found between the calculated frequency value and the actual frequency produced.

b) The waveform amplitude is generally not well stabilised so a good wave shape cannot be ensured.

c) It is difficult to design RC oscillator with variable frequency form as this would involve clouding of either 3 variable capacitors or 3 large variable resistors, and such components are not easily available.

Ans.9: RC oscillators are commanly used to generate sudio frequency and these oscillators can be fabriacted easily on the chip.

Ans.10: One due to series resonance and the second due to parallel resonace.

Ans.11: Crystal oscillator provide high frequency stability.

Ans.12: 2.89 KHz

Ans.13: 22.74 KHz

Ans.14: 505 pF and 10.1 pF.

Ans.15: 5 KHz

References and Suggested Readings

1. Basic Electronics and Linear Circuits b7y NN Bhargava, D C Kulshreshtha, and SC Gupta; published by McGrawHill Education.
2. www.Electronic-tutorials.ws
3. http://www.learnabout-electronics.org/Oscillators/osc31.php
4. Nicholson, Alexander M. Generating and transmitting electric currents U.S. Patent 2,212,845, filed April 10, 1918, granted August 27, 1940
5. *Bottom, Virgil E. (1981).* "A history of the quartz crystal industry in the USA". *Proc. 35th Frequency Control Symp. IEEE.*
6. *Marrison, Warren (1948).* "The Evolution of the Quartz Crystal Clock". *Bell System Technical Journal (AT&T) 27: 510–588.*
7. https://en.wikipedia.org/wiki/Crystal_oscillator
8. Electronic Devices and Circuit Theory by R.L. Boylestad and L. Nashelsky; published by Pearson Education.
9. Electronic Devices and Circuits by David A. Bellpublished by Oxford ; .University Press

UNIT-14
Number System

Structure of the Unit

14.0 Objectives

14.1 Introduction

14.2 Decimal Number System

14.3 Binary System

- 14.3.1 Binary to decimal Conversion
- 14.3.2 Decimal to Binary Conversion
- 14.3.3 Binary Arithmetic
- 14.3.4 Complement Representation

14.4 Octal Number System

- 14.4.1 Octal to Decimal Conversion
- 14.4.2 Decimal to Octal Conversion
- 14.4.3 Octal to Binary Conversion
- 14.4.4 Binary to Octal Conversion

14.5 Hexadecimal Number System

- 14.5.1 Hexadecimal to Decimal Conversion
- 14.5.2 Decimal to Hexadecimal Conversion
- 14.5.3 Conversion of Hexadecimal number into binary number
- 14.5.4 Conversion of binary number into hexadecimal number

14.6 Illustrative Examples

14.7 Self Learning Exercise

14.8 Summary

14.9 Glossary

4.10 Answers to Self Learning Exercise

14.11 Exercise

14.12 Answers to Exercise

References and Suggested Reading

14.0 Objectives

Students will able to learn about different number system which are used in digital electronics and devices.

14.1 Introduction

The electronic circuits have wide application in our daily life. Basically we deal with two types of signals in electronics, analog and digital. Analog systems deal with variables which are continuous in nature and digital variables which can assume by discrete values.

To deal which digital system the knowledge of number system is very important as these systems are based on digits. Switching circuits have two positions ON and OFF which can be represented by digits 1 and 0 respectively. Therefore these two digits are very important in digital system and number system based on these two digits is called binary number system. Similarly there are others radices, i.e. number bases, that a numbering system can use, are octal (base-8), decimal (base 10) and hexadecimal (base 16).

In this unit, we shall discuss different number systems viz. binary, octal, decimal and hexadecimal number system and their inter conversion.

14.2 Decimal Number System

This is well known number system in which there are ten digits 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9. Its radix or base is 10. A collection of these digits makes a number while in general has two parts: an integer and a fraction set apart by a decimal point (.). For example 1254.45, the integral number is 1254 and the fraction is 45. In the number on the left of decimal as we move towards the left the weightage increases by positive power of ten in order i.e. the first digit on the left of decimal has a weightage of 10^0, the second digit has a weightage of 10^1, third one has a weightage of 10^2, fourth one has a weightage of 10^3 etc. In the fractional number on the right of decimal the weightage of each digit decreases by negative power of 10 in order i.e. on the right of decimal the first digit has a weightage of 10^{-1}, the

second 10^{-2}, third 10^{-3} etc. In general we can express any number as $d_m d_{m-1} .. d_2 d_1 d_0$, (decimal point) $d_{-1}, d_{-2}, d_{-3}, \ldots d_{-n}$. So number in radix or base 'x' is written and interpreted as

$$d_m (x)^m + d_{m-1} (x)^{m-1} + ... + d_1 (x)^1 + d_0 (x)^0 + d_{-1} (x)^{-1} + d^{-2} (x)_{-2} \ldots + d_{-n} (x)^{-n}$$

↑ ↑

MSB LSB

Here (m+1) digits are to the left of decimal point and n digits are to the right of decimal point. The digit d_m at $(m+1)^{th}$ position to the left of decimal point has highest positional value (weightage) so this called most significant Bit (MSB). The digit d_{-n} at n^{th} position to the right of decimal point has minimum positional value (Weightage) so it is called least significant bit (LSB).

For example, a number 5917.421 in decimal number system can be interpreted as

$$5 (10)^3 + 9 (10)^2 + 1(10)^1 + 7 (10)^0 + 4 (10)^{-1} + 2 (10)^{-2} + 1 (10)^{-3}.$$

14.3 Binary System

In 1854 George Boole proposed a new algebra which was totally different from the usual one. It is a mathematical system of logic. In this system there are only two variables, 'true and false' or ON and OFF'. There variables are numerically represented by 1 and 0 respectively. So there are only two digits (0 and 1) in this number system and it is known as Binary number system. It uses only two digits to represent any number. It has a radix or base of 2. Thus a number 1101.101 in a binary number system world be interpreted as

$$1 (2)^3 + 1 (2)^2 + 0 (2)^1 + 1 (2)^0 + 1 (2)^{-1} + 0 (2)^{-2} + 1 (2)^{-3}$$

The decimal numbers and their corresponding binary number are given below in table 14.1.

Table 14.1

Decimal Number	Binary Number
0	0000
1	0001
2	0010
3	0011

4	0100
5	0101
6	0110
7	0111
8	1000
9	1001
10	1010
11	1011
12	1100
13	1101
14	1110
15	1111

14.3.1 Binary to Decimal Conversion

Some we need to convert binary number into decimal number. In the binary system also a binary point separates the integral and fractional parts of a number just like the decimal system. The weightage of positions on the left of the binary point increases by positive power of 2 in order i.e. the first number on the left of binary point has a weightage of 2^0, the second number has a weightage of 2^1, third number has 2^2 etc. The weightage of positions on the right of binary point in a binary number decreases by negative power of 2 in order i.e. the first number on the right of binary point has a weightage of 2^{-1}, second one 2^{-2}, third one 2^{-3} etc. For example we can convert binary number 1101.101 in decimal number like this

$$1101.101 = 1\times2^3+ 1\times2^2+0\times2^1+1\times2^0+1\times2^{-1}+0\times2^{-2}+1\times2^{-3}$$

whose value in decimal system will be

$$(1101.101)_2 = 8 + 04 + 0 + 1 + \frac{1}{2} + 0 + \frac{1}{8}$$

$$(1101.10)_2 = (13.625)_{10}$$

14.3.2 Decimal to Binary Conversion

A decimal number may consist of two parts integer part and fraction part. To convert the integer part of decimal number into the binary number it is divided by

2 in succession and each time the reminder is noted. The last remainder had the maximum weightage (MSB) and first one has the minimum weightage (LSB). To convert a fractional decimal number into the binary number, the fractional term is multiplied by 1, which gives a whole number and a fraction. The whole number may either be 0 or 1, this represents MSB of fraction part. The fractional part is again multiplied by 2. This process continues till the fractions part becomes zero. Thus successive number gives decimal bits starting from MSB to LSB. For example a decimal number 25.85 may convert in binary number as follow:

Conversion of integer part (whole Number) 25:



$\frac{25}{2}$	=	12	1	[LSB]
$\frac{12}{2}$	=	6	0	
$\frac{6}{2}$	=	3	0	
$\frac{3}{2}$	=	1	1	
$\frac{1}{2}$	=	0	1	[MSB]

The binary equivalent of 25 is obtained by writing the remainders from bottom to top (i.e. MSB to LSB). So $(25)_{10} = (11001)_2$

Conversion of fractional part 0.85:



Product	**Whole number**			
0.85× 2	=	1.7	1	[MSB]
0.7×2	=	1.4	1	
0.4×2	=	0.7	0	
0.8×2	=	1.6	1	
0.6×2	=	1.2	1	
0.2×2	=	0.4	0	[LSB]

So binary equivalent of fractional part (0.85) is

$(0.85)_{10} = (110110)_2$

So $(25.85)_{10} = (11001.110110)_2$

14.3.3 Binary Arithmetic

Binary numbers can be added, subtracted, multiplied and divided. Similar to that of decimal system.

(a)Binary addition : Rules for binary addition are as follows:

(i) 0+0 =0

(ii) 0 + 1= 1

(iii) 1 + 0 = 1

(iv) 1+1 = 10

1011	1101	10101
+ 101	+0111	+1010
10000	10100	11111

(b)Binary Subtraction : The rules for binary subtraction are as follows:

(i) 0 – 0 = 0

(ii) 1 – 0 = 1

(iii) 1 – 1 = 0

(iv) 10-1 = 1

Examples : (i) Subtract 001 from 101

```
  1 0 1
- 0 0 1
-------
  1 0 0
```

(ii) Subtract 001 from 110

```
  110
- 001
-----
  101
```

(iii) Subtract 101.10 from 1011.11

```
 1011.11
- 101.10
--------
 0110.01
```

(c) Binary multiplication: Rules for binary multiplication are as follow:

(i) $0 \times 0 = 0$

(ii) $0 \times 0 = 0$

(iii) $1 \times 0 = 0$

(iv) $1 \times 1 = 1$

Example (i) Multiply 1101 by 1010 **(ii)** Multiply 110.1 by 100.1

(i)

```
      1101
     ×1010
     ------
      0000
     1101×
    0000××
   1101×××
   --------
Total 10000010
```

(ii)

```
        110.1
       ×100.1
       ------
         1101
        0000×
       0000××
      1101×××
      --------
Total 11101.01
```

(d) Binary Division: Binary division is very similar to decimal division as examples of binary division given below.

Divide 100011 by 111

```
        101
111 ) 100011
      111
      -----
      00111
        111
        ---
        000
```

Here is divisor 111 and dividend 100011 quotient 101 and remainder 000.

14.3.4 Complement Representation

Since computer can recognize only the magnitude of numbers, some device to be incorporated to enable computers to distinguish between positive and negative numbers. To facilitate this, negative numbers are often stored in Complemented Form. It is known as 2's complement representation.

1. *1's Complement:*

The 1's complement of a binary number is the number that results when we complement each bit i.e. convert one to zero and zero to one.

If the input is $x_3\ x_2\ x_1\ x_0 = 1010$

the 1's complement is $\bar{x}_3\ \bar{x}_2\ \bar{x}_1\ \bar{x}_0 = 0101$

similarly, if the input is 1100 then the 1's complement is 0011.

2. *2's Complement:*

The 2's complement is the binary number that results when we add 1 to the 1's complement of binary number. So 2's complement = 1's complement +1.

So following the rule:

(i) 2's complement of binary number 1001 can be found as

1001 → 0110 (1's complement)

010+1 → 0111 (2's complement)

(ii) 2's complement of 1011 can bound as

1011 → 0100 (1's complement)

0100+1 → 0101 (2's complement)

Positive and Negative Numbers:

Before and After are synonymous with negative and positive; A idea to represent with number line of basic algebra as below:

101	110	110	111	111	000	000	001	001	010	010	011
1	0	1	0	1	0	1	0	1	0	1	0
-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6

0 marks the origin, positive numbers are on the right and negative numbers are on the left. Two points must be noted.

1. The MSB is the sign bit. 0 represents a (+) and 1 represents a (-) sign.
2. The negative numbers in the above diagram are the 2nd complement of the (+) numbers as it is clear from the following table:

Magnitude	Positive	Negative
1	0001	1111
2	0010	1110
3	0011	1101
4	0100	1100
5	0101	1011

6	0110	1010
7	0111	1001
8	1000	1000

Except the last entry '1000' the positive and negative numbers are 2's complements of each other. It is clear from table that taking 2's complement. It is equivalent to negative.

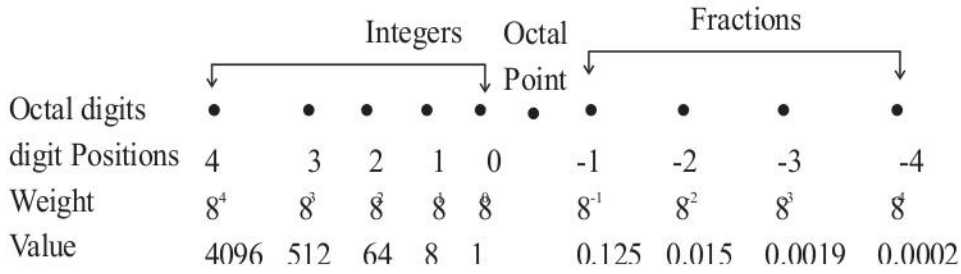
So we can summaries as follows:

1. Positive numbers always have a sign bit (MSB) of 0 and negative numbers always have a sign bit of 1.
2. Positive numbers are stored in sign Magnitude form.
3. Negative numbers are stored as 2's complements.
4. Taking the 2's complement is equivalent to a sign change.

14.4 Octal Number System

The octal number system is also popular number system in computers and microprocessors. The octal number system uses the digits from 0 to 7 (0, 1, 2, 3, 4, 5, 6, 7) and as such its radix or base is eight. Any number in octal number system can be expressed as

$$d_m(8)^m+d_{m-1}(8)^{m-1}+\ldots d_1(8)^1+d_0(8)^0 + d_{-1}(8)^{-1}+d_{-2}(8)^{-2}+\ldots+d_n(8)^{-n}$$



Following Table 14.2 shows octal numbers along with the equivalent decimal and binary numbers.

Decimal	Octal	Binary
0	0	000
1	1	001
2	2	010

3	3	011
4	4	100
5	5	101
6	6	110
7	7	111
8	10	1000
9	11	1001
10	12	1010
11	13	1011
12	14	1100
13	15	1101
14	16	1110
15	17	1111
16	20	10000

14.4.1 Octal to Decimal Conversion

In octal number the place value of each digit increases by a multiple of 8 as we move towards the left of octal point and on the right of octal point the place value decreases by a multiple of 8^{-1} successively.

For example a octal number can written as

$a_3\ a_2\ a_1\ a_0.\ a_{-1}\ a_{-2}\ a_{-3}$

The octal number can be interpreted into decimal number as

$$(a_3\ a_2\ a_1\ a_0\ .\ a_{-1}\ a_{-2}\ a_{-3})_8 = (a_3\times 8^3 + a_2 \times (8)^2 + a_1\ (8)^1 + a_0\ (8)^0 + a_{-1}\ (8)^{-1} + a_{-2}\ (8)^{-2} + a_{-3}\ (8)^{-3})_{10}$$

In above example each term is representing the weights of octal number or these are positional value. The octal to decimal conversion is done by multiplying each digit of octal number by its positional weight and adding all the products of octal digits and positional weights. Consider an octal number is $(4057.405)_8$ to convert this number in decimal number.

$$(4057.405)_8 = 4\times 8^3 + 0\times 8^2 + 5\times 8^1 + 7\times 8^0 + 4\times 8^{-1} + 0\times 8^{-2} + 5\times 8^{-3}$$

$$= 2048 + 0 + 40 + 7 + \frac{4}{8} + 0 + \frac{5}{\sqrt{512}}$$

$(4057.405)_8 = (2095.5097)_{10}$

So the octal number $(4057.405)_8$ is equivalent to $(2095.5097)_{10}$ decimal number.

14.4.2 Decimal to Octal Conversion

The conversion of decimal number to octal number is similar to decimal to binary conversion. As we know there are two parts in decimal number i.e. whole number (Integers) and fraction part.

To convert decimal Integer part in octal number the integer number is successively divided by 8 and its remainders recorded. The final octal result is obtained by assuming all the remainder from bottom to top. In this last remainder being the most significant bit (MSB) and the first remainder is least significant bit (LSB).

The conversion of fraction part of a decimal number into octal number, we multiply the fraction by 8. The result gives a integer and a new fraction. The integer can be any number from 0 to 7. Then again new fraction part multiplied by 8. Then again we get a integer and a new fraction part. The multiplication at each stage is done after reaoviye the whole number. This process continues till the fraction part becomes zero. The successive integer gives the digits star ting from first integer which has maximum weightage to last integrae number which has minimum weightage. To understand this procedure we consider a following example. Convert $(119.23)_{10}$ into its octal equivalent.

$(119.23)_{10}$ = 119 is integer part and 0.23 is fraction part first we convert (119)10 in octal number by successive division of 119 by 8.

Quotient	Remainder	
$\frac{119}{8} = 14$	7	[LSB]
$\frac{14}{8} = 1$	6	
$\frac{1}{8} = 0$	1	[MSB]

$(119)_{10} = (167)_8$

To convert fraction part (0.23) in octal by successive multiplication with 8 as below:

Product	Whole Number	Fraction
$0.23\times8 = 1.84$	1 [MSB]	0.84
$0.84\times8 = 6.72$	6	0.72
$0.72\times8 = 5.76$	5	0.76
$0.76\times8 = 6.08$	6 [LSB]	0.08

So $(0.23)_{10} = (0.1656)_8$.

So the octal equivalent of given decimal number 119.23 is 167.1656.

14.4.3 Octal to Binary Conversion

We can change each octal digit to its binary equivalent directly. Because the base of octal number 8 which is the cube of 2, So each digit in the octal number is replaced by its 3 bit binary equivalent. To get the 3-bit binary equivalent, first we will find the binary equivalent of octal digit as given in table 14.2. If it is not in 3 bits, then the zeros are placed before the binary equivalent to make it 3 bits. For example $(1)_8 = (1)_2 = (001)_2$, $(2)_8 = (10)_2 = (010)_2$ and $(3)_8 = (1)_2 = (011)_2$ etc. Now we take an example to convert octal number $(25.634)_8$ to its binary equivalent.

```
 2   5   6   3   4
 ↓   ↓ . ↓   ↓ . ↓
010 101 110 011 100
```

The digits are converted into their 3 bit binary equivalents Here the binary equivalents in 3 bits of 2 and 5 are 010 and 101 respectively so the binary equivalent of $(25)_8$ is $(010101)_2$. Fraction part's binary equivalents in 3 bits of 6, 3 and 4 are 110, 011 and 100 and 100 respectively so binary equivalent of $(.634)_8$ is $(.110011100)_2$.

So $(25.634)_8 = (010101.110011100)_2$

14.4.4 Binary to Octal Conversion

Conversion of binary to octal number is a reversal of octal to binary conversion. The whole number form groups of three bits starting at the binary point. Zeros are added at each end wherever necessary. Write the values of these 3-bit groups in the

form of octal numbers. The resultant number will be octal number like given below example. Convert the binary number 1101.0110111 to its equivalent octal number.

Binary Point

$$(1101.0110111)_2 = \underline{001}\ 101 . 011\ 011\ \underline{100}$$
$$\downarrow\ \downarrow\ .\ \downarrow\ \downarrow\ \downarrow$$
$$1\ 5\ 3\ 3\ 4$$

So $(1101.0110111)_2 = (15.334)_8$.

14.5 Hexa Decimal Number System

Hexadecimal number system has a radix or base of 16. The hexadecimal number has sixteen digits. These digits are represented by 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E and F. After reaching nine (9) in hexadecimal system, we count A, B, C, D, E and F. These alphabets A, B, C, D, E and F are used to represent 10, 11, 12, 13, 14 and 15 respectively. The hexadecimal number system is commonly used in the calculators these days. It is very convenient and extensively. It's also used in minicomputers to have memory organized into sets of bytes. One can convert hexadecimal number to other number systems like binary, octal and decimal and vice versa with help of table 14.3 given below.

Table 14.3

Decimal Numbers	Binary Numbers	Octal Numbers	Hexadecimal Numbers
0	0000	0	0
1	0001	1	1
2	0010	2	2
3	0011	3	3
4	0100	4	4
5	0101	5	5
6	0110	6	6
7	0111	7	7
8	1000	10	8

9	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F

14.5.1 Hexadecimal to Decimal Conversion

We can convert hexadecimal number into decimal number same as in binary or octal. The positional value of hexadecimal numbers on the left of hexadecimal point increases as a multiple as 16 in order and on the right of hexadecimal point the value decreases by factor $\frac{1}{16}$.

In general hexadecimal number can be interpreted into decimal number as below.

$$(N)_{16} = d_m (16)^m + d_{m-1} (16)^{m-1} + \ldots + d_1 (16)^1 + d_0 (16)^0 + d_{-1} (16)^{-1} + d_{-n} (16)^{-n}$$

It clear from equation that the conversion of hexadecimal number into decimal can be done by multiplying each hexa digit with its positional weight and then adding all the products of hexa digits and positional weight. We take an example to understand.

Conversion of $(2F59.2F)_{16}$ into decimal number

$$(2F59.2F)_{16} = 2 \times (16)^3 + F \times (16)^2 + 5 \times (16)^1 + 9 \times (16)^0 + 2 \times (16)^{-1} + F \times (16)^{-2}$$

$$= 2 \times 4096 + 15 \times 256 + 5 \times 16 + 9 \times 1 + 2\frac{1}{16} \times +15 \times \frac{1}{256} \qquad [\because F=15]$$

$$(2F59.2F)_{16} = (12121.1836)_{10}$$

14.5.2 Decimal to Hexadecimal Conversion

The method used for converting decimal number into hexadecimal number is same like conversion of binary or octal but only difference is that here integer part of decimal number is successive division by 16 to get quotient. The remainder is

recorded dividing again the quotient by 16. This process continues until the quotient becomes 0. Last remainder is MSB and first remainder is LSB.

The fraction part of decimal number is converted by multiplying the fractional part by 16. Then again we get a fraction part and a whole number. This new fraction is again multiplied by 16. The process is repeated till fraction part becomes zero.

We can understand the process by taking an example to convert a decimal number $(2162.14)_{10}$ into hexadecimal number.

$(2162.14)_{10}$ = 2162 is integer part and 14 is fraction part

$(2162)_{10}$ =

$\frac{2162}{16}$ = 135	2 [LSB]	
$\frac{135}{16}$ = 8	7	
$\frac{8}{16}$ = 0	8 [MSB]	

$(2162)_{10} = (872)_{16}$

Fraction Part

Product	Whole Number	Fraction
0.14×16 = 2.24	2	0.24
0.24×16 = 3.84	3	0.84
0.84×16 = 13.44	13=D	0.44
0.44×16 = 7.04	7.04	70.04

$(0.14)10 = (.2307)_{16}$

So $(216.14)_{10} = (872.23D7)_{16}$

14.5.3 Conversion of Hexadecimal Number into Binary Number

To convert hexadecimal number into binary number each hexadecimal number is replaced by 4 bits of binary equivalent. As binary number has a base of 2 whereas a hexadecimal has base of $16 = 2^4$. So each hexa digits is replaced by 4 bits of binary. In case there is not in 4 bits, then the zeroes are placed before the binary equivalent to make it 4 bits. For example

$(29AF.2B)_{16}$	=	2	9	A	F	.	2	B
Binary		↓	↓	↓	↓			↓
Equivalent		0010	1001	1010	111	.	0010	1011
(in 4 bits)								

So $(29AF.2B)_{16} = (0010100110101111.\ 0010\ 1011)_2$.

14.5.4 Conversion Of Binary Number Into Hexadecimal Number

To convert a binary number into hexadecimal number form groups of four binary bit starting from the binary point and on both sides of it. Then write the value of these groups in terms of equivalent hexadecimal numbers as given in table 14.3 the total number formed is hexadecimal number.

For example $(1110101.00101010)_2$

	=	0111	0101	.	0010	1010
					↓	↓
Equivalent Decimal Value		7	5	.	2	10
Equivalent Hexadecimal Value		7	5	.	2	A

So $(01110101.00101010)_2 = (75.2A)_{16}$

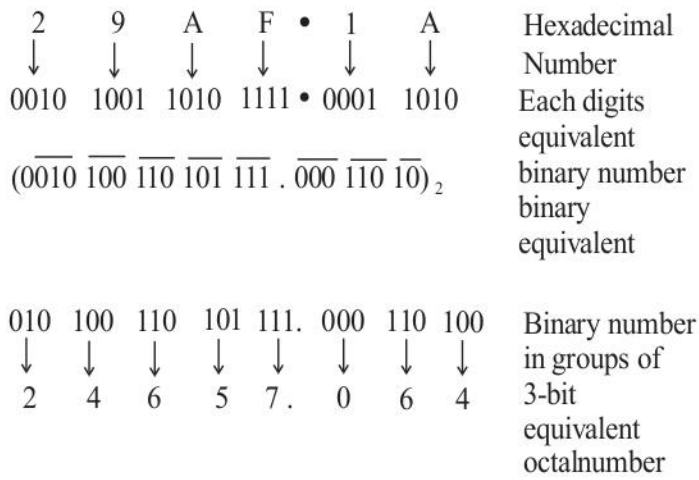
14.5.5 Hexadecimal to Octal Conversion

The conversion of hexadecimal number to octal number can be obtained by two steps.

Step First: It can be converted into binary number. In this step each digit in the hexadecimal number is replaced by its 4 bit binary equivalent. To get the 4-bit binary equivalent, first we will find the binary equivalent of digit. If it is not in 4 bits, then the zeroes are placed before the binary equivalent to make it of 4 bits.

Step Second: Now this equivalent binary number is grouped in 3 bits starting from the binary point to both side of it. Then there groups of 3-bits are converted in to their octal equivalent. Thus, the number obtained will represent the octal number. For example conversion of $(29AF.\ 1A)_{16}$ into octal as follow:

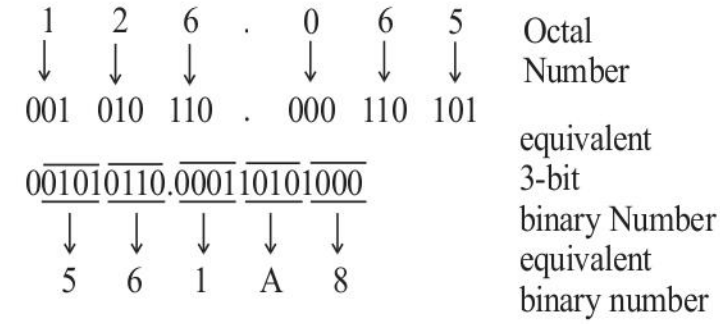
$(29AF.\ 1A)_{16} = (24657.064)_8$ Octal Number



14.5.6 Octal to Hexadecimal Conversion

To convert the octal number into hexadecimal number by First converting the given octal number into binary number by replacing each digit of octal number by its 3-bits binary equivalent. Now this binary equivalent is grouped in 4 bits starting from the binary point to both side of it. Then each of the groups of 4-bits are replaced by equivalent hexadecimal digit.

For example conversion of $(126.065)_8$ into hexa decimal by following process.



$(126.065)_8 = (56.1A8)_{16}$

14.6 Illustrative Examples

Example 14.1 Convert the decimal number 32.14 into binary number.

Solution: Decimal number = $(32.34)_{10}$

Integer Part = 32 and Fractional part = 0.34

Conversion of Integer Part 32 by dividing by 2

Quotient	Remainder	
$\frac{32}{2}=16$	0	LSB
$\frac{32}{2}=8$	0	
$\frac{8}{2}=4$	0	
$\frac{4}{2}=2$	0	
$\frac{2}{2}=1$	0	
$\frac{1}{2}=0$	1	MSB

So $(32)_{10}=(100000)_2$

Now the fractional Part (0.34) is converted into binary number by multiplying the fraction by 2 till fraction becomes zero.

Product	Whole Number	
$0.34\times2=0.68$	0	MSB
$0.68\times2=1.36$	1	
$0.36\times2=0.72$	0	
$0.72\times2=1.44$	1	
$0.44\times2=0.88$	0	LSB

So $(0.34)_{10}=(.01010)_2$

So that $(32.34)_{10}=(100000.01010)_2$

Example 14.2 Convert (1101.101)2 into decimal number.

Solution: The binary number can be converted into decimal by multiplying each binary bit by its positional weight. So

$(1101.101)2 = 1\times2^3+1\times2^2+0\times2^1+1\times2^0+1\times2^{-1}+0\times2^{-2}+1\times2^{-3}$

$= 8 + 4 + 0 + 1 + 0.50 + 0 + 0.125$

$= 13.625$

Example 14.3: Convert (119.23)10 into octal number.

Solution: A integer part of decimal number can converted into octal by successive division of Integer by 8 and fractional part by multiplying the fraction by 8. So

Conversion Integer Part 119

Quotient	Remainder	
$\frac{119}{8}=14$	7	(LSB)
$\frac{14}{8}=1$	6	
$\frac{1}{8}=0$	1	[MSB]

So (119)10 = (167)8

Now conversion of Fraction part = 0.23

Product	New fraction	Whole Number	
0.23×8 =1.84	0.84	1	[MSB]
0.84×8=6.72	0.72	6	
0.72×8=5.76	0.76	5	
0.76×8=6.08	0.08	6	[LSB]

So $(0.23)_{10} = (0.1656)_8$

So that octal equivalent to given decimal number 119.23 is 167.1656.

Example 14.4 Convert $(4057.15)_8$ Octal number into decimal number.

Solution: A octal number can be interpreted into decimal as

$$d.F. = (d_m(8)^m + d_{m-1}(8)^{m-1} + \ldots + a_0(8)^0) + (a_{-1}(8)^{-1} + a_{-2}(8)^{-2} + \ldots + a_{-n}(8)^{-2})$$

So $(4057.15)_8 = (4\times8^3+0\times8^2+5\times8^{-1}+7\times8^0)+(1\times8^1+5\times8^{-2})$

$$= 2048 + 0 + 40 + 7 + \frac{1}{8} + 5 \times \frac{1}{64}$$

$$= 2048 + 40 + 7 + 0.125 + 0.078$$

$$= (2095.203)10$$

Example-14.5 Convert (1526.23)8 into binary equivalent by 3-bit rule.

Solution: A octal number can converted into binary number by replacing each octal digit into its 3-bits binary equivalent.

So $(1526.23)_8$	=	1	5	2	6	.	2	3
		↓	↓	↓	↓		↓	↓
		001	101	010	110		010	011

So $(1526.23)_8 = (001101010110.010011)_2$

Example 14.6 Convert binary number $(11010.10101)_2$ into octal number.

Solution: A binary number can convert into octal number by grouping of three binary bits starting from binary point on both side.

	110	10	.	101	01
=	011	010	.	101	010
	↓	↓		↓	↓
	3	2		5	2

= (32.52)8

Example 14.7 Convert $(268)_{10}$ into its hexadecimal number.

Solution: A integer decimal number is converted into hexadecimal by successive dividing it by 16.

quotient	Remainder	
$\frac{268}{16} = 16$	12=C [LSB]	↑
$\frac{16}{16} = 1$	0	|
$\frac{1}{10} = 0$	1 [MSB]	|

So $(268)_{10} = (10C)_{16}$

Example 14.8 Convert $(8A3.D)_{16}$ into decimal number.

Solution: To convert a hexadecimal number into decimal number each digits of hexadecimal number is multiplied by its positional weightages.

$(8A3.D)1_6 = 8\times16^2 + A\times16^1 + 3\times16^0 + D\times16^{-1}$

A = 10 and D = 13

So $8\times16^2 + 10\times16^1 + 3\times16^0 + 13\times16^{-1}$

$= 2048 + 160 + 3 + 0.8125$

or $(8A3.D)_{16} = (2211.8125)_{10}$

Example 14.9 Convert $(B2F8.3B)_{16}$ into its binary equivalent.

Solution: A hexadecimal number is converted into binary number by replacing each hexadecimal digits by its 4-bit binary equivalent. So

$(B2F8.3B)_{16}$ = B 2 F 8 . 3 B

∵ B =11 and F = 15 So

	11	2	15	8	.	3	11
Binary	↓	↓	↓	↓		↓	↓
equivalent	1011	0010	1111	1000	.	011	1011

In group of 4-bit

So $(B2F8.3B)_{16} = (1011\ 0010\ 111\ 1000.0111011)_2$

Example 14.10 Convert the binary number $(110110011011)_2$ in (i) octal, (ii) decimal and (iii) hexadecimal number.

Solution: (i) Conversion in octal number by making groups of 3- digits.

= 110	110	011	011
↓	↓	↓	↓
6	6	3	3

(110110011011)2 = (6633)8

(ii) Conversion in decimal number

$= 1\times2^{11} + 1\times2^{10} + 0\times^{29} + 1\times2^{8} + 1\times2^{7}$

$+ 0\times2^{6} + 0\times2^{5} + 1\times2^{4} + 1\times2^{3} + 0\times2^{2} + 1\times2^{1} + 1\times2^{0}$

$= 2048+1024+0+256+128+0+0+16+8+0+2+1$

$= 3483$

(iii) To convert in hexadecimal number we make group of 4 bit

=	1101	1001	1011
	↓	↓	↓
	D	9	B

So $(110110011011)_2 = (D9B)_{16}$

14.7 Self Learning Exercise

Q.1 What is the base or radix of a number system?

Q.2 Write rules of addition in binary number system.

Q.3 Convert the following decimal numbers into binary numbers:

(a) $(109)_{10}$ (b) $(72.45)_{10}$

(c) $(53.75)_{10}$ and (d) $(29.375)_{10}$

Q.4 Define most significant bit (MSB)?

Q.5 Convert the hexadecimal number $(C146)_{16}$ to octal by first converting it to binary via 14 binary for one hexadecimal' rule and then using a similar rule to convert to octal.

14.8 Summary

In this unit we have discussed binary,octal and hexadecimal number system. Conversion from one number system into another number system is explained with suitable examples.

14.9 Glossary

Binary Number: Switching circuits have two positions ON and OFF which can be represented by digits 1 and 0 respectively. Therefore these two digits are very important in digital system and number system based on these two digits is called binary number system.

Hexadecimal number: This system has a radix or base of 16. The hexadecimal number has sixteen digits. These digits are represented by 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E and F.

14.10 Answers to Self Learning Exercise

Ans.1: It represents the number of digits/ bits used in a system As 10 in case of decimal number and 2 in case of binary system.

Ans.2: For the addition of binary numbers four rules are used

(i) $0 + 0 = 0$,

(ii) $0 + 1 = 1$,

(iii) 1 + 0 = 1 and

(iv) 1 + 1 = 10.

Ans.3: (a) $(1101101)_2$ (b) $(1001000.01110)_2$,

(c) $(110101.11)_2$ and (d) $(11101.011)_2$

Ans.4: The bit or digit which carries the highest weightage or positional value in that number is called most significant bit.

14.11 Exercise

Section A: Very Short Answer Type Questions

Q.1 How many digits are used in binary number system 9 write them.

Q.2 What is decimal equivalent of binary number 1001?

Q.3 Define octal number system.

Q.4 What is Hexadecimal equivalent of decimal number 12?

Q.5 Define Least significant bit (LSB).

Q.6 What is decimal equivalent of octal number $(65)_8$?

Q.7 Convert 2A to its equivalent decimal number?

Section B: Short Answer Type Questions

Q.8 Convert the following binary numbers into their decimal equivalents:

(a) $(101101)_2$, (b) $(10110)_2$

(c) $(1100.1011)_2$ and (d) $(0.10101)_2$

Q.9 Convert the following decimal numbers into octal numbers.

(a) $(119)_{10}$ (b) $(435)_{10}$ (c) $(47.23)_{10}$ (d) $(70.265)_{10}$

Q.10 Convert the following octal numbers into binary numbers.

(a) $(101)_8$ (b) $(217)_8$ (c) $(64.32)_8$ (d) $(216.23)_8$

Q.11 Draw a table for binary conversion of decimal numbers from 0 to 10.

Q.12 Draw a table for hexadecimal and binary number of decimal numbers from 0 to 15.

Q.13 Describe the methods of decimal number to Binary number conversion.

Q.14 Explain the methods of octal number to Binary number conversion.

Q.15 Explain the conversion method of decimal number into Hexadecimal number.

Q.16 Describe main steps for hexa decimal number to octal conversion.

Section C: Long Answer Type Questions

Q.17 Explain the methods of converting a decimal number into binary number and binary number into decimal number with the help of examples.

Q.18 What do you understand by octal number system? Describe the method of converting octal number into binary and decimal numbers?

Q.19 What do you understand by hexadecimal number system? Describe the method of hexa decimal number into decimal number system and binary number system.

Q.20 Explain the method of conversion of hexadecimal number into octal number and octal number into hexadecimal number.

Q.21 Convert the octal number $(15217)_8$ to hexadecimal by first converting it to binary via the "3 binary for one octal" rule and then using a similar rule to convert to hexadecimal.

Q.22 State the rules of binary addition, subtraction and multiplication with suitable examples.

Q.23 Convert the decimal number 3665 in (i) binary, (ii) octal and hexadecimal numbers.

14.12 Answers to Exercise

Ans.1: In binary number system two digits are used, which are 0 and 1.

Ans.2: Decimal equivalent is $= 1\times2^3 + 0\times2^2 + 0\times2^1 + 1\times2^0 = 8 + 0 + 0 + 1 = 9$.

Ans.3: A number which has radix or base of 8 and represented by digits 0, 1, 2, 3, 4, 5, 6 and 7.

Ans.4: Decimal number 12 is represented by 'C' in Hexadecimal number.

Ans.5: The bit or digit which carries minimum weightage or positional value in a number is called least significant bit.

Ans.6: $(65)_8 = 6\times8^1 + 5\times8^0 = 48 + 5 = (53)_{10}$

Ans.7: 2A is a hexadecimal number so equivalent decimal number is

$= 2 \times (16)^1 + A \times (16)^0$

$= 32 + 10 \times 1$

$= 32 + 10 = 42$

Ans.8: (a) $(45)_{10}$, (b) $(22)_{10}$,

(c) $(12.6875)_{10}$ and (d) $(0.65625)_{10}$

Ans.9: (a) $(167)_8$ (b) $(663)_8$ (c) $(57.1656)_8$ (d) $(106.2075)_8$

Ans.10: (a) $(1000001)_2$ (b) $(10001111)_2$

(c) $(110100.01101)_2$ (d) $(10001110.010011)_2$

Ans.23: $(3665)10 = (111001010001)_2 = (7121)_8 = (E51)_{10.}$

References and Suggested Readings

1. Nelson magor cooke et al, Basic mathematics for Electronics (7th Edition), Glencoe, Lake forest Ill., 1992
2. Wilson T. Price, Charlec Peselnick, Elements of data Processing mathematics (3rd edition), Holt, Rinehart and Winston 1987
3. Virendra Kumar, Digital Technology Principles and Practice (New Age International (P) Limited (2003)
4. S. L. Kakani, K. C. Bhandari, Electronics Theory and Application (3rd edition). New Age Int (P) Ltd. (2004)

UNIT-15
Boolean Algebra & Logic gates

Structure of the Unit

15.0 Objectives

15.1 Introduction

15.2 Positive and Negative Logic

15.3 Truth Table

15.4 Logic Gates: AND,OR, NOT & Buffer

15.5 Boolean Algebra

15.6 De Morgan's Theorems

15.7 Duality Principle

15.8 Logic Gates: NOR,NAND

15.9 Illustrative Examples

15.10 Self Learning Exercise-I

15.11 Logic Gates : XOR, XNOR

15.12 Illustrative Examples

15.13 Self Learning Exercise-II

15.14 Summary

15.15 Glossary

15.16 Answer to Self Learning Exercises

15.17 Exercise

15.18 Answers to Exercise

References and Suggested Readings

15.0 Objectives

In this unit, we will learn Boolean algebra and operations of basic logic gates on which digital electronics has been developed. Many commonly used devices such

as Computers, mobiles, digital calculators, digital cameras etc. are based on the principles of digital electronics.

15.1 Introduction

Analog signals are continuous signals (voltage or current) that can have any value in a given range. The most common digital signal is binary signal that can have two discrete levels (low and high). These two levels are known as on and off states in digital circuits. Binary digits 1 and 0 can be assigned with digital signal levels.

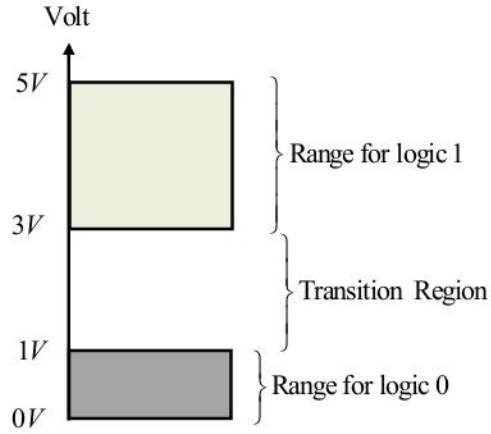
In 1854, George Boole developed binary number algebra which is known as Boolean algebra. In 1938, Claude Shannon applied the two valued Boolean algebra in switching circuits. The binary variables (0 and 1) are also called logical variables or Boolean variables. Boolean 0 and 1 represent the state or logic level. In this unit we study Boolean Algebra and different types of logic gates.

15.2 Positive and Negative Logic

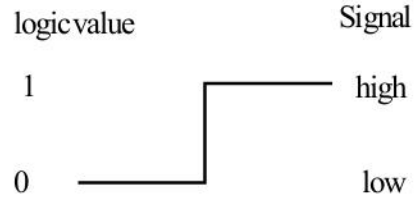
Digital signals have two discrete voltage levels (or values). These levels can be represented using binary variable logic 1 or logic 0. Each voltage level has limited (specified) range that has been shown below in figure for positive logic .

Positive Logic System:

For example, we may define say 0 to 1 volt particular range as low level and 3 Volts to 5 Volts particular range as high level.

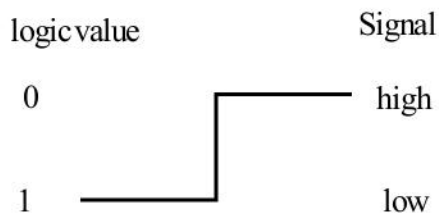


In this system low level is represented by logic 0 and high level is represented by logic 1.



Negative Logic System:

In this system low level is represented by logic 1 and high level is represented by logic 0.



Unless otherwise specified, positive logic system will be considered.

15.3 Truth Table

Logic operation can be expressed in the form of a truth table in which all the possible input combinations and their corresponding outputs are shown. For N inputs, there will be 2^N possible input combinations. Hence number of rows in a truth table will be 2^N. Truth table is a tabular representation of a logic function. The term literal is used for a variable or complement of the variable.

Binary variables are denoted by letters of the alphabets such as A,B,C,D, ...x,y,z,... etc.

15.4 Logic Gates: AND,OR , NOT & Buffer

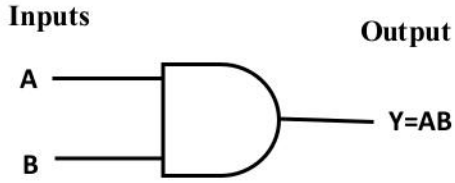
A gate is a digital (two state) circuit with one or more input signals but only one output signal.

AND Gate :-

It has two or more inputs and only one output. Suppose A and B are inputs and Y is output, then

$$\boxed{Y = A.B}$$

This is read as "Y equals A AND B". AND operator is denoted by a dot or by the absence of an operator.



AND Gate symbol

Boolean algebra of AND Gate:

$$A.B = Y$$

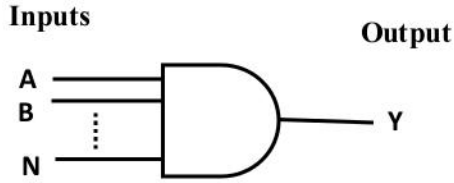
$$0.0 = 0,\ \ 0.1 = 0,\ 1.0 = 0,\ \ 1.1 = 1$$

Truth table of a 2 input AND Gate

Inputs		Output
A	B	Y=AB
0	0	0
0	1	0
1	0	0
1	1	1

For N inputs AND Gate, logical equation is given by

$$Y = A.B.C. .....N$$



OR Gate :-

It has two or more inputs and one output. Suppose A and B are inputs and Y is output, then

$$\boxed{Y = A + B}$$

The symbol + is used for OR operation.

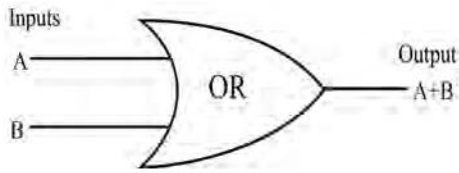
Y = A OR B

This is read as "Y equals A OR B"

Boolean Algebra of OR Gate

$$A + B = Y$$

$$0+0=0, \quad 0+1=1, \quad 1+0=1, \quad 1+1=1$$



OR Gate symbol

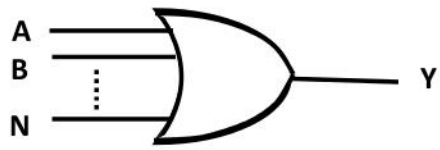
Truth table of a 2 input OR Gate:

Inputs		Output
A	B	Y=A+B
0	0	0
0	1	1

1	0	1
1	1	1

For N inputs OR Gate logical equation is given by

$$Y = A + B + C + ...N$$



NOT Gate :-

It has one input and one output. It is also known as inverter. Suppose A is input and Y is output, then

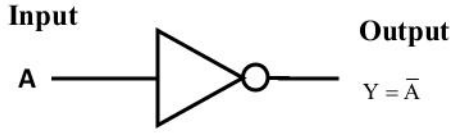
$$\boxed{Y = \overline{A}}$$

This can also be written as $Y = A'$

This is read as Y equals compliment of A. We can also read as Y equals NOT A .

NOT operation is denoted by a prime or by an overbar.

NOT Gate Symbol



An small circle is known as bubble.

It shows inversion operation in digital electronics.

Short symbol of NOT Gate is bubble as shown below



Boolean algebra of NOT operation:

$$\overline{0} = 1$$

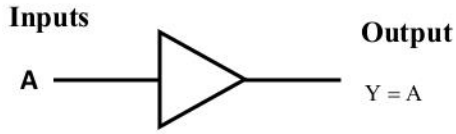
$$\overline{1} = 0$$

Truth table of NOT Gate:

Input	Output
A	$Y = \overline{A}$
0	1
1	0

Buffer :-

In Buffer circuit, binary value of output is equal to binary value of input. The triangle symbol is used for Buffer circuit.



Truth table of Buffer Gate

Input A	Output Y
0	0
1	1

Buffer operation is known as 'transfer' operation. Buffer can be used as electronic amplifier that increases the gain.

15.5 Boolean Algebra

(i) Commutative Law

$A + B = B + A$

$A.B = B.A$

(ii) Associative Law

$A+B+C = A+ (B+C) = (A+B)+C$

$ABC = A(BC) = (AB)C$

(iii) Distributive Law

$A(B+C) = AB + AC$

(iv) OR operation

$\boxed{A + 0 = A}$ Identity

$\boxed{A+1 = 1}$ Null element

$\boxed{A+A = A}$ Idempotence

$\boxed{A+\overline{A}=1}$ Compliments

(v) AND operation

$\boxed{A.0 = 0}$ Null element

$\boxed{A.1 = A}$ Identity

$\boxed{A.A = A}$ Idempotence

$\boxed{A.\overline{A}=0}$ Compliments

(vi) Involution

$\boxed{\overline{\overline{A}} = A}$

(vii) Absorption Rule

(a) $\boxed{A+AB = A}$

Proof: A+AB = A.1 + AB = A(1+B) = A.1 = A

(b) $\boxed{A+\overline{A}B = A+B}$

Proof: $A+\overline{A}B = A.1+\overline{A}B$

$= A(1+B)+\overline{A}B = A+AB+\overline{A}B$

$= A+(A+\overline{A})B = A+1.B = A+B$

(c) $\boxed{(A+B)\ (A+C) = A+BC}$

Proof: (A+B) (A+C) = A.A+ A.C+ B.A+ B.C

= A+AC+ BA + BC = A(1+C) +BA + BC

= A.1 + BA+ BC= A (1+B) + BC

= A.1 + BC= A+ BC

15.6 De Morgan's Theorems

These can be stated in two parts as:

(i) $\boxed{\overline{A+B} = \overline{A}.\overline{B}}$

The complement of the sum of variables is equal to the product of the complements of the variables. For more than two variables ,first part of the De Morgan's theorem can be written as

$$\overline{A+B+C} = \overline{A}.\overline{B}.\overline{C}$$

$$\overline{A+B+C+...+F} = \overline{A}.\overline{B}.\overline{C}....\overline{F}$$

(ii) $\boxed{\overline{A.B} = \overline{A}+\overline{B}}$

The complement of the product of variables is equal to sum of complements of the variables.

For more than two variables, second part of the theorem can be written as

$$\overline{ABC} = \overline{A}+\overline{B}+\overline{C}$$

$$\overline{ABC...F} = \overline{A}+\overline{B}+\overline{C}+....+\overline{F}$$

Hence De Morgan's theorems can be stated as

"Break the bar and interchange AND and OR operators."

De Morgan's First Theorem:



De Morgan's Second Theorem:



15.7 Duality Principle

The dual of an algebraic expression can be obtained in following steps:

(i) Interchange OR and AND operators.

(ii) Interchange 1's by 0's and 0's by 1's.

(iii) Variables and their complements should be left unchanged.

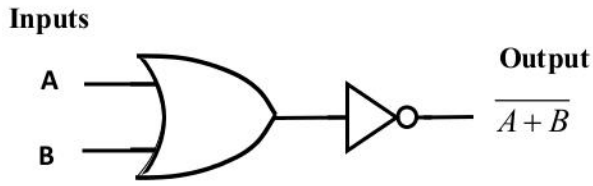
For example

	Dual
$A+1=1$	$A.0=0$
$A+0=A$	$A.1=A$
$A+\overline{A}=1$	$A\overline{A}=0$
$A+AB=A$	$A.(A+B)=A$

15.8 Logic Gates: NOR & NAND

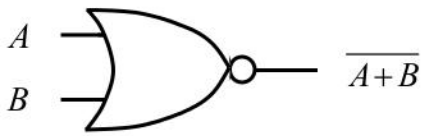
NOR Gate:

It is a combination of NOT gate and OR Gate



NOR Gate

We can also represent NOR Gate as shown below



NOR Gate Symbol

This is read as "Y equals NOT (A OR B)".

Truth table of NOR Gate

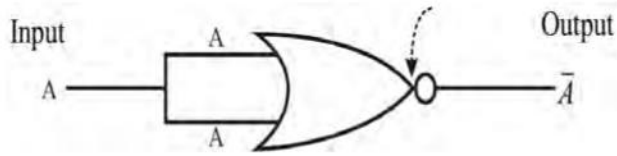
Inputs		Output
A	B	$Y=\overline{A+B}$
0	0	1
0	1	0
1	0	0
1	1	0

We can also express NOR operation as

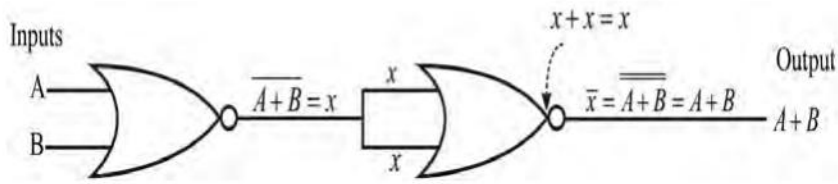
$Y=(A+B)' = A \downarrow B$

NOR Gate is known as universal Gate because AND, OR and NOT operations can be realized using NOR gates.

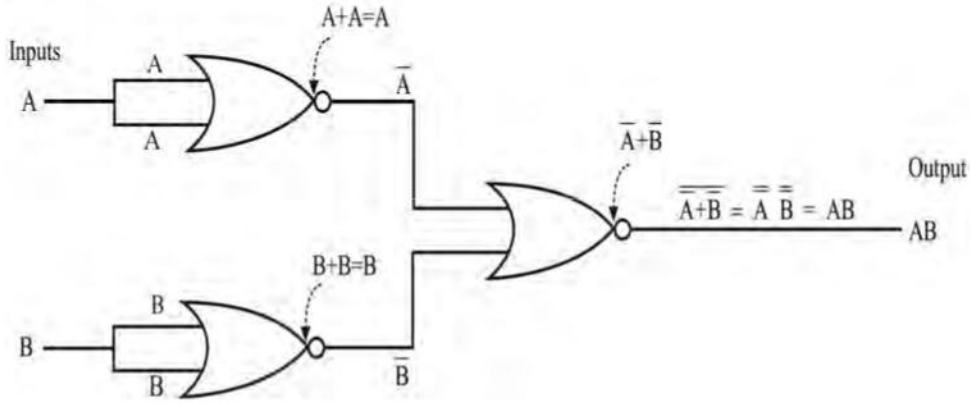
(i) Implementation of NOT Gate using NOR Gate:



(ii) Implementation of OR Gate using NOR Gates :



(iii) Implementation of AND Gate using NOR Gates:



NOR operation is commutative i.e. $A \downarrow B = B \downarrow A$

- NOR operation is not associative i.e.

$(A \downarrow B) \downarrow C \neq A \downarrow (B \downarrow C)$

$L.H.S.\ (A \downarrow B) \downarrow C = \overline{\overline{A+B}+C}$

$= \overline{\overline{A+B}}.\,\bar{C} = (A+B).\,\bar{C} = A\bar{C} + B\bar{C}$

$R.H.S. = A \downarrow (B \downarrow C)$

$= \overline{A + \overline{B+C}} = \bar{A}\ \overline{\overline{B+C}}$

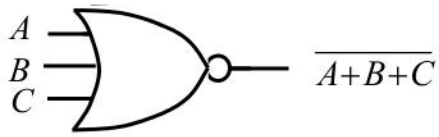
$= \bar{A}\,(B+C) = \bar{A}B + \bar{A}C$

$\therefore\ L.H.S. \neq R.H.S$

➢ We can define multiple NOR gates as

$A \downarrow B \downarrow C = \overline{A + B + C}$

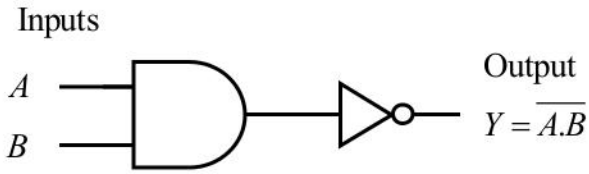
$Y = \overline{A + B + C}$



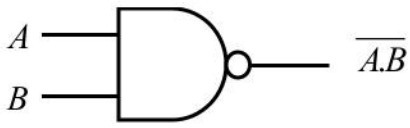
3 Input NOR Gate

NAND Gate:

It is a combination of NOT Gate and AND Gate



NAND Gate can be represented as



NAND Gate symbol

Truth Table of NAND Gate

Inputs		Output
A	B	$Y = \overline{AB}$
0	0	1
0	1	1
1	0	1
1	1	0

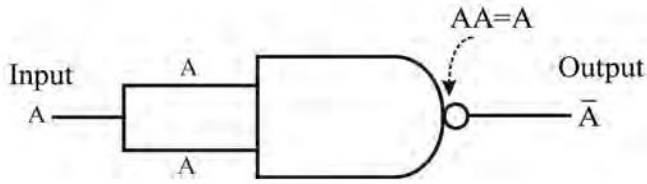
We can also write NAND operation as

$Y = (AB)' = A \uparrow B$

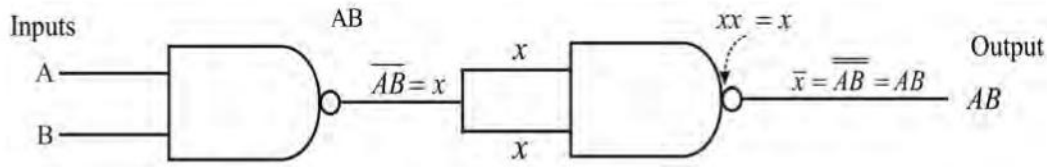
NAND Gate is known as universal Gate because AND, OR and NOT operations

can be realized using NAND Gates.

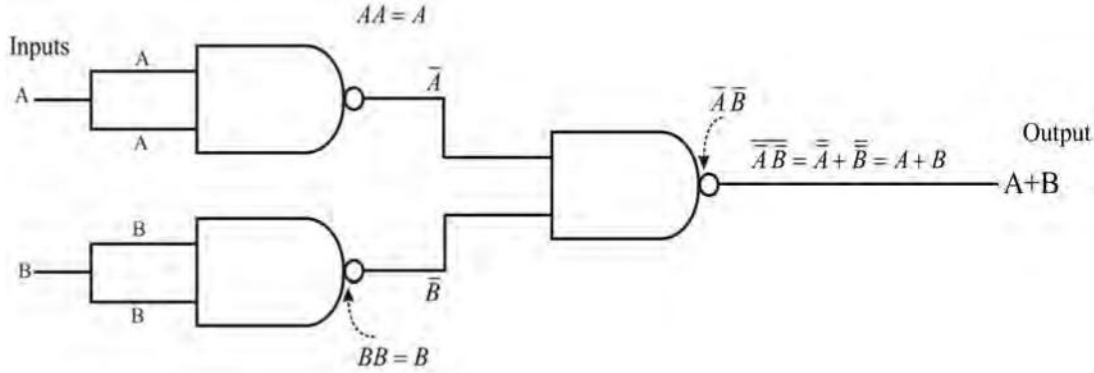
(i) Implementation of NOT Gate using NAND Gate:



(ii) Implementation of AND Gate using NAND Gates:



(iii) Implementation of OR Gate using NAND Gates:



- NAND operation is commutative

  $A \uparrow B = B \uparrow A$

- NAND operation is not associative

  $(A \uparrow B) \uparrow C \neq A \uparrow (B \uparrow C)$

  $L.H.S. = (A \uparrow B) \uparrow C$

  $= \overline{\overline{ABC}} = \overline{\overline{AB}} + \overline{C} = AB + \overline{C}$

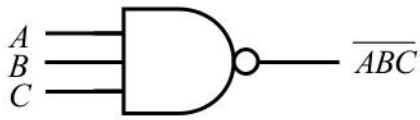
  $R.H.S. = A \uparrow (B \uparrow C)$

  $= \overline{A\,\overline{BC}} = \overline{A} + \overline{\overline{BC}} = \overline{A} + BC$

  Hence $L.H.S \neq R.H.S.$

➢ Multiple NAND Gate can be defined as

$A \uparrow B \uparrow C = \overline{ABC}$



3 Input NAND Gate

15.9 Illustrative Examples

Example 15.1 Show that negative logic OR Gate is equivalent to positive logic AND Gate.

Sol. Positive OR logic Gate truth Table is given below.

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

For Negative logic we will interchange 1 by 0 and 0 by 1, then we get converted truth table that is given below which is basically truth table of positive logic AND gate.

A	B	Y
1	1	1
1	0	0
0	1	0
0	0	0

Hence negative logic OR Gate is equivalent to positive logic AND Gate.

Similarly we can show the following equivalence:

Positive logic OR Gate ↔ Negative logic AND Gate

Positive logic AND Gate ↔ Negative logic OR Gate

Positive logic NOR Gate ↔ Negative logic NAND Gate

Positive logic NAND Gate ↔ Negative logic NOR Gate

15.10 Self Learning Exercise-I

Very Short Answer Type Questions

Q.1 Assume positive logic system. Express the following levels in binary form

(i) 0.5 V and 4.5 V

(ii) −0.4 V and −3.5 V

Q.2 Apply the duality theorem on the following logical equation

$(x + y)(\overline{x} + \overline{y}) = \overline{x}\, y + x\, \overline{y}$

Q.3 Show that $A + A\,B\,\overline{C}\,D = A$

Short Answer Type Questions

Q.4 Using three NOR Gates, implement the Boolean expression Y= (A+B) (C+D)

15.11 Logic Gates : XOR, XNOR

XOR Gate (Exclusive OR Gate) (EX-OR Gate):-

Suppose A and B are inputs and Y is output, then

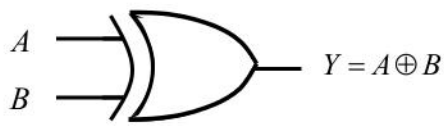
$$Y = A \oplus B$$

The function symbol for XOR operation is $\oplus$

$$A \oplus B = A\overline{B} + B\overline{A}$$

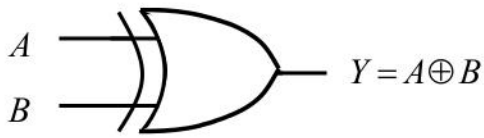
$$A \oplus B = (A + B)(\overline{AB}) = (A + B)(\overline{A} + \overline{B}) = A\overline{A} + A\overline{B} + B\overline{A} + B\overline{B} = A\overline{B} + B\overline{A}$$

$$A \oplus B = \overline{AB + \overline{A}\,\overline{B}} = \overline{AB}\;\overline{\overline{A}\,\overline{B}} = \overline{AB}\,(\overline{\overline{A}} + \overline{\overline{B}}) = \overline{AB}\,(A + B)$$



XOR Gate

We can also represent XOR Gate as shown below



Truth table for XOR Gate

Inputs		Output
A	B	$Y = A \oplus B$
0	0	0
0	1	1
1	0	1
1	1	0

Thus when both inputs are same, then output is zero.

When both the inputs are not same, then output is one.

XOR operation is both commutative and associative.

$A \oplus B = B \oplus A$

$A \oplus (B \oplus C) = (A \oplus B) \oplus C = A \oplus B \oplus C$

Truth table of $A \oplus B \oplus C$

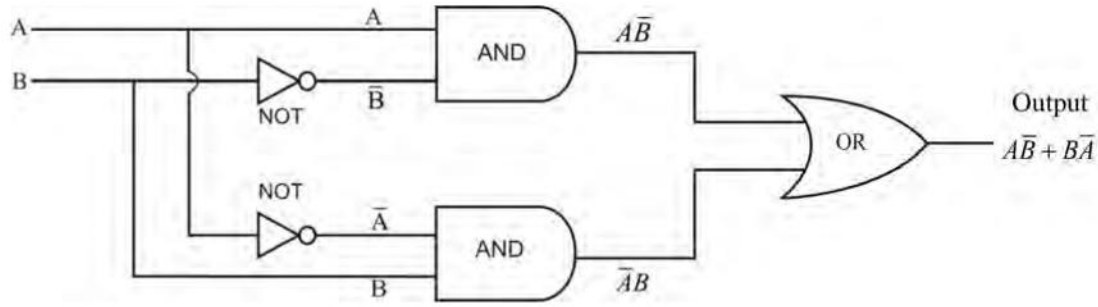
A	B	C	Number of 1's in word ABC	A⊕B	C	(A⊕B) ⊕ C
0	0	0	Zero (even)	0	0	0
0	0	1	One (odd)	0	1	1
0	1	0	One (odd)	1	0	1
0	1	1	Two (even)	1	1	0
1	0	0	One (odd)	1	0	1
1	0	1	Two (even)	1	1	0
1	1	0	Two (even)	0	0	0
1	1	1	Three (odd)	0	1	1

Thus number of 1's in word ABC is odd , then it has output 1.

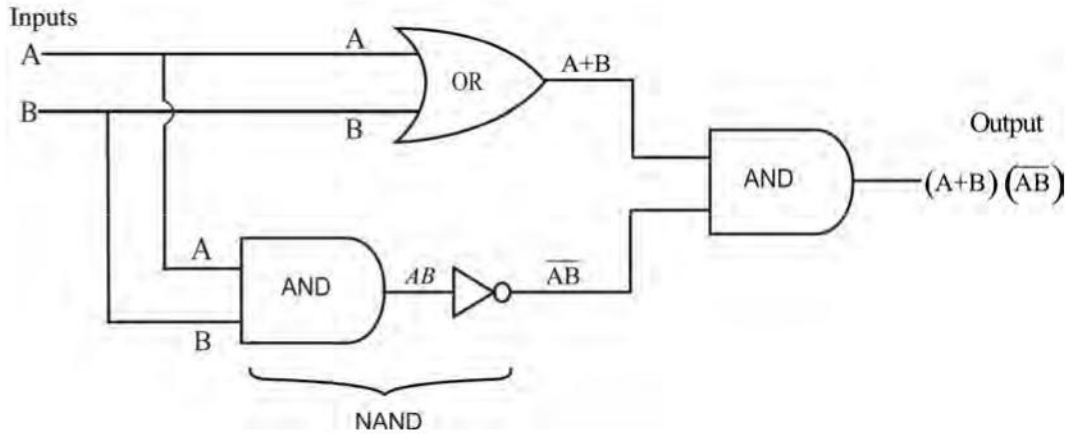
It is true irrespective of number of inputs at XOR.

Logic Circuit diagram

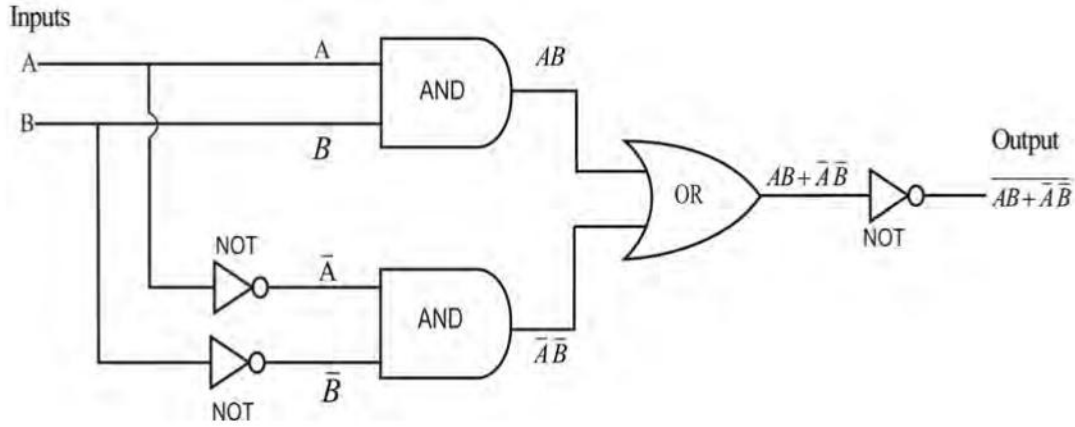
(i) Logic diagram for Y = $A \oplus B = A\overline{B} + B\overline{A}$



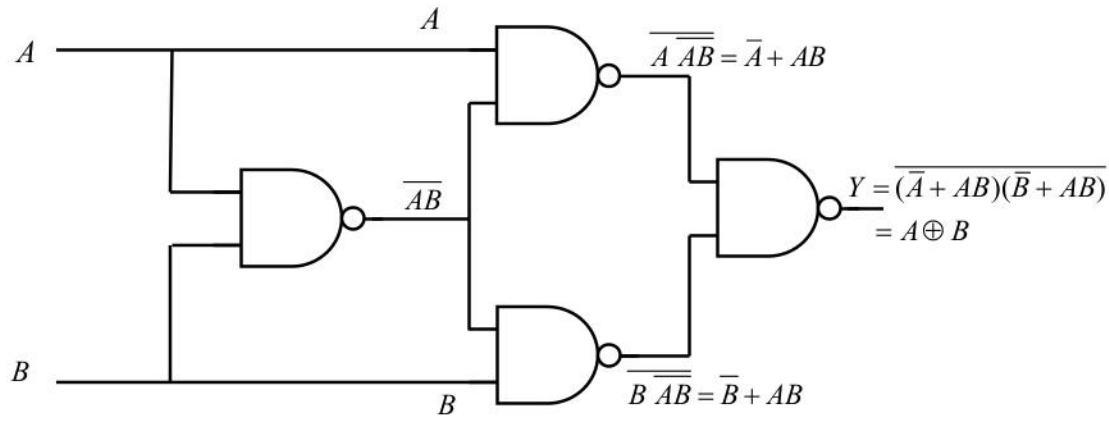
(ii) Logic diagram for Y = $A \oplus B$ = (A+B) ($\overline{AB}$)



(iii) Logic diagram for Y = $A \oplus B = \overline{AB + \overline{A}\overline{B}}$



(iv) Implementation of XOR Gate using NAND Gates only:

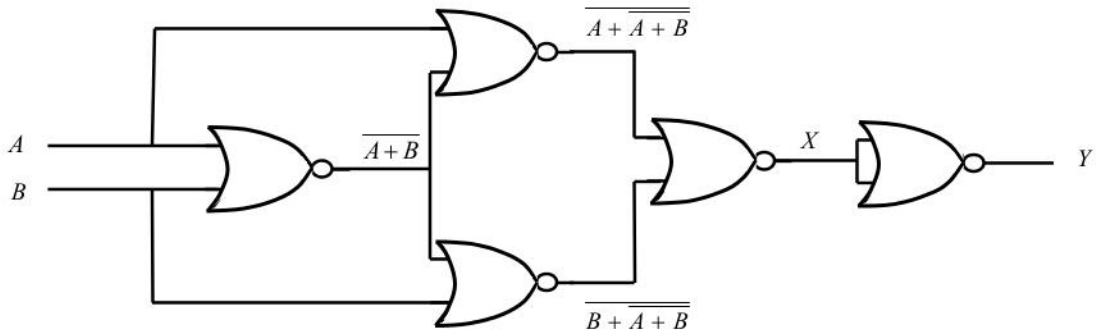


Here $Y = \overline{(\overline{A} + AB)(\overline{B} + AB)} = \overline{\overline{A} + AB} + \overline{\overline{B} + AB}$

$= \overline{\overline{A}}.\overline{AB} + \overline{\overline{B}}.\overline{AB} = A\,\overline{AB} + B\,\overline{AB}$

$Y = (A + B)\,(\overline{AB})$

(v) Implementation of XOR Gate using NOR Gates only:



$X = \overline{\overline{A + \overline{A+B}} + \overline{B + \overline{A+B}}}$

$= \overline{\overline{A}.(A+B) + \overline{B}\,(A+B)}$

$= \overline{\overline{A}A + \overline{A}B + \overline{B}\,A + \overline{B}\,B}$

$= \overline{\overline{A}\,B + \overline{B}\,A}$

$Y = \overline{X} = \overline{\overline{A}\,B + \overline{B}\,A}$

$Y = \overline{A}B + \overline{B}A = A \oplus B$

XNOR Gate (Exclusive NOR Gate)(EX-NOR Gate):-

If A and B are inputs and Y is output then

$Y = A \odot B$

Function symbol for XNOR operation is $\odot$

$$Y = A \text{ XNOR } B = \overline{A \text{ XOR } B}$$

$$= \overline{A\,\overline{B} + \overline{A}\,B} = \overline{A\,\overline{B}}.\overline{\overline{A}\,B}$$

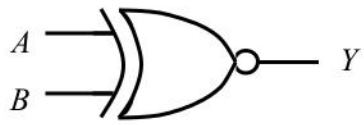
$$= (\overline{A} + \overline{\overline{B}}).(\overline{\overline{A}} + \overline{B}) = (\overline{A} + B)\ (A + \overline{B})$$

$$= \overline{A}\,A + AB + B\overline{B} + \overline{A}\,\overline{B}$$

$$Y = AB + \overline{A}\,\overline{B}$$

$$A \text{ XNOR } B = \overline{A \oplus B}$$

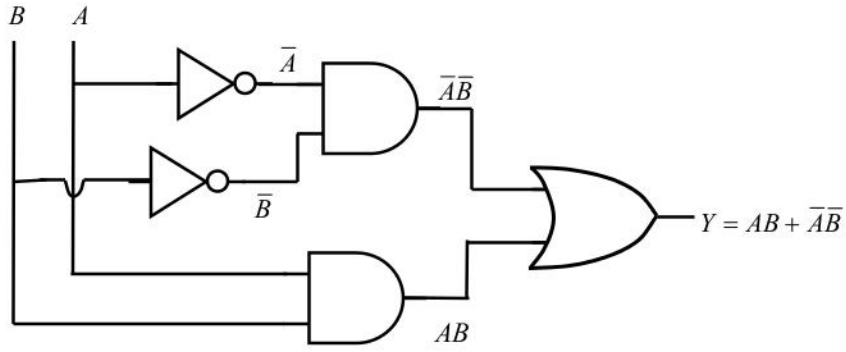
$$\boxed{A \odot B = \overline{A \oplus B}}$$



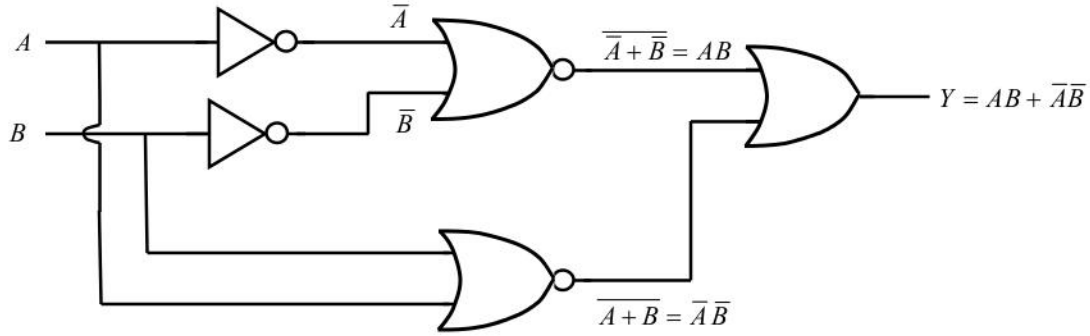
Truth table of XNOR Gate

Inputs		Output
A	B	Y
0	0	1
0	1	0
1	0	0
1	1	1

Logic Circuit



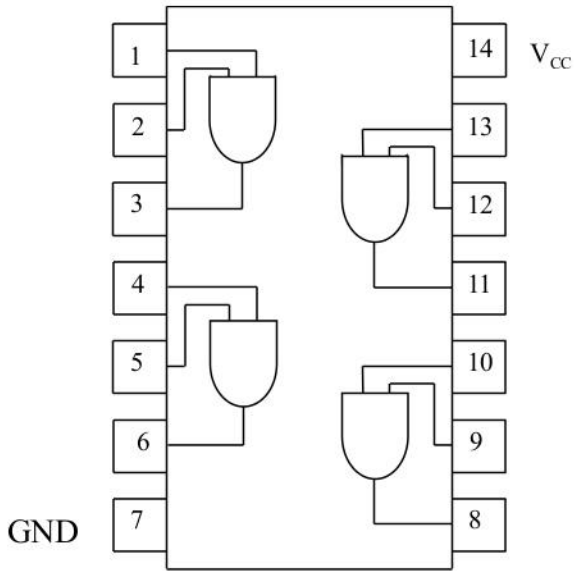
Alternate design of XNOR logic circuit



XNOR logic circuit can be implemented by various alternative logic designs.

Integrated Circuits (IC):-

There are integrated circuits (ICs) for logic functions. For example, 7408 IC chip is quadruple 2 input AND Gate is shown in figure.



Pin out diagram of IC 7408

Manufacturers' data catalogues have details of their pin connections etc.

Some of available IC Gates

IC No.	Description
7400	Quad 2 input NAND Gate
7402	Quad 2 input NOR Gate

7404	Hex inverter Gate
7408	Quad 2 input AND Gate
7432	Quad 2 input OR Gate
7486	Quad XOR Gate

Here Quad (Quadruple) and Hex mean that there are four and six gates respectively within the IC package.

15.12 Illustrative Examples

Example15.2 Show that $x \oplus \overline{y} = \overline{x} \oplus y = \overline{x \oplus y}$

Sol. $\because A \oplus B = A\overline{B} + B\overline{A}$

$\therefore x \oplus \overline{y} = x\overline{\overline{y}} + \overline{x}\,\overline{y} = xy + \overline{x}\,\overline{y}$

$\overline{x} \oplus y = \overline{x}\,\overline{y} + \overline{\overline{x}}\,y = \overline{x}\,\overline{y} + xy$

$\overline{x \oplus y} = \overline{x\,\overline{y} + y\,\overline{x}}$

$= \overline{x\,\overline{y}}\ \ \overline{y\,\overline{x}}$ [By De Morgan's theorems]

$= (\overline{x} + \overline{\overline{y}})\,(\overline{y} + \overline{\overline{x}}) = (\overline{x} + y)\,(\overline{y} + x)$

$= \overline{x}\,\overline{y} + \overline{x}\,x + y\,\overline{y} + xy = \overline{x}\,\overline{y} + 0 + 0 + xy$

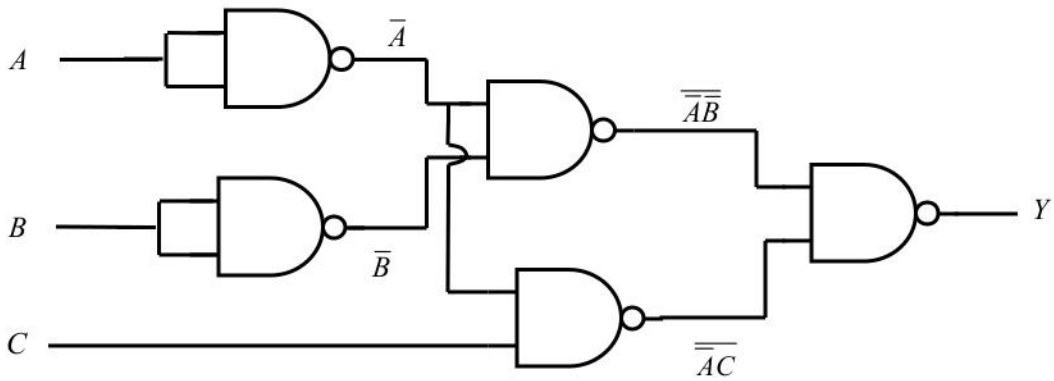
$= \overline{x}\,\overline{y} + xy$

Example15.3. Draw the logic diagram for Boolean function $Y = \overline{A}\,\overline{B} + \overline{A}\,C$ by NAND Gates only.

Sol. $Y = \overline{A}\,\overline{B} + \overline{A}\,C = \overline{\overline{\overline{A}\,\overline{B} + \overline{A}\,C}}$

Using De Morgan's theorem

$Y = \overline{\overline{\overline{A}\,\overline{B}}.\overline{\overline{A}\,C}}$

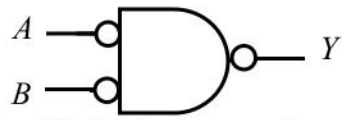


15.13 Self Learning Exercise-II

Very Short Answer Type Questions

Q.1 How many rows would a five variables logic function require ?

Q.2 A logic circuit is shown below

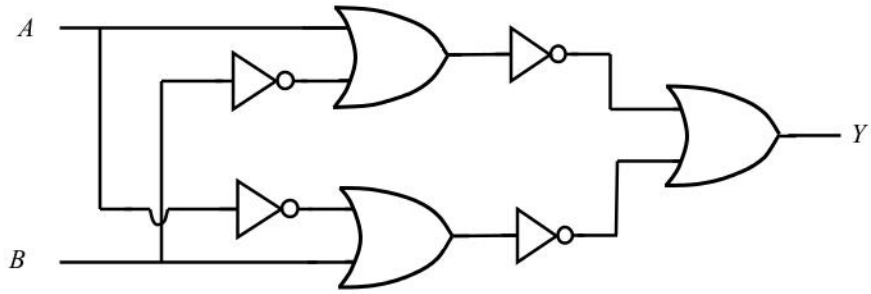


Identify the logic operation performed by the circuit.

Q.3 Show that $A \oplus B = A\overline{AB} + B\overline{AB}$

Short Answer Type Questions

Q.4 Write the logic function Y in terms of input variables.



15.14 Summary

Positive Logic System: In this system low level is represented by logic 0 and high level is represented by logic 1.

AND Gate: Output $Y = A.B$

OR Gate: Output $Y = A + B$

NOT Gate: Output $Y = \overline{A}$

De Morgan's Theorems: (i) $\overline{A+B} = \overline{A}.\overline{B}$ (ii) $\overline{A.B} = \overline{A} + \overline{B}$

NOR Gate: Output $Y = \overline{A+B}$

NAND Gate: Output $Y = \overline{AB}$

XOR Gate: $A \oplus B = A\overline{B} + B\overline{A} = (A+B)(\overline{AB}) = \overline{AB + \overline{A}\,\overline{B}}$

XNOR Gate: Output $Y = A \odot B = \overline{A \text{ XOR } B} = AB + \overline{A}\,\overline{B}$

15.15 Glossary

Binary Variables: The binary variables (0 and 1) are also called logical variables or Boolean variables.

IC: Integrated Circuit

15.16 Answer to Self Learning Exercises

Answer to Self Learning Exercise-I

Ans.1 (i)

Logic levels	Binary Form
0.5 V	0
4.5 V	1

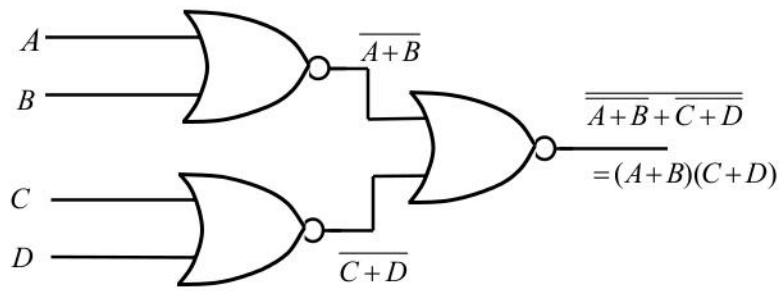
(ii)

Logic levels	Binary Form
−3.5 V	0
−0.4 V	1

Ans.2: $xy + \overline{x}\,\overline{y} = (\overline{x} + y)(x + \overline{y})$

Ans.3: $A + A\,B\,\overline{C}\,D = A\,(1 + B\,\overline{C}\,D) = A.1 = A$

Ans.4:



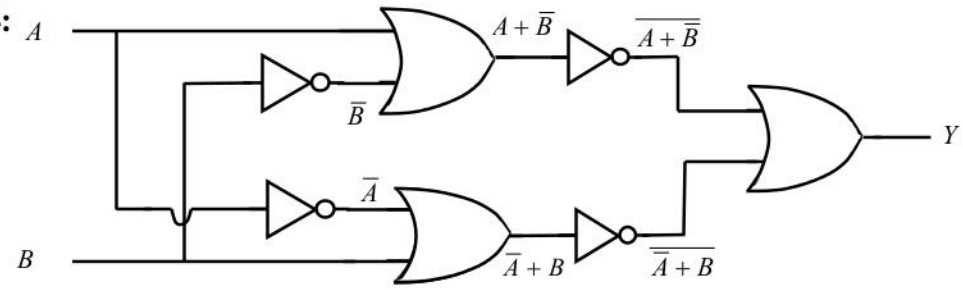
Answer to Self Learning Exercise-II

Ans.1: $2^5 = 32$

Ans.2: OR Gate $\quad Y = \overline{\overline{A}\,\overline{B}} = \overline{\overline{A}} + \overline{\overline{B}} = A + B$

Ans.3: $A\,\overline{AB} + B\,\overline{AB} = (A + B)\,\overline{AB} = A \oplus B$

Ans.4:



$$Y = \overline{\overline{A} + \overline{B}} + \overline{\overline{A} + B} \quad = \overline{A}\,\overline{\overline{B}} + \overline{\overline{A}}\,\overline{B}$$

$$Y = \overline{A}\,B + A\,\overline{B} = A \oplus B$$

15.17 Exercise

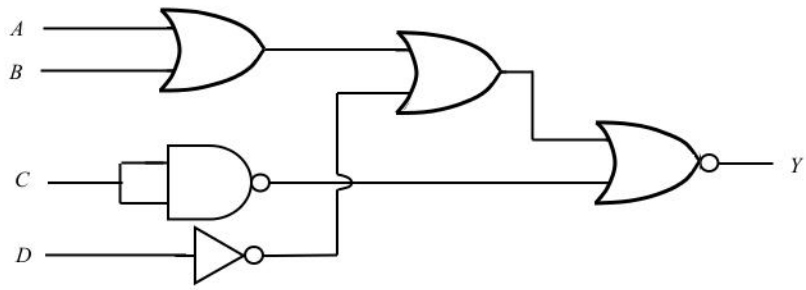
Section A: Very Short Answer Type Questions

Q.1 "Positive logic NAND Gate is equivalent to a Negative logic NOR Gate and Vice versa" Is this statement true ?

Q.2 State De Morgan's theorems.

Q.3 What do you mean by duality in Boolean algebra?

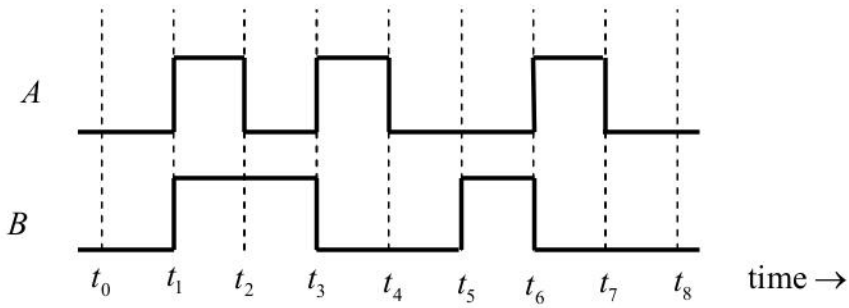
Q.4 Write the expression for Y in the given logic diagram.



Section B : Short Answer Type Questions

Q.5 Implement the NOR function using NAND Gates.

Q.6 The voltage waveform shown in figure are applied at the inputs of OR Gate. Draw the timing diagram of output waveform.

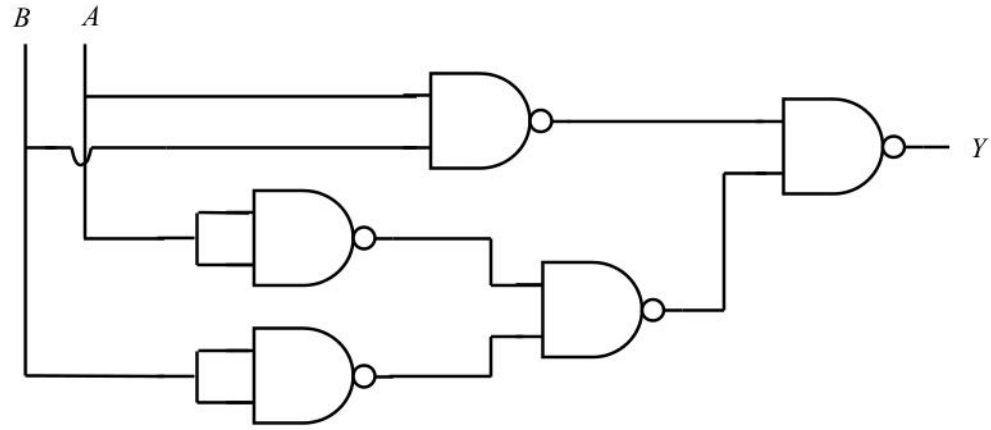


Q.7 Show that

$x \oplus 0 = x$, $x \oplus 1 = \overline{x}$, $x \oplus x = 0$, $x \oplus \overline{x} = 1$

Q.8 Use truth table to show that expression $A + \overline{B}A = A$ is true.

Q.9 Write the truth table for given logic circuit.



Section C: Long Answer Type Questions

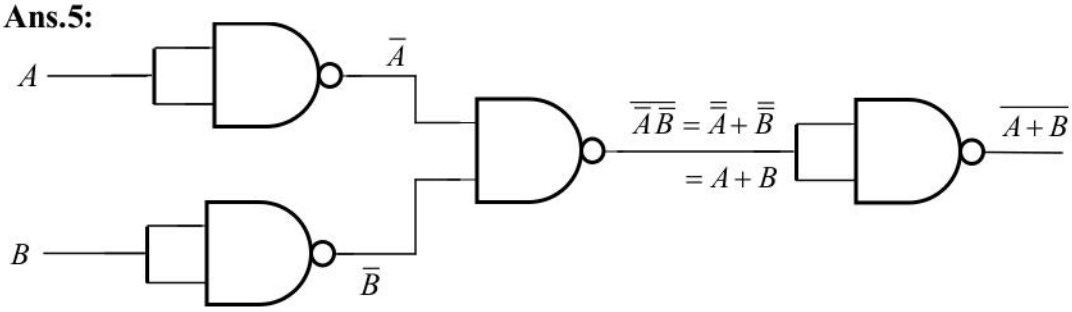
Q.10 How do you realize four inputs XOR Gate with the help of two inputs XOR Gates. Also construct the truth table.

15.18 Answers to Exercise

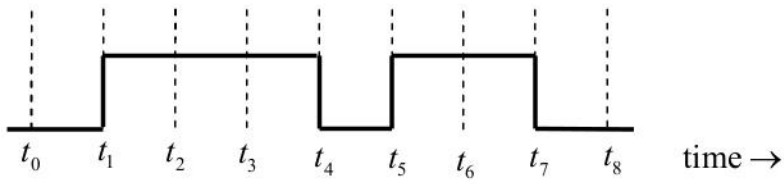
Ans.1: Yes.

Ans.4: $\overline{A + B + \overline{C} + \overline{D}} = \overline{A}\ \overline{B}\ C\ D$

Ans.5:



Ans.6: Timing diagram output waveform



Ans.7: (i) $x \oplus 0 = x.\overline{0} + 0.\overline{x} = x.1 + 0 = x$

(ii) $x \oplus 1 = x\,\overline{1} + \overline{x}.1 = x.0 + \overline{x} = \overline{x}$

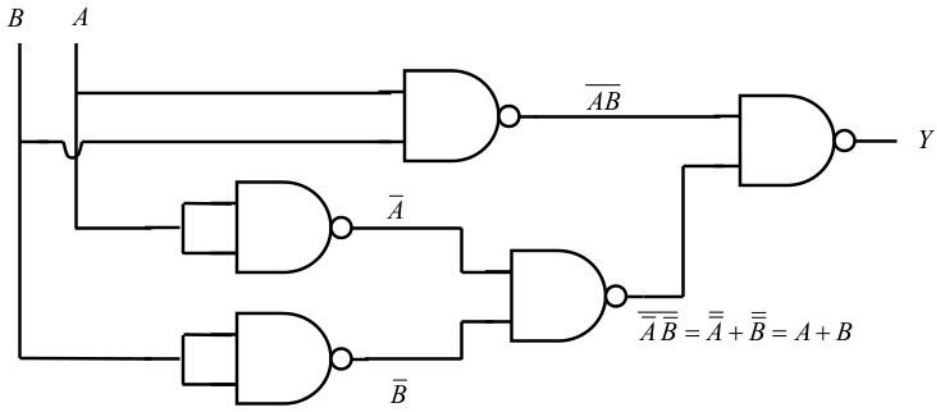
(iii) $x \oplus x = x\,\overline{x} + \overline{x}\,x = 0 + 0 = 0$

(iv) $x \oplus \overline{x} = x\,\overline{\overline{x}} + \overline{x}\,\overline{x} = x.x + \overline{x}.\overline{x} = x + \overline{x} = 1$

Ans.8:

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

Ans.9:



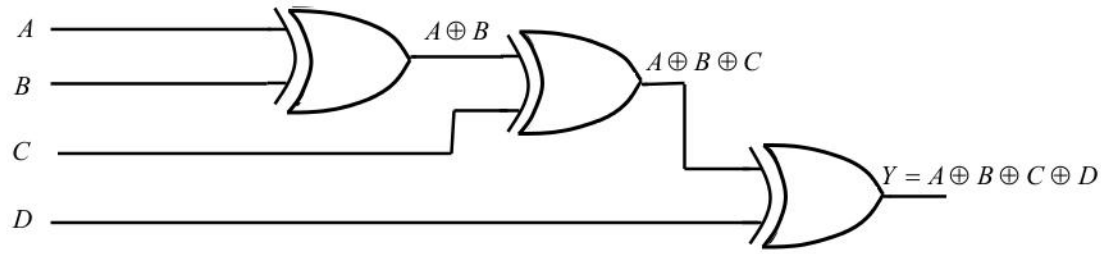
$$Y = \overline{(\overline{AB})\,(A + B)} = \overline{\overline{AB}} + \overline{A + B}$$

$$Y = AB + \overline{A}\,\overline{B} = A \odot B$$

Truth Table

Inputs		Output
A	B	Y
0	0	1
0	1	0
1	0	0
1	1	1

Ans.10:



$Y = \alpha \oplus \beta$ where $\alpha = A \oplus B = A\overline{B} + B\overline{A}$, $\beta = C \oplus D = C\,\overline{D} + \overline{C}\,D$

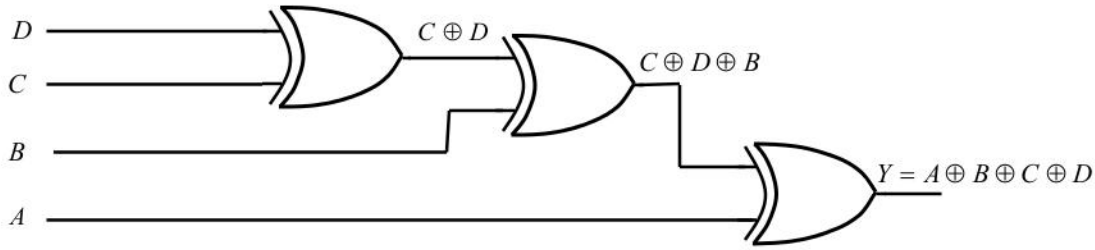
$Y = \alpha\,\overline{\beta} + \beta\overline{\alpha}$

$= (A\overline{B} + B\overline{A})\,\overline{(C\overline{D} + \overline{C}D)} + (C\overline{D} + \overline{C}D)\,\overline{(A\overline{B} + B\overline{A})}$

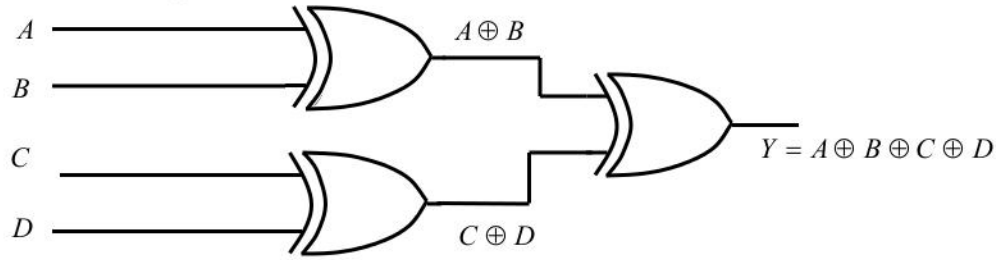
Truth table for $Y = A \oplus B \oplus C \oplus D$

S.No.	A	B	C	D	$A \oplus B$ α	$C \oplus D$ β	Y $\alpha \oplus \beta$
1	0	0	0	0	0	0	0
2	0	0	0	1	0	1	1
3	0	0	1	0	0	1	1
4	0	0	1	1	0	0	0
5	0	1	0	0	1	0	1
6	0	1	0	1	1	1	0
7	0	1	1	0	1	1	0
8	0	1	1	1	1	0	1
9	1	0	0	0	1	0	1
10	1	0	0	1	1	1	0
11	1	0	1	0	1	1	0
12	1	0	1	1	1	0	1
13	1	1	0	0	0	0	0
14	1	1	0	1	0	1	1
15	1	1	1	0	0	1	1
16	1	1	1	1	0	0	0

Alternate design



Alternate design



References and Suggested Readings

1. M. Morris Mano, Digital design (Fourth edition 2008), Pearson Education
2. M. Morris Mano, Digital logic and Computer Design (2002) Prentice - Hall India
3. Malvino and Brown, Digital computer Electronics (Third edition 2001), Tata McGraw Hill)
4. R.P. Jain, Modern Digital Electronics(Second edition 2002),Tata McGraw Hill.
5. Albert Paul Malvino,Donald P. Leach,Goutam Saha, Digital Principles and Applications (Eighth Edition 2014),McGraw Hill Education.
6. Jacob Millman Christos C. Halkias ,Integrated Electronics (Second Edition 2011) ,Tata McGraw Hill.

UNIT-16
Canonical & Standard Forms For Logical Expressions

Structure of the Unit

16.0 Objectives

16.1 Introduction

16.2 Sum-of-products (SOP) Form

16.3 Product-of Sums (POS) Form

16.4 Advantages of SOP and POS forms

16.5 Self- learning Exercise-I

16.6 Standard SOP and POS Forms

16.7 Standard SOP (SSOP) Form (Minterm)

16.8 Standard POS (SPOS) Form (Maxterm)

16.9 Interrelation between Minterms and Maxterms

16.10 K-Map with Two, Three and Four Variables

16.11 Self-Learning Exercise-II

16.12 Summary

16.13 Glossary

16.14 Answer to self Learning Exercises

16.15 Exercise

Reference and Suggested Readings

16.0 Objectives

This unit introduces you the sum-of product Form, conversion into SOP form, product of sum form and advantage of SOP and POS form. Basic idea in introducing standard SOP and POS form is that, we need there to reduce the Boolean expression. After completion of this unit, K-map with two, three and Four Variables can be understood .

16.1 Introduction

In previous unit, we have discussed different types of logical expression with the combination of literal and their output. The AND, OR and NOT are the basic logical expression, while NAND and NOR are the universal building blocks to perform any type of logical expression.

The Boolean functions can be expressed in their standard form. In standard form, the function may contain any number of literals (one or more). Hence, any Boolean function can be expressed in the following two forms,

(1) Sum-of-products form or SOP form.

(2) Product-of-sums form or POS form.

This type of standardization makes the evaluation, simple and implementable of Boolean expression is a systematic and easier way.

16.2 Sum of Products (SOP) Form

The sum-of-products (SOP) form is Boolean expressions that contain various AND terms of one or more literal. The each AND terms is called **product term** of SOP form. The sum expresses the Ring of these product terms. In other words, an SOP form can be implemented with one OR gate and two or more AND gates. The following logical expressions are the examples of SOP forms.

(1) $A + BC + \overline{A}\,BC$

(2) $AB + A\overline{B}\,C$

(3) $ABC + CDE + \overline{B}\,C\overline{D}$

(4) $AB + \overline{B}\,C\overline{D} + AC$

(5) $\overline{A}\,B + C + B\overline{C}$

If two or more product terms are summed in according to Boolean addition, the resulting expression is a sum-of-products.

The logic diagram(Fig16.1) of a SOP expression consists of a group of AND gates that followed by a single OR gate. It is shown in fig. (16.1). For this, each product term requires an AND gate except for a term with a single literal. The logical sum is formed with an OR gate whose inputs are the outputs of the different AND gates including the single literal. Also, it is assumed that all the input variables are

directly available in their complement form, so NOT gates are not included in the logic diagram.

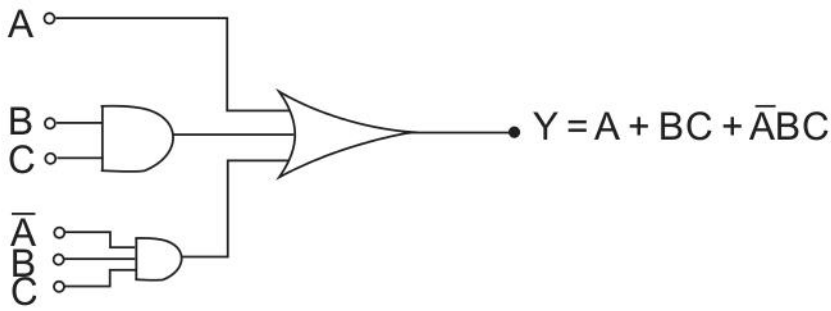


Fig. 16.1 Implementation of the SOP expression $Y = A + BC + \bar{A}BC$

The above logic diagram is called two-level implementation. It is to be noted that for an SOP expression, a single over bar cannot extend over more than one variable; however more than one variable in a term can have an over bar.

Conversion into SOP form

Any Boolean expression can be converting into sum-of-products (SOP) form by using Boolean algebra. The basic law of Boolean algebra i.e. (i) Associative law (ii) Absorption law (iii) Distributive law is used to convert the given expression into SOP form.

16.3 Product of Sums (POS) Form

The product-of-sums (POS) form is a Boolean expression that contains various OR terms of one or more literals. The each OR term is called sum term of POS form. The product expresses, the ANDing of these sum terms. In other words, an POS form can be implemented with one AND gate and two or more OR gates. The following logical expression is the examples of POS forms. If two or more sum terms are multiplied, the resulting expression is a product-of-sums.

(1) $A.(\bar{B}+C).(\bar{A}+B+\bar{C})$

(2) $(A+\bar{B}).(\bar{A}+\bar{B}+C)$

(3) $(A+B).(A+B+\bar{C}).(A+\bar{C})$

(4) $(A+B+C).(\bar{C}+D+E).(B+\bar{C}+D)$

(5) $A.(B+\bar{C}+D).(B+\bar{D})$

For this each sum term requires an OR gate except for a term with a single literal.

The logical product is formed with an AND gate, whose inputs are the outputs of the different OR gates including the single literal. Here, all the input is directly available. The description of POS form is similar to SOP form by inter changing the AND gate and OR gate with complemented variables. In a POS expression, a single over bar cannot extend over more than one variable, however more than one variable in a term can have on over bar.

16.4 Advantages of SOP and POS Forms

A Boolean expression may be expressed in a non-standard form neither SOP nor POS form. It requires the number of AND and OR gates to perform its logical diagram. As a result the gating level of the circuit is increased. The SOP and POS form are used to minimize the gating level of a circuit.

Ex. The expression Y = AB + C. (D+E)

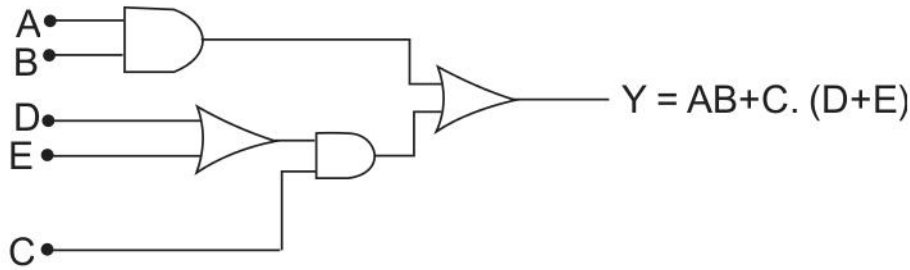


Fig. 16.2: Circuit Realization of Y = AB+C.(D+E)

This expression requires two and two OR gates. It is three level implementation by Boolean expression. Now,

$$Y = AB + C\,(D+E)$$

$$= AB + CD + CE$$

The above expression is in SOP form and it requires three AND and one OR gate. It is performed by two-level implementation.

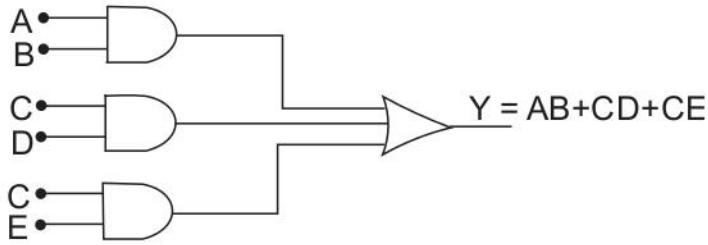


Fig. 16.3: Two level implementation of Y

16.5 Self Learning Exercise-I

Section A : Very Short Answer Type Questions

Q.1 Define sum of products (SOP) form.

Q.2 Define product of sums (POS) form.

Q.3 Give one Example of SOP form.

Section B : Short Answer Type Questions

Q.4 Give one example of POS form.

Q.5 Give the logical expression

$Y(A,B,C) = (A+BC) . (B+\overline{C}\,A)$

Convert thus equation in SOP form.

Q.6 Find the SOP form of the following expression

$Y (A,B,C) = (A+B) . (\overline{A}+C). (B+C)$

16.6 The Standard SOP and POS Forms

The Minterm and Maxterm

So for, we have discussed sum-of-products (SOP) expressions in which few of the product terms contain limited number of variables in the domain of logical expressions. Similarly, in product-of-sums (POS) expressions, the sum terms contain limited variables in the domain of logical expression. It means that either product term or sum terms in SOP or POS from do not contain all the variables in a given expression.

A standard logical expression is one in which all the variables in that domain must present in each of product or sum term. The standard logical expressions are very useful to construct the truth tables and Karnaugh-map simplification. Any nonstandard logical expression can be converted to the standard form by using Boolean theorems. The following two types of standard form of any logical expressions are used in K-map simplifications. The details of K-map are discussed in the next unit.

(i) The standard SOP (SSOP) form

(ii) The Standard POS (SPOS) form

16.7 The Standard SOP (SSOP) Form (Minterm

The standard SOP form or canonical SOP form is one in which each product term contains all the variables either in complement or uncomplement form and all the product term summed together. In a expression $\overline{A}BC + A\overline{B}D + AB\overline{C}\,\overline{D}$ made up by the variables A, B, C and D. However, the last product term contains all the literals, but first two product terms do not contain all the literals A, B and C. There should be forth literal D in the form of D or $\overline{D}$ for presenting the complete set of variables. Similarly in the second product form the missing variable is C, it should be there in the form of either C or $\overline{C}$, hence such type of expressions are called non standard SOP form.

(a) The MINTERM - In a standard or canonical sum-of-products form that contains all the variables of the function in either complemented or uncomplemented form than each product term known as a MINTERM. The minterms are also known as standard products or fundamental products. An n-variable logical expression has 2nd possible combinations of minterms. Each minterm, can be obtained by the AND operation of all the literals of that function in the domain. As we know that in a minterm, a variable is present either in complemented or uncomplemented form. For a minterm $\overline{A}.\overline{B}$ of two variable function. If Both A and B are in logical state 0 then.

The minterm $\overline{A}.\overline{B}$ is $\overline{0}.\overline{0}$ means 1.1 = 1 : and both A and B are in logical state 1 then the minterm $\overline{A}.\overline{B}$ is $\overline{1}.\overline{1}$ means 0.0 = 0. The minterm is represented by the symbol small m. The minterms of a n-variable functions can be represented by m_0, m_1, m_2, ….. mc_{2n-1}), The suffix (decimal code) indicates corresponding to the minterm combination.

The following table shows all the minterms of two and three variable logic functions.

$$Y = \sum m_i$$

$$= \sum m(0,1,2,.....) \text{ [m=minterm]}$$

$$= m_0 + m_1 + \ldots. m_{2n-1}$$

Table 1 : Minterms for two variables

A	B	Minerm	Minterm Notation
0	0	$\overline{A}\,\overline{B}$	m_0
0	1	$\overline{A}\,B$	m_1
1	0	$A\,\overline{B}$	m_2
1	1	AB	m_3

Table 2 : The three Variable minterms table

A	B	C	Minterm	Minterm Notation
0	0	0	$\overline{A}\,\overline{B}\,\overline{C}$	m_0
0	0	1	$\overline{A}\,\overline{B}\,C$	m_1
0	1	0	$\overline{A}\,B\,\overline{C}$	m_2
0	1	1	$\overline{A}\,BC$	m_3
1	0	0	$A\overline{B}\,\overline{C}$	m_4
1	0	1	$A\overline{B}C$	m_5
1	1	0	$AB\overline{C}$	m_6
1	1	1	ABC	m_7

(b) Characteristics of Minterm: - The important characteristic of a minterm is that it has the value 1 for only one combination of n-input variables. Since there are 2^n minterms for an n-variable function, only one minterm will have the value 1, while the remaining (2^n-1) minterms will have the value 0 for any arbitrary input combination. In a two variable function Y =A. $\overline{B}$, (A=1 and B=0) have value 1, while the other terms, $\overline{A}\,\overline{B}$, $\overline{A}\,B$ and AB have value 0. Similarly for three variable $\overline{A}\,BC$ (A=0, B=1, C=1) will have value 1 and all the remaining minterms will have value 0. So, an SOP expression is equals to 1 only if one or more of the product terms in that expression is equal to1. The standard SOP can be represents by Σm_i ; where i is the minterm notation.

(c) Conversion of SOP form into standard SOP (SSOP) form: Each product term in an SOP form that does not contain all the variables in function can be expanded to standard form to include all variables and their complements in the function. The following procedure is taken into account to convert an SOP form into standard SOP (SSOP) form by using Boolean theorem ($A+\overline{A}=1$).

Procedure (I)

(i) First of all write down all the product terms carefully as given in the expression.

(ii) If in any product term one or more variables are missing, expand this product term by multiplying it with the sum of each one of the missin variable and its complement ($A+\overline{A}=1$)

(iii) Solve the obtained function with the help of Boolean theorems.

(iv) Finally, drops out the repeating term . ($A+A=A$)

The alternative direct method can also be applicable to obtain standard SOP form (SSOP),

Procedure (II)

(i) First of all write down all the product terms carefully as given in the expression.

(ii) Then, put Xs in product terms where variables must be inserted to form a standard product term.

(iii) Now, replace the uncomplement variable by 1s and the complemented variable and by 0s.

(iv) Now, use all the possible combinations of Xs in product terms of 0s and 1s to get minterms.

(v) Finally, drops out the repeating terms.

(d) Four Variable logical expression (Minterm):- We know that for a n-input variables, there are 2^n possible input combinations. In case of Four input variables, the sixteen ($2^4=16$) possible combinations from 0000 to 1111 are available.

The corresponding standard products are from $\overline{A}\,\overline{B}\,\overline{C}\,\overline{D}$ to ABCD. Out of sixteen, standard products the particular is found in a quick way. It means if the input

variable is 0, then the variable is in its complemented form in that standard product. It can be understood as follows,

(i) For input 0001, the first three variables are in the complemented form, while the last variable is in its uncomplement form, So the resultant product term will be $\overline{A}\,\overline{B}\,\overline{C}\,D$.

(ii) Similarly, for 0100, the product term will be $\overline{A}\,B\,\overline{C}\,\overline{D}$ and so on.

The minterms for 4- variable expression are numbered as m_0, m_1, m_2 …… m_{14}, m_{15}. The logical output is the ORing of different minterms. The minterms are the ANDing of all the literals present in either complemented or uncomplemented form.

$Y(A, B, C, D) = Y = m_0 + m_1 + \ldots. + m_{14} + m_{15}$

$\Rightarrow \qquad Y = \Sigma m_i \ (0, 1, 2, \ldots 14, 15)$

The minterm starts form 0000 to 1111 in the different combinationof literals as given in the table 13.

Table - 3 : Minterms for four variables

Variable				Minterm	Minterm Notation
					m_i
A	B	C	D		
0	0	0	0	$\overline{A}\,\overline{B}\,\overline{C}\,\overline{D}$	m_0
0	0	0	1	$\overline{A}\,\overline{B}\,\overline{C}\,D$	m_1
0	0	1	0	$\overline{A}\,\overline{B}\,C\,\overline{D}$	m_2
0	0	1	1	$\overline{A}\,\overline{B}\,C\,D$	m_3
0	1	0	0	$\overline{A}\,B\overline{C}\,\overline{D}$	m_4
0	1	0	1	$\overline{A\,B}\,\overline{C}\,D$	m_5
0	1	1	0	$\overline{A}\,BC\,\overline{D}$	m_6
0	1	1	1	$\overline{A}\,BCD$	m_7
1	0	0	0	$A\,\overline{B}\overline{C}\overline{D}$	m_8
1	0	0	1	$A\,\overline{B}\overline{C}\,D$	m_9

1	0	1	0	$A\overline{B}C\overline{D}$	m_{10}
1	0	1	1	$A\overline{B}CD$	m_{11}
1	1	0	0	$AB\overline{C}\overline{D}$	m_{12}
1	1	0	1	$AB\overline{C}D$	m_{13}
1	1	1	0	$ABC\overline{D}$	m_{14}
1	1	1	1	ABCD	m_{15}

Each minterm is represented by m_i, where the subscript i is the decimal equivalent of the natural binary number corresponding to the minterm with uncomplemented variable taken as 1 and the complemented variable taken as 0.

(e) Logical Circuit of Minterms: In a Boolean expression, the different product terms are summed to form a standard sum-of-products (SSOP) expression. The corresponding logic circuit diagram of such minterms is accomplished by the different AND gates used for product terms and followed by a single OR gate. This is called AND-OR network.

16.8 The Standard POS Form (The Maxterm)

So far, we have seen POS Boolean expression in which few of the sum terms do not contain all the variables. The standard POS form or canonical product-of-sums form is one in which each sum term contains all the variables either in complemented or uncomplemented form and all the sum terms multiplied together. In a four variable expression $Y = (\overline{A}+\overline{B}+\overline{C})(\overline{A}+B+\overline{D}).(\overline{A}+\overline{B}+C+D)$ made up by the variables A, B, C and $\overline{D}$. However, the last sum term contains all the variables. But first two sum terms do not contain all the variables. The first sum term should be containing the forth literal D in the form of either D or $\overline{D}$ to obtain standard POS form. Similarly the second sum term should be contain the literal C in the form of either C or $\overline{C}$. Hence the above expression is a non-standard product-of-sums (POS).

On the other hand, if we see the logical expression

$Y(A,B,C,D)= (A+\overline{B}+C+\overline{D}).\ (A+B+\overline{C}+D).(\overline{A}+B+\overline{C}+\overline{D}).$

that possess all the variables. Each sum term contain all the four variables. Such a

logical expression are called standard or canonical product-of-sum (SPOS).

(a) The MAXTERM- In a domain of standard or canonical product-of-sums that contain all the variables of the function in either complemented or uncomplemented form, each sum term is known as a MAXTERM. The maxterms are also known as standard or fundamental sums. In a n-variable logical expression, there are 2_n possible combinations of maxterm hence, the number of possible combination for minterm and maxterm are same. The maxterm is represented by the symbol capital M.

$$Y = \text{IIM}_i \quad [\prod_i = \text{logical ANDing}]$$

$$= M\ (0, 1, 2, \ldots..) \ [M=\text{Maxterm}]$$

$$= M_0 . M_1 . \ldots\ldots\ldots M_{2^{n-1}}$$

The maxterms of a n-variable expression can be represented by M_0, M_1, M_2 …. $M_{2^{n-1}}$. The suffix indicates the decimal code corresponding to the Maxterm combination.

For a Maxterm Y = A + B, if both A and B are in logical state 0 then either A or B is in logical state 1, than the maxterm (A+B) is equal to 1. (0+1 = 1+0 = 1+1 = 1).

The following tables show all the possible maxterms corresponding to two and three variables logical functions.

Table- 4 : Two Variable Maxterms

A	B	Maxterm	Maxterm Notation
0	0	A+B	M_0
0	1	$A+\overline{B}$	M_1
1	0	$\overline{A}+B$	M_3
1	1	$\overline{A}+\overline{B}$	M_3

Table 5 : Three Variable Maxterms

A	B	C	Maxterm	Maxterm
0	0	0	A+B+C	M_0
0	0	1	$A+B+\overline{C}$	M_1

0	1	0	$A+\overline{B}+C$	M_2
0	1	1	$A+\overline{B}+\overline{C}$	M_3
1	0	0	$\overline{A}+B+C$	M_4
1	0	1	$\overline{A}+B+\overline{C}$	M_5
1	1	0	$A+B+C$	M_6
1	1	1	$\overline{A}+\overline{B}+C$	M_7

(b) Characteristics of Maxterm:- The important characteristics of a maxterm is that it has the value 0 for only one combination of n-input variables. Only one maxterm will have the value 0, while the remaining (2^n-1) maxterms will have the value 1 for any arbitrary input combination. Let us take an example,

$Y = \overline{A} + B$ (Maxterm)

If, A = 0 and B=1, then,

$Y = \overline{0}+1 = 1 + 1 = 1$

While the other maxterm are $(A+\overline{B})$, $(\overline{A}+\overline{B})$. (A+B) For above input (A=0, B = 1) the other maxterm will be,

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

And $Y = \overline{A} + \overline{B} = \overline{0}+\overline{1} = 1 + 0 = 1$

Also, $Y = A + B = 0 + 1 = 1$

Hence, for above combination, only maxterm $(A+\overline{B})$ has value 0, while remaining combination have value 1. Hence it is concluded that, a product-of-sums expression is equal to 0 only when one or more of the sum terms in the Boolean expression are equal to 0. The standard POS form can be represented by $\prod M_i$ where, i is the maxterm notation.

(c) Conversion of POS form into standard POS form: Each sum term in a POS form that does not contain all the variables in expression can be expended to standard form to include all variables in that function with their complements. The following procedures are taken to get canonical POS form by using Boolean theorem $\left(A.\overline{A}=0\right)$.

Procedure I :-

(i) First of all write down all the sum terms carefully as given in the expression.

(ii) If in any sum term one or more variables are missing, expand this sum term by adding it with the product of each one of the missing variable and its complement $\left(A.\overline{A}=0\right)$.

(iii) Solve the obtained function with the help of Boolean theorems.

(iv) Finally, drops out the repeating term. (AA=A).

The alternative direct method can also be applicable to obtain standard POS (SPOS) form or can be written in term of maxterm.

Procedure-II

(i) First of all write down all the sum terms carefully as given in the expression.

(ii) Then, put X's in sum terms, whose variables must be inserted to form a standard sum term.

(iii) Now replace the completed variables by 1s and the uncomplemented variables by 0s.

(iv) Now apply all combinations of Xs in terms of 0s and 1s to obtain the maxterms.

(v) Finally drops out the repeating terms.

(d) Logic Circuits of Maxterms

In a Boolean expression, the different sum terms are multiplied together to form a standard product-of sums expression. The corresponding logic circuit diagram of such maxterms is accomplished by the different OR gates used for sum terms and following by a single AND gate. This is called OR-AND network.

(e) Four variable logical expression (Maxterm)

As we know for four input variable, the total number of combinations are sixteen. These are from 0000 to 1111. The corresponding standard sums are from (A+B+C+D) ($\overline{A}+\overline{B}+\overline{C}+\overline{D}$). In such a case, if the input variable is in the state, then the variable in its uncomplemented form is in that standard sum. The 1 state of the variable represents the complemented form of the fundamental sum.

For better understanding, the following examples are given.

(i) For input 0001, the first three variables are in the form of their uncomplemented literal, while the last variable is in its complemented form. So the resultant maxterm will be $(A+B+C+\overline{D})$.

(ii) Similarly, for 0100, the fundamental sum term will be $(A+\overline{B}+C+D)$.

The maxterm for four-variable expression is numbered from M_0, M_1, …. M_{15}. The logical output is the ANDing of different maxterm. The maxterms are the ORing of all the literals present in either complementary or uncomplementary form as shown in table 17.

$Y(A, B, C, D) = M_0 . M_1 . M_2 \ldots . M_{15}$

$= \prod M_i = \prod M (0, 1, 2, \ldots\ldots 15)$

Table 6 : Maxterms for four variables

Variables				Maxterm	Maxterm Notation
A	B	C	D		
0	0	0	0	$A+B+C+D$	M_0
0	0	0	1	$A+B+C+\overline{D}$	M_1
0	0	1	0	$A+B+\overline{C}+D$	M_2
0	0	1	1	$A+B+\overline{C}+\overline{D}$	M_3
0	1	0	0	$A+\overline{B}+C+D$	M_4
0	1	0	1	$A+\overline{B}+C+\overline{D}$	M_5
0	1	1	0	$A+\overline{B}+\overline{C}+D$	M_6
0	1	1	1	$A+\overline{B}+\overline{C}+\overline{D}$	M_7
1	0	0	0	$\overline{A}+B+C+D$	M_8
1	0	0	1	$\overline{A}+B+C+\overline{D}$	M_9
1	0	1	0	$\overline{A}+B+\overline{C}+D$	M_{10}
1	0	1	1	$\overline{A}+B+\overline{C}+\overline{D}$	M_{11}
1	1	0	0	$\overline{A}+\overline{B}+C+D$	M_{12}
1	1	0	1	$\overline{A}+\overline{B}+C+\overline{D}$	M_{13}

1	1	1	0	$\overline{A}+\overline{B}+\overline{C}+D$	M_{14}
1	1	1	1	$\overline{A}+\overline{B}+\overline{C}+\overline{D}$	M_{15}

Each minterm is represented by M_i, where the subscript i is the decimal equivalent of the natural binary number corresponding to the maxterm with complemented variable taken as 1 and the uncomplemented variable taken as 0.

16.9 Interrelation Between Minterms and Maxterms

Since, the total number of minterms and maxterm for a n - input function are (2^n) the same. The corresponding decimal numbers are 0 to (2^n-1). In a three variable function where $2^3 = 8$ be the total number of minterms and maxterms. The 0 to 7 are the corresponding decimal number. Let a expression

$$Y(A, B, C) = m_0 + m_2 + m_4 + m_6$$

$$Y = (0, 2, 4, 6)$$

$$Y = \overline{A}\,\overline{B}\,\overline{C} + \overline{A}\;B\;\overline{C} + A\overline{B}\,\overline{C} + AB\overline{C}$$

The 0, 2, 4 and 6 are the minterms. For this expression, the 1, 3, 5 and 7 are the maxterms. So its maxterm representation will be,

$$Y = M_1 . M_3 . M_5 . M_7$$

$$= \prod M(1, 3, 5, 7)$$

$$Y(A, B, C) = (A+B+C).(A+\overline{B}+C) . (\overline{A}+B+C) . (\overline{A}+\overline{B}+C)$$

Hence, if a logical function is expressed in terms of minterm / maxterm, its maxterm / minterm representation can be obtained by using the following complementary characteristic. This can be well understood in the following 4-input function.

$$\text{If, } Y(A, B, C, D) = m_0 + m_3 + m_6 + m_7 + m_{10} + m_9 + m_{12} + m_{14}$$

$$= \sum m(0,3,6,7,10,12,14)$$

$$\text{Then } Y(A, B, C, D) = M_1 . M_2 . M_4 . M_5 . M_8 . M_9 . M_{11} . M_{13} . M_{15}$$

$$= \prod M(1, 2, 4, 5, 8, 9, 11, 13, 15)$$

16.10 K-map with Two, Three & Four Variables

Four two input variable A and B, there are four possible combinations. Each of these combination is in standard SOP form an hence called a minterm or in

case of standard POS form it is called maxterm. The presence of a minterm in a expression shows the output of the logic is assumed as logic 1, while absence shows the logic 0. In addition to this, the presence of a maxterm shows the output of logic as logic 0, while absence shows the logic 1. There are four, nine and sixteen cells corresponding to two, three and four variables K-map.

(1) Mapping of standard SOP and POS expressions:-

The K-map of two, three and four variable is designed in a specific way. In this variables have been marked as A, B, C, and D, and the binary number formed by these variable are taken as AB, ABC, and ABCD for two, three and four variables respectively. The decimal code corresponding to the particular combination of variables is shown in following K-maps. In the following K-maps, the variables of all possible combinationis marked in which the first bit corresponds to the first variable and the second bit corresponds to second variable and so on.

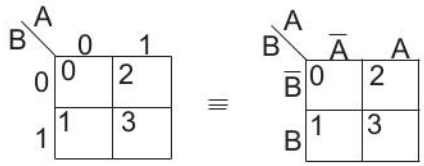


Fig. 16.4: K-map for two-variables

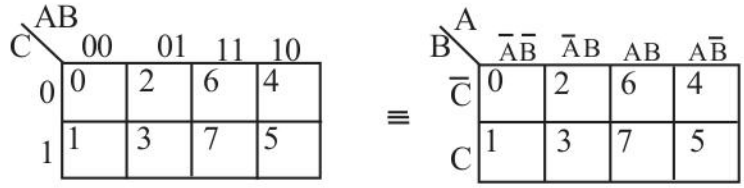


Fig. 16.5: K-map for three variables

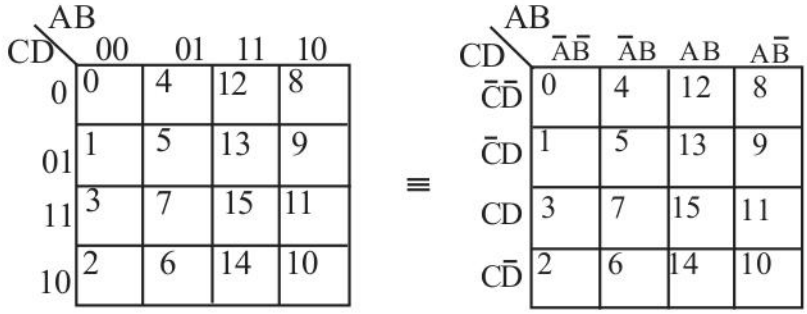


Fig. 16.6: K-map for four variables

For more understanding about K-map, it is required to design a K-map in term of minterms or maxterms. The following diagram shows the K-map in terms of

standard sum-of-products form or, minterms and standard product of sums form or maxterms.

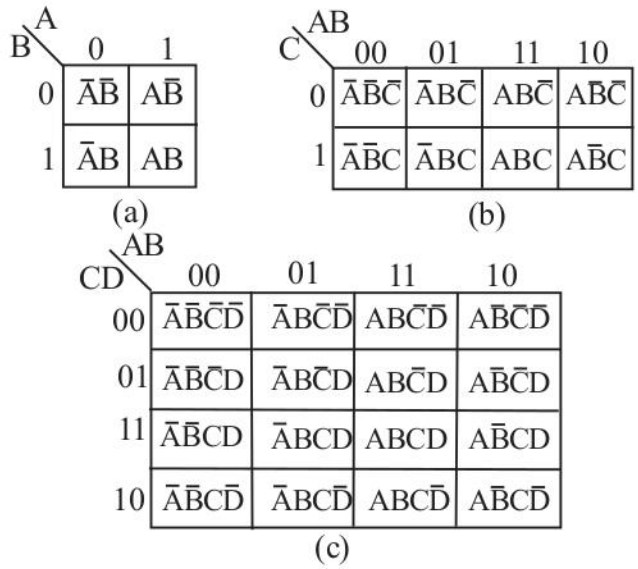


Fig. 16.7 : K-map in terms of minterm for (a) Two variable (b) Three variables and (c) Four variables.

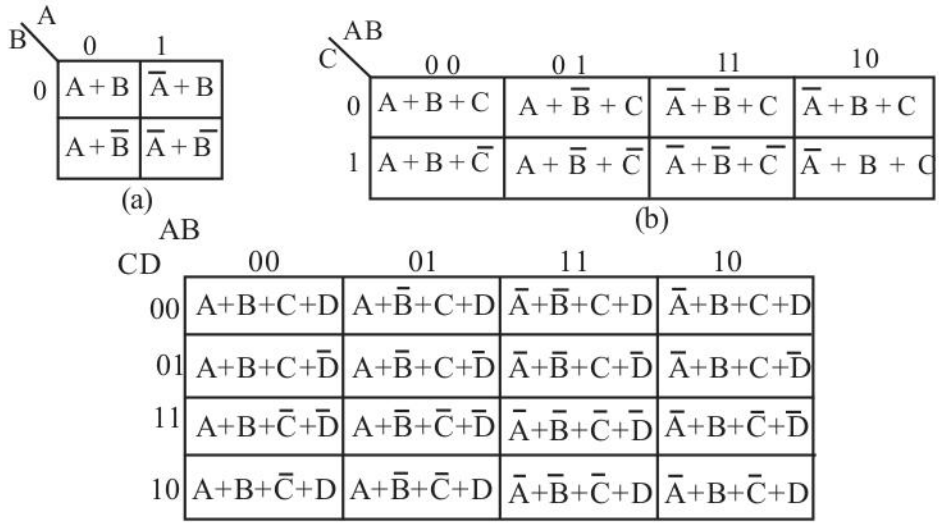


Fig. 16.8 : K-map in terms of maxterms for (a) two variables (b) three variables (c) four variables

16.11 Self Learning Exercise - II

Section A: Very Short Answer Type Questions

Q.1 Convert the following expression into SSOP form:

$Y = \overline{A}\,\overline{B}C + \overline{A}\;B\;\overline{C} + \overline{A}BC + AB\overline{C}$

Q.2 Express the following Boolean expression Y (A,B,C) = $A+\overline{B}C$ in standard SOP form.

Q.3 Express the following Boolean expression

$Y (A, B, C) = A + \overline{B}C$

in standard POS form.

Section B: Short Answer Type Questions

Q.4 Express the following expression in its standard POS form:

$Y (A, B, C) = (A+\overline{B}).(B+C) . (A+\overline{C})$

Q.5 Find the standard SOP form of the following function.

Y (A, B, C, D) = AB + ACD

Q.6 Given the logical expression

$Y (A, B, C) = (A+BC) . (B+\overline{C}\,CA)$

Write the above standard SOP and POS equations in the form of minterm and Maxterm notations.

16.12 Summary

Sum of product (SOP) form of Boolean expression is also defined as ORing of the AND operated variables. e.g. $Y = A\overline{B} + A\overline{B}C + AC.$ Product of sum (POS) form is also defined as Anding of OR operated variables e.g.

$Y = (A+\overline{B})(B+C)(\overline{A}+B+C)$ for n- variable expression, the number of minterms or maxterm are 2^n. Minterm has value 1 only for one combination of n input variables. Standard SOP form is logical sum of minterm, for which the function has the value 1. It can be shown as $Y = \sum_m(1,2,3)$ maxterm has value 0 only for one combination of n input variables. Standard POS form is logical product of maxterm for which function the value 0. It can be shown as $Y = \prod M\ (1, 2, 3)$

16.13 Glossary

Minterm: Also known as the standard product or canonic product term. This is a

term such as $A\overline{B}CD$ or $\overline{A}BC\overline{D}$ or $AB\overline{C}\,\overline{D}$ etc. where each variable is used once and once only.

Maxterm: Also known as the standard sum or canonic sum term. This is a term such as $\left(A+B+\overline{C}+\overline{D}\right)$ or $\left(\overline{A}+B+C+D\right)$ etc. where each variable is used once and once only.

Standard Sum of Products form: Also known as the minterm canonic form or Canonic sum function. A function in the form of the "Sum" (OR) of minterms. eg. $f(A, B, C, D) = AB\overline{C}\,D + \overline{A}\,BCD + ABCD$.

Standard Product of Sums form: Also known as the maxterm canonic form or canonic product function. A function in the form of the "Product" (AND) of maxterms. e.g. $f(A, B, C\ D) = \left(A+\overline{B}+C+\overline{D}\right)\left(\overline{A}+B+\overline{C}+\overline{D}\right)$.

Sum of Products: Also known as the normal sum function. A function in the form of the "SUM" of normal product terms.

e.g. : $f(A, B, C, D) = \overline{A}+\overline{D}+\overline{A}+CD+BC+B\overline{C}\,\overline{D}$

Product of Sums: Also known as the normal product function. A function in the form of the "Product" of normal sum terms.

e.g : $f(A, B, C, D) = \left(\overline{B}+C\right)\left(\overline{A}+B+D\right)\left(B+D\right)\left(\overline{A}+B\right)$

16.14 Answer to Self Learning Exercises

Answers to Self Learning Exercise-I

Ans.1: The sum of products (SOP) form is a Boolean expression that contains various AND terms of one or more literal. The each and terms is called product terms of SOP form. The sum expresses the ORing of these product terms. In other words, an SOP form can be implemented with one OR gate and two or more AND gates.

Ans.2: The product of sums (POS) form is a Boolean expression that contains various OR terms of one or more literals. The each OR terms is called sum term of POS form. The product express the ANDing of there sum terms. In other words, a POS form can be implemented with one gate and two or more OR gates.

Ans.3: $A + BC + \overline{A}\,BC$

Ans.4: $(A+B)\,.\,(A+B+\overline{C}).\,(A+\overline{C})$

Ans.5: $Y = AB + A\overline{C} + BC$

Ans.6: $Y = AC + \overline{A}\,B + BC.$

Answers to Self Learning Exercise-II

Ans.1: $Y = \sum_m(1,2,3,6)$

Ans.2: $Y = ABC + AB\overline{C} + A\overline{B}\,C + A\overline{B}\,\overline{C} + \overline{A}\,\overline{B}\,C$

Ans.3: $Y = (A+\overline{B}+C).\,(A+\overline{B}+\overline{C}).\,(A+B+C)$

Ans.4: $Y = (A+\overline{B}+C).\,(A+\overline{B}+\overline{C}).\,(A+B+C).\,(\overline{A}+B+C)\,(A+B+\overline{C})$

Ans.5: $Y = ABCD + ABC\overline{D} + AB\overline{C}\,D + AB\overline{C}\,\overline{D} + A\overline{B}\,CD$

Ans.6: $Y = \sum m(3,4,6,7)$ and $Y = \prod M(0,1,2,5)$

16.15 Exercise

Section A :Very Short Answer Type Questions

Q.1 What is a minterm?

Q.2 Define maxterm.

Q.3 What do you understand by standard SOP form?

Q.4 What do you understand by standard POS form?

Q.5 What is the advantage of SOP and POS form?

Section B: Short Answer Type Questions

Q.6 Find the standard POS form of a Boolean function F= $\sum m(0,1,5,7,8,9,12)$ given in standard SOP form.

Q.7 What is the reduced expression of $Y = AB + \overline{A}\,\overline{C} + A\overline{B}\,C\,(AB+C)$

Q.8 Evaluate the following SOP expression and find its equivalent POS Expression.

$\overline{A}\,\overline{B}\,\overline{C} + \overline{A}\overline{B}\overline{C} + \overline{A}\,BC + A\overline{B}\,C + ABC$

Q.9 Express the following function into minterm notation.

$Y(A, B, C, D) = D$

Q.10 Find the reduced form of the following expression.

$Y(A, B, C) = \overline{A}\,\overline{B}\,\overline{C} + \overline{A}B\overline{C} + A\overline{B}\,\overline{C} + AB\overline{C}$

Section C: Long Answer Type Questions

Q.11 Convert the following Boolean expression into standard POS form.

$(A + \overline{B} + C)(\overline{B} + C + \overline{D})(A + \overline{B} + \overline{C} + D)$

Q.12 Determine the truth table for the following standard POS expression:

$(A + B + C)(A + \overline{B} + C)(\overline{A} + B + \overline{C})(\overline{A} + B + \overline{C})(\overline{A} + \overline{B} + C)$

Q.13 Convert the following Boolean expression into standard SOP form:

$A\overline{B}\,C + \overline{A}\,\overline{B} + AB\overline{C}\,D$

Q.14 Convert the following SOP Boolean Expression in POS form.

$Y = AB\overline{C} + \overline{A}\,BC + \overline{A}\,B\overline{C} + \overline{A}\,\overline{B}\,\overline{C}$

And draw its logical diagram.

Q.15 Draw logical diagram of following expression using only NAND gate.

$Y = \overline{A}\,\overline{B}C + \overline{A}B\overline{C} + \overline{A}BC + AB\overline{C}$

And convert it in standard SPOS form.

Reference and Suggested Readings

1. G. K. Kharate, Digital Electronics, Oxford University Press, 2013.
2. Dr. S. S. Rawat and Ashok Nagawat, Digital Electronic Circuits, College Book House, 2010.
3. Dr. S. N. Dolia, P. K. Sharma and Gaurav Saxena, Digital Electronics and circuit, Genius Publications, 2009.
4. R. P. Jain, Modern Digital Electronics, S. Chand and company, 2013.

UNIT-17
Karnaugh Map (K-Map)

Structure of the Unit

17.0 Objectives

17.1 Introduction

17.2 Mapping of SOP or POS expressions.

17.3 Minimization of SOP or POS expression

17.4 Self Learning Exercise-I

17.5 Mapping of k-map from Truth table

17.6 Don't care combination

17.7 Self Learning Exercise-II

17.8 Summary

17.9 Glossary

17.10 Answers to self Learning Exercises

17.11 Exercise

References and Suggested Readings

17.0 Objectives

This unit introduces the Mapping of SOP and POS expressions, minimization of SOP and POS expression with different variables, and mapping of k-map from truth table. Karnaugh's Map is also useful techniques for reducing the Boolean expression. After completion of this unit, you should be in a position to don't care combination using k-map technique.

17.1 Introduction

In the previous unit we have discussed about the standard form of (SOP or POS) any n-variable logical expressions. Also, these standard forms are known as minterm and maxterm respectively. In this unit we shall discussed about the minimization of a Boolean expression. The karnaugh map or K-map is the most

popular method for minimization. This is a graphical method in which each minterm or maxterm has a specific cell. Again, the K-map is basically graphical representation of the Boolean function (expression) and its minimization is based on the Boolean theorems as discussed in the previous unit.

The K-map can be designed for both the standard SOP (Minterm) and standard POS (Maxterm) logical expression. As we know that there are 2n cells, corresponding to their possible combinations. Hence a specific cell location is fixed for each of minterm or maxterm. The K-map can also designed by using truth table. K-map for two, three and four variables has four, nine and sixteen cells respectively. The simplification or minimization is based on the basic Boolean rule AB + AB' = A in case of minterm representation. The following processes are to be taken in account to minimize the logical (n variable) expression by using K-map.

(A) Mapping of SOP or POS expressions

(B) Minimization of SOP or POS expressions.

17.2 Mapping of SOP or POS Expressions

(a) Mapping of two variable SOP and POS expression

There are four squares or cell in a two variable K-map. Each of cells has a single or unique minterm. In a cell the '1' indicates that the corresponding minterm is included in the output expression, while 'o' indicates that corresponding minterm does not appear in the expression for output. Let us take an example as to map the expression $Y = A\overline{B} + \overline{A}B$. The minterm presented in the given expression are m_2 and m_1. So it can be written as

$Y = m_1 + m_2 = \Sigma m\,(1, 2) = 01 + 10$

Hence, the K-map corresponds to the expression is,

B \ A	0	1
0	0	1
	1	0

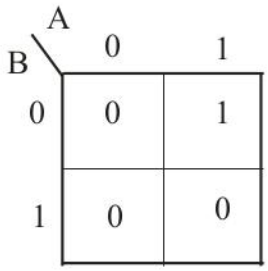
For mapping POS expression on the K-map, the 0s are placed in the cell

corresponding to maxterm which are available in the expression. The 1s are placed in the cell corresponding to the maxterm which are not available in the expression. Since each term of canonical produce of sums expression is known as a maxterm. For two variable, there are four possible maxterm i.e. M_0, M_1, M_2, and M_3 and their expression are A+B, A+$\overline{B}$, $\overline{A}$ +B and, $\overline{A}+\overline{B}$ respectively. Let us take an example as to map the expression Y = (A+B). $\left(\overline{A}+B\right).\left(A+\overline{B}\right)$. In this expression, the M_0, and M_1 and M_2 are present, so,

$$Y = M_0 . M_1 . M_2$$

$$= \Pi M\ (0, 1, 2)$$

Hence, the K-map corresponds to this expression is,



(b) Mapping of three variable SOP and POS expression

A three variable K-map has eight cells and each cell represent the minterm (SOP) or maxterm (POS). The eight possible combinations corresponds to minterm, m_0, m_1, m_2, m_3, m_4, m_5, m_6 and m_7 are $\overline{A}\overline{B}\overline{C}$, $\overline{A}\overline{B}C$, $\overline{A}B\overline{C}$, $\overline{A}BC$, $A\overline{B}C$, $A\overline{B}\overline{C}$, $AB\overline{C}$ and ABC respectively. In case of POS form, the eight possible combinations corresponds to maxterm, M_0, M_1, M_2, M_3, M_4, M_5, M_6 and M_7 are (A+B+C), (A+B+$\overline{C}$), (A+$\overline{B}$+C), (A+$\overline{B}$+$\overline{C}$), ($\overline{A}$+B+C), ($\overline{A}$+B+$\overline{C}$), ($\overline{A}$+$\overline{B}$+C) and ($\overline{A}+\overline{B}+\overline{C}$) respectively. The marking of minterm and maxterm is similar to those of two variables K-map. (0s marked for unavailable minterm and 1s marked for available minterm, while 0s marked for available maxterm and 1s marked for unavailable maxterm.). The decimal number of the cell for maxterm and minterm representation is the same, now, let us take an example of each case.

$$Y = \overline{A}\,B\,\overline{C} + \overline{A}\overline{B}\,C + \overline{A}\overline{B}\overline{C} + A\,B\,\overline{C} + A\,B\,C.$$

The available minterms are $\overline{A}\,B\,\overline{C}$ = 010 = m_2; $\overline{A}\overline{B}\,C$ = 001 = m_1, = 000 $\overline{A}\overline{B}\overline{C}$ = m_0, $AB\,\overline{C}$ = 110 = m_6 and ABC = 111= m_7.

Hence, the K-map will be,

$$Y = m_0 + m_1 + m_6 + m_7 = \Sigma m\ (0, 1, 6, 7)$$

C \ AB	00	01	11	10
0	1	1	1	0
1	1	0	1	0

For $Y = (\overline{A} + \overline{B} + C).\ (\overline{A} + B + \overline{C}).\ (A + \overline{B} + C).\ (A + B + C)$

$= M_6.\ M_5.\ M_2.\ M_0$

$= \Pi M\ (0, 2, 5, 6)$

So, the K-map (POS)

C \ AB	00	01	11	10
0	0	0	0	1
1	1	1	1	0

(c) Mapping of four variable SOP and POS expression

A four variable K-map has sixteen cells and each cell of map represents either maxterm (POS) or minterm (SOP). The binary number of rows and columns are the gray code. The conditions of A and B along the column are marked by the binary number along the top of the map. Similarly, the conditions of C and D along the row are marked by the binary number along the left side of the map.

The sixteen possible combinations correspond to minterms, $m_0, m_1, m_2, \ldots\ldots m_{15}$ are $\overline{A}\overline{B}\overline{C}\overline{D}$, $\overline{A}\overline{B}\overline{C}D$, ….. ABCD respectively. For POS form the maxterm $M_0, M_1,$ ….. M15 are (A+B+C+D), $(A+B+C+\overline{D})$ …. $\left(\overline{A} + \overline{B} + \overline{C} + \overline{D}\right)$ respectively. Now, if we want to map the following expression

$$Y = \Sigma m\ (0, 1, 3, 5, 8, 11, 14)$$

then, the K-map will be,

CD \ AB	00	01	11	10
00	0 1	4 0	12 0	8 1
01	1 1	5 1	13 0	9 0
11	3 1	7 1	15 0	11 1
10	2 0	6 0	14 1	10 0

For, Y = ΠM (2, 6, 7, 9, 11, 13, 14, 15), the K-map will be,

CD \ AB	00	01	11	10
00	0 1	4 1	12 1	8 1
01	1 1	5 1	13 0	9 0
11	3 1	7 0	15 0	11 0
10	2 0	6 0	14 0	10 1

17.3 Minimization of SOP or POS Expression

After mapping the K-map, it is required to minimize the expression.

(a) (i) Two variable SOP

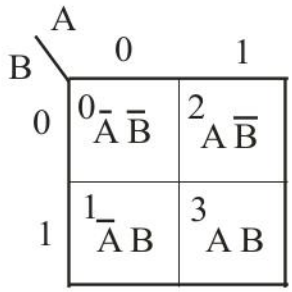
In order to minimize a given Boolean expression by using karnaugh map, one should have to look carefully for adjacent (neighboured) cells having 1s. It means that the two or more adjacent minterms, when combine together gives the elimination of one or more variables. Two adjacent cells are said to be combined to each other when their minterms differ by single variable. Let us take an example in

case of two variable K-map the minterms m_0 and m_2 i.e. $\overline{A}\,\overline{B}$ and $A\overline{B}$ differ by single variable A ($\overline{A}$ in m_0 & A in m_2). Since $\overline{B}$ is common in both the m_0 and m_2, hence the minterms m_0 & m_2 may be combined to form a pair that eliminates the variable A as given below.

$$Y = m_0 + m_2 = \overline{A}\,\overline{B} + A\overline{B}$$

$$= (\overline{A} + A) \,.\, B$$

$$= 1 \,.\, \overline{B} = \overline{B} \qquad \therefore \text{ (Elimination of variable A)}$$



Further, the following groups are possible to form a pair and eliminate a variable,

(i) The minterm m_0 and m_1 can be combined together to form a pair and yield to = $m_0 + m_1 = \overline{A}\,\overline{B} + \overline{A}\,B = \overline{A} = Y_1$ [Fig. 17.1 (a)]

(ii) The minterm m_0 and m_2 can be combined together to form a pair and yield to $(m_0, m_2) = m_0 + m_2 = A\overline{B} + A\overline{B} = \overline{B} = Y_2$ [Fig. 17.1 (b)

(iii) The minterm m_1 and m_3 can be combined together to form a pair and yield to $(m_1, m_3) = m_1 + m_3 = \overline{A}\,B + A\,B = B = Y_3$ [Fig. 17.1 (c)]

(iv) The minterm m_2 and m_3 can be combined together to form a pair and yield to $(m_2, m_3) = m_2 + m_3 = A\overline{B} + A\,B = A = Y_4$ [Fig. 17.1 (d)]

Here, it is to be noted that the minterm m_0 ($\overline{A}\,\overline{B}$) and the minterm m_3 (AB) does not form a pair. It is also applicable in care of minterm m_1 ($\overline{A}\,B$) and the minterm m_3 (AB) does not form a pair. It is also applicable in care of minterm m_1 ($\overline{A}\,B$) and minterm m_2 ($A\overline{B}$). However, all the four minterm can be combined to form a Quad and yield to the output as logic 1, as give below, [Fig. 17.1 (e)]

$$Y_5 = \Sigma m\,(0, 1, 2, 3) = m_0 + m_1 + m_2 + m_3$$

$$= \overline{A}\,\overline{B} + \overline{A}\,B + A\overline{B} + AB$$

$= \overline{A}(\overline{B}+B) + A(\overline{B}+B)$ $\therefore$ (Elimination of B)

$= \overline{A} + A = 1$ $\because$ (Elimination of A)

Now, it is clear that the two adjacent pair can be combined to each other and form a quad that eliminates both the variables. The combination of two or more neighboured minterm to form a pair or quad that yield to eliminate the one or both the variables is called grouping of cells in K-map. The following figure shows the possibilities of grouping in the two variables K-map.

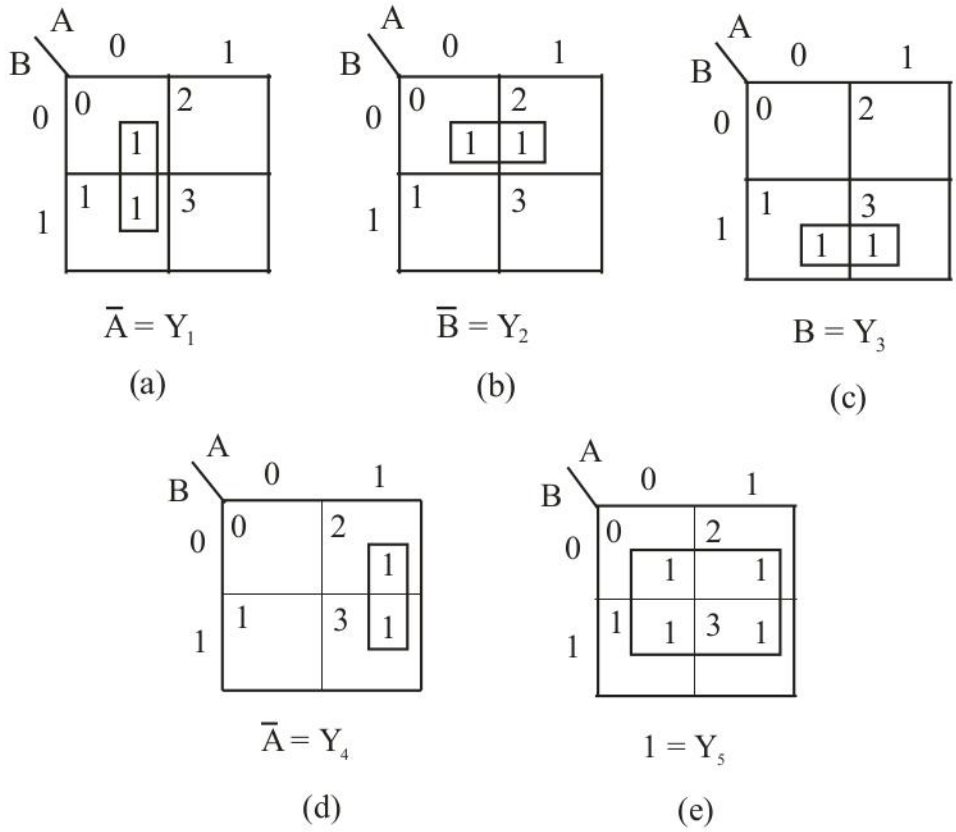


Fig. 17.1: Grouping of minterms for two variable K-map

The Y_1, Y_2, Y_3, Y_4 and Y_5 are the reduced output depending upon the grouping of different cells.

(ii) **Two variable POS: -** As we see in the previous article that groups are form the pairs and quads. These are the direct evidence of minimization by using Boolean theorems. Instead to see 1s in SOP expression, here we see carefully 0s in the K-map to form a pair or quad. Here we take an example for two variable K-map, the maxterm M_0 and M_2 i.e. (A+B) and ($\overline{A}$ +B) are differ by one variable A (A in M_0

and $\overline{A}$ in M_2). Since B is common in both the M_0 and M_2, hence the maxterms M_0 and M_2 may be combined to form a pair that eliminates the variable A. In the following manner,

$Y = M_0 . M_2$

$= (A+B) . (\overline{A} + B)$

$= A\overline{A} + AB + B\overline{A} + BB$ (Elimination of A)

$= 0 + B(A + \overline{A}) + B$

$= 0 + B = B$

Further, the following groups are possible to form a pair and eliminate a variable.

B \ A	0	1
0	0 $A+B$	2 $\overline{A}+B$
1	1 $A+\overline{B}$	3 $\overline{A}+\overline{B}$

(i) In maxterm M_0 and M_1 can be combined together to form a pair and yields to = $M_0 . M_1 = (A+B) . (A+\overline{B}) = A = Y_1$ [Fig. 17.2 (a)]

(ii) In maxterm M_0 and M_2 can be combined together to form a pair and yields to = $(M_0, M_2) = (A+B) . (\overline{A} + B) = B = Y_2$ [Fig. 17.2 (b)]

(iii) In maxterm M_1 and M_3 can be combined together to form a pair and yields to $(M_1 . M_3) = (\overline{A} + B) . (\overline{A} + \overline{B}) = \overline{B} = Y_3$ (Fig. 17.2 (c)]

(iv) In maxterm M_2 and M_3 can be combined together to form a pair and yields to = $(M_2, M_3) = (\overline{A} + B) . (\overline{A} + \overline{B}) = \overline{A} = Y_4$ [Fig. 17.2 (d)]

The maxterms M_0 and M_3 and M_1 and M_2 does not form a pair, However all the maxterms M_0, M_1, M_2 and M_4 can be combined to form a quad and yields to the output as logic 0 as given below. [Fig. 17.2 (e)]

$Y_5 = \prod M (0, 1, 2, 3)$

$= M_0 . M_1 . M_2 . M_3$

$= (A + B) . (A + \overline{B}) . (\overline{A} + B) . (\overline{A} + \overline{B})$

$=(A.A. + A\overline{B} + BA + B\overline{B})\,(\overline{A}\;\overline{A} + \overline{A}\;\overline{B} + B\overline{A} + B\overline{B})$

$=[A + A(B + \overline{B})].[\overline{A} + \overline{A}\;(B+B)]$

$=(A+A).(\overline{A} + \overline{A})$ $\quad \because$ (Elimination of B)

$=A.\overline{A} = 0$ $\quad \because$ (Elimination of A)

Similar to SOP minimization, the pairs and quads can directly, be eliminates the variables in POS minimization. In addition to this, here one important observation is that, the combination of maxterms (in the form of pairs or quads) gives the complementary output as in case of the combination of minterms. The following figure shows the possible grouping in POS for two variable K-map.

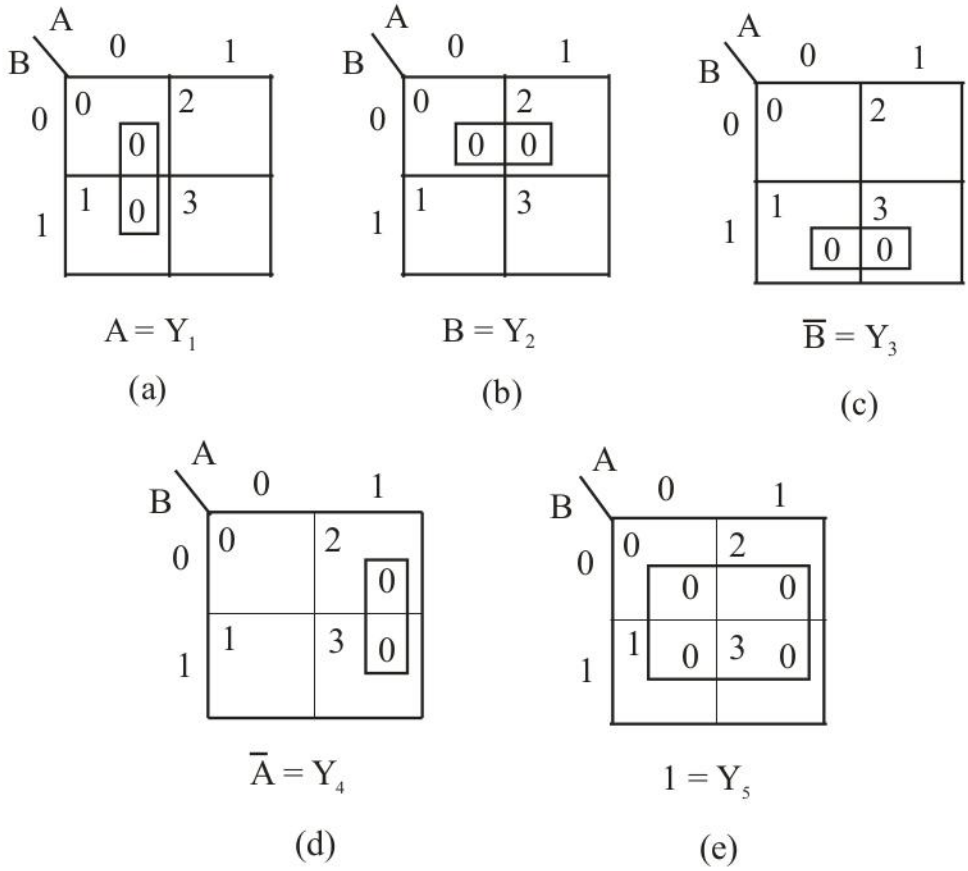


Fig. 17.2 Grouping of Maxterms for two variables K-map

The Y_1, Y_2, Y_3, Y_4 and Y_5 are the reduced output depending upon the grouping of different cells.

(b) **Three variable SOP & POS: -** In three variable K-map, the number of pairs and quads may be increased depending upon the minterm or maxterms present in a

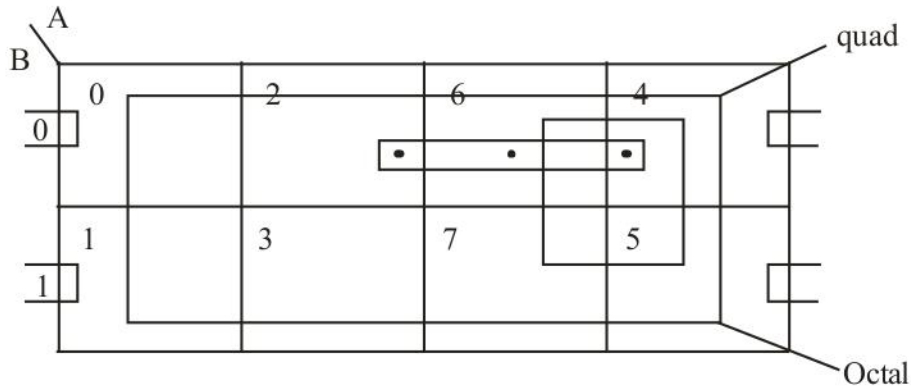
given expression. To minimize the given expression. First prepare the K-map and than see the grouping in the form of adjacent 1s (SOP) and 0s (POS). In a three variable K-map, there are eight cells and each minterm or maxterm have all the three variables. When two minterm of maxterm combine together, they will form a pair. It results to the elimination of one variable. Combination of four minterms or maxterm results the elimination of two variables. The following possibilities are in three variables K-map.

(i) **Pair: -** A group of two adjacent 1s (SOP) or 0s (POS)

(ii) **Quad: -** A group of four adjacent 1s (SOP) or 0s (POS)

(iii) **Octal: -** A group of eight adjacent 1s (SOP) or 0s (POS).

It is to be noted that grouping of variables is either horizontal or vertical but never be diagonal.



The pair formed by the combination of minterm or maxterm (0, 4), (1, 5) (Leftmost & right most) are known as fold back pairs.

Simplification with pair, quads and octals, (SOP)

(a) Pair: Let in a expression m_2 and m_6 minterms are combine to each other than,

$$Y = \Sigma m\ (2, 6)$$

$$Y = m_2 + m_6$$

$$= \overline{A}\, B \overline{C} + AB \overline{C}$$

$$= (\overline{A} + A)\, B \overline{C}$$

$$Y = B \overline{C} \qquad \because \text{(Elimination of A)}$$

The other pairs can be simplifying in the same way.

(b) **Quad:** Let in a expression the minterms m_0, m_1, m_4 and m_5 are combine together. Since it is an open folded back quad. So,

$$Y = \Sigma m\,(0, 1, 4, 5)$$

$$Y = m_0 + m_1 + m_4 + m_5$$

$$= \overline{A}\,\overline{B}\,\overline{C} + \overline{A}\,\overline{B}\,C + A\,\overline{B}\,\overline{C} + A\,\overline{B}\,C$$

$$= \overline{A}\,\overline{B}\,(\overline{C}+C) + A\,\overline{B}\,(\overline{C}+C) \qquad \because \text{(Elimination of A)}$$

$$= \overline{A}\,\overline{B} + A\,\overline{B}$$

$$= (\overline{A}+A)\,\overline{B} = \overline{B}$$

Hence, it is clear that a quad always gives a single variable. The other quads can be simplified in the same way.

(c) **Octal:** In which all the minterms combine together and gives the output of expression as logic 1. It gives the elimination of all the three variables.

$$Y = \Sigma m\,(0, 1, 2, 3, 4, 5, 6, 7)$$

$$= m_0 + m_1 + m_2 + m_3 + m_4 + m_5 + m_7$$

$$= \overline{A}\,\overline{B}\,\overline{C} + \overline{A}\,\overline{B}\,C + \overline{A}\,B\,\overline{C} + \overline{A}\,BC + A\,\overline{B}\,\overline{C} + A\,\overline{B}\,C + AB\,\overline{C} + ABC$$

$$= \overline{A}\,\overline{B}\,(\overline{C}+C) + \overline{A}\,B(\overline{C}+C) + A\,\overline{B}\,(\overline{C}+C) + AB\,(\overline{C}+C)$$

$$= \overline{A}\,\overline{B} + \overline{A}\,B + A\,\overline{B} + AB$$

$$= \overline{A}\,(\overline{B}+B) + A\,(\overline{B}+B)$$

$$= \overline{A} + A = 1$$

Simplification of logical expression with pairs, quads and octals (POS)

(a) **Pair:** Let in an expression M_2 and M_6 maxterms combine together then one variable will be omitted from the expression as follows,

$$Y = \Pi M\,(2,6)$$

$$= M_2 . M_6$$

$$= (A + \overline{B} + C).\,(\overline{A} + \overline{B} + C)$$

$$= A\,\overline{A} + A\,\overline{B} + AC + \overline{B}\,\overline{A} + \overline{B}\,\overline{B} + \overline{B}\,C + C\,\overline{A} + C\,\overline{B} + CC$$

$$= 0 + A\,\overline{B} + AC + \overline{A}\,\overline{B} + \overline{B}\,C + \overline{A}\,C + \overline{B}\,C + C + \overline{B}$$

$$=A\overline{B} + AC + \overline{A}\,\overline{B} + \overline{B}\,(1+C) + \overline{A}\,C + C\,(\overline{B}+1)$$

$$=A\overline{B} + AC + (\overline{A}+1)\,\overline{B} + (\overline{A}+1)\,C$$

$$=\overline{B}\,(1+A) + (1+A)\,C$$

$$=\overline{B} + C$$

(b) **Quad:** In the same way the quad formed by the maxterms, M_0, M_1, M_4, M_5 gives the expression in one variable (Elimination of two variable)

$$Y = \prod M\,(0, 1, 4, 5)$$

$$=(A + B + C)\,.\,(A+B+\overline{C})\,.\,(\overline{A}+B+C).\,(\overline{A}+B+\overline{C})$$

$$=B$$

The other pairs and quads can be solving in the same way.

(c) **Octal:** To form a octal, all the either maxterms combine together and give the result as logic 0, here, all the three variables eliminates.

$$Y = \prod M\,(0, 1, 2, 3, 4, 5, 6, 7)$$

$$=M_0\,.\,M_1.\,M_2.\,M_3.\,M_4.\,M_5.\,M_6.\,M_7$$

$$=0$$

The output of the expression with the combination of minterms (SOP) is complimentary to the combination of maxterms (POS)

(c) Four Variable SOP & POS

There are sixteen cells in a four variable K-map. The possibility of pairs, quads and octals is more. But again it depends upon the minterms or maxterms available in a given expression. The grouping of adjacent cells to form a pair, quad or octal is similar as discussed for two or three variable K-map. Since, in a four variable K-map there are four variables A, B, C, D and the minterms or maxterm have all the variables in either complemented or uncomplemented form. The pair reduced the expression by one variable, the quad reduced the expression by two variables and octal reduced the expression by three variables. In a special care, while all the minterms are present in the expression that leads to the output as logic 1 and when all the maxterm are present the output of the expression goes to logic 0.

Here, the number of folebd back pairs, quads and octals is also more.

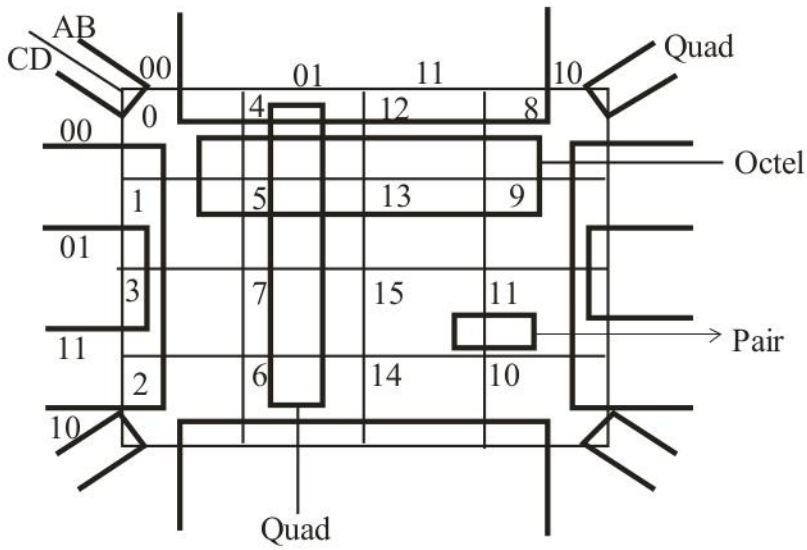


Fig. 17.3: Various possibilities of pairs, quads of octal in a four variable K-map.

In a three variable K-map, there is only one possibility of octal (all the minterm or maxterms combine together). Instead of this, in a four variable K-map. There are eight possibilities to form an octal. They are as follows.

Simplifications of four variables K-map (SOP & POS)

(a) **Pair -** A large number of possibilities to form a pair in four variables K-map. One pair of minterm or maxterm reduced the expression by one variable.

Let (SOP) $Y = \Sigma m\,(0, 1)$

$= m_0 + m_1$

$= \overline{A}\,\overline{B}\,\overline{C}\,\overline{D} + \overline{A}\,\overline{B}\,\overline{C}\,D$

$= \overline{A}\,\overline{B}\,\overline{C}\,(\overline{D} + D)$

$= \overline{A}\,\overline{B}\,\overline{C}$ (Elimination of D)

For POS $Y = \pi M\,(0, 1)$

$= M_0 . M_1$

$= (A + B + C + D).(A + B + C + \overline{D})$

$= (A + B + C) + D\,\overline{D}$

(b) **Quad:** A quad reduced the expression by two variables.

Let (SOP) $Y = \Sigma m\,(0, 1, 2, 3)$

$= m_0 + m_1 + m_2 + m_3$

$= \overline{A}\,\overline{B}\,\overline{C}\,\overline{D} + \overline{A}\,\overline{B}\,\overline{C}D + \overline{A}\,\overline{B}CD + \overline{A}\,\overline{B}C\overline{D}$

$= \overline{A}\,\overline{B}\,\overline{C}\,(\overline{D} + D) + \overline{A}\,\overline{B}C\,(D + \overline{D})$

$= \overline{A}\,\overline{B}\,\overline{C} + \overline{A}\,\overline{B}C$

$= \overline{A}\,\overline{B}\,(\overline{C} + C) = \overline{A}\,\overline{B}$

For POS $Y = \prod M\,(0, 1, 2, 3)$

$= M_0 . M_1 . M_2 . M_3 = A + B$

(c) **Octal:** An octal reduced the expression by three variables.

Let (SOP) $Y = \Sigma m\,(0, 1, 3, 2, 4, 5, 7, 6)$

$= m_0 + m_1 + m_3 + m_2 + m_4 + m_5 + m_7 + m_6$

$= \overline{A}\,\overline{B}\,\overline{C}\,(\overline{D} + D) + \overline{A}\,\overline{B}C\,(\overline{D} + D) + \overline{A}B\overline{C}\,(\overline{D} + D) + \overline{A}BC\,(\overline{D} + D)$

$= \overline{A}\,\overline{B} + \overline{A}B = \overline{A}$

For POS, $Y = \prod M\,(0,1,3,2,4,5,7,6)$

$= M_0 . M_3 . M_2 . M_4 . M_5 M_7 . M_6$

$= A$

For all minterms (SOP)

$Y = \prod M\,(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15)$

$= m_0 + m_1 + \ldots m_{15}$

$Y = 1$

For all maxterms (POS)

$Y = \Sigma M\,(0, 1, 2, \ldots\ldots 15)$

$= M_0 . M_1 \ldots\ldots M_{15}$

$Y = 0$

17.4 Self Learning Exercise-I

Q.1 Which in not a minimization technique for logical expression?

Q.2 Write the two variable of K-map in terms of minterm.

Q.3 Write the three variables for K-map.

Q.4 Write the four variables of K-map in terms of minterm.

Q.5 Write the four variables of K-map in terms of maxterm.

Q.6 Write the two variables of K-map in terms of maxterm.

17.5 Mapping of K-map from Truth table

For a given truth table, the output of logical state 1 corresponds to various combination of variables is available. These combinations are also available in K-map. So for mapping a K-map by using given truth table we have to mark 1 in the corresponding cell number in case of SOP expression. The remaining cells are leave blank or marks with 0 (zero). The pairs, quad and octals are then marked by using adjacent ones in SOP K-map.

For POS K-map, the available maxterm is marked with 0s in the K-map.

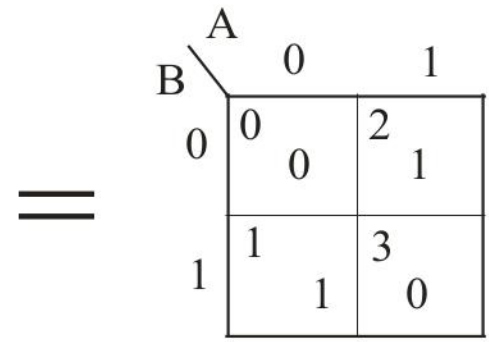
$$Y = A\overline{B} + \overline{A}B \text{ (SOP)}$$

The cell number 0 and 3 are marked with 0 and cell number 1 and 2 marked with 1.

Let $Y = A\overline{B} + \overline{A}B$ (SOP)

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

Truth table



K-map

$$Y = (A + \overline{B}).(\overline{A} + B) \text{ (POS from)}$$

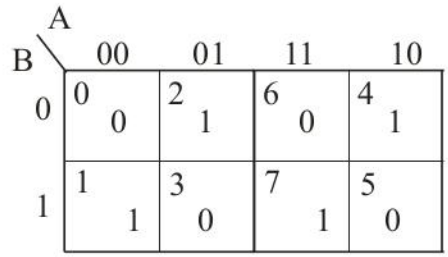
As we know that the minterm is marked with 1s and maxterm is marked with 0s.

For, $Y = \overline{A}\overline{B}C + \overline{A}B\overline{C} + A\overline{B}\overline{C} + ABC$ (SOP)

$$Y(A, B, C) = m_1 + m_2 + m_4 + m_7$$

Table - 1 Truth Table and K-map (Three Variables)

Cell No.	A	B	C	Y
0	0	0	0	0
1	0	0	1	1
2	0	1	0	1
3	0	1	1	0
4	1	0	0	1
5	1	0	1	0
6	1	1	0	0
7	0	1	1	1



K-Map

Pos form of the above expressing will be,

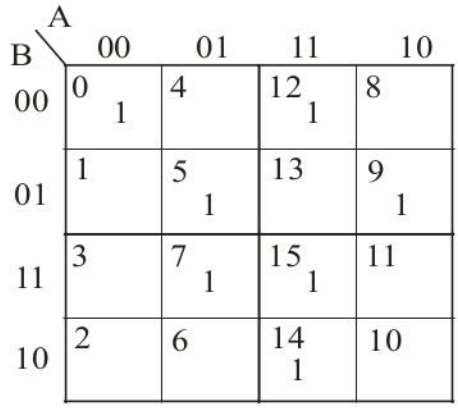
$$Y = (A+B+C) \cdot (A+\overline{B}+\overline{C}) \cdot (\overline{A}+B+\overline{C}) \cdot (\overline{A}+\overline{B}+C)$$

These maxterms will be marked with zeros. For the following truth table we need to enter either zeros or ones. In a K-map, it is understood that if only ones are entered, then remaining cells are empty and they all are treated as zeros. If only zeros are entered, then remaining cells are empty and they all are treated as ones.

Table: 2 Truth Table and K-map (Four variables)

Cell No.	Inputs				Output
	A	B	C	D	Y
0	0	0	0	0	1
1	0	0	0	1	0
2	0	0	1	0	0
3	0	0	1	1	0
4	0	1	0	0	0
5	0	1	0	1	1
6	0	1	1	0	0
7	0	1	1	1	1
8	1	0	0	0	0
9	1	0	0	1	1
10	1	0	1	0	0
11	1	0	1	1	0
12	1	1	0	0	1
13	1	1	0	1	0
14	1	1	1	0	1
15	1	1	1	1	1

=



K-Map

The value of output Y (either 0 or 1) for each cell number or row of truth table must be entered in the corresponding cells of the K-map.

Simplification by K-map

As we discussed earlier when one or more variables appear in both the complemented and uncomplemented form within a group, than these variables can be omitted. However following steps are to be taken in account to simplify the expression using K-map.

Step -1: First of all construct a K-map for a given number of variables.

Step-2: Now, enter the ones in those cells for which the combination is available in the expression.

Step-3: Check the map and marked these ones which cannot combined to form a group. These are singles.

Step-4: Now, look upon the adjacent ones that grouped in the form of pair, quad or octal.

Step-5: Use only minimum number of groups, for the first sees about octal, then quad and finally pair. It is possible of over lapping of groups for common ones.

Step-6: Now omit or eliminate the repeating groups.

Step-7: At least write down the expression of logical sums for each group.

17.6 Don't Care Combination

Certain logical expressions, in which the combinations of variables are not specified and hence these unspecified combinations of variables are called don't care combinations. Apart from the completely or precisely minterms of maxterms, there can be some minterms or maxterm for which the output is of no important consequence. Therefore the don't care combinations are stands for incompletely specified combinations of variables. The don't care conditions or combinations are generally denoted either by d (small D) or x (cross) or ϕ (Phi). For simplification using K-map, the don't care combinations are treated as logic 1, if it is useful to reductions of the expression or otherwise it may be treated as logic 0 and left behind in a sum-of-products (SOP) K-map. Further these don't' care may be treated as logic 0, if it is useful to reduction of the expression or otherwise may be treated

as logic 1 in a product of sums (POS) K-map.

Now, it is clear that the don't care combination may be taken either 1 or 0 depending upon the simplicity of the process. An SOP expression with don't care can be converted into a POS expression and vice-versa. The following example can help you about better understanding don't care combinations.

Example:- Write the logical expression for the following truth table and (a) simplify it using K-map (b) Draws the logic diagram of the result.

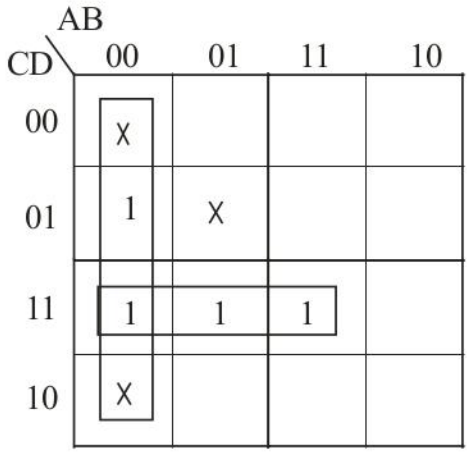
Cell No.	**Variables**				**Output**
	A	**B**	**C**	**D**	**Y**
0	0	0	0	0	×
1	0	0	0	1	1
2	0	0	1	0	×
3	0	0	1	1	1
4	0	1	0	0	0
5	0	1	0	1	×
6	0	1	1	0	0
7	0	1	1	1	1
8	1	0	0	0	0
9	1	0	0	1	0
10	1	0	1	0	0
11	1	0	1	1	1
12	1	1	0	0	0
13	1	1	0	1	0
14	1	1	1	0	0
15	1	1	1	1	1

Solution: The logical expression of the above truth table will be, Y (A, B, C, D) = $\Sigma m\,(1, 3, 7, 11, 15) + \Sigma d\,(0, 2, 5)$

Because the minterm m_1, m_3, m_7, m_{11} and m_{15} are specified completely, while minterms m_0, m_2 & m_{15} are a specified, so they are don't care combinations. So the K-map of the above expression will be constructed by marking in the specified minterms and the corresponding cell numbers are 1, 3, 7, 11 and 15. Also the cell number corresponding to don't care combinations as 0, 2, 5 are marked with × (cross). By assuming these crosses as 1, there are two quads. First is formed by cells 3, 7, 11 and 15 second is formed by cells 0, 1, 2 and 3. Now all the ones have been covered and hence, the reduced expression will be,

Quad 1

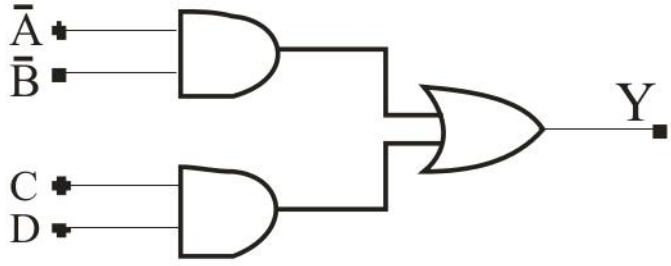
$$Y(A, B, C, D) = m_0 + m_1 + m_2 + m_3$$
$$= \overline{A}\,\overline{B}$$



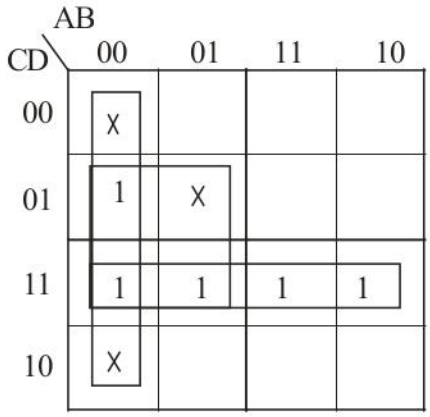
And Quad 2

$$Y(A, B, C, D) = m_3 + m_7 + m_{11} + m_{15}$$
$$= C\,D$$

Hence, $Y = \overline{A}\,\overline{B} + CD$ ….. (1)



The above problem can also be solved as,



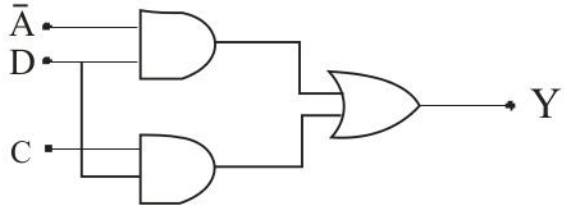
Quad 1 $Y(A, B, C, D) = m_1 + m_3 + m_5 + m_7$

$= \overline{A}\ D$

Quad 2 $Y(A, B, C, D) = m_3 + m_7 + m_{11} + m_{15}$

$= CD$

so, $Y = \overline{A}\,D + CD$ …. (2)



Both the results (1) and (2) are possible, it depends upon the selection of quads.

17.7 Self Learning Exercise-II

Q.1 Plot the following logical expression on K-map

$Y(A, B, C, D) = \Sigma m\ (0, 2, 5, 7, 10, 13, 15)$

Q.2 Simply the following logical expression.

$Y(A, B, C, D) = \Sigma m\ (3, 4, 5, 7, 9, 13, 14, 15)$by using K-map method.

Q.3 Simplify $Y(A, B, C, D) = \Sigma m\ (0, 1, 3, 7)$

Q.4 How many cell in a 4 - variable K-map?

Q.5 How many n- variable of K-map?

Q.6 Simplify $Y = \Sigma m\ (0,3)$

17.8 Summary

K-map is diagram, made up of squares (Cells) to minimize the Boolean expression. For n-variable K-map, the decimal codes to corresponding cells are given from 0 to (2^n-1). To simplifying the K-map, the looping of two adjacent cells can be done in form of pair, quad or octal.

17.9 Glossary

K-Map: A diagram consisting of a rectangular array of squares each representing a different combination of the variables of a Boolean function.

Boolean Variable : A Boolean variable may be regarded as a numerical variable with a single binary digit which can store only two values.

Overlapping Groups: Groups of ones or zeroes in a K-map that share cells.

Truth Table: A truth table tabulates for a particular algebraic expression/ Logic operation, the output values corresponding to all possible input combinations.

17.10 Answers to Self Learning Exercises

Answers to Self Learning Exercise-I

Ans.1: Gate realization

Ans.2:

B \ A	0	1
0	$\bar{A}\,\bar{B}$	$A\,\bar{B}$
1	$\bar{A}\,B$	$A\,B$

Ans.3:

C \ AB	00	01	11	10
0	0	2	6	4
1	1	3	7	5

Ans.4:

CD \ AB	00	01	11	10
00	$\bar{A}\bar{B}\bar{C}\bar{D}$	$\bar{A}B\bar{C}\bar{D}$	$AB\bar{C}\bar{D}$	$A\bar{B}\bar{C}\bar{D}$
01	$\bar{A}\bar{B}\bar{C}D$	$\bar{A}B\bar{C}D$	$AB\bar{C}D$	$A\bar{B}\bar{C}D$
11	$\bar{A}\bar{B}CD$	$\bar{A}BCD$	$ABCD$	$A\bar{B}CD$
10	$\bar{A}\bar{B}C\bar{D}$	$\bar{A}BC\bar{D}$	$ABC\bar{D}$	$A\bar{B}C\bar{D}$

Ans.5:

CD \ AB	00	01	11	10
00	A+B+C+D	A+$\bar{B}$+C+D	$\bar{A}$+$\bar{B}$+C+D	$\bar{A}$+B+C+D
01	A+B+C+$\bar{D}$	A+$\bar{B}$+C+$\bar{D}$	$\bar{A}$+$\bar{B}$+C+$\bar{D}$	$\bar{A}$+B+$\bar{C}$+$\bar{D}$
11	A+B+$\bar{C}$+$\bar{D}$	A+$\bar{B}$+$\bar{C}$+$\bar{D}$	$\bar{A}$+$\bar{B}$+$\bar{C}$+$\bar{D}$	$\bar{A}$+B+$\bar{C}$+$\bar{D}$
10	A+B+$\bar{C}$+D	A+$\bar{B}$+$\bar{C}$+D	$\bar{A}$+$\bar{B}$+$\bar{C}$+D	$\bar{A}$+B+$\bar{C}$+$\bar{D}$

Ans.6:

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

Answers to Self Learning Exercise-II

Ans.1:

CD \ AB	$\overline{A}\,\overline{B}$	$\overline{A}B$	AB	$A\overline{B}$
$\overline{C}\,\overline{D}$	0 1	4	12	8
$\overline{C}D$	1	5 1	13 1	9
$C D$	3	7 1	15 1	11
$C\overline{D}$	2 1	6	14	10 1

Ans.2: $Y = \overline{A}\,CD + \overline{A}\,B\overline{C} + A\overline{C}\,D + ABC$

Ans.3: $Y = \overline{A}\,\overline{B} + BC$

Ans.4: 16 Cells

Ans.5: 2^n Cells

Ans.6: $\overline{A}\,\overline{B} + AB$

17.11 Exercise

Section A : Very Short Answer Type Questions

Q.1 What do you mean by K-map?

Q.2 Which symbol is not use for don't care combination?

Q.3 What do you understand by don't care combination?

Q.4 What is use of K-map in circuit designing?

Q.5 Define redundant group in a given K-map?

Section B: Short Answer Type Questions

Q.6 Simplify the following expression by using K-map.

$Y(A, B, C, D) = \Sigma m\,(1, 3, 5, 7, 8, 9, 12, 13)$

Q.7 Plot the following logical expression

$Y(A, B, C, D) = ABCD + A\overline{B}\,\overline{C}\,\overline{D} + A\,\overline{B}\,C + AB.$

on a four variable K-map. Obtain its simplified expression.

Q.8 Plot the following logical expression on K-map

$Y(A, B, C, D) = \prod M (0, 1, 4, 7, 11, 13, 14)$

Q.9 Simplify the expression

$Y(A, B, C, D) = \prod M (0, 1, 4, 5, 6, 8, 9, 12, 13, 14)$

using K-map method.

Section C: Long Answer Type Questions

Q.10 Discuss the grouping of maxterms or minterms in K-map.

Q.11 Explain the following terms

(i) Pair (ii) Quad (iii) Octal

Q.12 Make a K-map for the following functions

$Y = ABC + \overline{A}BC + \overline{B}\,\overline{C}$

Q.13 Given the logical expression

$Y = ABC + B\overline{C}D + \overline{A}BC$

(a) Make a truth table.

(b) Simplify using K-map

(c) Realize Y using NAND gates only.

Q.14 Simplify the expression

$Y(A, B, C, D) = \Sigma m (1, 5, 10, 11, 12, 13, 15)$

using K-map method.

References and Suggested Readings

1. G. K. Kharate, Digital Electronics, Oxford University Press, 2013.
2. Dr. S. S. Rawat and Ashok Nagawat, Digital Electronic Circuits, College Book House, 2010.
3. Dr. S. N. Dolia, P. K. Sharma and Gaurav Saxena Digital Electronics and circuit, Genius Publications, 2009.
4. R. P. Jain, Modern Digital Electronics, S. Chand and Company, 2013.

UNIT-18
Digital Logic Families

Structure of the Unit

18.0 Objectives

18.1 Introduction

18.2 Characteristics of Digital IC's

18.3 Register Transistor Logic (RTL)

18.4 Diode Transistor Logic (DTL)

18.5 Self-Learning Exercise-I

18.6 Introduction of Unipolar logic (MOS)

18.7 CMOS Logics

18.8 TTL Circuits

18.9 Self - learning Exercise-II

18.10 Summary

18.11 Glossary

18.12 Answer to self learning Exercises

18.13 Exercise

References and Suggested Readings

18.0 Objectives

In this unit we study the basics of Digital logic families. The definition of IC's is given at the beginning. You are introduced to the Digital IC's and characteristics of Digital IC's. Basic ideas on introduction of unipolar logics are also covered in this unit. After completion of this unit you should be in a position to explain the CMOS logic and the TTL Circuits.

18.1 Introduction

The digital IC's are categorized as Small Scale integration (SSI) with fewer as a part of the chip, Medium scale integration (MSI) with 10 to 100 gates per chip and

large Scale Interaction (LSI) with more than 100 to 1000 gates per chip. The very large scale interaction (VLSI) has several 1000 gates per chip. IC's make digital systems more reliable by reducing the number of interconnections from one device to another device on the printed circuit board. IC's cannot handle very large currents or Voltages because heat generated in such small spaces would cause temperature to rise beyond acceptable limits.

Digital IC's are classified not only by their level of interaction or logic operation but also by the specific circuit technology. The technology is known as digital logic family. The classification of digital ICs shown in figure.

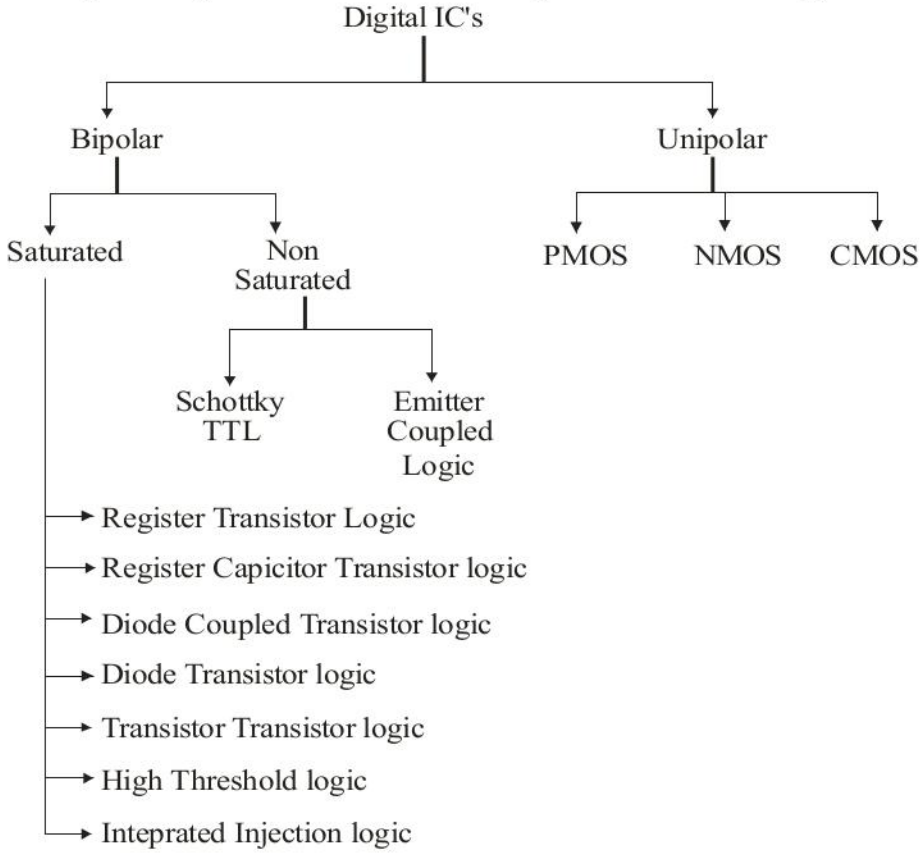


Figure: 18.1 : Classification of Digital IC's

18.2 Characteristics of Digital IC's

With the widespread use of IC's in digital systems and with the development of various technologies for the fabrication of IC's it has become necessary to be families with the characteristic of IC logic families and their relative advantages and disadvantages. Digital ICs are classified their according to the complexity of

the circuit as the relative number of individual basic gates. The various characteristic of digital ICs used to compare their performances are:

(1) Speed of operation

(2) Power dissipation

(3) Figure of merit

(4) Fan-out

(5) Current and voltage parameters

(6) Noise immunity

(7) Operating temperature range

(8) Power supply requirements and

Speed of Operation : The speed of operation of a digital circuit is specified in terms of the propagation delay time. The input and output waveforms of a logic gate is given below. The delay times are measured between the 50%. Voltage level of input and output waveforms. These are two delay times:

t_{PHL} = When the output from the high state to Low state.

t_{PLH} = When the output goes from the low state to High State.

The Propagation delay time of the logic gate is taken as the average of these two delay times.

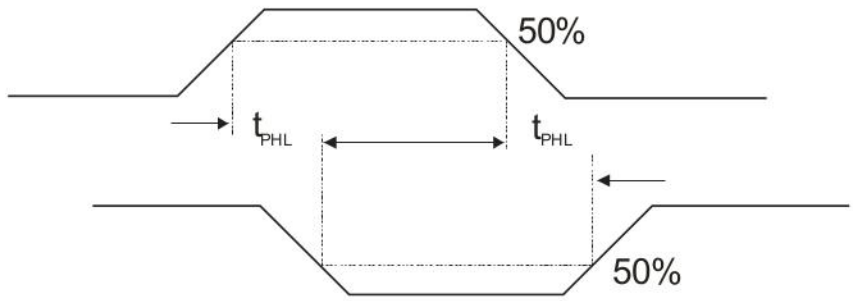


Figure 18.2: Input and output voltage waveforms to define, propagation delay times.

Power Dissipation: This is refers as the power dissipated in an IC. It is determined by the current ICC, that is draws from the VCC supply and the value of its given by VCC × ICC. The ICC is the average value of ICC (0) and ICC (1). This power is specified in milliwatts.

Figure of merit: The figure of merit of a digital IC is defined as the product of speed and power. The speed is specified in terms of propagation delay time expressed in nanoseconds.

Figure of merit = Propagation delay time (ns) × power (mw).

It is specified in pico joules (ns × mw = PJ)

A law value of speed power product is desirable. In a digital circuit, If it is desired to have high speed, i.e. law propagation delay, then there is a corresponding increase in the power dissipation and vice versa.

Fan out: When the output of any logic gate is connected to one or more standard loads. i.e. the other gates of same logic family, it crates the loading on the driver gate. There will be a certain limit to the number of loads that can be driven by output current of the driver gate. This limit of the loads is called Fan-out. It is also known as loading because the output of a gate can supply a limited amount of current. If the loading of the gate is increased them there may be a possibility of overloading and it will not operate properly. Which input that can be connected to the output of the gate. It is expressed by number. It can be calculated from the amount of current available at the output of the gate and the current needed at the input of each load. For example, the circuit shown in Figure 18.3 is in high level state.

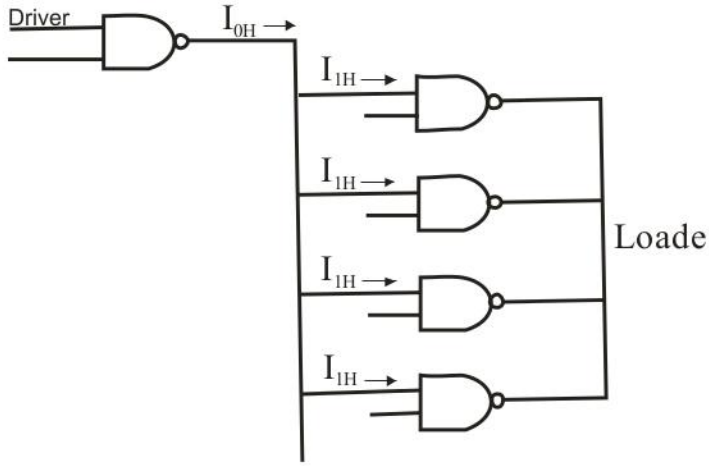


Figure 18.3 High Level State Circuit

The output of the driver is connected to the input of other gates. It is working as the current source (I_{0H}) for the loadly gates. Each input of the load need I_{1H} current

to operate properly. The Fan-out of the circuit is calculated from the ratio of $\frac{I_{0H}}{I_{1H}}$.

If I_{0H} and I_{1H} are 300μA and 30μA respectively, fan out will be

$$\frac{I_{0H}}{I_{1H}} = \frac{200\mu A}{30\mu A} = 10$$

Which indicate that we can't connect more than 10 loading gates?

Current and voltage parameters: The following currents and voltages are specified which are very useful in the design of digital system.

High-level input voltage, V_{IH}: This is the minimum input voltage which is recognized by the gate as logic 1.

Low-level input voltage, V_{IL}: This is the maximum input voltage which is recognized by the gate as logic 0.

High-level output voltage, V_{OH}: This is the minimum voltage available at the output corresponding to logic 1.

Low-level output Voltage, V_{OL}: This is the maximum voltage available at the output corresponding to logic 0.

High -level input current, I_{IH}: This is the minimum current which must be supplied by a driving source corresponding to 1 level voltage.

Low-level input current, I_{IL}: This is the minimum current which must be supplied by a driving source corresponding to 0 level voltages.

High-level output current, I_{OH}: This is the maximum current which the gate can sink in 1 level.

Low-level output current, I_{OL}: This is the maximum current which the gate can sink in 0 level.

High-level supply current, $I_{CC\ (1)}$: This is the supply current when the output of the gate is at logic 1.

Low-level supply current, $I_{CC\ (0)}$: This is the supply current when the output of the gate is at logic (0).

The current directions are illustrated in Fig.

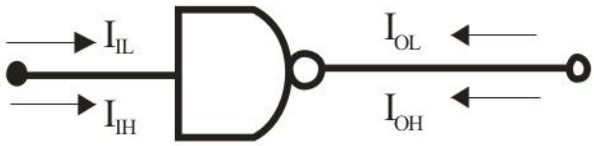


Figure : 18.4 (a)

Noise immunity : The input and output voltage defined above are shown in stray electric and magnetic fields may induce unwanted, known as noise, on the connecting wires between logic circuits. This may cause the voltage at the input to a logic circuit to drop below VIH or rise above VIL and may produce underived operation.

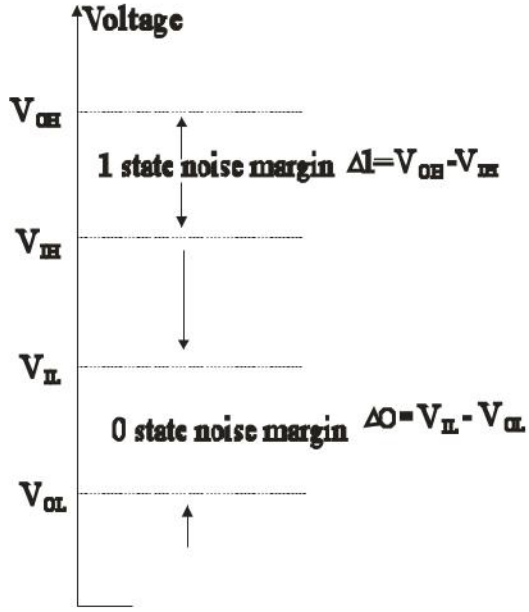


Figure: 18.4 (b) Noise Margin States

The circuits ability to tolerate noise signals is referred to as the noise immunity a quantitative measure of which is called noise margin Noise margin are illustrated in Fig. 18.4

The noise margins defined above are referred to as do noise margins. For high speed IC's, a pulse width of a few microseconds is extremely long in compare on to the propagation only time of the circuit and therefore, may be treated as d.c. as for as the response of the logic circuits is concerned. As the noise pulse width decreases and approaches the propagation delay time of the circuit, the pulse duration is too short for the circuit to respond. Under this condition, large pulse amplitude would be required to produce a change in the circuit output. This means

that a logic circuit can effectively tolerate large noise amplitude if the noise is of very short duration. This is referred to as or noise margin and is substantially greater than the dc noise margin.

Operating temperature: The temperature range in which an IC functions Property must be known. The accepted temperature range is: 0^0 to 70^0C for consumer and industrial applications and -55^0 C to 125^0 C for military purposes.

Power supply Requirements: The supply voltage (s) and the amount of power required by an IC are important characteristics required to choose the proper, power purpose.

18.3 Register Transistor Logic (RTL)

The resister-transistor logic was the most popular form of logic in common use before the development of IC's. RTL circuits consist of resisters and transistors and were the earliest logic family to be integrated. Although RTL has become obsolete how, because of its simplicity and for historical reason, it is proper to devote some attenuation to it and introduce some of the important concepts, useful for all types of gates, through this. The basic RTL gate is a NOR gate as shown fig. 18.5. For the sale of simplicity, a two input NOR gate driving N similar gates is shown in the number of inputs. The number of input terminals is referred to as the Fan-in.

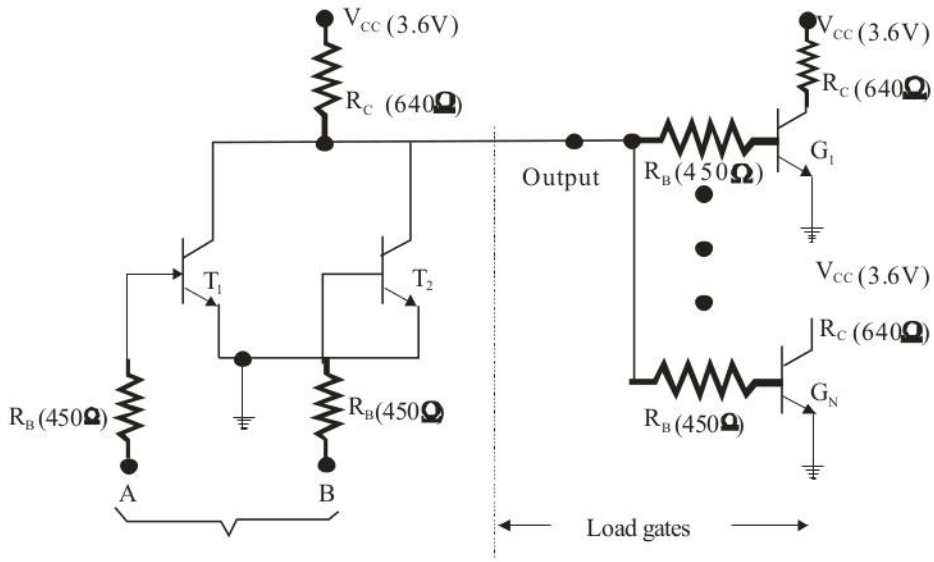


Figure 18.5 : A 2- input RTL NOR gate driving N similar gates

Logic operation : Inputs representing the logic levels are applied at A and B terminals. The voltage corresponding to LOW level should be low enough to drive the corresponding transistor to cut-off similarly the input voltage corresponding to HIGH level should be high enough to drive the corresponding transistor to saturation.

If both the inputs are LOW, transistors and T_2 are cutoff and the output is HIGH. A high level on any input will drive the corresponding transistor to saturation causing the output to go LOW. The LOW (0) level output voltage is $V_{CE, Sat}$ of a transistor (~0.2v) and HIGH (1) level output voltage depends on the number of gates connected to the output this causes the output voltage to be variable and is a deciding factor for the Fan-out of the gate.

Loading considerations: If all the inputs to the gate are Low, the output is HIGH and if the gate is not driving any other gate, i.e. no load is connected, the output Voltage will be slightly less than V_{cc} (there is voltage drop across the common collector resistor due to I_{co} of T_1 and T_2).

When N similar gates are being driven, the load will be equivalent to resistor of value 450/N ohms in serious with a voltage source of 0.8V 450/N (being the voltage between base and emitter of a transistor is saturation). The relevant portion of the circuits is shown in Fig. 18.6.

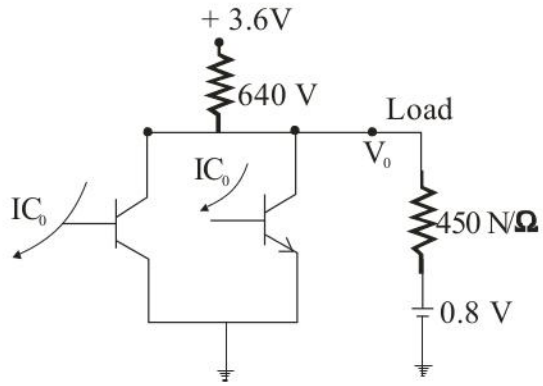


Figure 18.6 A circuit illustrating the equivalent circuit at the input of the load gates.

The base current for each load transistor is

$$I_B = \left[\frac{36-0.8}{640+\frac{450}{V}}\right]\frac{I}{N} = \frac{2.8}{640N+450}$$

The collector current for the load transistor in saturation is

$$I_{c,sat} = \frac{3.6 - 0.2}{640} = 5.31mA$$

The value of N must satisfy the following relation

$$h_{EF}.IB \geq I_c, sat$$

For N = S. I_B = 0.767 mA. Therefore, h_{FE} must be greater than 7.

Noise Margins: When the output is in a state (i.e. logic 0) there the output voltage V_0 = 0.2V. If this voltages increases to cut in voltage of transistor i.e. 0.5V then the load transistor may conduct which can cause multifunctioning of the circuit? Hence the noise margin of RTL will be 0.5-0.2 = 03V.

Similarly when the output is in logic 1 State, the logic 1 noise margin depends upon the number of gates being driven. For N = 5.

$$\frac{V_0 - 36}{640} + \frac{V_0 - 0.8}{450/N} = 0$$

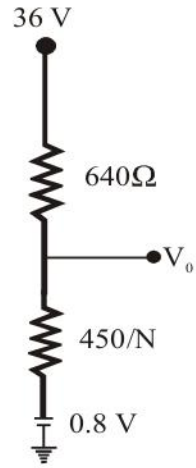
For N = 5

$$\frac{V_0 - 36}{640} + \frac{V_0 - 0.8}{450/S} = 0$$

$$\frac{V_0 - 36}{640} + \frac{V_0 - 0.8}{90} = 0$$

$$V_0\left(\frac{1}{90} + \frac{1}{640}\right) = \frac{3.6}{640} + \frac{0.8}{90}$$

V_0 = 1.14 V



18.4 Diode Transistor Logic (DTL)

The diode transistor Logic, is somewhat more complex then RTL but because of its greater Fan-out and improved noise margins it has replaced RTL. Its main disadvantage is slower speed and because of this it was modified and emerged as transistor logic (TTL) which is the most popular logic family today, as for as small and medium scale ICs are concerned. Although has completely replaced DTL for historical reason as well as for better appreciation of TTL circuit, it is worth. While discussing the details of DTL.

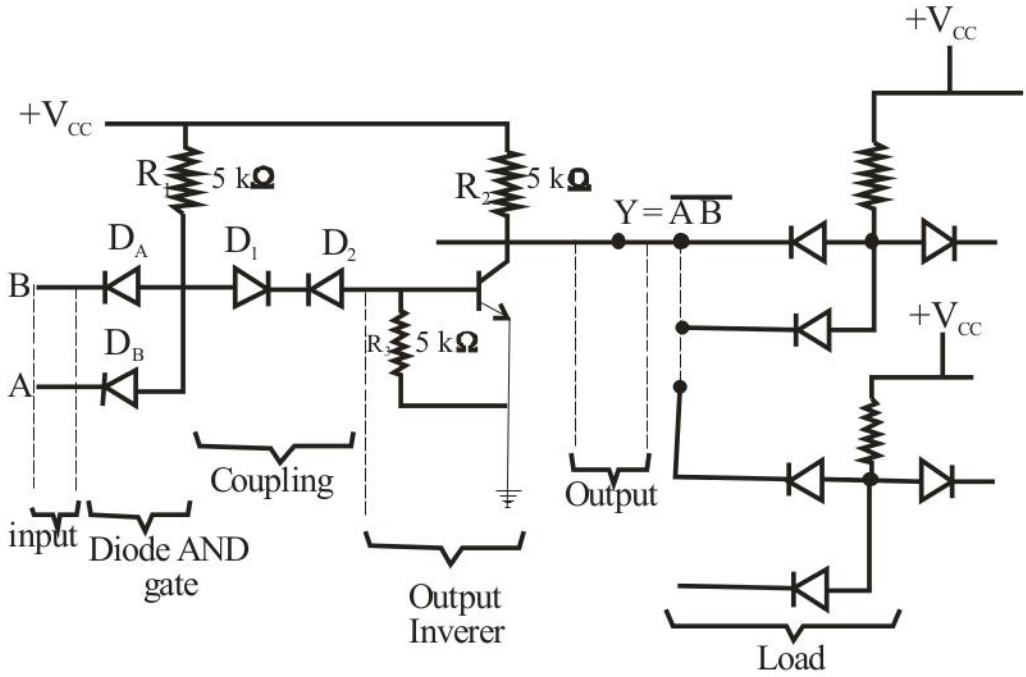


Figure 18.7 : Circuit of DTL as NAND gate.

DTL circuits using discrete components was made using diodes and a transistor inverter (NOT), which was modified for integrated circuit implementation as shown in figure 18.8.

Operation of DTL NAND gate: The basic DTL gate is a NAND gate, a-3-input NAND gate driving N similar gates is shown in figure 18.7. Input diodes D_A, D_B and D_C conduct through the resistor R_1, if the corresponding input is in the Low state, while corresponding to HIGH state the diode is nonconducting. Therefore, if at least one of the inputs is Low, the diode connected to this input conducts and the voltage V_p at point P is one diode drop above the low level voltage at the input. The voltage V_p should be such as to keep transistor in cut-off. Therefore, the output of transistor is V_{cc}. On the other hand, if all the three inputs are in HIGH state, the input diodes are cut-off and consequently current flowing from V_{cc} through R should be sufficient to drive transistor in saturation. Therefore, the output of transistor is $V_{CE'sat.}$

If we consider the voltage corresponding to logic 1 and 0 as V_{cc} and V_{cesat} respectively, this circuit performs NAND operations. The following example illustrates the loading (Fan-out) considerations and the noise margins.

Modified DTL: The Fan-out of DTL gate can be improved by increasing the base current to the output transistor. It is possible by replacing one of the diodes as in the coupling circuit by transistor. The modified form of DTL is shown in Fig. 18.8.

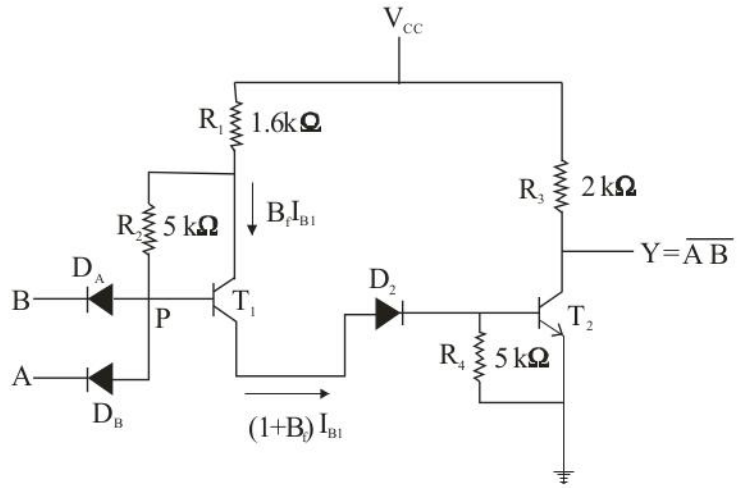


Figure : 18.8

The Simple DTL shown in Fig and modified DTL have same circuit but the only difference is in the coupling circuit. The serious connection of two coupling diodes will exist since the base emitter function of transistor T_1 acts and diode D_1. The circuit is designed so that the transistor T_1 operates in the active region and serves as a current amplifier. When the inputs A and B are high (5V), the transistors T_1 and T_2 will turn on.

The voltage at point P will be $V_{BE1\,(ON)} + V_{D2\,(on)} + V_{BE2\,(sat)} = 0.7+0.7+0.7 =$ 2.1V. The input diodes D_A and D_B will get reverse biased.

Therefore, the current will flow through resister R_2. The base collector junction of transistor T_1 will come in active region and the emitter current of transistor T_1 will be $(1+B_f)\ I_{B1}$. This amplified current will flow to the output transistor stage of the DTL gate. The increased amount current at the base of transistor T_2 will increase the loads connected to the output of the modified DTL gate and transistor T_2 remains in saturation so the fan-out capability of the circuit can be improved.

18.5 Self Learning Exercise-I

Very Short Answer Type Questions

Q.1 Write the full form of JC.

Q.2 Define RTL?

Q.3 What are Diode transistor logic?

Short Answer Type Questions

Q.4 What are the characteristics of digital IC's.

Q.5 A gate draws 2.8 μA when its output is HIGH and 3.6μwhen its output is LOW. Find out the power dissipation when the voltage supply Vcc is 5v and the gate is operated on 50% duty cycle?

Q.6 Write the formula of the average power dissipation of the Digital logic circuit.

18.6 Introduction of Unipolar Logic

The field-effect transistor (FET) is a unipolar transistor since its operation depends on the flow of only one type of carrier. There are two types of field-effect transistors, the junction field-effects transistor (JFET) and the metal-oxide semi-conductor (MOS). MOS transistor can be fabricated in less area than bipolar transistor.

The MOSFETs are generally used in logic circuits but JEFTs are used in linear or analog circuits. MOSFETs have many applications in logic gates, registers, microprocessor and memory arrays because these can be fabricated easily and can have many circuits on a clip.

These are also active switching elements. It means the switching action is basically the same as BJT i.e. they function ideally as open or closed switches according to the input. The difference is in construction and internal operation From BJT. It has high input resistance which draws very low current, so the power dissipation will be low but the switching speed of FET is lower than BJT.

Construction MOSFET: In the metal oxide semiconductor FET, there are three terminals or layer of metal, oxide and semiconductors. There are basically two type of MOSFET:

(1) n-channel MOSFET

(2) P- channel MOSFET

These n-type and p-type MOSFET further classified according to the working of

MOSFET, as:

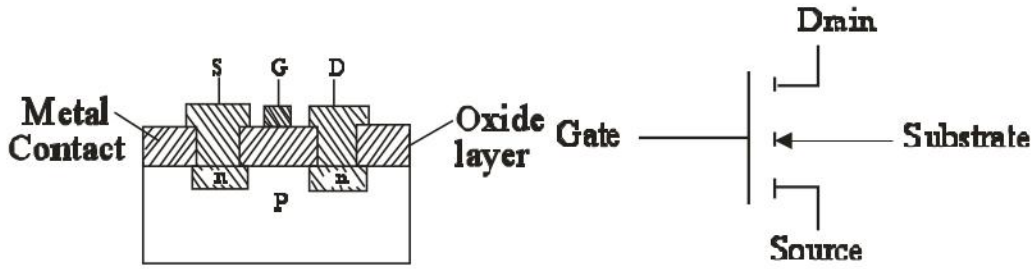
(a) n - channel, enhancement type

(b) P-channel, enhancement type

(c) n-channel, depletion type

(d) P-channel, depletion type.

The MOSFET has three terminals i.e. source, drain and gate. The basic structure of n-channel enhancement type is shown.



(a) Structure **(b) Circuit Symbol**

Figure : 18.9 : Structure or circuit symbol of MOSFET.

In n- channel MOSEFT the P-type silicon semiconductor channel material is called substrate i.e. doped n-type impurities. These two region from source and drain terminals of MOSFET. Then the silicon wafer is coated with an oxide layer (uses silicon dioxide) except the opening leads to source and drain. Then the metal is laid in the opening leads to the source and drain and the oxide layer between the Source and drain. The region between the two-n-type regions is called a channel. The metal between the source and drain is known as a gate. The input voltage applied at this point controls the flow of current is the MOSFET.

Working of MOSFET: - For an n-channel MOS, supply voltage a function of the threshold voltage V_T. The low level ranges anywhere from zero to V_T, and the high level range from V_T to V_{DD}. The n - channel gates usually employs positive logic.

The figure 18.10 shows a MOSFET driver with a passive load resister (R_D). MOSFET acts as switches between saturation and cutoff. When V_{in} is low, MOSFET is cutoff and hence V_{out} is high. When V_{in} is high, MOSFET is saturated and hence V_{out} is low.

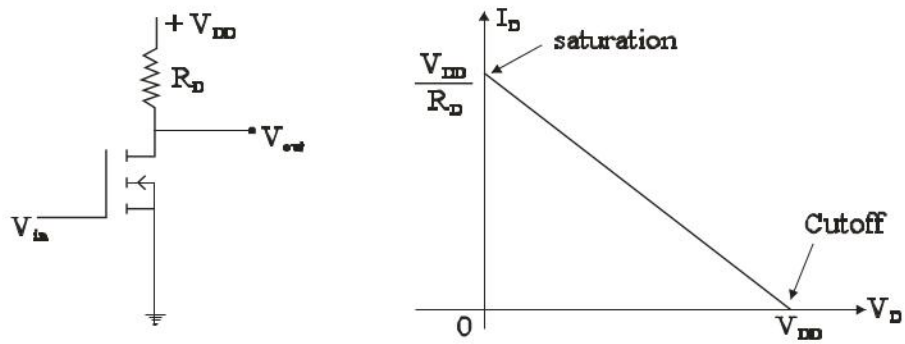


Figure 18.10: MOSFET driver with passive load and its load line

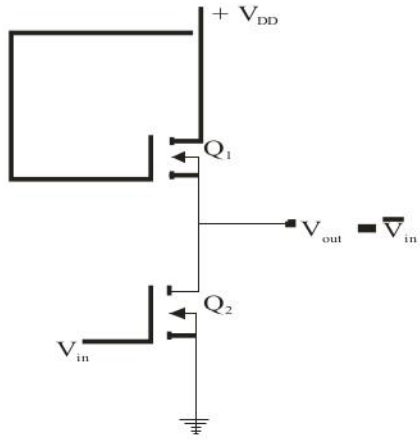


Figure 18.11: MOSFET driver with active load

In MOS technology, MOSFET fabrication is easier than resisters. For this reason, the resistor R_0 in figure 18.10 is replaced by an n-channel MOSFET as shown in fig. 18.11. The gate of the upper MOSFET is connected to the drain; this MOSFET always conducts and behaves like the resister R_0. The upper MOSFET has a resistance ten times greater than that of the lower MOSFET. When the input voltage is low (below V_T), Q_2 turns off. Since Q, is always ON, the output voltage is at about V_{DD}. When the input voltage is high (above V_T), Q_2 turns ON. Current flows from V_{DD} through the active load resister Q_1 and into Q_2. Thus, the circuit behaves as an inverter. The geometry of the two MOS device must be such that the resister of Q_2 when conductivity is much less than the resistance of Q_1 to maintain the output Y at a voltage below V_T.

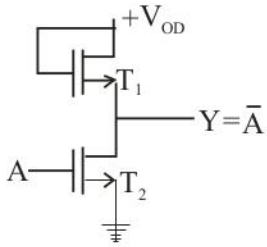
Characteristic of MOS logic

MOS logic families are slower in operating speed, reprove much less power, have a better noise margin and higher fan-out.

1. **Noise margin -** The noise margin of MOS logic around IV.
2. **Power drain: -** MOS logic circuit draw small amount of power because of the relatively large resistance of being used.
3. **Fan out: -** The fan-out capabilities of MOS logic would be virtually unlimited because of the extremely high input resistance at each MOSFET input. MOS logic can easily operate at a fan-out of 50.
4. **Operating speed: -** A typical N-MOS NAND gate has a propagation delay of 50 ns. The combination of large Rout and large cited serves to increase switching time.

NMOS Inverters (NOT Gate):

The logic circuit of NMOS inverter is shown below.

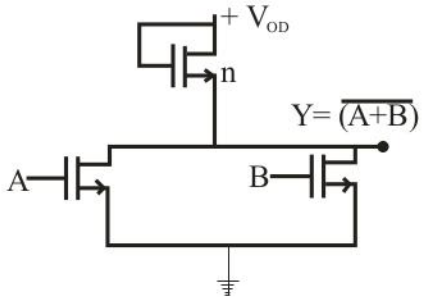


Input	Output
A	Y
0	1
1	0

Figure 18.12: NMOS inverter

Here two NMOS are attached with each other, the T_1 acts as a load resister and T_2 as a active device. The input is applied to the gate of T_2. When the input at the T_2 is high (logic) i.e. voltage above V_T is applied at A then T_2 will be on and current flow from V_{DD} to load T_1 and T_2. So output 1 will be 0, and when the input at T_2 is low (logic 0) i.e. the gate voltage U_T is low, Then T_2 turns OFF and output 1 will be at high level or logic 1. It is because T_0 is always on and the output voltage at 1 will be through V_{DD}.

NMOS NOR Gate: The logic circuit of NMOS NOR gate is shown below.

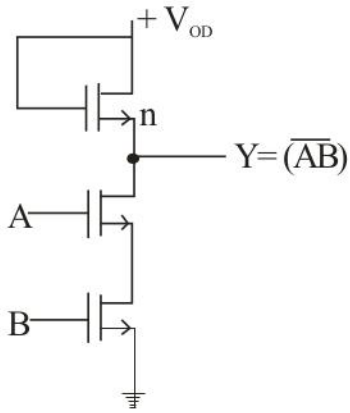


Inputs		Output
A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

Figure 18.13 NMOS NOR gate

In the NMOS NOR gate logic there are one load NMOS are connected in series with parallel combination of two NMOS A and B. According to the logic if the two inputs A and B are less than V_T i.e. logic 0, then the channels are not created in either of the lower MOSFETs and output will be high level i.e. logic. 1 Now if the input of at least one of the lower MOSFET is greater than, V_T i.e. logic Y will be at low level i.e. logic 0. The operation of NMOS NOR gate is summarized in the table which gives the output $Y = \left(\overline{A+B}\right)$.

NMOS NAND Gate: The logic circuit of NMOS NAND gate is shown below.



Inputs		Output
A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

Figure 18.14: NMOS NAND Gate

Here one n-channel MOS is connected parallel, with two n-channel MOS serial connection. If at least one input voltage is less than V_T i.e. logic, then the corresponding MOSFET will be OFF and output Y will be high (logic 1) i.e. the voltage will be V_{DD} at output Y. If both the inputs are greater than V_T then n-channel will be formed. If the on resistance of these two lower n-channel MOSFETs is much less than MOSFET load resister, the output Y will be nearly zero i.e. logic 0. All operation is shown in Truth table.

18.7 CMOS Logic

A complement or MOSFET (CMOS) is obtained by connecting a p-channel and an n-channel MOSFET in source with drains tied together and the output is taken at the common drain. Input is applied at the common gate formed by connecting the two gates together in a CMOS p-channel and n-channel enhancement MOS devices are fabricated on the same chip, which makes its fabrication more complicated and reduces the packing density. But because of negligibly small power consumption, CMOS is ideally suited for balladry operated systems. The n-channel MOS will

conduct when the gate to source voltage will be negative either of the devices will be OFF if the gate to source voltage will be zero.

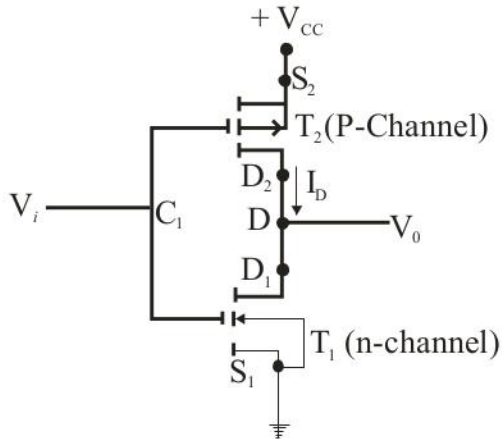


Figure 18.15 A CMOS Switch

CMOS Inverter: The inverter is basic to the CMOS gate. A Basic inverter connection is shown in Fig. 18.16.

The driver is transistor Q_2 which is the n-channel, and Q_1 the p-channel device acts as the

load. Notice that drains are connected together to provide the output and the source and substrate are connected together. The source of p-channel device is connected to $+V_{DD}$ and the source of n-channel is connected to ground. The gates of the two devices are connected together as a common input. It is important to note that the output voltage is equal to the supply voltage and the current flows through the circuit only during switching of the input voltage from one level to the other.

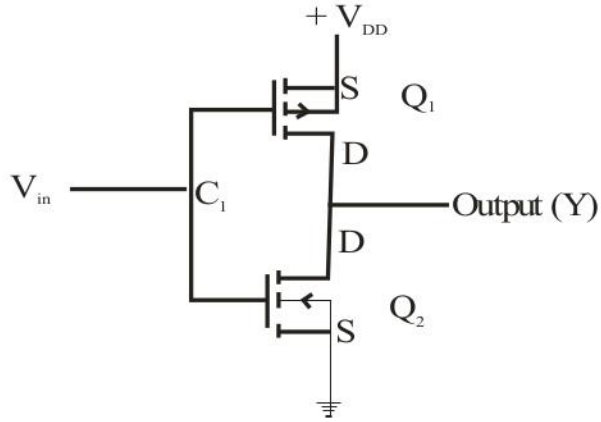


Figure 18.16: CMOS Inverter

Operation : When Vin is LOW, Q_2 is OFF but Q_1 is ON thus means the output voltage is HIGH. On the other hand, when V_{in} is HIGH, Q_2 is ON and Q_1 is OFF, In this case, the output voltage is LOW.

Since the output voltage is always opposite in phase to the input voltage, the circuit acts as an inverter.

The operation of CMOS transistors can be described by the following rules:

n-channel MOSFETs are turned ON by a positive gate voltage.

p- channel MOSFETs are turned ON by a negative gate voltage.

The CMOS inerter can be modified to build other CMOS logic circuits. A CMOS circuit is ideal in a number of ways. First it needs extremely low power to operate the circuit. Since either one of the MOS devices is OFF when the input is in LOW or HIGH state, only the leakage current in the order of nanoampers. Flow through the circuit and the power dissipation of the CMOS devices is typically in the range of nanowatts. This low power consumption is the reason for the popularity of CMOS devices in pocket calculators, digital very calculators, digital very watches, and portable microcomputers.

Fan-out for CMOS circuits is dually infinite since no loading occurs when it is connected to the gate of enhancement MOSFET. Practical values of Fan-Out greater than 50 are typical.

Propagation delay of CMOS gate is typically about 25 to 100 ns depending on the particular device. It increases with greater load capacitance.

CMOS NAND Gate (74C00): A two input NAND gate which consists of two p-type units in parallel and two n-type units in series in shown in figure 18.17. Notice that Q_1 and Q_2 from one complementary connection: Q_3 and Q from another.

If both inputs are HIGH, both p-channel transistors turn OFF and both n-channel transistor turn ON. The output has low impedance to ground and produces a LOW state. If any input is LOW, the associated n-channel transistor is turned OFF and the associated p-channel transistor is turned ON. The output is coupled to V_{DD} and goes to the HIGH state. This functions as a logic NAND gate. The 74C00 is a quad 2-input NAND gate.

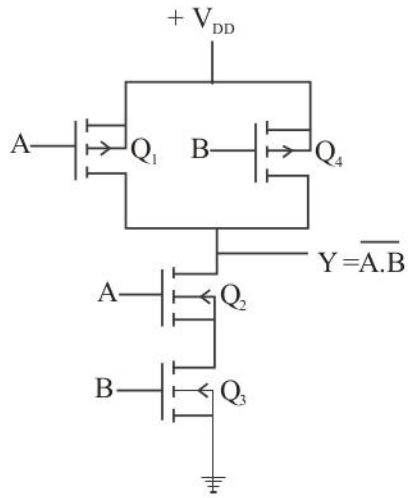


Figure 18.17 CMOS NAND Gate

To produce the positive AND function, the output of the CMOS NAND gate can be connected to a CMOS inverter.

CMOS NOR Gate: A two-input CMOS NOR gate using a pair of PMOS transistors (Q_1 and Q_2) and NMOS transistors (Q_3 and Q_4) is shown in figure 18.18. The two inputs, A and B either of the inputs can turn ON the PMOS or NMOS device connected to it.

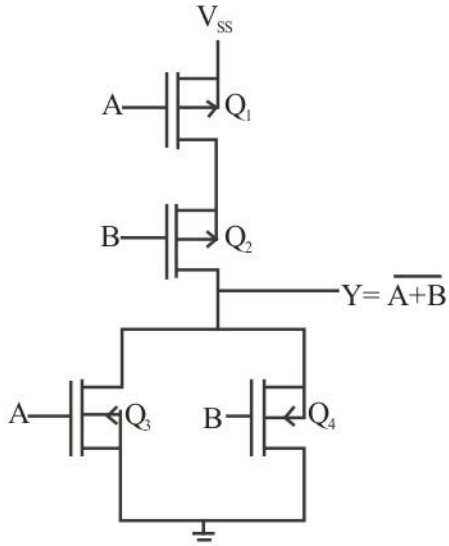


Figure 18.18: CMOS NOR Gate

When both inputs are LOW, both PMOS devices are driven ON and both NMOS devices OFF. The output is coupled to V_{DD} and goes to the HIGH state. If any input is HIGH, the associated p-channel transistor is turned OFF and the associated n-channel transistor turns ON. This connects the output to the ground causing a

LOW output. Thus this circuit functions as a NOR Gate. A CMOS or Gate can be not formed by combining the output of the CMOS NOR gate with a CMOS inverter.

18.8 Transistor Transistor Logic (TTL)

To improve the switching speed of DTL NAND gate, the circuit is modified and the modified circuit is called transistor transistor logic (TTL) in which the diodes are replaced by multi-emitter transistor. Since the speed is inversely proportional to the propagation delay, the propagation delay of the transistor circuit that goes into saturation depends on two factors.

1. Storage Time 2. RC time constant

If the storage time decreases, the propagation delay decreases which improves the speed. Similarly, by reducing the values of resistance, the RC time constant decreases which decrease the propagation delay. On the other hand by reducing the value of resistance, the power dissipation increases because the current drawn by the resistance is more.

There are several sub-families or series of TTL technology. The commercial TTL IC's have a number designation that starts with 74 and follows with a suffix that identifies the series.

TTL NAND Gate (7400 Series)

The basic circuit of the TTL NAND gate is shown in Fig. 18.19.

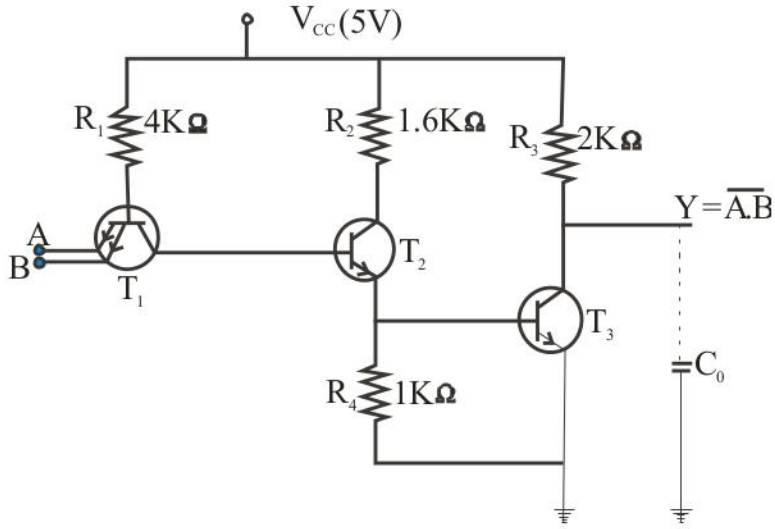


Fig. 18.19 TTL NAND Gate

The Difference between DTL NAND circuit and TTL NAND circuit are as follows:

1. The diodes D_A and D_B are replaced by the multi-emitter NPN transistor T_1. Since the multi-emitter transistor has small area than the diodes it replaces, so the wafer size can be reduced. The number of emitters represents the number of inputs. The base emitter junction of the multi-emitter transistor acts as diodes D_A and D_B.
2. The base collector junction of transistor T_1 acts as the coupling diode D_1 in the DTL circuit.
3. Similarly, the base emitter junction of transistor T_2 replaces the coupling diode D_2.
4. The final stage i.e. the transistor inverter circuit is same in DTL and TTL circuits.

Operation of TTL NAND Gate

Operation of TTL NAND gate can be understood by applying three conditions out of which the third condition differentiates the operation of TTL from DTL. Under this condition the operation of TTL NAND gate is faster in comparison to the operation of DTL NAND gate. To understand its operation, it is assumed that $V_{BE\ (ON)} = V_{BE\ (sat)} = 0.2V$. The analysis of these three conditions is as follows:

Condition 1.

When both the inputs are low (0.2V) or logic 0, the emitter base junction of transistor T_1 becomes forward biased making the voltage at the base of transistor T_1 equal to $V_{in} + V_{in} + V_{BE1\ (sat)} = 0.2 + 0.2 + 0.7 = 1.1V$ as shown in fig. 18.20.

It allows the flow of base current I_1. The voltage at the base of transistor T_1 is insufficient to support the base collector junction of transistor T_1 plus the base emitter junction drop of transistor T_2 and T_3.

Hence, transistor T_2 and T_3 get cut off and the output voltage at Y will be high i.e. + 5V. Therefore, the current will be equal to approximately zero.

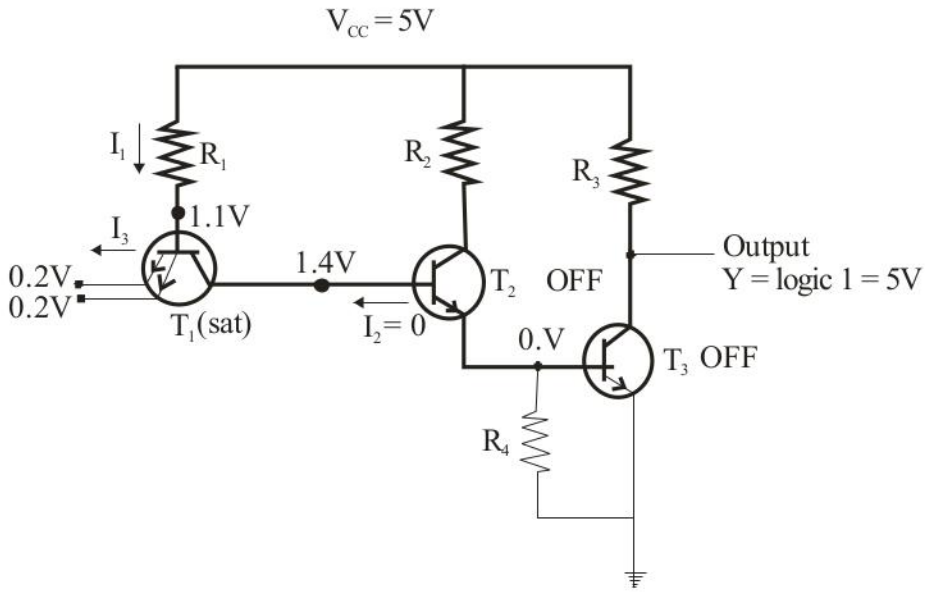


Fig. 18.20 When both inputs are at logic 0

Condition 2.

When both the inputs i.e. A and B are high at 5 V as shown in fig. 18.21 the base collector junction of transistor T_1 and base emitter junctions of transistors T_2 and T_3 will be forward biased.

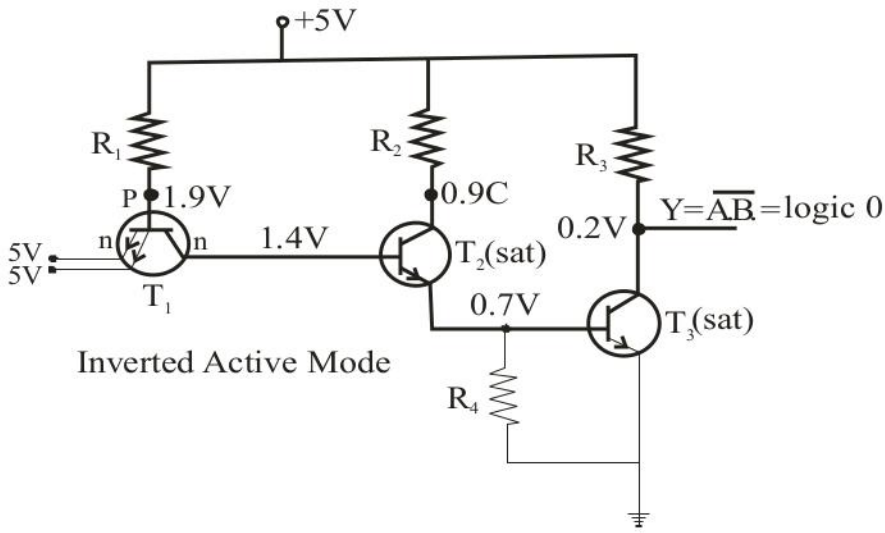


Fig. 18.21 TTL NAND when all inputs are high

So, the transistors T_2 and T_3 get saturated and output Y i.e. V_{CE3} (sat) reaches 0.2V which corresponds to logic 0. Since transistors T_2 and T_3 get saturated, the voltage at the base of transistor T_3 will be V_{BE3} (sat) = 0.7V and voltage at the base of T_2 will be $V_{BE3(sat)} + V_{BE2(sat)}$ = 0.7+ 0.7 = 1.4V. Since the base collector junction of

transistor T_1 is forward biased, the voltage at the base of T_1 will become $V_{BE3\ (sat)}$ + $V_{BE2\ (sat)}$ + V_{BC1} (ON) = 0.7+0.7+0.5 =1.9 V. Under this condition the input base emitter junction gets reverse biased as all the inputs are at 5V. Therefore, base emitter junction of T_1 is forward biased which means T_1 operates in the inverted active mode. So, the current flowing through the emitter of T_1 is very small.

Condition 3.

Since the base collector junction of transistor T_1 acts as pn-junction diode, when one of the inputs goes to low, the transistor T_1 switches to saturation from its inverted active mode and there will be a transistor effect which reduces the propagation delay of the gate as shown in fig. 18.22

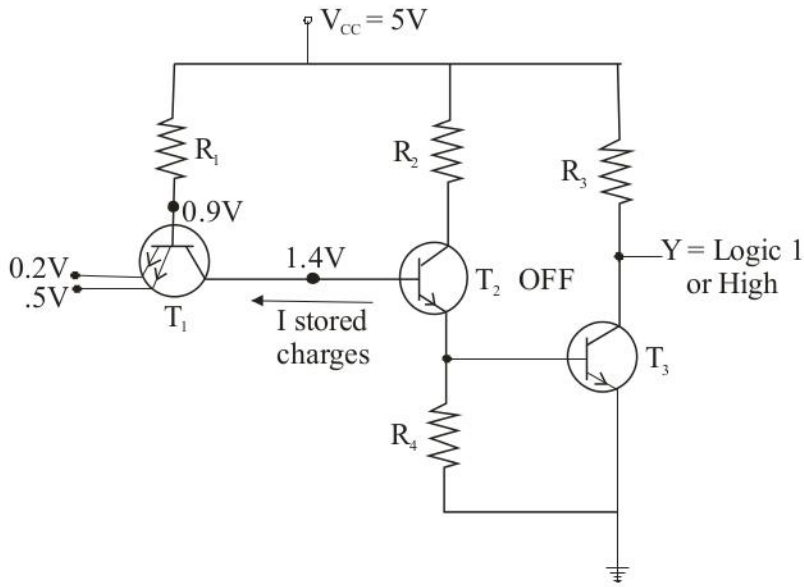


Fig. 18.22

When one of the inputs goes to low at 0.2V, the transistors T_2 and T_3 switch to cut off mode but these transistors cannot come out of saturation until the stored charges are removed from their base. As one input goes to low or 0.2V, the base of transistor T_1 quickly goes to 0.9V and the corresponding P-N junction becomes forward biased while the collector remains at 1.4V. At the time of switching, the base collector junction of transistor T_1 is reverse biased by 1.4-0.9 = 0.5V and base emitter junction is forward biased. So the transistor T_1 goes into active mode which allows a large collector current to flow. This collector current corresponds to the removal of the stored charges of transistors T_2 and T_3. As the stored charges are removed, the transistor T_1 enters into saturation. The removal of stored charges

affects the switching speed of transistors T_2 and T_3 from saturation to cut off. As the switching speed of transistor increases, the propagation delay decreases. So, TTL is much faster than DTL. This speed can also be improved by decreasing the resistance R_3. But by decreasing R_3, the power dissipation increases. The TFL circuits are differentiated by their series as mentioned in table previously.

18.9 Self Learning Exercise - II

Very Short Answer Type Questions

Q.1 What is the most suitable IC family for low power consumption?

Q.2 What is the fastest bipolar IC family?

Q.3 What is standard T7L gate, average power dissipation?

Short Answer Type Questions

Q.4 Write the truth table of NMOS NOR Gate?

Q.5 What is unipolar IC?

Q.6 Draw the circuit diagram of NMOS NAND Gate.

18.10 Summary

Modern digital systems use integrates circuits because of the following advantages.

(i) Overall size of digital system gets reduced, because IC's consist of a large number of digital circuits.

(ii) The cost of ICs is very low.

(iii) Power consumption is less.

(iv) They have high noise margin.

(v) High reliability and operation speed is high.

The logic families are classified into two types Bipolar and unipolar logic families. RTL and DTL families are no more used for new systems because of their low speed, high power dissipation and low fan-out. TTL is the most popular general purpose logic family. It is available in seven different series with a wide range of operating speed, power dissipation and fan-out. MOS devices occupy a very small fraction of silicon chip area in comparison to bipolar devices and require very small power. CMOS has the lowest speed power product and requires

very small power.

18.11 Glossary

Logic family : The set of compatible IC's with the same logic levels and same supply voltages, fabricated to perform various logic function.

Unipolar logic family : Logic family having unipolar devices like MOSFETs as its key element.

Bipolar logic family : Logic family with transistors and diodes as its key element.

8.12 Answer to Self Learning Exercises

Answer to Self Learning Exercise-I

Ans.1: IC stand for 'Integrated circuit'.

Ans.2: Each input is connected to the base of the transistor through a resistor. The collectors of all the transistors are tied together at the output.

Ans.3: DTL circuit using discrete components was made using discrete components using diodes and a transistor inverter.

Ans.4: See section 18.2

Ans.5: 16μW.

Ans.6: $p_{\triangle}(avg) = V_{cc}\left(\frac{I_{CCH} + I_{CCL}}{2}\right)$

Answer to Self Learning Exercise-II

Ans.1: CMOS

Ans.2: ECL

Ans.3: 10 MW

Ans.6:

Input		Output
A	B	Y
0	0	1
0	1	0
1	0	0

1	1	0

Ans.4: Unipolar IC are(i) PMOS(ii) NMOS(iii) CMOS.

18.13 Exercise

Section A : Very Short Answer Type Questions

Q.1 Define Noise immunity.

Q.2 What do you mean by fan-out or Fan in condition?

Q.3 What is the difference between current Sinking and current sourcing logic circuits?

Q.4 What do you understand by propagation delay.

Section B : Short Answer Type Questions

Q.5 Explain the meaning of logic family.

Q.6 Write a brief note on MOS digital IC's.

Q.7 Explain the working DTL.

Q.8 Explain TTL - NAND gate.

Section C : Long Answer type Questions

Q.9 Explain the working DTL.

Q.10 What are the advantage and disadvantage of the TTL logic family in comparison to MOS Family?

Section C: Long Answer Type Questions

Q.11 Discuss the circuit of RTL NOR gate.

Q.12 Draw the CMOS NOR gate and explain its operation.

References and Suggested Readings

1. Modern Digital Electronics, R.P. Jain, 2e. Tata McGraw Hill Publishing Company Limited. New Delhi.
2. Digital Electronics, G. K. Kharate, Oxford University Press.
3. Digital Electronics & Circuits, S. N. Dolia, Genius Publication, New Delhi.

UNIT-19
Arithmetic Logic Units

Structure of the Unit

19.0 Objectives

19.1 Introduction

19.2 1-Bit Binary Adders

- 19.2.1 Half Adder
- 19.2.2 Full Adder
- 19.2.3 Parallel Binary (Ripple Carry) Adder Circuit

19.3 Binary Subtractor

- 19.3.1 Binary Subtraction of two Bits
- 19.3.2 Half Binary Subtraction
- 19.3.3 Full Binary Subtraction
- 19.3.4 N-Bit Binary Parallel Subtractor

19.4 Digital (Binary) Comparators

- 19.4.1 Equality Comparator
- 19.4.2 Magnitude Comparator

19.5 Parity Generators and Checkers

- 19.5.1 Parity Generators
- 19.5.2 Parity Checkers

19.6 Illustrative Examples

19.7 Self Learning Exercise

19.8 Summary

19.9 Glossary

19.10 Exercise

References and suggested Readings

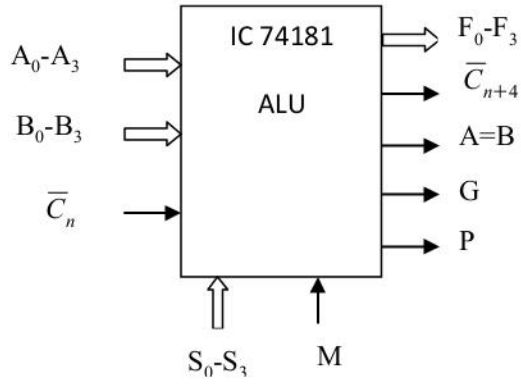
19.0 Objectives

In this unit, we study about the ALU (Arithmetic and Logic Units) as it is part of CPU (Central Processing Unit). It does all processes related to arithmetic and logic operations that need to be done on instruction words. In some microprocessor architectures, the ALU is divided into the arithmetic unit (AU) and the logic unit (LU). Therefore, it becomes necessary to understand the basic operation units, for examples adder, subtractor, multiplier, comparator etc., we study the gate logic designs and their operations through the Boolean expressions. Thus, in the following sections, the logic gate circuits of Binary Adder, Binary Subtractor, Comparator and Parity Checker have been discussed.

19.1 Introduction

An arithmetic and logic unit (ALU) is a combinational circuit consists of input variables, logic gates, and output variables. The arithmetic operations performed by an ALU are addition subtraction, multiplication, division and complement etc., which are executed by AND, OR, EXOR gates etc. The logic gates accept signals from the inputs and generate signals from the outputs. Both input and output data are represented by binary signals. For n inputs variables, there are 2^n possible combinations of binary input values. An input variable may appear in two wires, supplying normal and complement forms to the input of the circuit.

The ALU is the number crunching part of the computer, this implies that not only logic operations and logic as well. The ALU can also shift numbers in several different ways. The ALU can also compare two numbers to determine the equality and inequality relationships and also detects simple errors. The ALU's are available in the form of IC 74181, 74381, 40181 etc. Most commonly used ALU is IC 74181, which is a high speed 4-bit parallel ALU. The functional diagram of 74181 is shown.



where the descriptions of pins are

$A_0 - A_3$ & $B_0 - B_3$: 4-bit binary input

$F_0 - F_3$: 4-bit binary data output

$S_0 - S_3$: Select output

$\overline{C}_n$: Active low carry input

$\overline{C}_{n+4}$: Active low carry output

A=B: Output for equality

G: Carry generated output

P: Carry propagated output

M: Mode control

ALU can perform total 16 arithmetic operations. S_0-S_3 inputs are used to choose any one logical or arithmetic operation. The mode control signal M decides that whether the operations are arithmetic or logical, i.e. M = 0 for arithmetic operation and M = 1 for logical operation. When M = 0, all internal carries are enabled and when M = 1, all internal carries are disabled. $\overline{C}_{n+4}$ indicates the sign of the output in the subtraction operation. G and P outputs are used when number of IC74181 circuits are used cascaded along with IC 74182. Function table of IC74181 ALU for M = 0 and M = 1 are given below:

Select Inputs				Logic function for (M=1)	Logic Functions for (M =0)	
S_3	S_2	S_1	S_0	$\overline{C}_n$	$\overline{C}_{n=1}$	$\overline{C}_{n=0}$
0	0	0	0	$F = \overline{A}$	$F = A$	$F = A + 1$
0	0	0	1	$F = \overline{A + B}$	$F = A + B$	$F = (A + B) + 1$
0	0	1	0	$F = \overline{A}.B$	$F = A + \overline{B}$	$F = (A + \overline{B}) + 1$
0	0	1	1	$F = 0$	$F = -1$ (2's complement)	$F = 0$
0	1	0	0	$F = \overline{A}\,\overline{B}$	$F = A + A\overline{B}$	$F = A + A\overline{B} + 1$
0	1	0	1	$F = \overline{B}$	$F = (A + B) + A\overline{B}$	$F = (A + B) + A\overline{B} + 1$

0	1	1	0	$F = A \oplus B$	$F = A - B - 1$	$F = A - B$
0	1	1	1	$F = A\overline{B}$	$F = A\overline{B} - 1$	$F = A\overline{B}$
1	0	0	0	$F = \overline{A} + B$	$F = A + AB$	$F = A + AB + 1$
1	0	0	1	$F = \overline{A \oplus B}$	$F = A + B$	$F = A + B + 1$
1	0	1	0	$F = B$	$F = (A + \overline{B}) + AB$	$F = (A + \overline{B}) + AB + 1$
1	0	1	1	$F = AB$	$F = AB - 1$	$F = AB$
1	1	0	0	$F = 1$	$F = A + A$	$F = A + A + 1$
1	1	0	1	$F = A + \overline{B}$	$F = (A + B) + A$	$F = (A + B) + A + 1$
1	1	1	0	$F = A + B$	$F = (A + \overline{B}) + A$	$F = (A + \overline{B}) + A + 1$
1	1	1	1	$F = A$	$F = A - 1$	$F = A$

19.2 1-Bit Binary Adders

The digital circuits which are used to add two or more numbers is called binary adder. The simple addition consists of four possible elementary operations; 0 + 0 = 0, 0 + 1 = 1, 1 + 0 = 1, 1 + 1 = 10. The last operation sum's length is 2. The higher significant digit is called *carry*. The carry obtained from two bits is added to the next higher-order pair of significant bits. A combinational circuit that performs the addition of two bits is called *half-adder*. One that performs the addition of three bits (two significant and a carry) is *full-adder*.

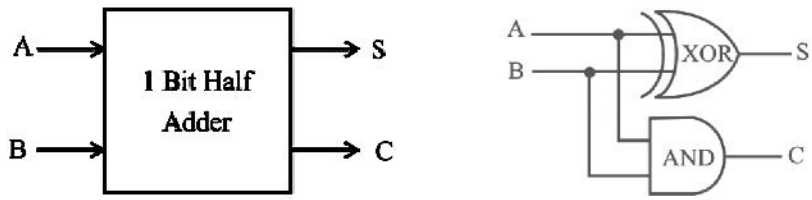
19.2.1 Half Adder

To understand about the half adder explicitly, first we should know what is an adder? *Adder circuit is a combinational digital circuit that is used for adding two numbers. A typical adder circuit produces a sum bit (denoted by S) and a carry bit (denoted by C) as the output*. Typically adders are realized for adding binary numbers but they can be also realized for adding other formats like binary coded decimal, XS-3 etc. Besides addition, adder circuits can be used for a lot of other applications in digital electronics like address decoding, table index calculation etc. Adder circuits are of two types: Half adder and Full adder. Here we will focus first at half adders.

Half adder is a combinational arithmetic circuit that adds two numbers and

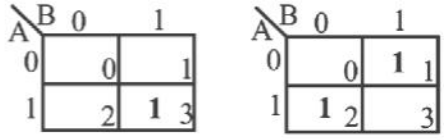
produces a sum bit (S) and carry bit (C) as the output. If A and B are the input bits, then sum bit (S) is the X-OR of A and B and the carry bit (C) will be the AND of A and B. From this it is clear that a half adder circuit can be easily constructed using one X-OR gate and one AND gate. Half adder is the simplest of all adder circuit, but it has a major disadvantage. The half adder can add only two input bits (A and B) and has nothing to do with the carry if there is any in the input. So if the input to a half adder have a carry, then it will be neglected and adds only the A and B bits. That means the binary addition process is not complete and that's why it is called a half adder. The truth table, schematic representation and XOR//AND realization of a half adder are shown in the figure below.

Decimal Equivalents	Inputs		Outputs	
	A	B	Sum (S)	Carry (C)
0	0	0	0	0
1	0	1	1	0
2	1	0	1	0
3	1	1	0	1



Block Diagram and Logic Circuit of Half Adder

The logic expression for sum and carry are obtained from K-map, as



K-map for Carry K-map for Sum

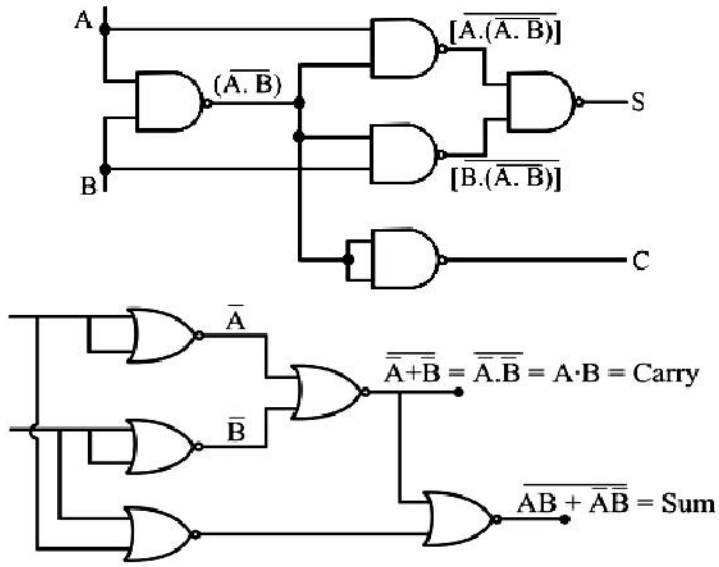
In truth table, the sum output is 1 when *AB=01* and *AB=10*. Therefore, the expression for sum is $S = \overline{A}B + A\overline{B} = A \oplus B$.

Similarly, the logic expression for Carry output can be expressed as Sum of the product expression by summing up the input combinations for which the carry

is equal to 1. Therefore, in truth table, the carry is 1 when *AB=11*. Thus the expression is $C = AB$.

The sum output corresponds to a logic Ex-OR function while the carry output corresponds to an AND function. Hence, the Half adder circuit can be implemented using Ex-OR and AND gates.

NAND gates or NOR gates can be used for realizing the half adder in universal logic and the relevant circuit diagrams are shown in the figure below.



Half Adder using (a) NAND, (b) NOR logic gates

Disadvantages of half adder:

Number $A = A_1 \quad A_0$; Number $B = B_1 \quad B_0$,

Addition of A and B number is given as

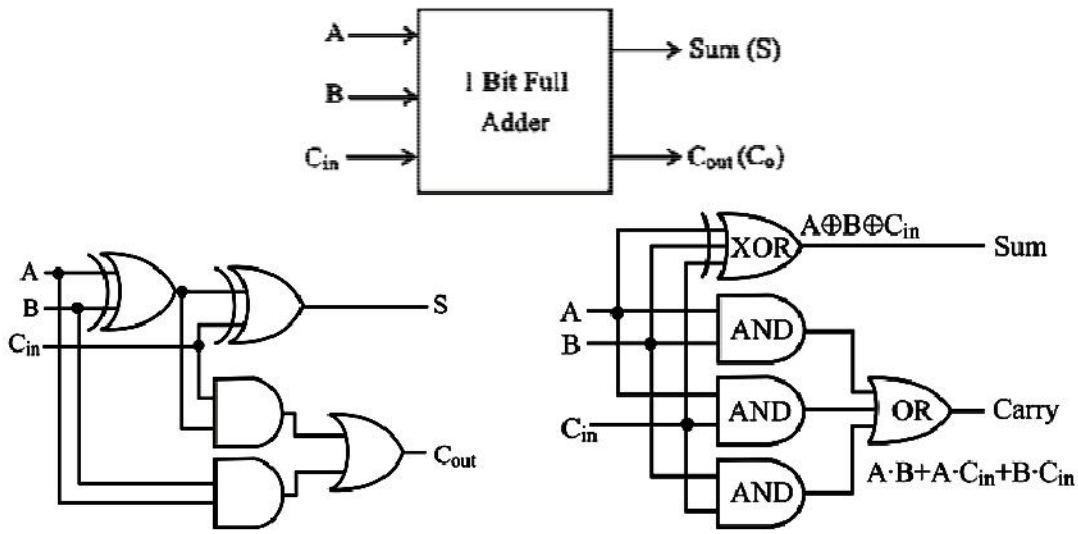
$$\begin{array}{lll} & A_1 & A_0 \\ + & B_1 & B_0 \\ + & C_0 & \leftarrow \text{Carry generated from the previous addition } (A_0 + B_0) \\ \hline & S_1 & S_0 \end{array}$$

It is observed that a half adder may add A_0 and B_0 to produce S_0 and C_0 but the addition of next three bits A_1, B_1 and C_0 is not possible by using half adder and thus restricted the use of half adder in practice.

19.2.2 Full Adder

In order to rectify the drawbacks of the half adder, a full adder was developed. Full adder is a logic circuit that adds two input operand bits plus a Carry in bit and outputs a Carry out bit and a sum bit. The Sum out (S_o) of a full adder is the XOR of input operand bits A, B and the Carry in (C_{in}) bit. Truth table and schematic of a 1 bit Full adder is shown below.

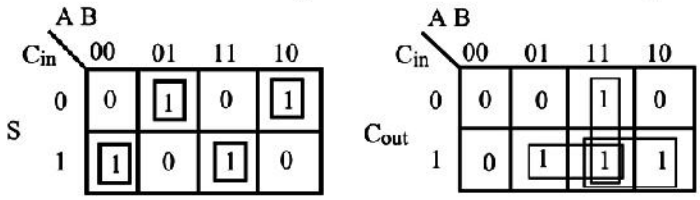
Inputs			Outputs	
A	B	C_{in}	S	S_o /C_o
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1



Block diagram and logic circuits of Full Adder

There is a simple trick to find results of a full adder. Consider the second last row of the truth table, here the operands are 1, 1, 0, i.e. (A, B, C_{in}). Add them together i.e. 1+1+0 = 10 . In binary system, the number order is 0, 1, 10, 11……. and so the result of 1+1+0 is 10 just like we get 1+1+0 =2 in decimal system. 2 in the decimal system corresponds to 10 in the binary system. Swapping the result "10" will give S=0 and C_{out} = 1 and the second last row is justified. This can be applied to any row in the table.

The logic expression for sum and carry are obtained from K- map as



$$S = \overline{A}\,\overline{B}\,C_{in} + \overline{A}\,B\,\overline{C}_{in} + A\,\overline{B}\,\overline{C}_{in} + ABC_{in}$$

$$= C_{in}(\overline{A}\,\overline{B} + AB) + \overline{C}_{in}(\overline{A}\,B + A\,\overline{B})$$

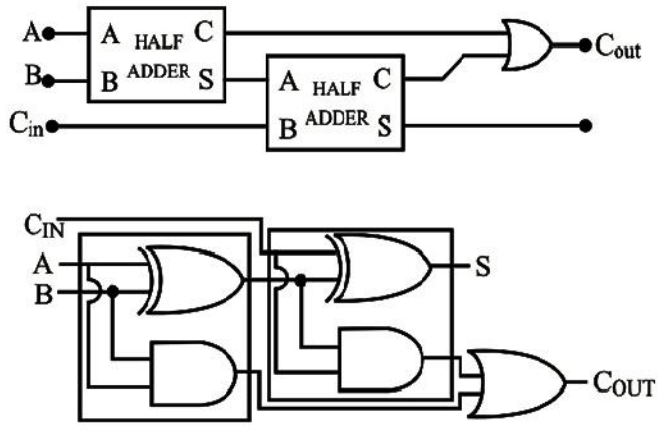
$$= \overline{C}_{in}\left(\overline{A \oplus B}\right) + \overline{C}_{in}\left(A \oplus B\right)$$

$$= \left(A \oplus B\right) \oplus C_{in}$$

and $C_o = AB + AC_{in} + BC_{in}$

Full Adder Using Two Half Adder Circuits:

A Full adder can be made by combining two half adder circuits together *(a half adder is a circuit that adds two input bits and outputs a sum bit and a carry bit).*



Full adder using two half adders

To verify that the above circuit acts as a full adder, we now work out the following algebra:-

The expression for the sum output is

$$S = (A \oplus B) \oplus C_{in} = A \oplus B \oplus C_{in},$$

which is just the same as for sum output for full adder.

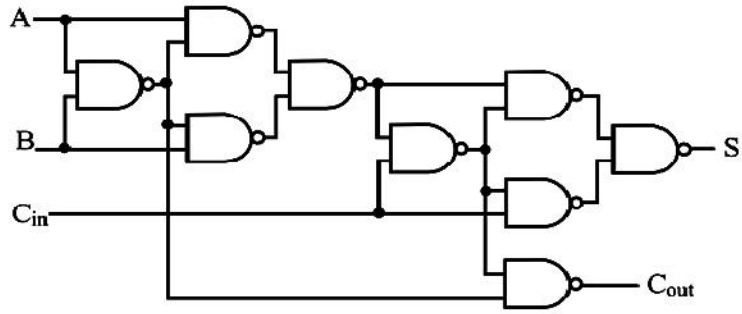
The expression for the carry output is

$$\begin{aligned} C_o &= (A \oplus B)C_{in} + AB \\ &= (\overline{A}\,B + A\,\overline{B})C_{in} + A\,B \\ &= \overline{A}\,B\,C_{in} + A\,\overline{B}\,C_{in} + AB(1 + C_{in}) \\ &= BC_{in}(A + \overline{A}) + A\overline{B}C_{in} + AB \\ &= BC_{in} + A\overline{B}C_{in} + AB(1 + C_{in}) \\ &= BC_{in} + AB + AC_{in}(B + \overline{B}) \\ &= BC_{in} + AB + AC_{in} \end{aligned}$$

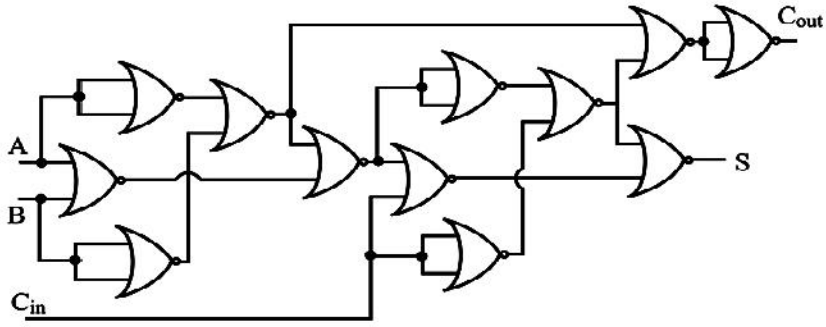
This expression is same as that for a full adder.

Full Adder Using NAND or NOR Logic Gates:

Alternatively the full adder can be made using NAND or NOR logic. These schemes are universally accepted and their circuit diagrams are shown below.



Full adder using NAND logic gates

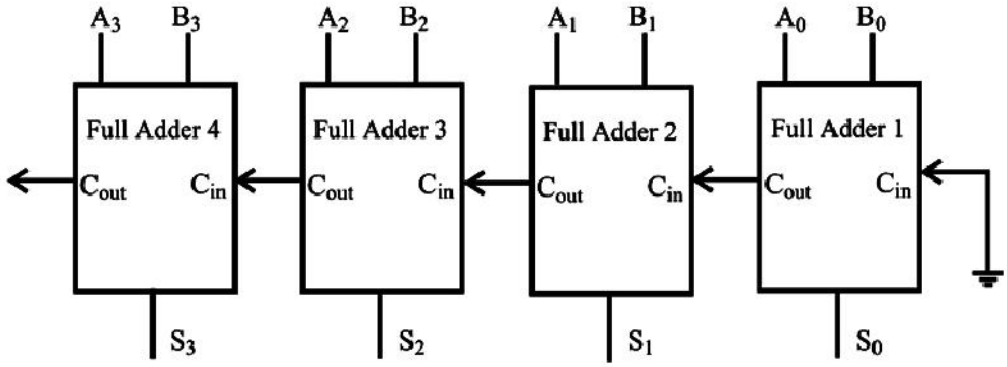


Full adder using NOR logic gates

19.2.3 Parallel Binary (Ripple Carry) Adder Circuit

From above discussion, we now understand that single 1-bit binary adders can be constructed from basic logic gates. But if we need to add together two n-bit numbers, then n number of 1-bit full adders need to be connected or "cascaded". The addition of multi-bit numbers can be achieved using several full-adders. For example, suppose we want to "add" together two 4-bit numbers, the two outputs of the first full adder will provide the first place digit sum (S) of the addition plus a carry-out bit that acts as the carry-in digit of the next binary adder. The second binary adder in the chain also produces a summed output (the 2nd bit) plus another carry-out bit and we can keep adding more full adders to the combination to add larger numbers, linking the carry bit output from the first full binary adder to the next full adder, and so on. Since all the bits of augend and addend are fed into adder circuits simultaneously and the additions in each position are taking place at the same time, such circuit is known parallel adder.

In a 4-bit ripple carry adder (binary adder), as shown in figure below, the input to each full adder will be A_i, B_i and C_i and the outputs will be S_i and C_{i+1}. The carry output of the lower order stage is connected to the carry input of the next higher order stage.



Thus, this type of circuit is called ripple carry adder, results in a sum ($S_3S_2S_1S_0$) and a carry output (C_{out}).

One main disadvantage of "cascading" together 1-bit **binary adders** to add large binary numbers is that if inputs A and B change, the sum at its output will not be valid until any carry-input has "rippled" through every full adder in the chain because the MSB (most significant bit) of the sum has to wait for any changes from the carry input of the LSB (less significant bit). Consequently, there will be a

finite delay before the output of the adder responds to any change in its inputs resulting in a accumulated delay. When the size of the bits being added is not too large (4 or 8 bits), or the summing speed of the adder is not important, this delay may not be important. But, however, when the size of the bits is larger, like 32 or 64 bits, or summation is required at a very high clock speed, this delay may become prohibitively large with the addition processes not being completed correctly within one clock cycle. This unwanted delay time is called Propagation delay. Also another problem called "overflow" occurs when an n-bit adder adds two parallel numbers together whose sum is greater than or equal to 2^n. However, there are several methods to reduce this delay.

One of the methods is to generate the carry-input signals directly from the A and B inputs rather than using the ripple arrangement above. This then produces another type of binary adder circuit called a *Carry Look Ahead Binary Adder* where the speed of the parallel adder can be greatly improved using carry-look ahead logic.

The advantage of carry look ahead adders is that the length of time a carry look ahead adder needs in order to produce the correct SUM is independent of the number of data bits used in the operation, unlike the cycle time a parallel ripple adder needs to complete the SUM which is a function of the total number of bits in the addend.

4-bit full adder circuits with carry look ahead features are available as standard IC packages in the form of the TTL 4-bit binary adder 74LS83 or the 74LS283 and the CMOS 4008 which can add together two 4-bit binary numbers and generate a SUM and a CARRY output as shown.

19.3 Binary Subtractor

The Binary Subtractor is another type of combinational arithmetic circuit, which is just opposite to the Binary Adder, where the **binary subtractor** is a decision making circuit that subtracts two binary numbers from each other, for example, X – Y to find the resulting difference between the two numbers. To the analogy of Binary Adder which produces a SUM (S) and a CARRY (C) bit when two binary numbers are added together, the *binary subtractor* produces a DIFFERENCE, (D) by using a BORROW bit, (B) from the previous column. Then obviously, the operation of subtraction is the opposite to that of addition. The

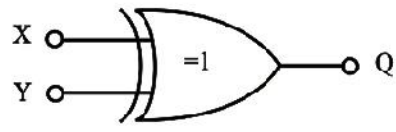
subtraction of one binary number from another is exactly the same idea as that for subtracting two decimal numbers but as the *binary number system* is a Base-2 numbering system which uses "0" and "1" as its two independent digits, large binary numbers which are to be subtracted from each other are therefore represented in terms of "0's" and "1's".

19.3.1 Binary Subtraction of Two Bits

Binary subtraction can take many forms but the rules for subtraction are the same whichever method is used. As binary notation only has two digits, thus subtracting "0" from "0" or "1" leaves the result unchanged as 0-0 = 0 and 1-0 = 1. Subtracting "1" from "1" results in "0", but subtracting "1" from "0" requires a borrow. In other words "0 – 1" requires a borrow as shown below.

0	1	1	(borrow)1 → 0
– 0	– 0	– 1	– 1
0	1	0	1

For the above simple problem, 1-bit subtraction, if the borrow bit is ignored then the result are same as that of an Exclusive-OR gate. To avoid any confusion in notations for subtractor, we will label the two input bits as X for the minuend and Y for the subtrahend because between a binary subtractor input labelled, B and the resulting borrow bit output from the binary subtractor also being labelled, B. Then the resulting truth table is the difference between the two input bits of a single binary subtractor is given as:

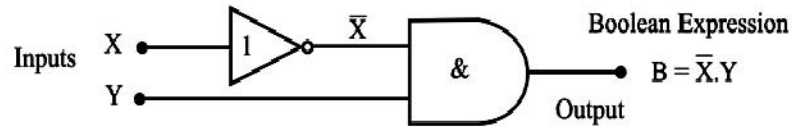


Y	X	Q (Output)
0	0	0
0	1	1
1	0	1
1	1	0

As with the Binary Adder, the difference between the two digits is only a "1" when these two inputs are not equal as given by the Ex-OR expression. However, we need an additional output to produce the borrow bit when input A = 0 and B = 1 for which, unfortunately, there are no standard logic gates that will produce an output

for this particular combination of X and Y inputs.

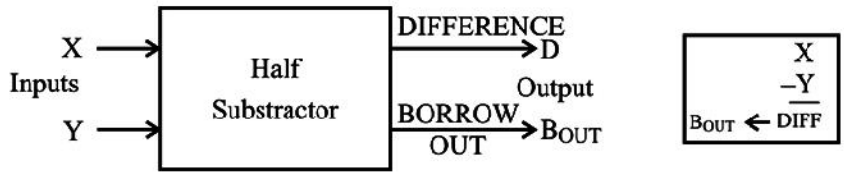
Since, it is known that an AND Gate produces an output "1" when both of its inputs X and Y are "1" (HIGH) so if we use an inverter or NOT Gate to complement the input X before it is fed to the AND gate, we can produce the required borrow output when X = 0 and Y = 1 as shown below.



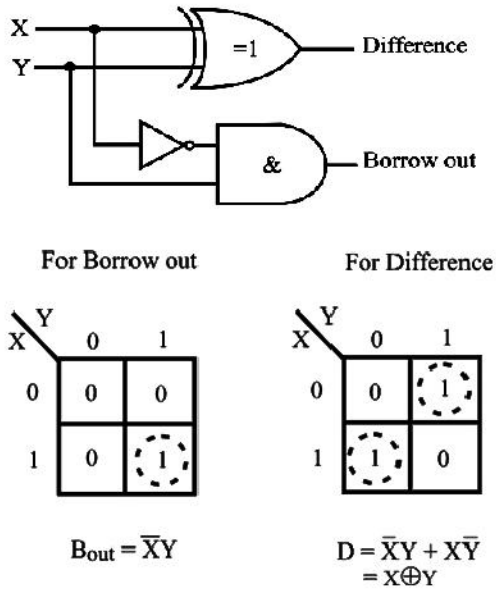
Then by combining the Exclusive-OR gate with the NOT-AND combination results in a simple digital binary subtractor circuit known commonly as the *Half Subtractor*.

19.3.2 Half Binary Subtractor

A half subtractor is a logical circuit that performs a subtraction operation on two binary digits. The half subtractor produces a sum and a borrow bit for the next stage.



Block Diagram of Half Subtractor



Logic circuit and K-map simplification of Half Subtractor

Truth Table			
Y	X	Difference (D)	Borrow-Out (B_{OUT})
0	0	0	0
0	1	1	0
1	0	1	1
1	1	0	0

As it is seen from the truth table of the half subtractor, the DIFFERENCE (D) output is the result of the Exclusive-OR gate and the Borrow-out (Bout) is the result of the NOT-AND combination. Then the Boolean expression for a half subtractor will be

For the DIFFERENCE (D) bit:

$$D = X \ \ XOR \ Y = X \oplus Y$$

For the BORROW (B_{OUT}) bit

$$B = NOT - X \ \ AND \ Y = \overline{X}.Y$$

On making the comparison of Boolean expressions of the half subtractor with a half adder, we can see that the two expressions for the SUM (adder) and DIFFERENCE (subtractor) are exactly the same and so they should be because of the Exclusive-OR gate function. The two Boolean expressions for the binary subtractor BORROW is also very similar to that for the adders CARRY. Then all that is needed to convert a half adder to a half subtractor is the inversion of the minuend input X.

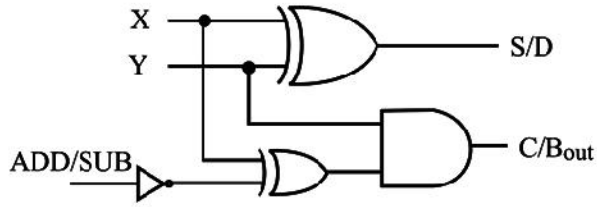
One major disadvantage of the *Half Subtractor* circuit (when used as a binary subtractor) is, that there is no provision for a "Borrow-in" from the previous circuit when subtracting multiple data bits from each other. Then we need to produce what is called a "full binary subtractor" circuit to take into account this borrow-in input from a previous circuit.

Corollary: Half Adder/Substractor

The half adder and half subtractor have only a very small difference in designing. A single circuit for half adder and half subtractor can be designed by using

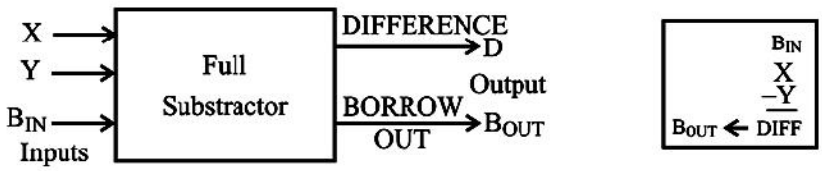
additional input signal like ADD/SUB. Therefore, the circuit start to act as a half adder if the ADD/SUB input is high otherwise it will act as half subtractor if the ADD/SUB input is low. The circuit is shown below, consists of X-OR gate to act as a inverter.

ADD/SUB = 0 for half adder and =1 for full adder

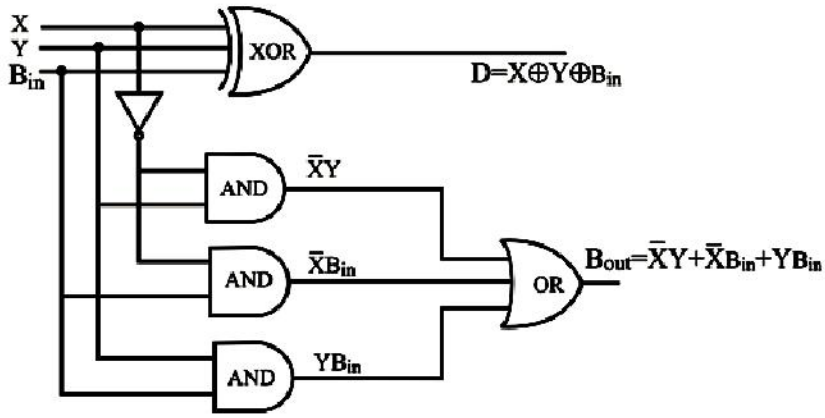


19.3.3 Full Binary Subtractor

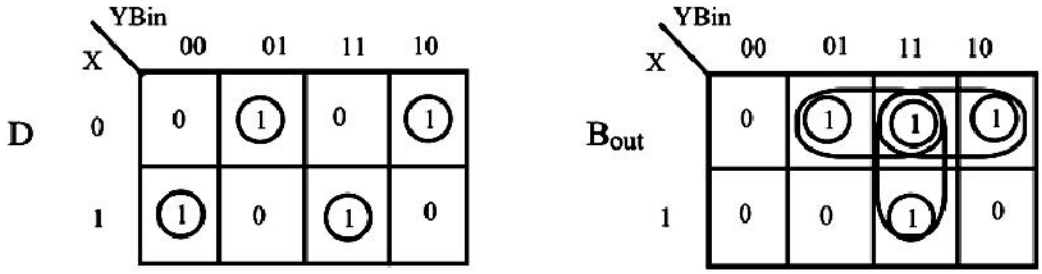
The main difference between the **Full Subtractor** and the **Half Subtractor** circuit is that a full subtractor has three inputs. The two single bit data inputs X (minuend) and Y (subtrahend) the same as before plus an additional *Borrow-in* (B_{IN}) input to receive the borrow generated by the subtraction process from a previous stage as shown below.



Then the combinational circuits of a "full subtractor" perform the operation of subtraction on three binary bits producing outputs for the difference "D" and borrow "B_{OUT}".



Logic Circuit for Full Subtractor



K-map for Full Subtractor

X	Y	B_{in}	Difference (D)	Borrow Out (B_{OUT})
0	0	0	0	0
0	1	0	1	1
1	0	0	1	0
1	1	0	0	0
0	0	1	1	1
0	1	1	0	1
1	0	1	0	0
1	1	1	1	1

Then the Boolean expression for a full subtractor is as follows:

For the DIFFERENCE (D) bit:

$$D = (\overline{X}.\overline{Y}.B_{IN}) + (\overline{X}.Y.\overline{B}_{IN}) + (X.\overline{Y}.\overline{B}_{IN}) + (X.Y.B_{IN})$$

which can be written as

$$\begin{aligned} D &= \overline{X}(\overline{Y}B_{IN} + Y\overline{B}_{IN}) + X(YB_{IN} + \overline{Y}\overline{B}_{IN}) \\ &= \overline{X}(Y \oplus B_{IN}) + X(\overline{Y \oplus B_{IN}}) \\ &= X \oplus Y \oplus B_{IN} \\ &= (X \ XOR \ Y) \ XOR \ B_{IN} \end{aligned}$$

and for the BORROW OUT (B_{OUT}) bit:

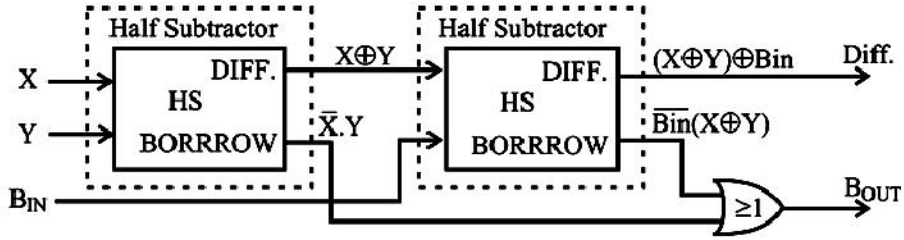
$$B_{OUT} = (\overline{X}.\overline{Y}.B_{IN}) + (\overline{X}.Y.\overline{B}_{IN}) + (X.\overline{Y}.\overline{B}_{IN}) + (X.Y.B_{IN})$$

which reduces to

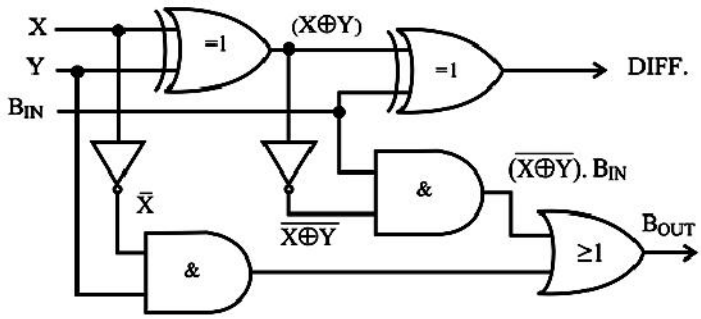
$$B_{OUT} = \overline{X} \ AND \ Y \ OR(\overline{X \ XOR \ Y})B_{IN} = X.Y + (\overline{X \oplus Y})B_{IN}$$

Full Subtractor Using Two Half Subtractor Circuits:

Similar to the binary adder circuit, the full subtractor can also realized as two half subtractors connected together, with the first half subtractor passing its borrow to the second half subtractor as follows.



As the full subtractor circuit above represents two half subtractors cascaded together, the truth table for the full subtractor will have eight different input combinations as there are three input variables, the data bits and the *Borrow-in*, B_{IN} input. Also includes the difference output, D and the Borrow-out, B_{OUT} bit.



Now, we will derive the proof that the above circuit act as full subtractor:

The expression for Difference "D" output is

$$D = (X \oplus Y) \oplus B_{IN} = X \oplus Y \oplus B_{IN},$$

which is the same as for full subtractor.

The expression for the Borrow output "B_{OUT}" is

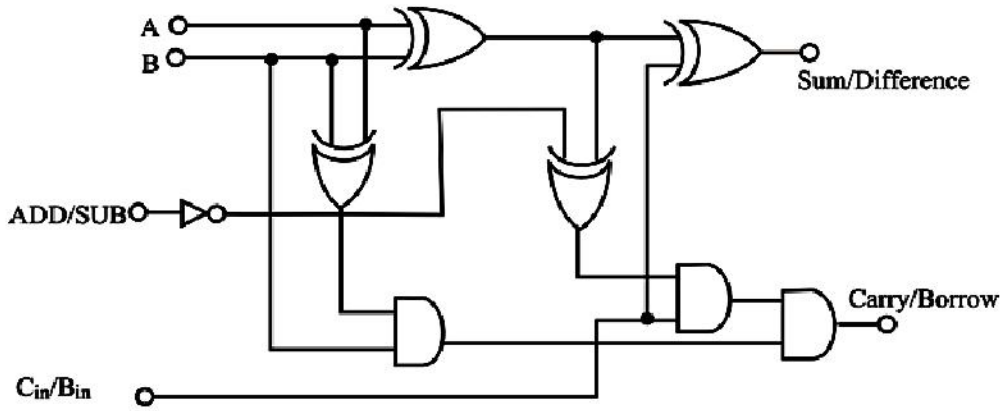
$$\begin{aligned} B_{OUT} &= (X \oplus Y)B_{IN} + \overline{X}Y \\ &= (\overline{\overline{X}Y \oplus X\overline{Y}})B_{IN} + \overline{X}Y \\ &= (\overline{XY} \oplus XY)B_{IN} + \overline{X}Y \\ &= \overline{X}\,\overline{Y}B_{IN} + XYB_{IN} + \overline{X}Y(1 + B_{IN}) \\ &= \overline{X}\,\overline{Y}B_{IN} + YB_{IN}(X + \overline{X}) + \overline{X}Y \\ &= \overline{X}\,\overline{Y}B_{IN} + YB_{IN} + \overline{X}Y \\ &= \overline{X}\,\overline{Y}B_{IN} + YB_{IN} + \overline{X}Y(1 + B_{IN}) \end{aligned}$$

$$= \overline{X}B_{IN}(Y + \overline{Y}) + YB_{IN} + \overline{X}Y$$
$$= \overline{X}B_{IN} + YB_{IN} + \overline{X}Y$$

This is just similar to the expression of Borrow-out for the full subtractor. Thus the above circuit acts as a full subtractor.

Corollary: Full Adder/Subtractor

In similarly fashion, full adder and full subtractor have only a minor difference. It is possible to construct a common circuit, shown below, for both full adder and full subtractor with an additional input ADD/SUB. It consists of Ex-OR gate to act as a inverter when ADD/SUB is low.



19.3.4 N-Bit Binary Parallel Subtractor

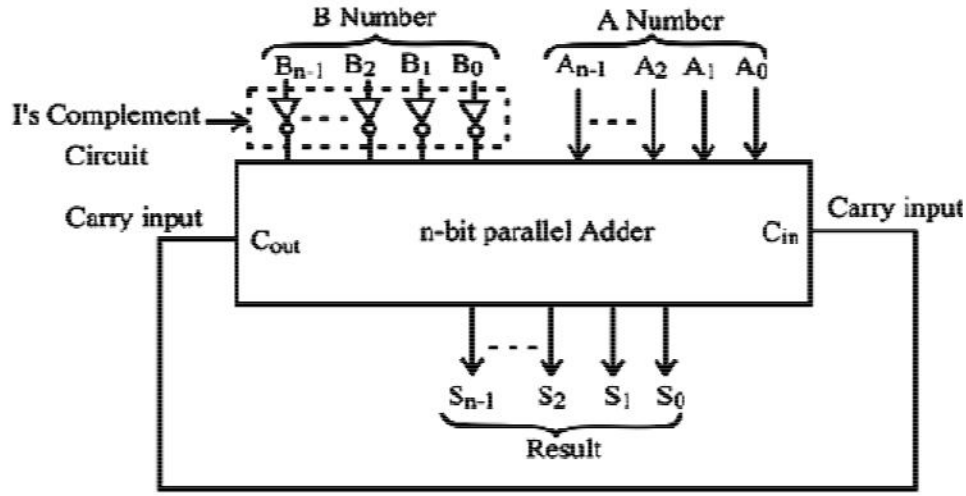
We can also have n number of 1-bit full binary subtractor connected or "cascaded" together to subtract two parallel n-bit numbers from each other. It is already known that the only difference between a full adder and a full subtractor was the inversion of one of the inputs. So by using an n-bit adder and n number of inverters (NOT Gates), the process of subtraction becomes an addition as we can use two ways in carrying out the subtractor, namely, 1's and 2's complement:

A – B =A+(-B)=A + 1's and 2's complement of B

(A) Using 1's Complement Method

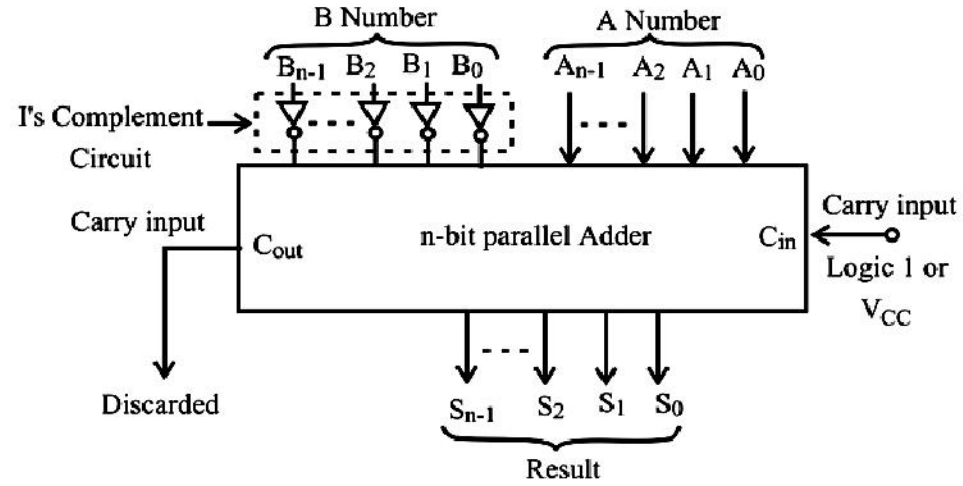
The following block diagram for n-bit parallel subtractor using 1's complement method is shown , where the 1's complement of B is added to A using binary adder and the inverters are connected at input B to obtain 1's complement. In this method, after addition, if a Carry is generated then it is added to the LSB of sum

via connecting Carry out (C_{out}) to Carry input (C_{in}). If A>B, then (C_{out} = 0) and the result is positive while if A<B then (C_{out} = 1) and the result is negative.



(B) Using 2's Complement Method

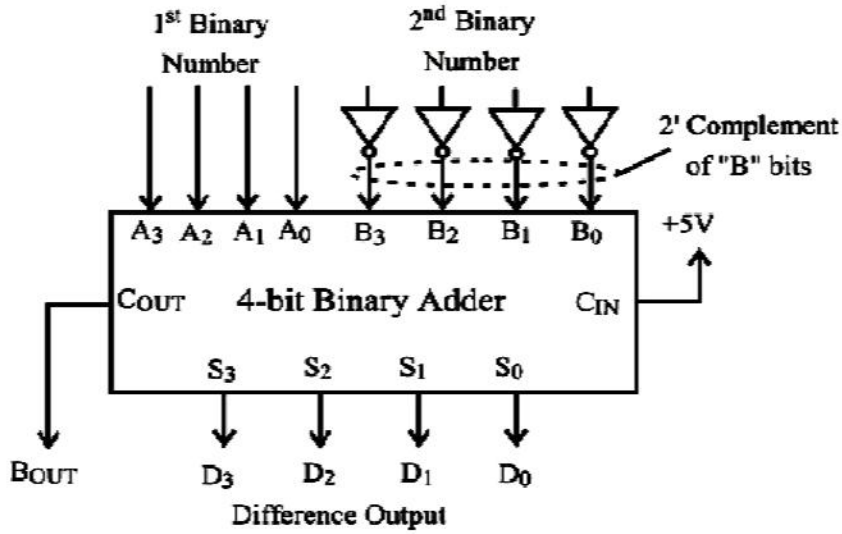
The block diagram of n-bit parallel subtractor using 2's complements method is shown below. Here the designing and operations are same except the Carry output (C_{out}) is discarded.



Then we can use a 4-bit full-adder ICs such as the 74LS283 and CD4008 to perform subtraction simply by using two's complement on the subtrahend, B inputs as X – Y is the same as saying, X + (-Y) which equals X plus the two's complement of Y.

If we wanted to use the 4-bit adder for addition once again, all we would need to do is set the carry-in (C_{IN}) input LOW at logic "0". Because we can use the 4-bit adder IC such as 74LS283 as a full-adder or a full-subtractor they are available as a

single adder/subtractor circuit with a single control input for selecting between the two operations.



19.4 Digital (Binary) Comparators

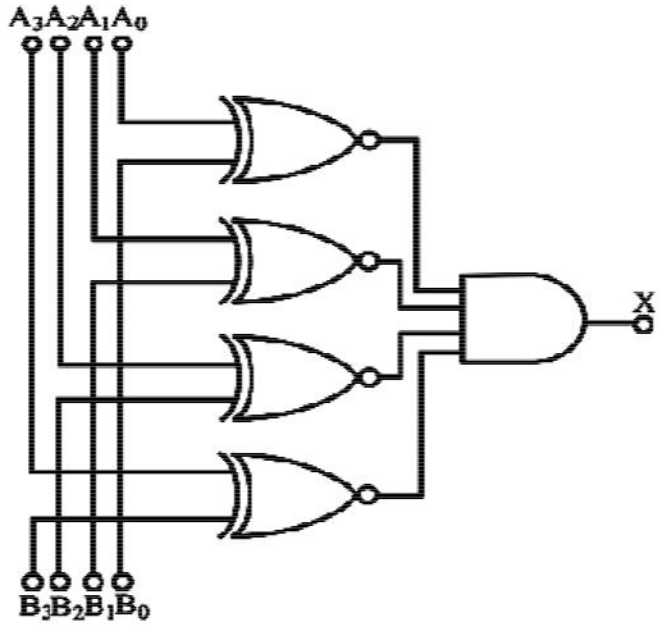
Binary comparators are also called digital comparators or logic comparators, which are used for testing whether the value represented by one binary word is greater than, less than, or equal to the value represented by another binary word. These comparators are also called combinational logic circuits because these are constructed from basic AND, NOR and NOT gates that compare the digital signals present at their input terminals and produce an output depending upon the condition of those inputs, like, along with being able to addition and subtraction, we should be able to compare them and determine whether the value of input A is greater than, smaller than or equal to the value at input B etc. The digital comparator operates on the principles of *Boolean Algebra*. There are two main types of d**igital comparator** available and these are.

- **Equality/Identity Comparator** – an *Equality/Identity Comparator* has only one output terminal for when A = B either "High" A = B = 1 or "Low" A = B = 0
- **Magnitude Comparator** – a *Magnitude Comparator* has three output terminals, one each for equality, A = B greater than, A > B and less than A < B

19.4.1 Equality Comparator

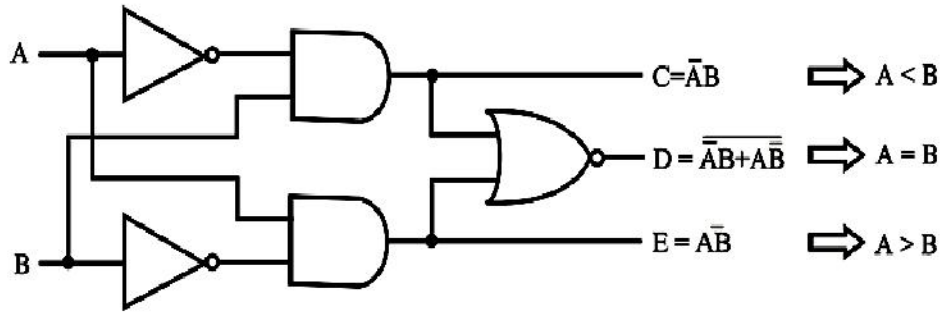
The logic circuit of an equality comparator is shown in figure below. It is the simplest multi-bit logic comparator, and can be used for such circuits as

electronic locks and security devices where a binary password consisting of multiple bits is input to the comparator to be compared with another preset word. In the circuit, a logic 1 will be present at the output if the two input words match, otherwise the output remains at 0. Therefore there is only one input combination that is correct, and the more bits the input word possesses, the more possible wrong combinations there are. The circuit of the equality comparator consists of an exclusive NOR (XNOR) gate per pair of input bits. If the two inputs are identical (both 1s or both 0s) an output of logic 1 is obtained. The outputs of the XNOR gates are then combined in an AND gate, the output of which will be 1, only when all the XNOR gates indicate matched inputs.



19.4.2 Magnitude Comparator

Actually, the purpose of a **digital comparator** is to compare a set of variables or unknown numbers, A (A_1, A_2, A_3, …. A_n, etc) against that of a constant or unknown value such as B (B_1, B_2, B_3, …. B_n, etc) and produce an output condition or flag depending upon the result of the comparison. For example, a magnitude comparator of two 1-bits (A and B) inputs would produce the following three output conditions when compared to each other, A>B, A=B, and A<B. It is useful if we want to compare two variables and want to produce an output when any of the above three conditions are achieved. To understanding this, an example, consider the simple 1-bit comparator below:



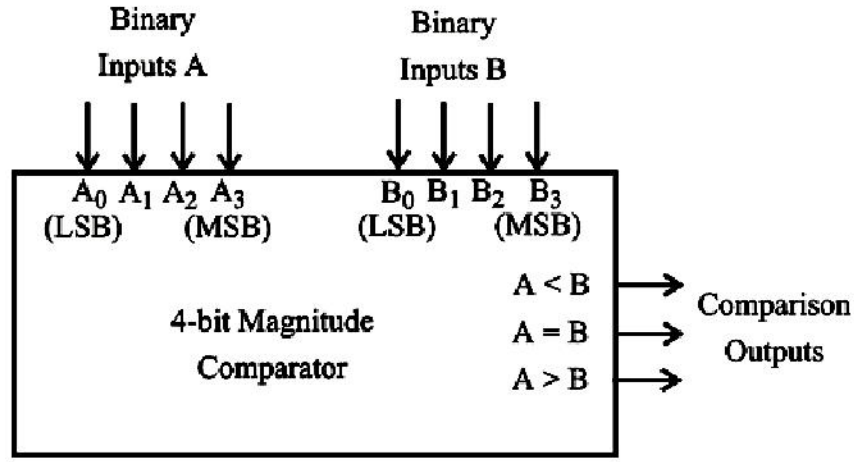
Then the operation of this comparator is given by the following Truth Table.

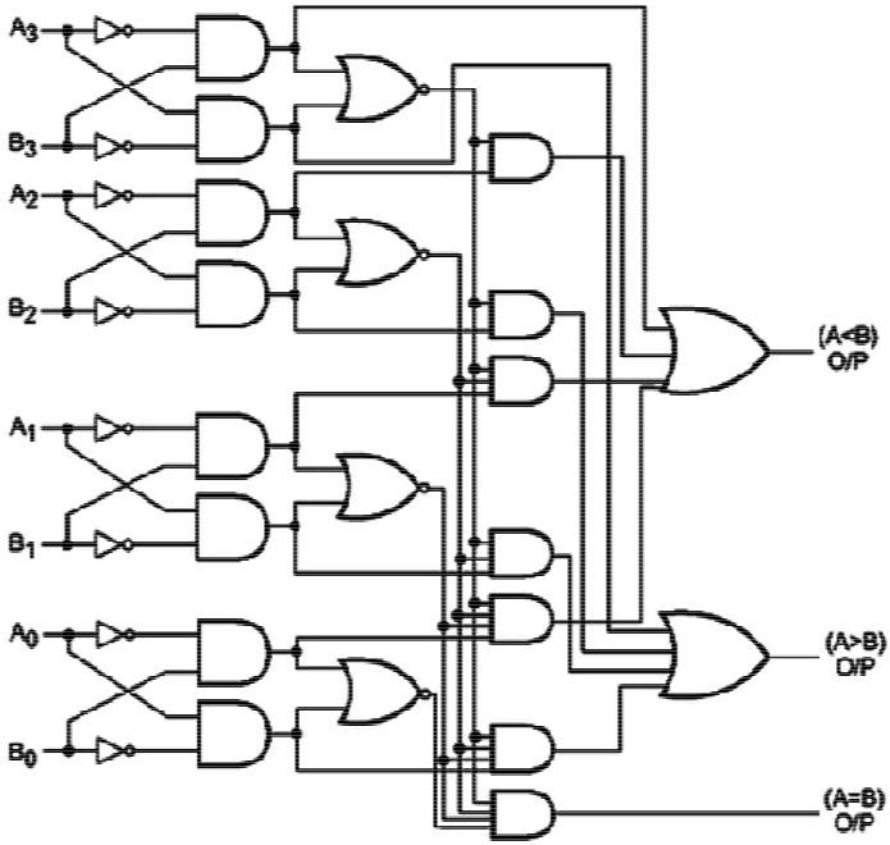
INPUTS		OUTPUTS		
B	A	A>B	A=B	A<B
0	0	0	1	0
0	1	1	0	0
1	0	0	0	1
1	1	0	1	0

Two distinguish features have been found from the above truth table. (i) The circuit does not distinguish between either two "0" or two "1"'s as an output A = B is produced when they are both equal, either A = B = "0" or A = B = "1" and (ii) the output condition for A = B resembles that of a commonly available logic gate, the Ex-NOR function (equivalence) on each of the n-bits giving: Q = A $\oplus$ B

Digital comparators are actually designed with the use of Exclusive-NOR gates for comparing their respective pairs of bits. When two binary or BCD values or variables against each other are compared, we compare the "magnitude" of these values, a logic "0" against a logic "1" which is where the term *Magnitude Comparator* comes from. As well as comparing individual bits, we can design larger bit comparators by cascading together. Multi-bit comparators can be constructed to compare whole binary (BCD words) to produce an output if one word is larger, equal to or less than the other.

The 4-bit Magnitude Comparator is a good example, where two 4-bit words are compared to each other and produce the relevant output with one word connected to inputs A and the other to be compared against connected to input B as shown here.

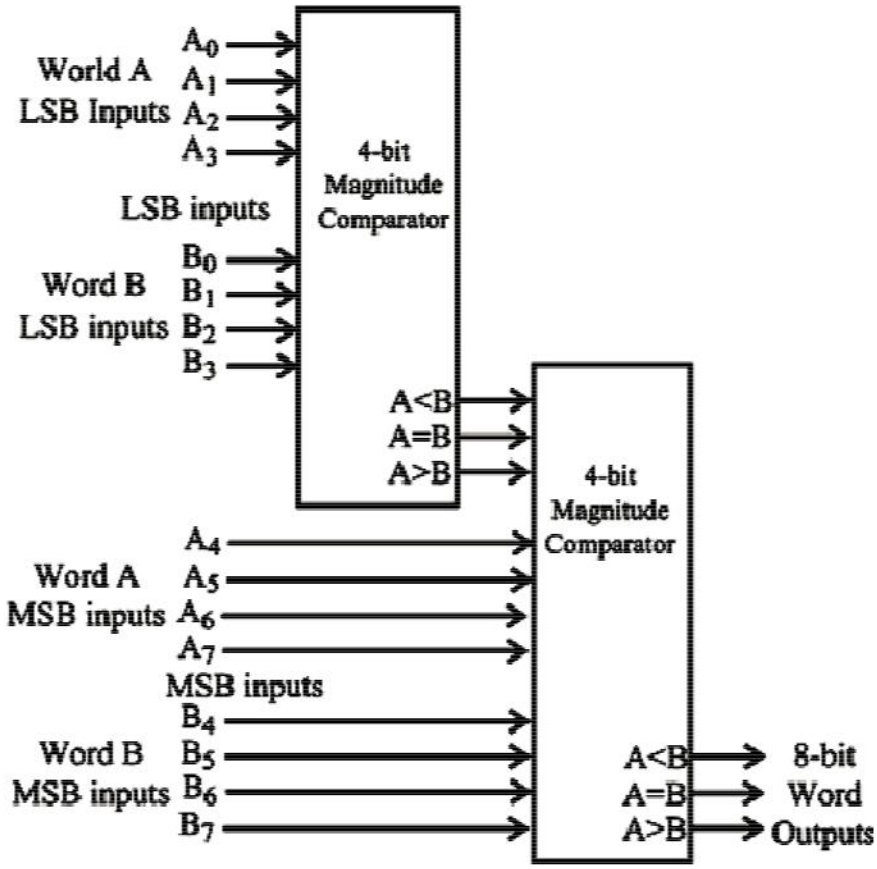




Commercially some digital comparators such as the TTL 74LS85 or CMOS 4063 are 4-bit magnitude comparator whose have additional input terminals that allow more individual comparators to be "cascaded" together to compare words larger than 4-bits with magnitude comparators of "n"-bits being produced.

When comparing large binary or BCD numbers, the comparator starts by comparing the highest-order bit (MSB) first. If equality exists, A = B then it

compares the next lowest bit and so on until it reaches the lowest-order bit, (LSB).



If equality still exists then the two numbers are defined as being equal. If inequality (either $A > B$ or $A < B$) is found, the relationship between the two numbers is determined and the comparison between any additional lower order bits stops. *Digital Comparator* are widely used in Analogue-to-Digital converters, (ADC) and Arithmetic Logic Units, (ALU) to perform a variety of arithmetic operations.

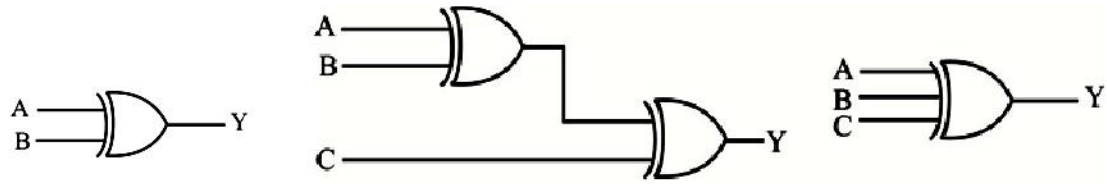
19.5 Parity Generators and Checkers

Let us first understand that what is a parity bit? The parity generating technique is error detection technique for the data transmission. In digital systems, when binary data is transmitted and processed, data may be subjected to noise so that such noise can alter 0s (of data bits) to 1s and 1s to 0s.

Hence, **parity bit** is added to the word containing data in order to make number of 1s either even or odd. Thus it is used to detect errors, during the transmission of binary data .The message containing the data bits along with parity bit is transmitted from transmitter node to receiver node. At the receiving end, the

number of 1s in the message is counted and if it doesn't match with the transmitted one, then it means there is an error in the data.

A parity generator is a combinational logic circuit that generates the parity bit in the transmitter, while a circuit that checks the parity in the receiver is called parity checker. The sum of the data bits and parity bits can be even or odd. In even parity, the added parity bit will make the total number of 1s an even amount whereas in odd parity the added parity bit will make the total number of 1s odd amount. The basic principle used in parity circuits is "that sum of odd number of 1s is always 1 and sum of even number of 1s is always zero" and such error detection and correction is done by using Ex-OR gates (because we know that Ex-OR gate produce zero output when there are even number of inputs).



Two Bits Three Bits

A combined circuit of parity generators and parity checkers are commonly used in digital systems to detect the single bit errors in the transmitted data word.

19.5.1 Parity Generators

It is a combinational circuit which accepts an n-1 bit stream data and generates the additional bit that is to be transmitted with the bit stream. This additional or extra bit is termed as a parity bit. This parity bit is of two types; even parity and odd parity.

(i) *Even Parity Bit Scheme*: In the data stream, the parity bit is '**0**' if there are **even number of 1s** and the parity bit is '**1**' if there are **odd number of 1s**.

(ii) *Odd Parity Bit Scheme*: In the data stream, the parity bit is '**1**' if there are **even number of 1**s and the parity bit is '**0**' if there are **odd number of 1s**.

Now we discuss in brief about the both even and odd parity generators.

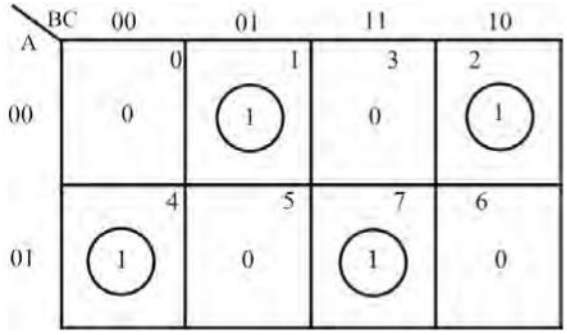
Even Parity Generator:

Consider that a 3-bit message is to be transmitted with an even parity bit. Let the three inputs A, B and C are applied to the circuits and output bit is the parity bit P. The total number of 1s must be even, to generate the even parity bit P.

In the following truth table of even parity generator in which 1 is placed as parity bit in order to make all 1s as even when the number of 1s in the truth table is odd.

3-Bit Inputs			Even Parity Bit Generator (P)
A	B	C	Y
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

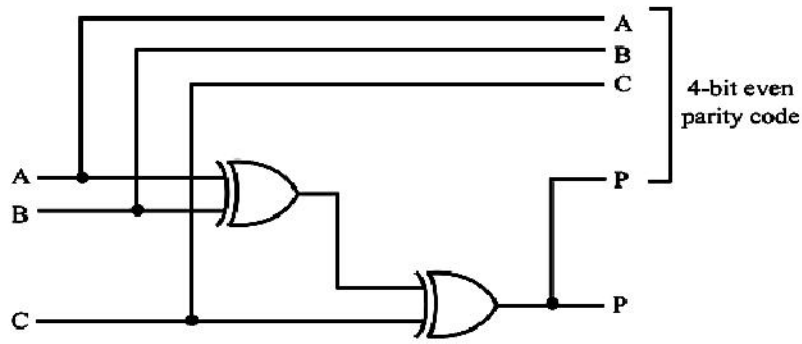
Truth table of the 3-bits even parity generators is simplified by using K-map as



The simplified expression can be written from the above truth table as

$$
\begin{aligned}
P &= \overline{A}\,\overline{B}\,C + \overline{A}\,B\overline{C} + A\overline{B}\,\overline{C} + ABC \\
&= \overline{A}(\overline{B}C + B\overline{C}) + A(\overline{B}\,\overline{C} + BC) \\
&= \overline{A}(B \oplus C) + A(\overline{B \oplus C}) \\
&= A \oplus B \oplus C
\end{aligned}
$$

The above Boolean algebra can be realized by using of two EX-OR gates as shown below, showing that three bit message along with the parity generated is transmitted to the receiving end where the parity checkers checks whether any error is present or not.

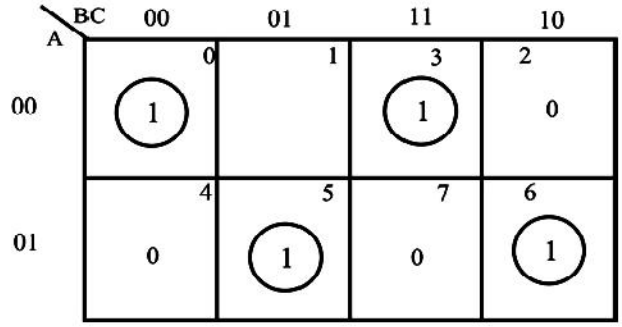


Odd Parity Generators:

Let us consider that the 3-bit data is to be transmitted with an odd parity bit. The three inputs are A, B and C and P is the output parity bit. The total number of bits must be odd in order to generate the odd parity bit. In the following truth table, 1 is placed in the parity bit in order to make the total number of bits odd when the total number of 1s in the truth table is even

3-Bit Inputs			Odd Parity Bit Generator (P)
A	B	C	Y
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	0

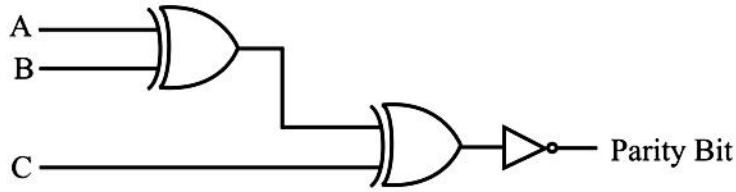
Truth table of 3-bits odd parity generators is simplified by using K- maps as



The Boolean algebra for this odd parity generator can be written as

$$P = A \oplus B \; Ex - NOR \; C$$

This expression can be implemented by using two EX-OR gate and one NOT gate in order to have 3-bit odd parity generator, as shown below.



In the above logic circuit, two inputs are applied at one Ex-OR gate, and the output of this Ex-OR gate and third input is applied to the EX-NOR gate, so that odd parity bit is produced.

19.5.2 Parity Checkers

It is a logic circuit that checks for possible errors in the transmission. This circuit can be an even parity checker or odd parity checker depending on the type of parity generated at the transmission end.

If the logic circuit is used as *even parity checker* then the number of input bits must always be even, and when a parity error occurs, the 'sum even' output goes low and 'sum odd' output goes high. On the other hand, if this logic circuit is used as an *odd parity checker*, the number of input bits should be odd, but if parity error occurs, the 'sum odd' output goes low and 'sum even' output goes high.

Even Parity Checkers:

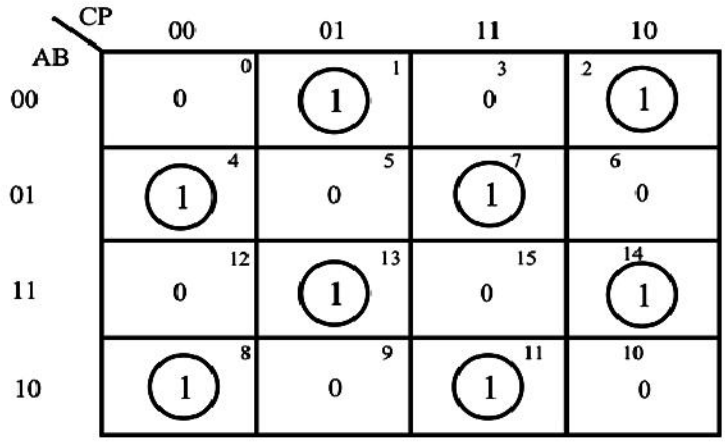
From even parity generator, three input message along with even parity bit is generated at the transmitting end. These 4 bits are applied as input to the parity checker circuit which checks the possibility of error in the data. Since the data is transmitted with even parity, four bits received at circuit must have an even number of 1s. If any error occurs, the received message consists of odd number of 1s. The output of the parity checker is denoted by C_P (Parity Error Check).

The below table shows the truth table for the even parity checker in which $C_P = 1$ if the error occurs, and $C_P = 0$ if no error occurs.

4-Bit Inputs	Parity Error Check

A	B	C	P	C_P
0	0	0	0	0
0	0	0	1	1
0	0	1	0	1
0	0	1	1	0
0	1	0	0	1
0	1	0	1	0
0	1	1	0	0
0	1	1	1	1
1	0	0	0	1
1	0	0	1	0
1	0	1	0	0
1	0	1	1	1
1	1	0	0	0
1	1	0	1	1
1	1	1	0	1
1	1	1	1	0

This truth table can be simplified using K- maps as

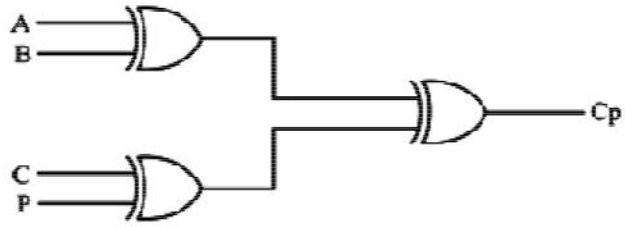


$$C_P = \overline{A}\,\overline{B}(\overline{C}\,P + C\,\overline{P}) + \overline{A}\,B(\overline{C}\,\overline{P} + C\,P) + A\,B(\overline{C}\,P + C\,\overline{P}) + A\,\overline{B}(\overline{C}\,\overline{P} + C\,P)$$

$$= \overline{A}\,\overline{B}(C \oplus P) + \overline{A}\,B(\overline{C \oplus P}) + AB(C \oplus P) + A\overline{B}(\overline{C \oplus P})$$

$$= (\overline{A}\,\overline{B} + A\,B)(C \oplus P) + (\overline{A}\,B + A\,\overline{B})(\overline{C \oplus P}) = (A \oplus B) \oplus (C \oplus P)$$

This logic expression for the even parity checker can be implemented by using three Ex-OR gates as shown below. If the received message consists of five bits, then one more Ex-OR gate will be required for the even parity checking.



Odd Parity Checker:

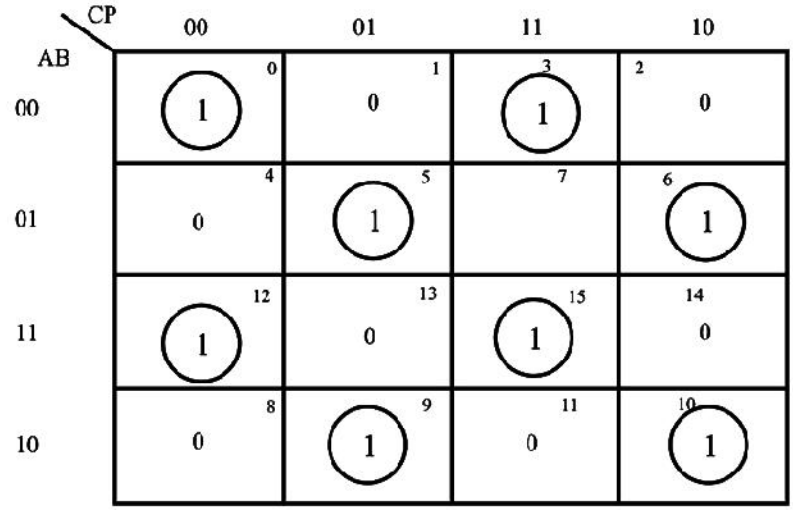
Similar to the even parity checker, a three bit message along with odd parity bit is transmitted at the transmitting end. Odd parity checker circuit receives these 4 bits and checks whether any error are present in the data. If the total number of 1s in the data is odd, then it indicates no error, whereas if the total number of 1s is even then it indicates the error since the data is transmitted with odd parity at transmitting end.

In the truth table for odd parity checker where $\mathbf{C_P = 1}$ if the 4-bit message received consists of **even number of 1s**, the error occurs and $\mathbf{C_P = 0}$ if the message contains **odd number of 1s**, no error occurs.

4-Bit Inputs				Parity Error Check
A	B	C	P	C_P
0	0	0	0	1
0	0	0	1	0
0	0	1	0	0
0	0	1	1	0
0	1	0	0	0
0	1	0	1	1
0	1	1	0	1
0	1	1	1	0
1	0	0	0	0

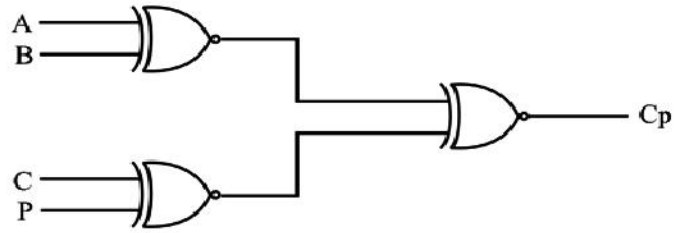
1	0	0	1	1
1	0	1	0	1
1	0	1	1	0
1	1	0	0	1
1	1	0	1	0
1	1	1	0	0
1	1	1	1	1

The expression for parity error check in the above truth table is simplified using K-maps as



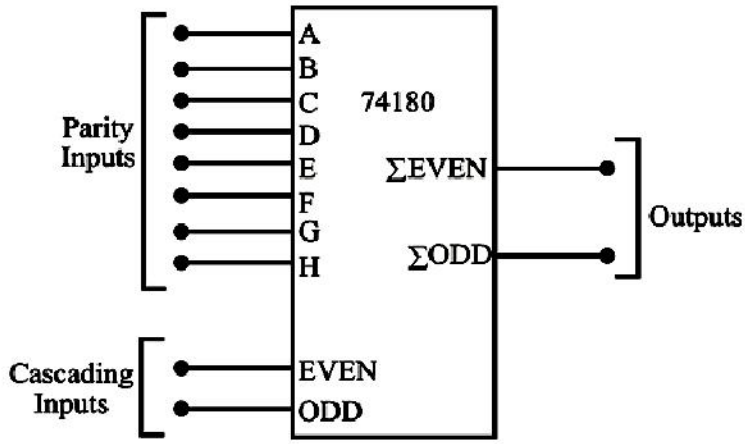
The odd parity checker can be designed by using three Ex-NOR gates as shown below, for which the Boolean algebra is written as

$$C_P = (A \textit{ Ex-NOR } B) \textit{ Ex-NOR } (C \textit{ Ex-NOR } P)$$



Therefore, the most commonly used and standard type of parity generator/checker IC is 74180, though there are various types of parity generator /checker ICs are available with different input configurations such as 5-bit, 4-bit, 9-bit, 12-bit, etc. IC-74180 is a 9-bit parity generator or checker, widely used to detect errors in high

speed data transmission or data retrieval systems. The figure below shows the pin diagram of 74180 IC.

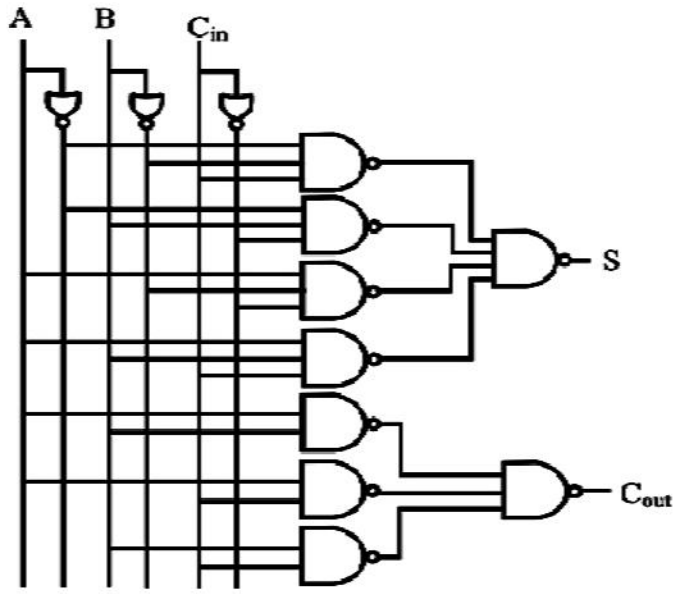


From A to H, there are eight parity inputs and two cascading inputs, for which there are two outputs even sum and odd sum. In implementing generator or checker circuits, unused parity bits must be tied to logic zero and the cascading inputs must not be equal.

19.6 Illustrative Examples

Example19.1: Construct the full adder using NAND gates and write the Boolean expression.

Sol.: The full adder using NAND gates only can be realized as



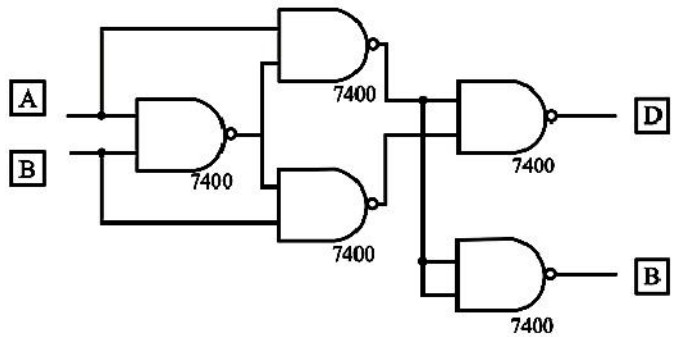
The expression for Sum (S) and Carry out (C_{out}) are as follows:

$$
\begin{aligned}
S &= A \oplus B \oplus C_{in} \\
&= (\overline{A \oplus B})C_{in} + (A \oplus B)C_{in} \\
&= (\overline{A}\,\overline{B} + AB)C_{in} + (\overline{A}\,B + A\overline{B})C_{in} \\
&= \overline{A}\,\overline{B}C_{in} + ABC_{in} + \overline{A}BC_{in} + A\overline{B}C_{in} \\
&= \overline{\overline{A}\,\overline{B}C_{in} + ABC_{in} + \overline{A}BC_{in} . + A\overline{B}C_{in}} \\
&= \overline{\overline{A}\,\overline{B}C_{in}} . \overline{ABC_{in}} . \overline{\overline{A}BC_{in}} . \overline{A\overline{B}C_{in}}
\end{aligned}
$$

$$
\begin{aligned}
C &= AB + AC_{in} + BC_{in} \\
&= \overline{AB + AC_{in} + BC_{in}} = \overline{\overline{AB}} . \overline{\overline{AC_{in}}} . \overline{\overline{BC_{in}}}
\end{aligned}
$$

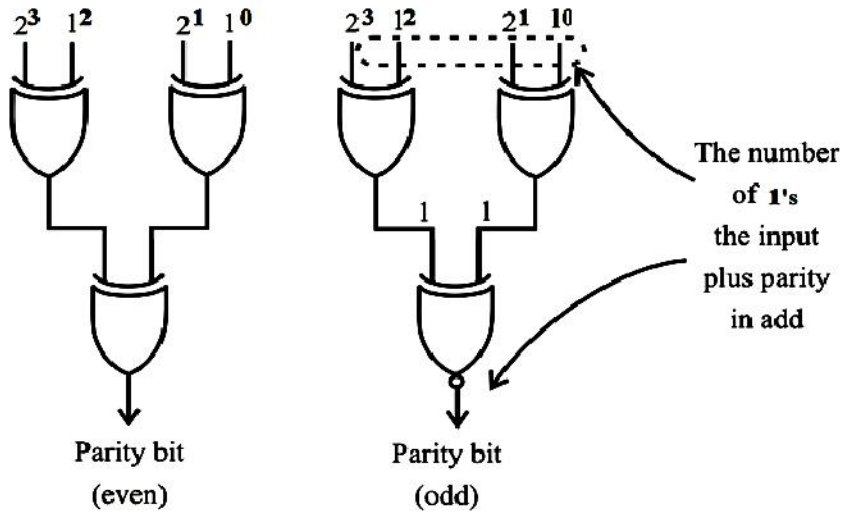
Example 19.2: Construct the half subtractor using NAND gates only.

Sol.:



Example 19.3: Give one example of an even and of an odd parity generator.

Sol.: Making use of Ex-OR logic gates, 4-bit even and odd parity generators are designed shown below.



19.7 Self Learning Exercise

Q.1 What is full adder and differentiate with half adder?

Q.2 Realize the half adder using NAND gates only.

Q.3 What do you mean by magnitude comparator?

Q.4 Draw the complete logic circuit of a 4- bit magnitude comparator using combinational circuit which can compare two binary numbers $R=R_4R_3R_2R_1$ and $S = S_4S_3S_2S_1$ and produce the output when R is greater than S or not.

19.8 Summary

The ALU is the number crunching part of the computer. It provides several different arithmetic and logic operations. A half adder is realized as a basic adder circuit. By itself it cannot perform addition of numbers of more than 1 bit. Full adders may be constructed from two half adders and an OR gate. The other possibilities have been also explored to design the adders or subtractors. The equality comparison is the easiest to implement as compare to inequality comparisons. The parity generators is the circuit creates the parity bits. Errors can be detected by checking the parity of binary words or subsets of bits of word.

19.9 Glossary

ALU: Arithmetic Logic Unit

Half Adder : A combinational circuit that performs the addition of two bits is called *half-adder*.

Full Adder : One that performs the addition of three bits (two significant and a carry) is *full-adder*.

19.10 Exercise

Q.1 What do you mean by subtraction of binary numbers? Design a full subtractor and implement its Boolean algebra by using suitable logic gates.

Q.2 Implement a full adder using two half adders. Draw the circuit and derive the expressions for sum and carry.

Q.3 Construct the full subtractor using the NAND gates only and derive the Boolean expression.

Q.4 Design a full subtractor using basic gates.

Q.5 Design a full adder using NAND gates only for both Sum and Carry outputs and use K- maps for reducing the expression generated from truth table.

References and Suggested Readings

1. S. Salivahanan, S. Arivazhagan; *Digital Circuits and Design* (2009), Vikas Publishing House Pvt. Ltd., New Delhi.
2. M. Morris Mano, Michael D. Ciletti; *Digital Design* (2013), Pearson Publishing Company.
3. A. P. Godse and D. A. Godse; Digital *Electronics and Logical Design*, Technical Publications, Pune.
4. *G.K. Kharate; Digital Electronics, (2013), Oxford University Press, New Delhi.*
5. *Amit Rathi, K.M. Bakwad; Digital Electronics (2010), Jain Publications, Jaipur*
6. *www.electronics-tutorials.ws › Combinational Logic*
7. *www.circuitstoday.com/*

UNIT-20
Encoders-Multiplexers, Decoder-Demultiplexers

Structure of the Unit

20.0 Objectives

20.1 Introduction

20.2 Multiplexer

20.3 Encoder

20.4 Self Learning Exercise-I

20.5 Demultiplexer and Decoder

20.6 Self Learning Exercise-II

20.7 Summary

20.8 Glossary

20.9 Answers to Self Learning Exercises

20.10 Exercise

20.11 Answers to Exercise

References and Suggested Readings

20.0 Objectives

After interacting with the material presented here students will be able to understand concept of

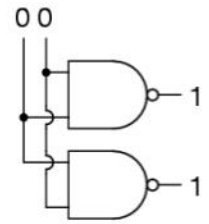
- Encoders

- Multiplexers
- Decoder
- Demultiplexers

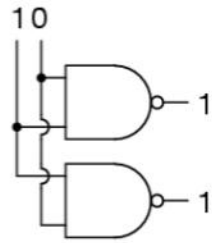
20.1 Introduction

The term "combinational" comes to us from mathematics. In mathematics a combination is an unordered set, which is a formal way to say that nobody cares which order the items came in. Most games work this way, if you rolled dice one at a time and get a 2 followed by a 3 it is the same as if you had rolled a 3 followed by a 2. With combinational logic, the circuit produces the same output regardless of the order the inputs are changed. Let's follow how two NAND gate works if we provide them inputs in different orders.

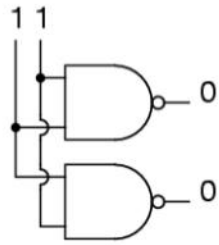
We begin with both inputs being 0.



We then set one input high.



We then set the other input high.



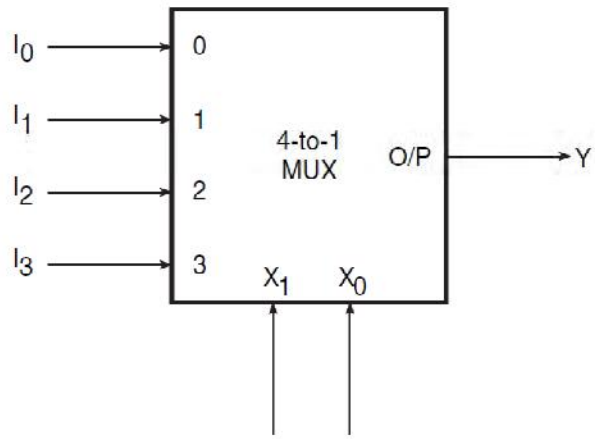
So NAND gates do not care about the order of the inputs, and you will find the same true of all the other gates covered up to this point (AND, XOR, OR, NOR, XNOR, and NOT).

Examples of combinational logic circuits are decoder, encoder,…etc.

20.2 Multiplexer

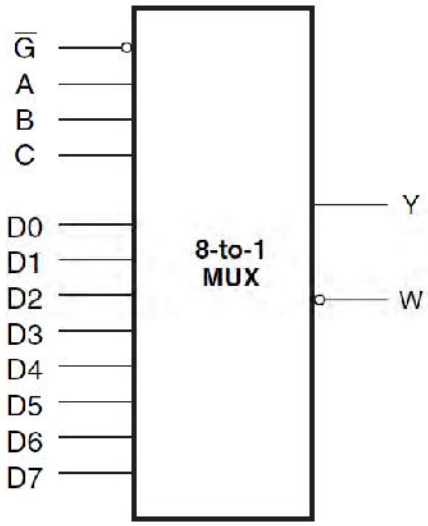
A multiplexer or MUX, also called a data selector, is a combinational circuit with more than one input line, one output line and more than one selection line. There are some multiplexer ICs that provide complementary outputs. Also, multiplexers in IC form almost invariably have an ENABLE or STROBE input, which needs to be active for the multiplexer to be able to perform its intended function. A multiplexer selects binary information present on any one of the input lines, depending upon the logic status of the selection inputs, and routes it to the output line.

If there are n selection lines, then the number of maximum possible input lines is 2^n and the multiplexer is referred to as a 2^n-to-1 multiplexer or $2^n \times 1$ multiplexer.



X_1	X_0	Y
0	0	I_0
0	1	I_1
1	0	I_2
1	1	I_3

Figures respectively show the circuit representation and truth table of a basic 4-to-1 multiplexer. Figure show the circuit representation and function table of 8-to-1 multiplexers. The 8-to-1 multiplexer is IC type number 74151 of the TTL family. It has an active LOW ENABLE input and provides complementary outputs.

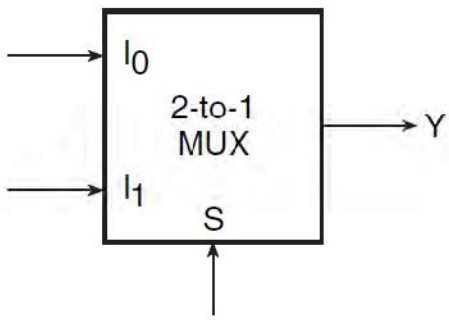


Inputs				Output	
Select			Enable		
C	B	A	$\overline{G}$	Y	W
X	X	X	H	L	H
L	L	L	L	D0	D0
L	L	H	L	D1	D1
L	H	L	L	D2	D2
L	H	H	L	D3	D3
H	L	L	L	D4	D4
H	L	H	L	D5	D5
H	H	L	L	D6	D6
H	H	H	L	D7	D7

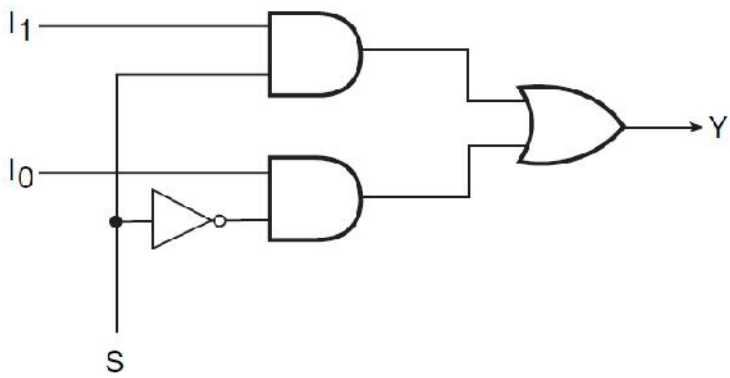
$\overline{G}$: ENABLE input
A, B, C : Select inputs
D0-D7 : Data inputs
Y,W : outputs

Inside the Multiplexer

We will briefly describe the type of combinational logic circuit found inside a multiplexer by considering the 2-to-1 multiplexer in Fig.a, the functional table of which is shown in Fig. b . Figure c shows the possible logic diagram of this multiplexer. The circuit functions as follows:

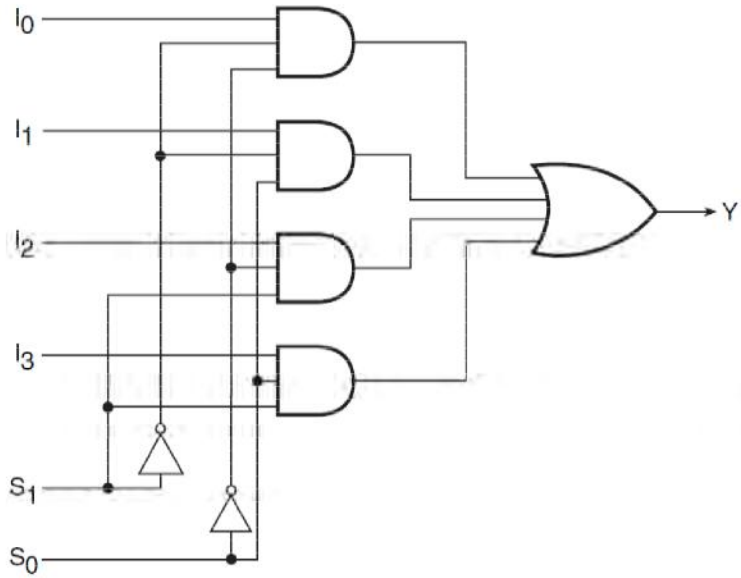


S	Y
0	I_0
1	I_1



- For S = 0, the Boolean expression for the output becomes $Y = I_0$.
- For S = 1, the Boolean expression for the output becomes $Y = I_1$.

Thus, inputs I_0 and I_1 are respectively switched to the output for S = 0 and S = 1.



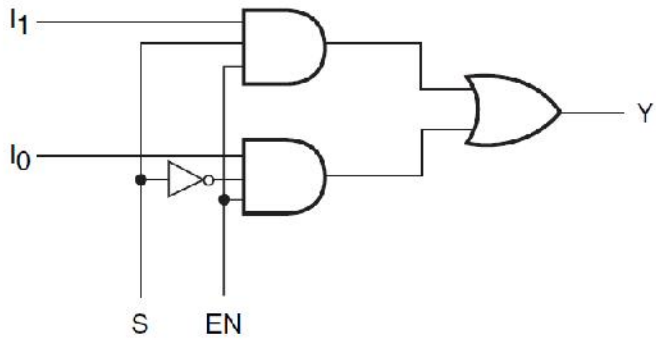
S_1	S_0	Y
0	0	I_0
0	1	I_1
1	0	I_2
1	1	I_3

Extending the concept further, Fig. shows the logic diagram of a 4-to-1 multiplexer. The input combinations 00, 01, 10 and 11 on the select lines respectively switch I_0, I_1, I_2 and I_3 to the output. The operation of the circuit is governed by the Boolean function.

$$Y = I_0.\overline{S}_1.\overline{S}_0 + I_1.\overline{S}_1.S_0 + I_2.S_1.\overline{S}_0 + I_3.S_1.S_0$$

As outlined earlier, multiplexers usually have an ENABLE input that can be used to control the multiplexing function. When this input is enabled, that is, when it is in logic '1' or logic '0' state, depending upon whether the ENABLE input is active HIGH or active LOW respectively, the output is enabled. The multiplexer functions normally.

When the ENABLE input is inactive, the output is disabled and permanently goes to either logic '0' or logic '1' state, depending upon whether the output is uncomplemented or complemented.

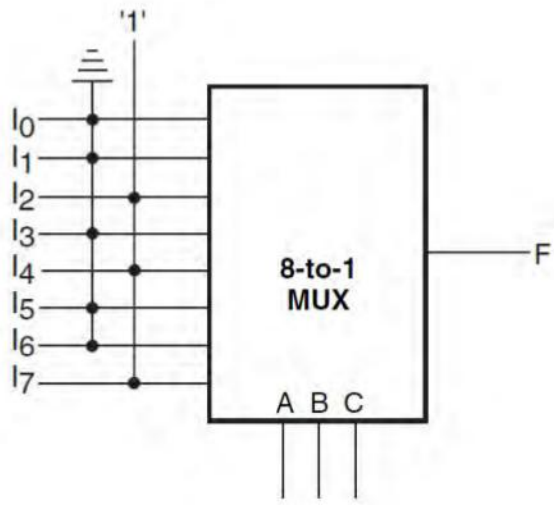


S	EN	Y
X	0	0
0	1	I_0
1	1	I_1

Figure shows how the 2-to-1 multiplexer can be modified to include an ENABLE input. The functional table of this modified multiplexer is also shown in Fig.

Implementing Boolean Functions with Multiplexers:

One of the most common applications of a multiplexer is its use for implementation of combinational logic Boolean functions. The simplest technique for doing so is to employ a 2^n-to-1 MUX to implement an n-variable Boolean function. The input lines corresponding to each of the minterms present in the Boolean function are made equal to logic '1' state. The remaining minterms that are absent in the Boolean function are disabled by making their corresponding input lines equal to logic '0'.



As an example, Fig. shows the use of an 8-to-1 MUX for implementing the Boolean function given by the equation

$$f(A,B,C) = \sum 2,4,7$$

In terms of variables A, B and C, above equation can be written as follows:

$$f(A,B,C) = \overline{A}.B.\overline{C} + A.\overline{B}.\overline{C} + A.B.C$$

As shown in Fig., the input lines corresponding to the three minterms present in the given Boolean function are tied to logic '1'. The remaining five possible minterms absent in the Boolean function are tied to logic '0'.

However, there is a better technique available for doing the same. In this, a 2^n-to-1 MUX can be used to implement a Boolean function with n + 1 variables. The procedure is as follows. Out of n +1 variables, n are connected to the n selection lines of the 2^n-to-1 multiplexer. The left-over variable is used with the input lines. Various input lines are tied to one of the following: '0', '1', the left-over variable and the complement of the left-over variable. Which line is given what logic status can be easily determined with the help of a simple procedure. The complete procedure is illustrated for the Boolean function given by equation.

It is a three-variable Boolean function. Conventionally, we will need to use an 8-to-1 multiplexer to implement this function. We will now see how this can be implemented with a 4-to-1 multiplexer. The chosen multiplexer has two selection lines. The first step here is to determine the truth table of the given Boolean function, which is shown below.

Minterm	*A*	*B*	*C*	*f*(*A*,*B*,*C*)
0	0	0	0	0
1	0	0	1	0
2	0	1	0	1
3	0	1	1	0
4	1	0	0	1
5	1	0	1	0
6	1	1	0	0
7	1	1	1	1

In the next step, two of the three variables are connected to the two selection lines, with the higherorder variable connected to the higher-order selection line. For instance, in the present case, variables B and C are the chosen variables for the

selection lines and are respectively connected to selection lines S_1 and S_0. In the third step, a table of the type shown below is constructed.

	I_0	I_1	I_2	I_3
$\bar{A}$	0	1	**2**	3
A	**4**	5	6	**7**
	A	0	A	A

Under the inputs to the multiplexer, minterms are listed in two rows, as shown. The first row lists those terms where remaining variable A is complemented, and second row lists those terms where A is uncomplemented. This is easily done with the help of the truth table.

The required minterms are identified or marked in some manner in this table. In the given table, these entries have been highlighted. Each column is inspected individually. If neither minterm of a certain column is highlighted, a '0' is written below that. If both are highlighted, a '1' is written. If only one is highlighted, the corresponding variable (complemented or uncomplemented) is written. The input lines are then given appropriate logic status. In the present case, I_0, I_1, I_2 and I_3 would be connected to A, 0, A and A respectively.

It is not necessary to choose only the leftmost variable in the sequence to be used as input to the multiplexer. Any of the variables can be used provided the implementation table is constructed accordingly. In the problem illustrated above, A was chosen as the variable for the input lines, and accordingly the first row of the implementation table contained those entries where 'A' was complemented and the second row contained those entries where A was uncomplemented. If we consider C as the left-out variable, the implementation table will be as shown below.

	I_0	I_1	I_2	I_3
$\bar{C}$	0	**2**	**4**	6
C	1	3	5	**7**
	0	$\bar{C}$	$\bar{C}$	C

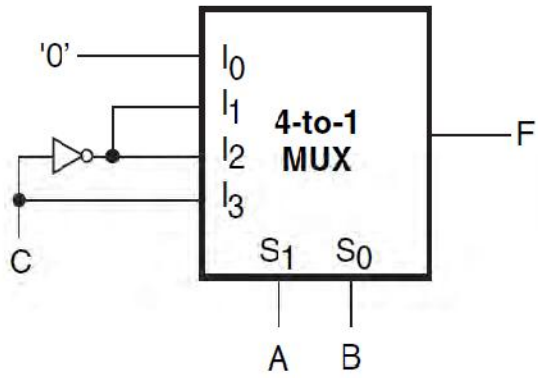


Figure shows the hardware implementation.

Cascading Multiplexer Circuits:

There can possibly be a situation where the desired number of input channels is not available in IC multiplexers. A multiple number of devices of a given size can be used to construct multiplexers that can handle a larger number of input channels. For instance, 8-to-1 multiplexers can be used to construct 16-to-1 or 32-to-1 or even larger multiplexer circuits. The basic steps to be followed to carry out the design are as follows:

1. If 2^n is the number of input lines in the available multiplexer and 2N is the number of input lines in the desired multiplexer, then the number of individual multiplexers required to construct the desired

multiplexer circuit would be 2^{N-n}.

2. From the knowledge of the number of selection inputs of the available multiplexer and that of the desired multiplexer, connect the less significant bits of the selection inputs of the desired multiplexer to the selection inputs of the available multiplexer.

3. The left-over bits of the selection inputs of the desired multiplexer circuit are used to enable or disable the individual multiplexers so that their outputs when ORed produce the final output.

20.3 Encoders

An *encoder* is a multiplexer without its single output line. It is a combinational

logic function that has 2^n (or fewer) input lines and n output lines, which correspond to n selection lines in a multiplexer. The n output lines generate the binary code for the possible 2^n input lines. Let us take the case of an octal-to-binary encoder. Such an encoder would have eight input lines, each representing an octal digit, and three output lines representing the three-bit binary equivalent. The truth table of such an encoder is given below.

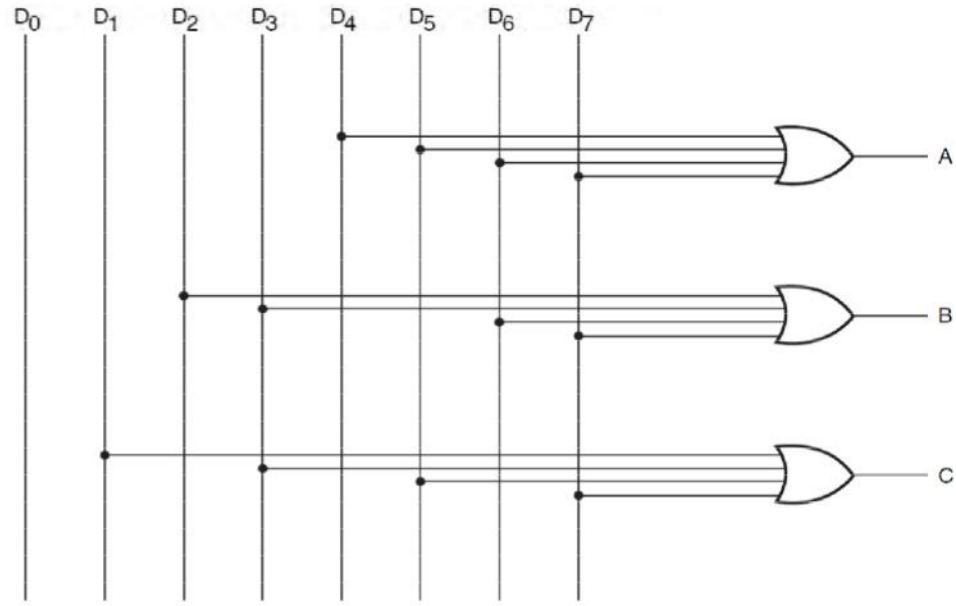
D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7	A	B	C
1	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	1
0	0	1	0	0	0	0	0	0	1	0
0	0	0	1	0	0	0	0	0	1	1
0	0	0	0	1	0	0	0	1	0	0
0	0	0	0	0	1	0	0	1	0	1
0	0	0	0	0	0	1	0	1	1	0
0	0	0	0	0	0	0	1	1	1	1

In the truth table, D_0 to D_7 represent octal digits 0 to 7. A, B and C represent the binary digits.

The eight input lines would have $2^8 = 256$ possible combinations. However, in the case of an octal-to-binary encoder, only eight of these 256 combinations would have any meaning. The remaining

combinations of input variables are 'don't care' input combinations. Also, only one of the input lines at a time is in logic '1' state.

Figure shows the hardware implementation of the octal-to-binary encoder described by the truth table . This circuit has the shortcoming that it produces an all 0s output sequence when all input lines are in logic '0' state. This can be overcome by having an additional line to indicate an all 0s input sequence.



Priority Encoder:

A *priority encoder* is a practical form of an encoder. The encoders available in IC form are all priority encoders. In this type of encoder, a priority is assigned to each input so that, when more

than one input is simultaneously active, the input with the highest priority is encoded. We will illustrate the concept of priority encoding with the help of an example. Let us assume that the octal to- binary encoder described in the previous paragraph has an input priority for higher-order digits.

Let us also assume that input lines D_2, D_4 and D_7 are all simultaneously in logic '1' state. In that case, only D_7 will be encoded and the output will be 111. The truth table of such a priority encoder will then be modified to what is shown below.

D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7	A	B	C
1	0	0	0	0	0	0	0	0	0	0
X	1	0	0	0	0	0	0	0	0	1
X	X	1	0	0	0	0	0	0	1	0
X	X	X	1	0	0	0	0	0	1	1
X	X	X	X	1	0	0	0	1	0	0
X	X	X	X	X	1	0	0	1	0	1
X	X	X	X	X	X	1	0	1	1	0
X	X	X	X	X	X	X	1	1	1	1

Looking at the last row, it implies that, if $D_7 = 1$, then, irrespective of the logic status of other inputs, the output is 111 as D7 will only be encoded.

20.4 Self Learning Exercise -I

Q.1 We have an eight-line to three-line priority encoder circuit with D_0, D_1,D_2,D_3,D_4,D_5,D_6 and D_7 as the data input lines. the output bits are A (MSB), B and C (LSB). Higher-order data bits have been assigned a higher priority, with D_7 having the highest priority. If the data inputs and outputs are active when LOW, determine the logic status of output bits for the following logic status of data inputs:

(a) All inputs are in logic '0' state.

(b) D_1 to D_4 are in logic '1' state and D_5 to D_7 are in logic '0' state.

(c) D_7 is in logic '0' state. The logic status of the other inputs is not known.

Q.2 What is the working of a combinational logic.

Q.3 Draw truth tables for the following:

(a) an 4-to-1 multiplexer with active LOW inputs and an active LOW ENABLE input;

(b) an eight-line to three-line priority encoder with active LOW inputs and outputs and an active LOW ENABLE input.

20.5 Demultiplexers and Decoders

A *demultiplexer* is a combinational logic circuit with an input line, 2^n output lines and n select lines. It routes the information present on the input line to any of the output lines. The output line that gets the information present on the input line is decided by the bit status of the selection lines. A *decoder* is a special case of a demultiplexer without the input line.

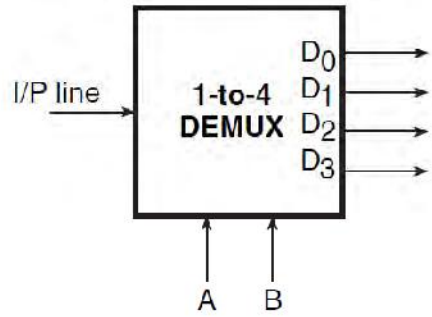


Figure shows the circuit representation of a 1-to-4 demultiplexer.

I/P	Select		O/P			
	A	B	D0	D1	D2	D3
1	0	0	1	0	0	0
1	0	1	0	1	0	0
1	1	0	0	0	1	0
1	1	1	0	0	0	1

Figure shows the truth table of the demultiplexer when the input

line is held HIGH. A decoder, as mentioned earlier, is a combinational circuit that decodes the information on n input lines to a maximum of 2^n unique output lines.

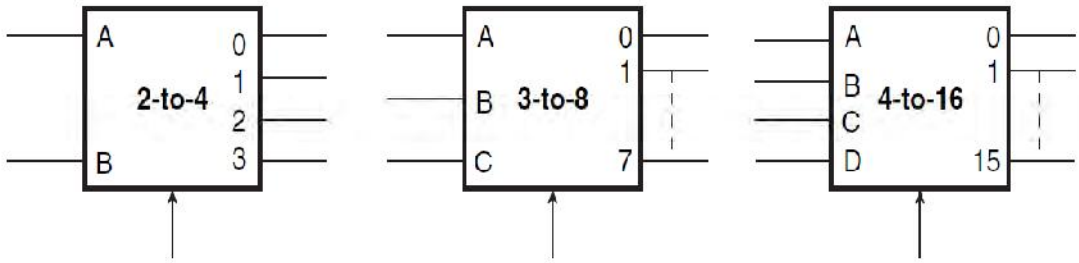
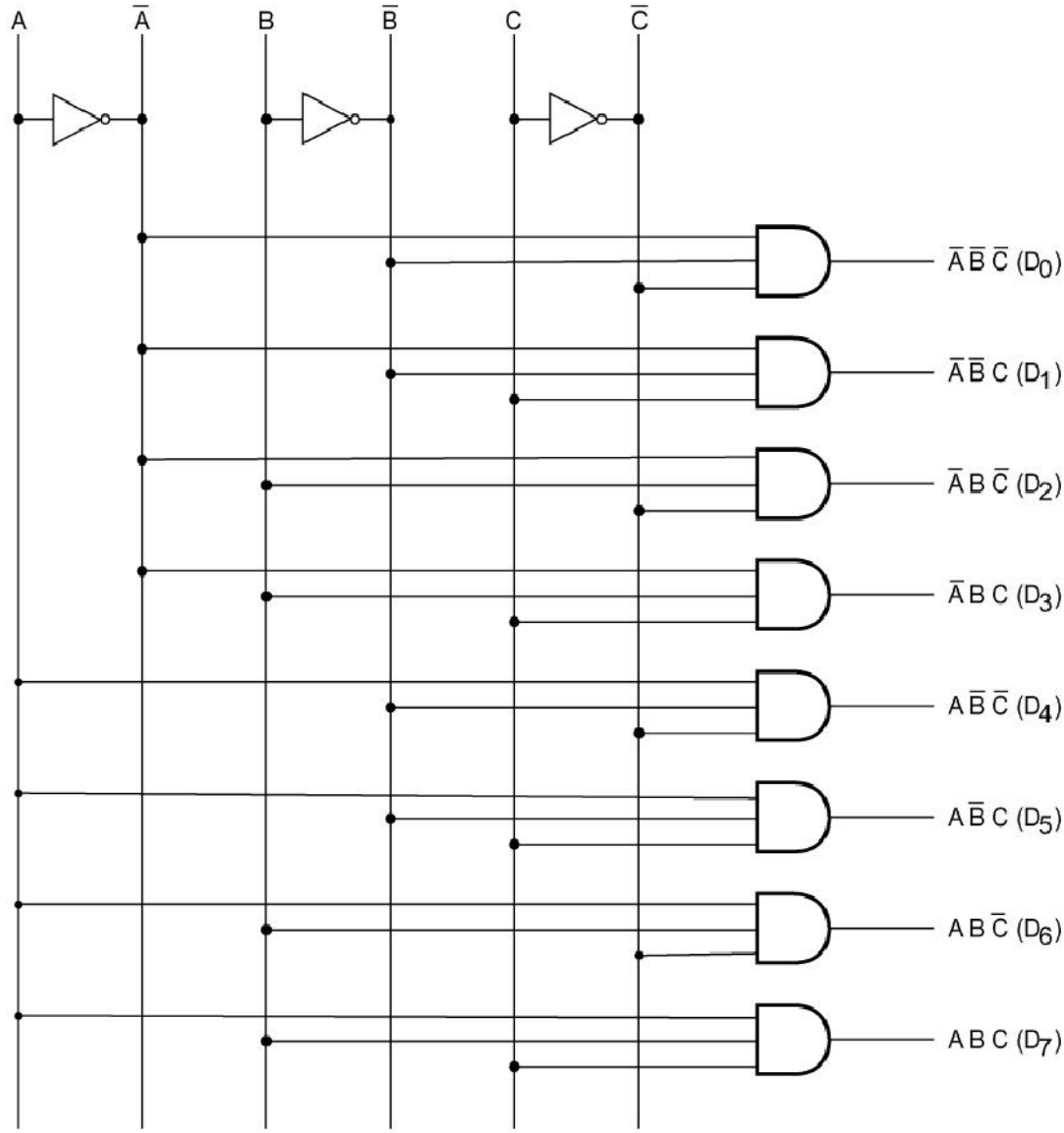


Figure shows the circuit representation of 2-to-4, 3-to-8 and 4-to-16 line decoders. If there are some unused or 'don't care' combinations in the n-bit code, then there will be fewer than 2^n output lines. As an illustration, if there are three input lines, it can have a maximum of eight unique output lines. If, in the three-bit input code, the only used three-bit combinations are 000, 001, 010, 100, 110 and 111 (011 and 101 being either unused or don't care combinations), then this decoder will have only six output lines. In general, if n and m are respectively the numbers of input and output lines, then $m \leq 2^n$.

A decoder can generate a maximum of 2^n possible minterms with an n-bit binary code. In order to illustrate further the operation of a decoder, consider the logic circuit diagram in Fig.



This logic circuit, as we will see, implements a 3-to-8 line decoder function. This decoder has three inputs designated as A, B and C and eight outputs designated as D_0, D_1, D_2, D_3, D_4, D_5, D_6 and D_7.

INPUTS			OUTPUTS							
A	B	C	D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7
0	0	0	1	0	0	0	0	0	0	0
0	0	1	0	1	0	0	0	0	0	0
0	1	0	0	0	1	0	0	0	0	0
0	1	1	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0
1	0	1	0	0	0	0	0	1	0	0
1	1	0	0	0	0	0	0	0	1	0
1	1	1	0	0	0	0	0	0	0	1

From the truth table given along with the logic diagram it is clear that, for any given input combination, only one of the eight outputs is in logic '1' state. Thus, each output produces a certain minterm that corresponds to the binary number currently present at the input. In the present case, D_0, D_1, D_2, D_3, D_4, D_5, D_6 and D_7 respectively represent the following minterms:

$$D_0 \rightarrow \overline{A}.\overline{B}.\overline{C}, D_1 \rightarrow \overline{A}.\overline{B}.C, D_2 \rightarrow \overline{A}.B.\overline{C}, D_3 \rightarrow \overline{A}.B.C, D_4$$
$$\rightarrow \overline{A}.\overline{B}.\overline{C}, D_5 \rightarrow A.\overline{B}.C, D_6 \rightarrow A.B.\overline{C}, D_7 \rightarrow A.B.C$$

Implementing Boolean Functions with Decoders:

A decoder can be conveniently used to implement a given Boolean function. The decoder generates the required minterms and an external OR gate is used to produce the sum of minterms.

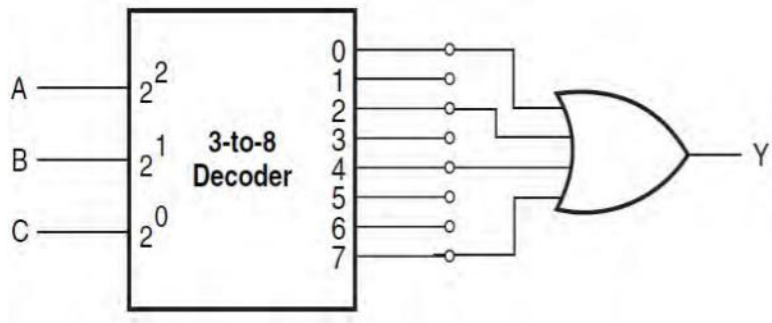


Figure shows the logic diagram where a 3-to-8 line decoder is used to generate the Boolean function given by the equation

$$Y = A.\overline{B}.\overline{C} + \overline{A}.B.\overline{C} + A.B.C + \overline{A}.\overline{B}.\overline{C}$$

In general, an n-to-2^n decoder and m external OR gates can be used to implement any combinational circuit with n inputs and m outputs. We can appreciate that a Boolean function with a large number of minterms, if implemented with a decoder and an external OR gate, would require an OR gate with an equally large number of inputs. Let us consider the case of implementing a four-variable Boolean function with 12 minterms using a 4-to-16 line decoder and an external OR gate. The OR gate here needs to be a 12-input gate. In all such cases, where the number of minterms in a given Boolean function with n variables is greater than $2^n/2$ (or 2^{n-1}), the complement Boolean function will have fewer minterms. In that case it would be more advantageous to do NORing of minterms of the complement Boolean function using a NOR gate rather than doing ORing of the given function using an OR gate. The output will be nothing but the given Boolean function.

Cascading Decoder Circuits:

There can possibly be a situation where the desired number of input and output lines is not available in IC decoders. More than one of these devices of a given size may be used to construct a decoder that can handle a larger number of input and output lines. For instance, 3-to-8 line decoders can be used to construct 4-to-16 or 5-to-32 or even larger decoder circuits. The basic steps to be followed to carry out the design are as follows:

1. If n is the number of input lines in the available decoder and N is the number of input lines in the desired decoder, then the number of individual decoders required to construct the desired decoder circuit would be 2^{N-n}.

2. Connect the less significant bits of the input lines of the desired decoder to the input lines of the available decoder.

3. The left-over bits of the input lines of the desired decoder circuit are used to enable or disable the individual decoders.

4. The output lines of the individual decoders together constitute the output lines, with the outputs of the less significant decoder constituting the less significant output lines and those of the higher–order decoders constituting the more significant output lines.

20.6 Self Learning Exercise- II

Q.1 Draw circuit representation of a 1-to-4 demultiplexer and truth table. when the input line is held HIGH.

Q.2 Define demultiplexer.

Q.3 Write down the Boolean function of 3-to-8 line decoder.

20.7 Summary

In this chapter we discussed in detail various types of decoder, encoder, multiplexer and demultiplexer.

20.8 Glossary

Encode: To express given information by means of a code. **Encoder** An encoder is an electro-mechanical device that translates mechanical motion (such as position, velocity, acceleration, speed, direction) into electrical signals.

Decoder: - A receiver that allows for control of the model or device in which it is installed. Each decoder is assigned (by the DCC system) an individual address that lets it receive only the information (digital packet) sent by the base station to it and not any of the other decoders on the track rails. There are typically two types of decoders.

Multiplexer: A device that combines two or more signals into one output.

Demultiplexer: A module that separates two or more signals previously combined by compatible multiplexing equipment.

20.9 Answer To Self Learning Exercises

Answer To Self Learning Exercise-I

Ans.1: (a) Since all inputs are in logic '0' state, it implies that all inputs are active.

Since D_7 has the highest priority and all inputs and outputs are active when LOW, the output bits are A = 0, B = 0 and C = 0.

(b) Inputs D_5 to D_7 are the ones that are active. among these, D_7 has the highest priority. Therefore, the output bits are A = 0, B = 0 and C = 0.

(c) D_7 is active. Since D_7 has the highest priority, it will be encoded irrespective of the logic status of other inputs. Therefore, the output bits are

A = 0, B = 0 and C = 0.

Ans.2: Combinational logic circuit produces the same output regardless of the order the inputs are changed.

Ans.3: (a)Section 20.2 (b)Section 20.3

Answer To Self Learning Exercise-II

Ans.1: Section 20.5

Ans.2: A demultiplexer is a combinational logic circuit with an input line, 2^n output lines and n select lines. It routes the information present on the input line to any of the output lines. The output line that gets the information present on the input line is decided by the bit status of the selection lines.

Ans.3: $Y = A.\overline{B}.\overline{C} + \overline{A}.B.\overline{C} + A.B.C + \overline{A}.\overline{B}.\overline{C}$

20.10 Exercise

Q.1 Implement the three-variable Boolean function $F = \overline{A}.C + A.\overline{B}.C + A.B.\overline{C}$

using (i) an 8-to-1 multiplexer and (ii) a 4-to-1 multiplexer.

Q.2 Write short note on

1. Decoder 2. Encoder 3. Multiplexer 4. Demultiplexer

Q.3 What is a multiplexer circuit? Briefly describe one or two applications of a multiplexer?

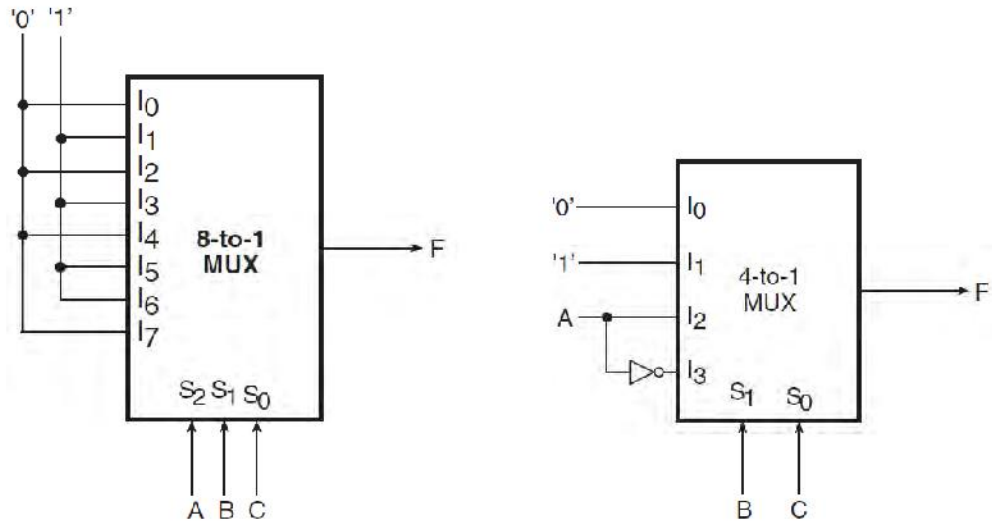
Q.4 What is an encoder? How does a priority encoder differ from a conventional encoder?

Q.5 What is a demultiplexer and how does it differ from a decoder? Can a decoder be used as a demultiplexer? If yes, from where do we get the required input line?

Q.6 Briefly describe how we can use a decoder optimally to implement a given Boolean function?

20.11 Answer To Exercise

Ans.1: (i) (ii)



Reference and Suggested Readings

1. Digital Fundamentals by T. L. Floyd, (2005).
2. Schaum's Outline Series of Digital Principles by R. L.Tokheim, (1994).
3. Digital Systems – Principles and Applications by R. J. Tocci, (2006).
4. Practical Digital Electronics by N. P. Cook, (2003).
5. Fundamentals of Digital Logic and Microcomputer Design by M.Rafiquzzaman, (2005).
6. Logic and Computer Design Fundamentals by M. Morris Mano, and C. R. Kime, (2003).

UNIT-21
Flip Flops

Structure of the Unit

21.0 Objectives

21.1 Introduction

21.2 Synchronous and Asynchronous Operation

21.3 Various types of Flip-flops and their Behavior

21.4 Self Learning Exercise-I

21.5 *Edge-Triggered Flip-flops*

21.6 *Pulse-Triggered (Master-Slave) Flip-flops*

21.7 Self Learning Exercise-II

21.8 Summary

21.9 Glossary

21.10 Answers to Self Learning Exercises

21.11 Exercise

References and Suggested Readings

21.0 Objectives

After interacting with the material presented here students will be able to understand

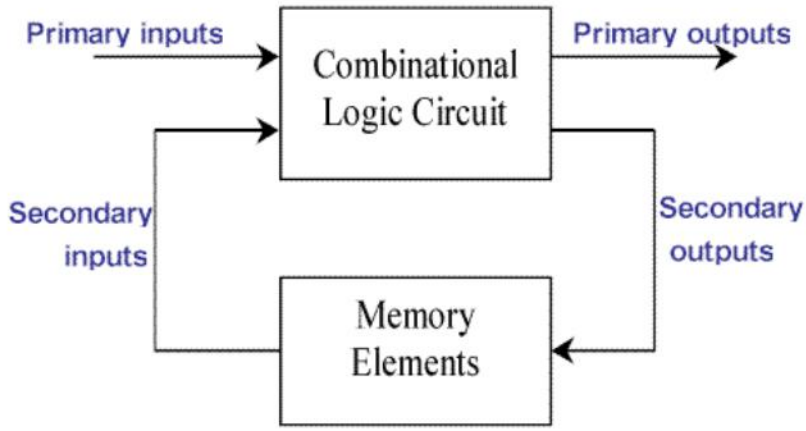
1. Synchronous and Asynchronous Operation,
2. Various types of Flip-flop Behavior
3. Edge triggered flip flops
4. *Pulse-Triggered (Master-Slave) Flip-flops*

21.1 Introduction

Combinational logic refers to circuits whose output is strictly depended on the present value of the inputs. As soon as inputs are changed, the information about

the previous inputs is lost, that is, combinational logics circuits have no memory. In many applications, information regarding input values at a certain instant of time is required at some future time. Although every digital system is likely to have combinational circuits, most systems encountered in practice also include memory elements, which require that the system be described in terms of sequential logic. Circuits whose outputs depends not only on the present input value but also the past input value are known as sequential logic circuits. The mathematical model of a sequential circuit is usually referred to as a sequential machine.

A general block diagram of a sequential circuit is shown below in Figure.



The diagram consists of combinational circuit to which memory elements are connected to form a feedback path. The memory elements are devices capable of storing binary information within them. The combinational part of the circuit receives two sets of input signals:

- one is primary (coming from the circuit environment) and
- secondary (coming from memory elements).

The particular combination of secondary input variables at a given time is called the present state of the circuit. The secondary input variables are also known as the state variables.

The block diagram shows that the external outputs in a sequential circuit are a function not only of external inputs but also of the present state of the memory elements. The next state of the memory elements is also a function of external inputs and the present state. Thus a sequential circuit is specified by a time sequence of inputs, outputs, and internal states.

21.2 Synchronous and Asynchronous Operation

Sequential circuits are divided into two main types:

1. synchronous
2. asynchronous

Their classification depends on the timing of their signals.

Synchronous sequential circuits change their states and output values at discrete instants of time, which are specified by the rising and falling edge of a free-running clock signal. The clock signal is generally some form of square wave as shown in Figure below.

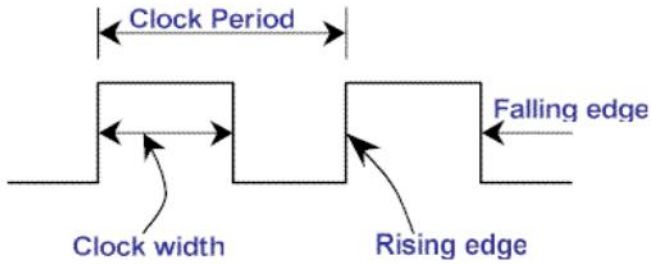


Figure 1 Clock Signal

From the diagram you can see that the clock period is the time between successive transitions in the same direction, that is, between two rising or two falling edges.

State transitions in synchronous sequential circuits are made to take place at times when the clock is making a transition from 0 to 1 (rising edge) or from 1 to 0 (falling edge). Between successive clock pulses there is no change in the information stored in memory.

The reciprocal of the clock period is referred to as the clock frequency. The clock width is defined as the time during which the value of the clock signal is equal to 1. The ratio of the clock width and clock period is referred to as the duty cycle. A clock signal is said to be active high if the state changes occur at the clock's rising

edge or during the clock width. Otherwise, the clock is said to be active low. Synchronous sequential circuits are also known as clocked sequential circuits.

The memory elements used in synchronous sequential circuits are usually flip-flops. These circuits are binary cells capable of storing one bit of information. A flip-flop circuit has two outputs, one for the normal value and one for the complement value of the bit stored in it. Binary information can enter a flip-flop in a variety of ways, a fact which give rise to the different types of flip-flops. For information on the different types of basic flip-flop circuits and their logical properties, see the previous tutorial on flip-flops.

In asynchronous sequential circuits, the transition from one state to another is initiated by the change in the primary inputs; there is no external synchronization. The memory commonly used in asynchronous sequential circuits is time-delayed devices, usually implemented by feedback among logic gates. Thus, asynchronous sequential circuits may be regarded as combinational circuits with feedback. Because of the feedback among logic gates, asynchronous sequential circuits may, at times, become unstable due to transient conditions. The instability problem imposes many difficulties on the designer. Hence, they are not as commonly used as synchronous systems.

21.3 Various types of Flip-flops and their Behavior

Since memory elements in sequential circuits are usually flip-flops, it is worth summarizing the behavior of various flip-flop types before proceeding further.

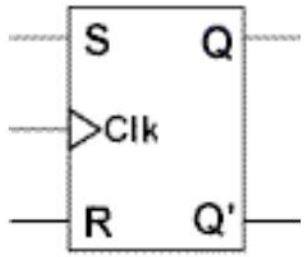
All Flip Flops (ffs) can be divided into four basic types: **SR**, **JK**, **D**, and **T**.

- The **SR** flip flop has two inputs, S (set) and R (reset), that set or reset the output Q when asserted.
- The **JK** flip flop has two inputs, J and K just like the S and R. However, when both J and K are asserted at the same time, the JK flip flop changes its state.
- The **D** flip flop has one input D (data) which sets the flip flop when D=1 and resets it when D=0.
- The **T** flip flop has one input T (toggle) which forces the flip-flop to change states when T=1.

They differ in the number of inputs and in the response invoked by different value of input signals. The four types of flip-flops are defined below.

1. SR Flip Flop:

Symbol

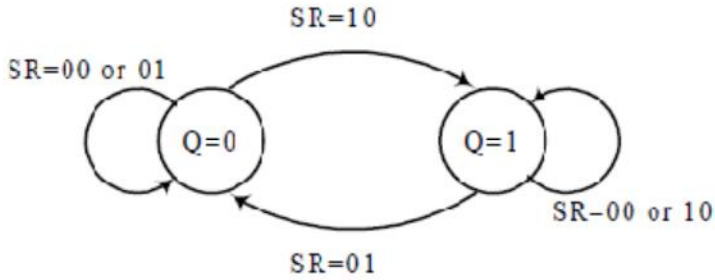


Note that The triangle indicates that the flip flop is triggered by the rising edge.

Characteristic Table

S	R	Q(next)
0	0	Q
0	1	0
1	0	1
1	1	NA

State Diagram/ Characteristic Equation



Q(next) = S + R'Q

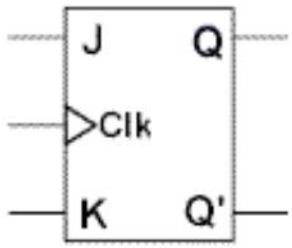
SR = 0

Excitation Table

Q	Q(next)	S	R
0	0	0	X
0	1	1	0
1	0	0	1
1	1	X	0

2. JK Flip Flop:

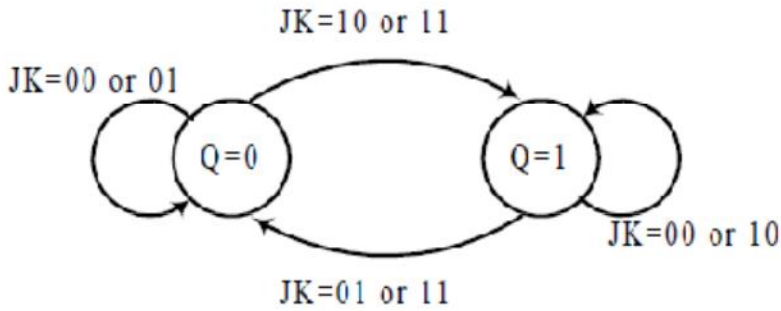
Symbol



Characteristic Table

J	K	Q(next)
0	0	Q
0	1	0
1	0	1
1	1	Q'

State Diagram/ Characteristic Equation



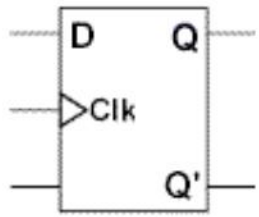
Q(next) = JQ' + K'Q

Excitation Table

Q	Q(next)	J	K
0	0	0	X
0	1	1	X
1	0	X	1
1	1	X	0

3. D Flip Flop:

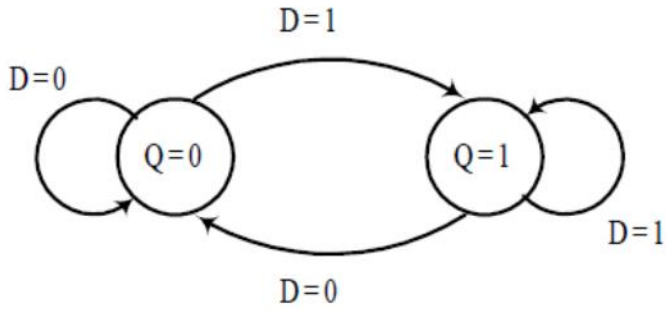
Symbol



Characteristic Table

D	Q(next)
0	0
1	1

State Diagram/ Characteristic Equation



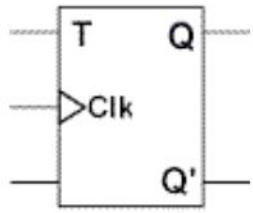
Q(next) = D

Excitation Table

Q	Q(next)	D
0	0	0
0	1	1
1	0	0
1	1	1

4. T Flip Flop:

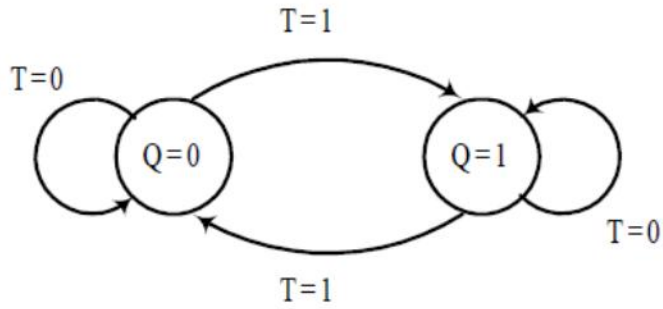
Symbol



Characteristic Table

T	Q(next)
0	Q
1	Q'

State Diagram/ Characteristic Equation



Q(next) = TQ' + T'Q

Excitation Table

Q	Q(next)	T
0	0	0
0	1	1
1	0	1
1	1	0

Each of these flip-flops can be uniquely described by its graphical symbol, its characteristic table, its characteristic equation or excitation table. All flip-flops have output signals Q and Q'.

The characteristic table defines the state of each flip-flop as a function of its inputs and previous state. Q refers to the present state and Q(next) refers to the next state after the occurrence of the clock pulse. The characteristic table for the RS flip-flop shows that the next state is equal to the present state when both inputs S and R are equal to 0. When R=1, the next clock pulse clears the flip-flop. When S=1, the flip-flop output Q is set to 1. The equation mark (?) for the next state when S and R are both equal to 1 designates an indeterminate next state.

The characteristic table for the JK flip-flop is the same as that of the RS when J and K are replaced by S and R respectively, except for the indeterminate case. When both J and K are equal to 1, the next state is equal to the complement of the present state, that is, Q(next) = Q'.

The next state of the D flip-flop is completely dependent on the input D and independent of the present state.

The next state for the T flip-flop is the same as the present state Q if T=0 and complemented if T=1.

The characteristic table is useful during the analysis of sequential circuits when the value of flip-flop inputs are known and we want to find the value of the flip-flop output Q after the rising edge of the clock signal. As with any other truth table, we can use the map method to derive the characteristic equation for each flip-flop, which are shown in the third column of Table 1.

During the design process we usually know the transition from present state to the next state and wish to find the flip-flop input conditions that will cause the required transition. For this reason we will need a table that lists the required inputs for a

given change of state. Such a list is called the *excitation table*. There are four possible transitions from present state to the next state. The required input conditions are derived from the information available in the characteristic table. The symbol NA in the table represents a "don't care" condition, that is, it does not matter whether the input is 1 or 0.

21.4 Self Learning Exercise -I

Q.1 Define sequential logic circuits.

Q.2 When a clock signal is said to be active high?

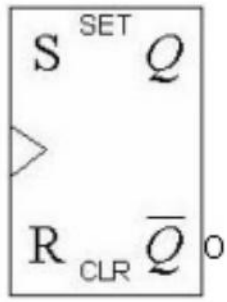
Q.3 What is Clock period?

Q.4 Write down a short note on –

1. SR Flip Flop
2. JK Flip Flop

21.5 Edge-Triggered Flip-flops

An edge-triggered flip-flop changes states either at the positive edge (rising edge) or at the negative edge (falling edge) of the clock pulse on the control input. The three basic types are introduced here: S-R, J-K and D.



- Positive edge-triggered (without bubble at Clock input): S-R, J-K, and D.
- Negative edge-triggered (with bubble at Clock input): S-R, J-K, and D.

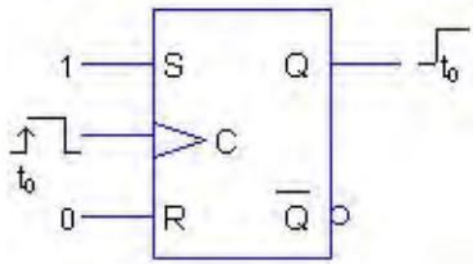
The S-R, J-K and D inputs are called synchronous inputs because data on these inputs are transferred to the flip-flop's output only on the triggering edge of the clock pulse. On the other hand, the direct set (SET) and clear (CLR) inputs are

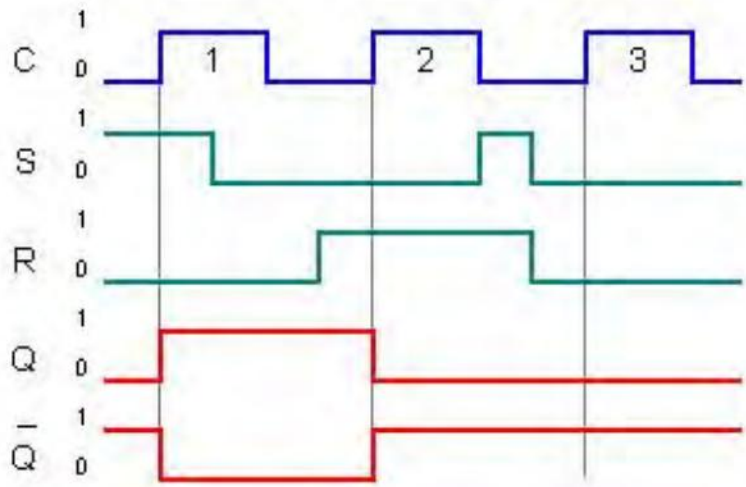
called asynchronous inputs, as they are inputs that affect the state of the flip-flop independent of the clock. For the synchronous operations to work properly, these asynchronous inputs must both be kept LOW.

1. Edge-triggered S-R flip-flop:

The basic operation is illustrated below, along with the truth table for this type of flip-flop. The operation and truth table for a negative edge-triggered flip-flop are the same as those for a positive except that the falling edge of the clock pulse is the triggering edge.

Inputs			Outputs		
S	R	C	Q	Q'	Comments
0	0	↑	Q	Q'	No change
0	1	↑	0	1	RESET
1	0	↑	1	0	SET
1	1	↑	?	?	Invalid





Note that the S and R inputs can be changed at any time when the clock input is LOW or HIGH (except for a very short interval around the triggering transition of the clock) without affecting the output. This is illustrated in the timing diagram below:

2. Edge-triggered J-K flip-flop:

The J-K flip-flop works very similar to S-R flip-flop. The only difference is that this flip-flop has NO invalid state. The outputs toggle (change to the opposite state) when both J and K inputs are HIGH. The truth table is shown below.

Inputs			Outputs		
J	K	C	Q	Q'	Comments
0	0	↑	Q	Q'	No change
0	1	↑	0	1	RESET
1	0	↑	1	0	SET
1	1	↑	Q'	Q	Toggle

3. Edge-triggered D flip-flop:

The operation of a D flip-flop is much simpler. It has only one input addition to the clock. It is very useful when a single data bit (0 or 1) is to be stored. If there is a HIGH on the D input when a clock pulse is applied, the flip-flop SETs and stores a 1. If there is a LOW on the D input when a clock pulse is applied, the flip-flop RESETs and stores a 0. The truth table below summarizes the operations of the positive edge-triggered D flip-flop. As before, the negative edge-triggered flip-flop works the same except that the falling edge of the clock pulse is the triggering edge.

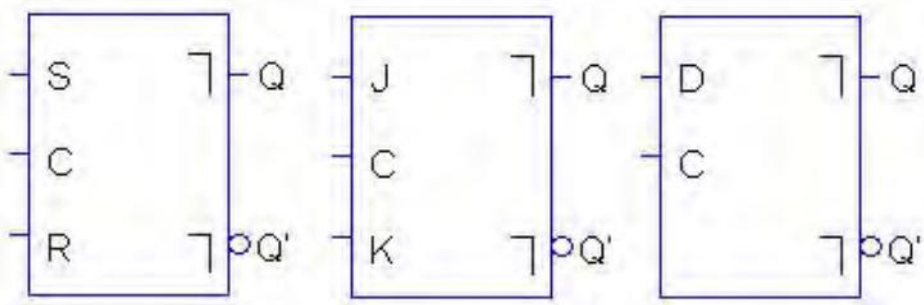
Inputs		Outputs		
D	C	Q	Q'	Comments
0	↑	0	1	RESET
1	↑	1	0	SET

21.6 Pulse-Triggered (Master-Slave) Flip-flops

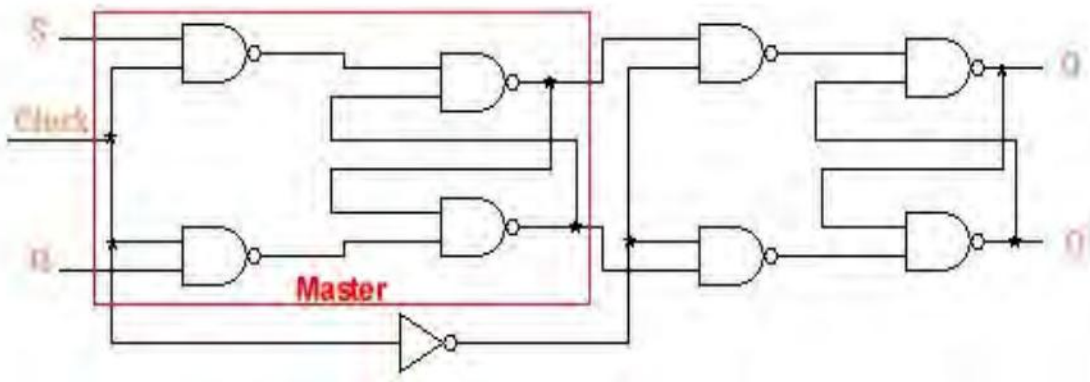
The term pulse-triggered means that data are entered into the flip-flop on the rising edge of the clock pulse, but the output does not reflect the input state until the

falling edge of the clock pulse. As this kind of flip-flops are sensitive to any change of the input levels during the clock pulse is still HIGH, the inputs must be set up prior to the clock pulse's rising edge and must not be changed before the falling edge. Otherwise, ambiguous results will happen.

The three basic types of pulse-triggered flip-flops are S-R, J-K and D. Their logic symbols are shown below. Notice that they do not have the dynamic input indicator at the clock input but have postponed output symbols at the outputs.



The truth tables for the above pulse-triggered flip-flops are all the same as that for the edge-triggered flip-flops, except for the way they are clocked. These flip-flops are also called Master-Slave flip-flops simply because their internal constructions are divided into two sections. The slave section is basically the same as the master section except that it is clocked on the inverted clock pulse and is controlled by the outputs of the master section rather than by the external inputs. The logic diagram for a basic master-slave S-R flip-flop is shown below.



21.7 Self Learning Exercise –II

Q.1 Define pulse-triggered Flip Flop.

Q.2 Draw the input and output waveforms of a negative edge triggered SR Flip Flop.

Q.3 Explain the working of a positive edge triggered D type Flip Flop.

21.8 Summary

In this chapter we presented memory devices called Flip-Flops. We consider using Flip-Flops in the presence of a clock signal. The clock signal causes the Flip-Flops to sample the value of the input towards the end of a clock cycle and output the sampled value during the next clock cycle. Flip-Flops play a crucial role in bounding the segments of time during which signals may be instable.

In a sense, Flip-Flops and combinational circuits have opposite roles. Combinational circuits compute interesting Boolean functions but increase uncertainty (namely, lengthen segments of time during which signals may be instable). Flip-Flops, one the other hand, output the same value that is fed as input but they limit uncertainty.

21.9 Glossary

Binary: The basic counting system used in computer logic. Two values are available - 0 and 1. A zero is normally represented by a 0 Volt signal, and a one by a voltage of approximately 5 Volts - these levels are dependent upon the type of logic used

Binary Code: A coding scheme that communicates information by using a series of "1s" and "0s" that are represented, respectively, by the digital "ON" and "OFF" states

Scan Design :A design technique in which the internal registers or flip-flops of a circuit can be chained, to allow an external circuit to easily read and write their contents. When internal memory elements are not directly accessible from the circuit's outside pins, testing is difficult because their state is unknown. With scan design, a signal reconfigures the elements into a "scan chain" and their contents can be read and if desired, altered.

Shift Register: Two or more bistable elements (flip-flops) connected in series. With each tick of the clock, the output of stage n is shifted to stage n+1. Applications include clock or signal delays, delay lines, linear-feedback shift registers

Clear: An input to a latch or FLIP FLOP used to make $Q = 0$.

Clear State : The $Q = 0$ state of a flip-flop.

Clock: Digital signal in the form of a rectangular pulse train or a square wave.

Clock Skew: Arrival of a clock signal at the clock inputs of different flip-flops at different times as a result of propagation delays.

Clock Transition Times: Minimum rise and fall times for the clock signal transitions used by a particular IC, specified by the IC manufacturer.

Clocked D Flip-Flop: Type of flip-flop in which the *D* (data) input is the synchronous input.

Clocked Flip-Flops: Flip-flops that have a clock input.

Clocked J-K FLip-Flop: Type of flip-flop in which inputs *J* and *K* are the synchronous inputs.

Trigger: Input signal to a flip-flop or one-shot that causes the output to change states depending on the conditions of the control signals.

Truth Table: Logic table that depicts a circuit's output response to the various combinations of the logic levels at its inputs.

21.10 Answer To Self Learning Exercise

Answer To Self Learning Exercise -I

Ans.1: Circuits whose outputs depends not only on the present input value but also the past input value are known as sequential logic circuits.

Ans.2: A clock signal is said to be active high if the state changes occur at the clock's rising edge or during the clock width. Otherwise, the clock is said to be active low.

Ans.3: Clock period is the time between successive transitions in the same direction, i.e., between two rising or two falling edges.

Ans.4: Sections 21.3

Answer to Self Learning Exercise -II

Ans.1: The term pulse-triggered means that data are entered into the flip-flop on the rising edge of the clock pulse, but the output does not reflect the input state until the falling edge of the clock pulse.

Ans.2: Section 21.5

21.11 Exercise

Q.1 Describe basically how data are clocked into a pulse triggered, master slave flip flop.

Q.2 How does the operation of an asynchronous input differ from that of a synchronous input.

Q.3 Draw the truth table of the SR Flip Flops.

Reference and Suggested Readings

1. Linear Integrated circuits by D.Roy Choudhury, 2015.
2. Power Electronics by P.C.Sen, 1987.
3. Integrated circuits by K. R. Botkar, 1993.
4. Digital Principles and Applications by D. P. Leach and A. P. Malvino, 1975.
5. Modern Digital Electronics By- R.P.Jain, 1984.
6. Digital Electronics By R. J.Tocci, 1980

UNIT-22
Registers & Counters

Structure of the Unit

22.0 Objectives

22.1 Introduction

22.2 Registers

22.3 Shift registers

22.4 Parallel access shift registers

22.5 Self Learning Exercise-I

22.6 Counter

22.7 Ripple counter

22.8 Synchronous counter

22.9 Ring counter

22.10 Self Learning Exercise-II

22.11 Summary

22.12 Glossary

22.13 Answers to Self Learning Exercises

22.14 Exercise

References and Suggested Readings

22.0 Objectives

After interacting with the material presented here students will be able to

understand

1. Concept of registers and counters
2. Working of Shift and Parallel access shift registers and
3. Detailed description of synchronous, asynchronous counter and Ring counter

22.1 Introduction

The ideas in combinational circuits and sequential methods when brought together as one system gives sequential building blocks, usually in the form or registers and counters.

A register is a set of flip-flops with combinational logic to implement state transitions that allow information to be stored and retrieved from them. In the simplest form, a flip-flop is a one-bit register.

A counter is simply a register with combinational logic to implement counting, that is it is possible to retrieve the contents, add or subtract one to the contents, and then store it back into the register in one operation.

22.2 Registers

In the last chapter we introduced Flip Flops that can be used as storage elements. A flip-flop stores one bit of information. When a set of n flip-flops is used to store n bits of information, such as an n-bit number, we refer to these flip-flops as a register. A common clock is used for each flip-flop in a register, and each flip-flop operates as described in the previous chapter. The term register is merely a convenience for referring to n-bit structures consisting of flip-flops.

22.3 Shift Register

In the previous chapter we explained that a given number is multiplied by 2 if its bits are shifted one bit position to the left and a 0 is inserted as the new least-significant bit. Similarly, the number is divided by 2 if the bits are shifted one bit-position to the right. A register that provides the ability to shift its contents is called a *shift register*.

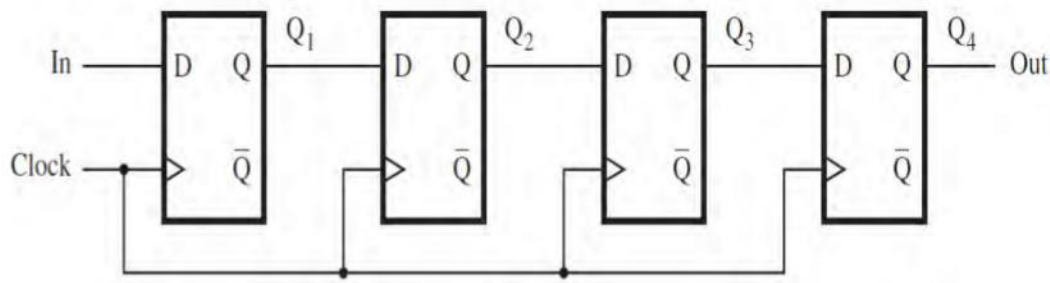


Figure shows a four-bit shift register that is used to shift its contents one bit position to the right. The data bits are loaded into the shift register in a serial fashion using the *In* input. The contents of each flip-flop are transferred to the next flip-flop at each positive edge of the clock. An illustration of the transfer is given in table, which shows what happens when the signal values at *In* during eight consecutive clock cycles are 1, 0, 1, 1, 1, 0, 0, and 0, assuming that the initial state of all flip-flops is 0.

	In	Q_1	Q_2	Q_3	Q_4=out
t_0	1	0	0	0	0
t_1	0	1	0	0	0
t_2	1	0	1	0	0
t_3	1	1	0	1	0
t_4	1	1	1	0	1
t_5	0	1	1	1	0
t_6	0	0	1	1	1
t_7	0	0	0	1	1

To implement a shift register, it is necessary to use either edge-triggered or master-slave flip-flops. The level-sensitive gated latches are not suitable, because a change in the value of *In* would propagate through more than one latch during the time when the clock is equal to 1.

22.4 Parallel-Access Shift Register

In computer systems it is often necessary to transfer *n*-bit data items. This may be done by transmitting all bits at once using *n* separate wires, in which case we say that the transfer is performed in *parallel*. But it is also possible to transfer all bits using a single wire, by performing the transfer one bit at a time, in *n* consecutive clock cycles. We refer to this scheme as *serial* transfer. To transfer an *n*-bit data item serially, we can use a shift register that can be loaded with all *n* bits in parallel (in one clock cycle). Then during the next *n* clock cycles, the contents of the register can be shifted out for serial transfer. The reverse operation is also needed. If bits are received serially, then after *n* clock cycles the contents of the register can be accessed in parallel as an *n*-bit item.

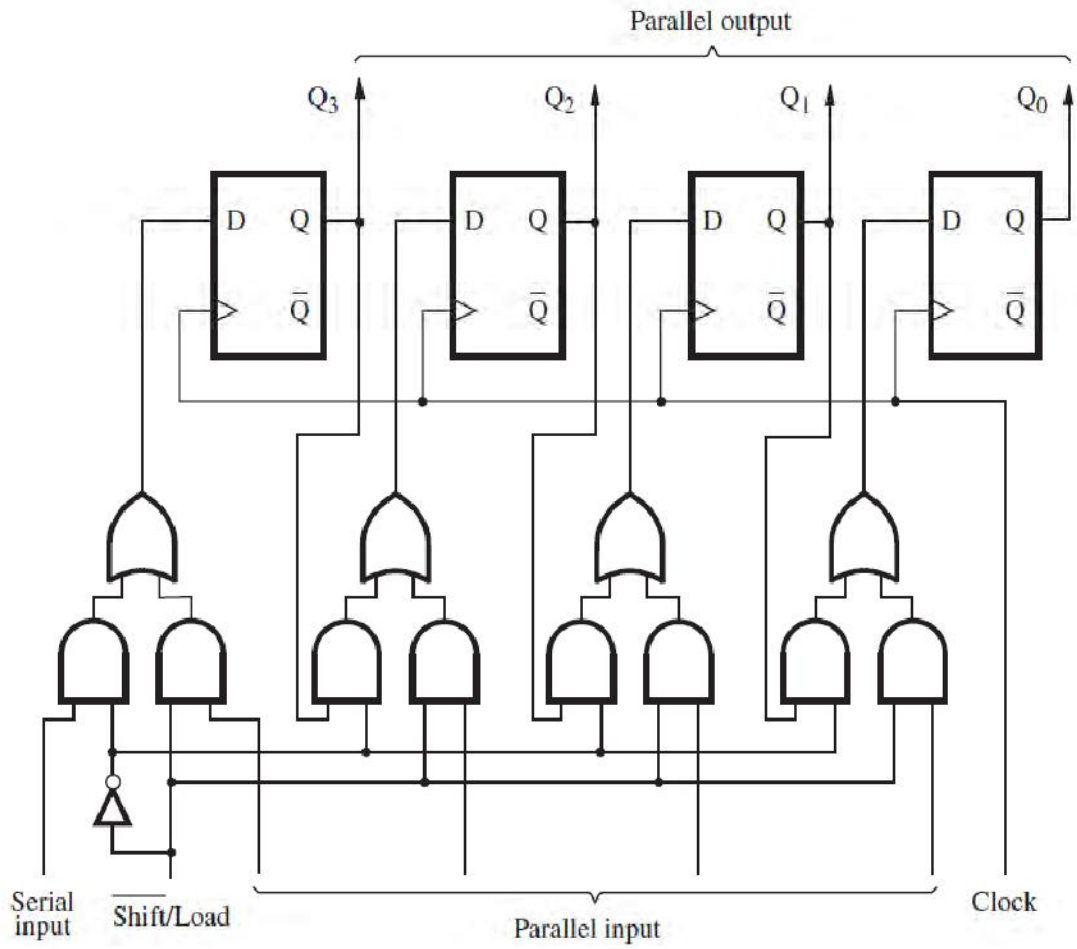


Figure shows a four-bit shift register that allows the parallel access. Instead of using the normal shift register connection, the *D* input of each flip-flop is

connected to two different sources. One source is the preceding flip-flop, which is needed for the shift register operation. The other source is the external input that corresponds to the bit that is to be loaded into the flip-flop as a part of the parallel-load operation. The control signal *Shift*/*Load* is used to select the mode of operation.

- If *Shift*/*Load* = 0, then the circuit operates as a shift register.
- If *Shift*/*Load* = 1, then the parallel input data are loaded into the register.

In both cases the action takes place on the positive edge of the clock. In Figure we have chosen to label the flip-flops outputs as Q3, . . . ,Q0 because shift registers are often used to hold binary numbers. The contents of the register can be accessed in parallel by observing the outputs of all flip-flops. The flip-flops can also be accessed serially, by observing the values of Q0 during consecutive clock cycles while the contents are being shifted. A circuit in which data can be loaded in series and then accessed in parallel is called a series-to-parallel converter. Similarly, the opposite type of circuit is a parallel-to-series converter. The circuit in Figure can perform both of these functions.

22.5 Self Learning Exercise -I

Q.1 What is register?

Q.2 Draw a simple shift register.

Q.3 Write down the condition to select the mode of operation in registers.

Q.4 Brief note on

1. Shift registers.
2. Parallel-Access Shift Register

22.6 Counters

Now we examine special types of addition and subtraction operations, which are used for the purpose of counting. In particular, we want to design circuits that can increment or decrement a count by 1. Counter circuits are used in digital systems

for many purposes. They may count the number of occurrences of certain events, generate timing intervals for control of various tasks in a system, keep track of time elapsed between specific events, and so on. Counters can be implemented using the adder/subtractor circuits and the registers. However, since we only need to change the contents of a counter by 1, it is not necessary to use such elaborate circuits. Instead, we can use much simpler circuits that have a significantly lower cost. We will show how the counter circuits can be designed using T and D flip-flops.

22.7 Asynchronous Counters (Ripple Counter)

The simplest counter circuits can be built using T flip-flops because the toggle feature is naturally suited for the implementation of the counting operation.

Up-Counter with T Flip-Flops:

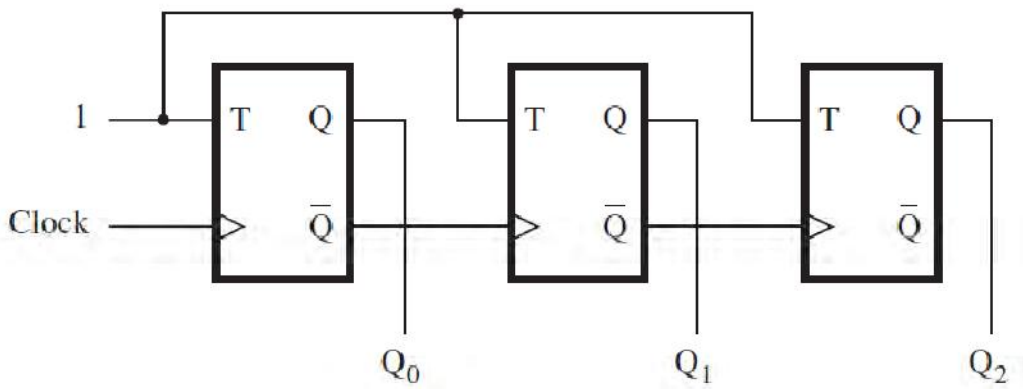


Figure gives a three-bit counter capable of counting from 0 to 7. The clock inputs of the three flip-flops are connected in cascade. The *T* input of each flip-flop is connected to a constant 1, which means that the state of the flip-flop will be reversed (toggled) at each positive edge of its clock. We are assuming that the purpose of this circuit is to count the number of pulses that occur on the primary input called *Clock*. Thus the clock input of the first flip-flop is connected to the *Clock* line. The other two flip-flops have their clock inputs driven by the $\overline{Q}$ output of the preceding flip-flop. Therefore, they toggle their state whenever the preceding flip-flop changes its state from Q = 1 to Q = 0, which results in a positive edge of the $\overline{Q}$ signal.

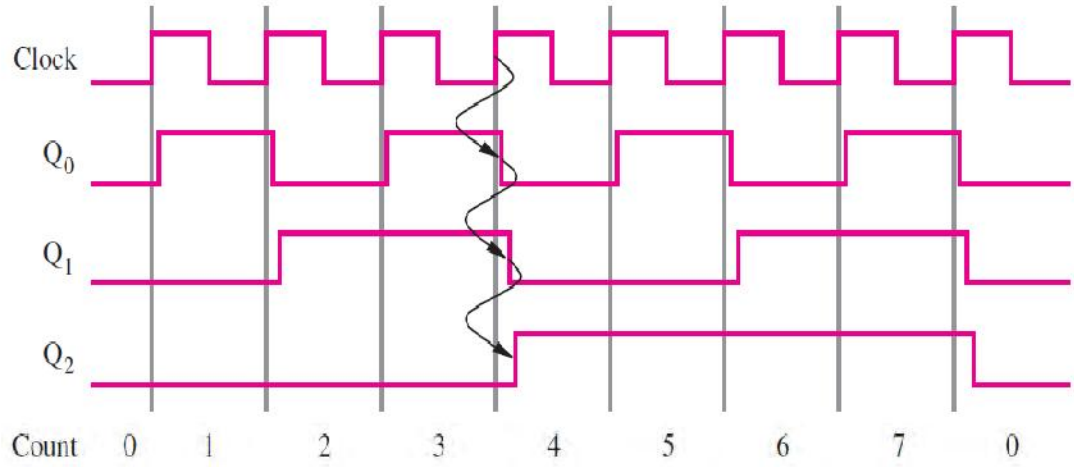
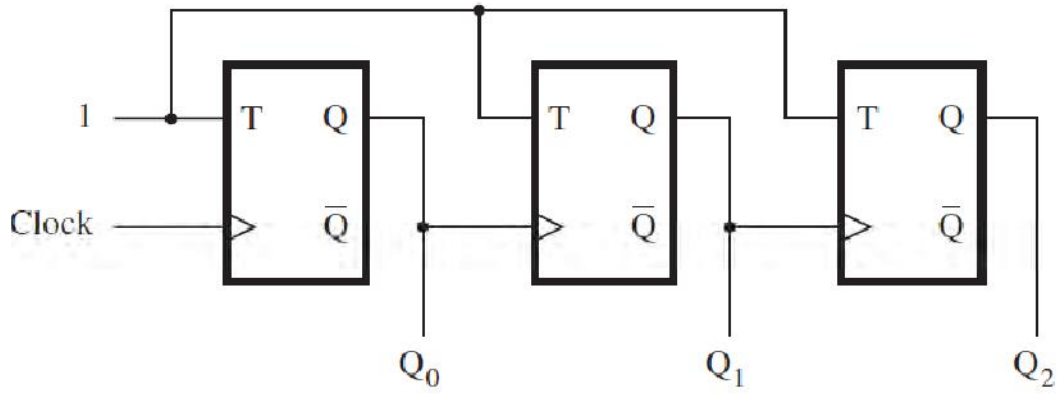


Figure shows a timing diagram for the counter. The value of Q_0 toggles once each clock cycle. The change takes place shortly after the positive edge of the *Clock* signal. The delay is caused by the propagation delay through the flip-flop. Since the second flip-flop is clocked by Q_0, the value of Q_1 changes shortly after the negative edge of the $\overline{Q}_0$ signal.

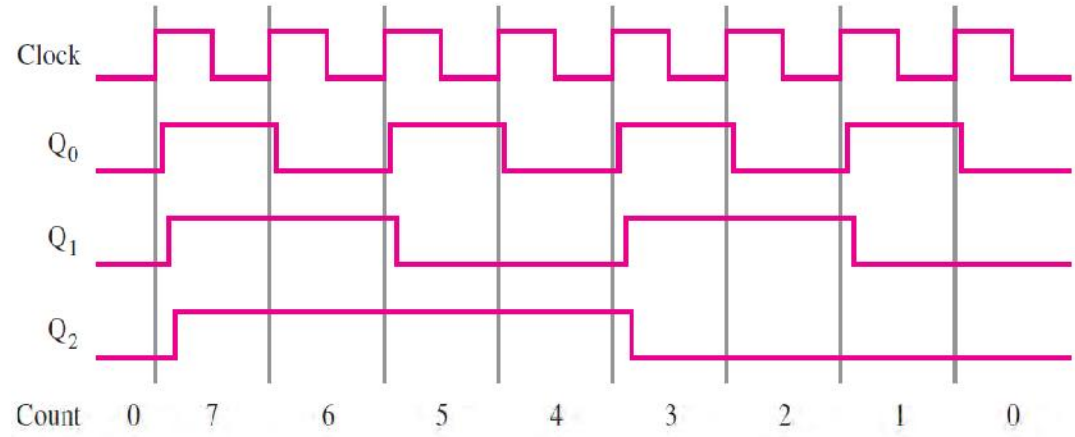
Similarly, the value of Q_2 changes shortly after the negative edge of the Q_1 signal. If we look at the values $Q_2Q_1Q_0$ as the count, then the timing diagram indicates that the counting sequence is 0, 1, 2, 3, 4, 5, 6, 7, 0, 1, and so on. This circuit is a modulo-8 counter. Because it counts in the upward direction, we call it an *up-counter*.

The counter has three *stages*, each comprising a single flip-flop. Only the first stage responds directly to the *Clock* signal; we say that this stage is *synchronized* to the clock. The other two stages respond after an additional delay. For example, when *Count* = 3, the next clock pulse will cause the *Count* to go to 4. As indicated by the arrows in the timing diagram, this change requires the toggling of the states of all three flip-flops. The change in Q_0 is observed only after a propagation delay from the positive edge of *Clock*. The Q_1 and Q_2 flip-flops have not yet changed; hence for a brief time the count is $Q_2\,Q_1\,Q_0 = 010$. The change in Q_1 appears after a second propagation delay, at which point the count is 000. Finally, the change in Q_2 occurs after a third delay, at which point the stable state of the circuit is reached and the count is 100.

Down-Counter with T Flip-Flops:



A slight modification of the up counter circuit is presented in Figure. The only difference is that in Figure the clock inputs of the second and third flip-flops are driven by the Q outputs of the preceding stages, rather than by the $\bar{Q}$ outputs. The timing diagram, given in below Figure, shows that this circuit counts in the sequence 0, 7, 6, 5, 4, 3, 2, 1, 0, 7, and so on. Because it counts in the downward direction, we say that it is a *down-counter*.



It is possible to combine the functionality of the up and down counter circuits to form a counter that can count either up or down. Such a counter is called an *up/downcounter*.

22.8 Synchronous Counters

The asynchronous counters are simple, but not very fast. If a counter with a larger number of bits is constructed in this manner, then the delays caused by the

cascaded clocking scheme may become too long to meet the desired performance requirements. We can build a faster counter by clocking all flip-flops at the same time, using the approach described below.

Synchronous Counter with T Flip-Flops:

Clock cycle	Q_2	Q_1	Q_0
0	0	0	0
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1
8	0	0	0

Table shows the contents of a three-bit up-counter for eight consecutive clock cycles, assuming that the count is initially 0. Observing the pattern of bits in each row of the table, it is apparent that bit Q_0 changes on each clock cycle. Bit Q_1 changes only when $Q_0 = 1$. Bit Q_2 changes only when both Q_1 and Q_0 are equal to 1. In general, for an n-bit up-counter, a given flip-flop changes its state only when all the preceding flip-flops are in the state Q = 1. Therefore, if we use T flip-flops to realize the counter, then the T inputs are defined as

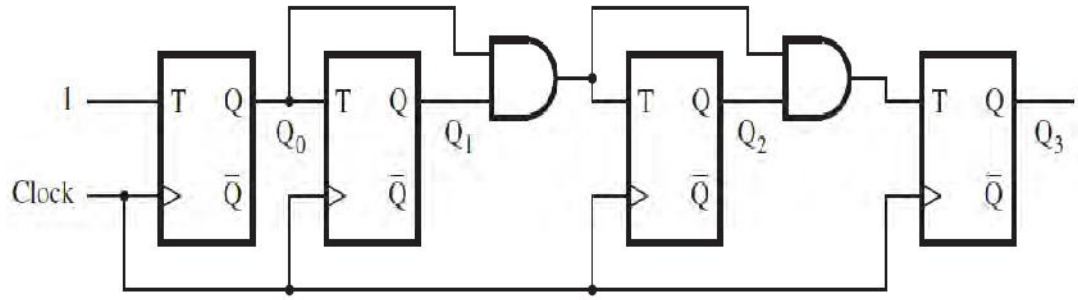
$$T_0 = 1$$

$$T_1 = Q_0$$

$$T_2 = Q_0 Q_1$$

$$T_3 = Q_0 Q_1 Q_2$$
$$\vdots$$
$$T_n = Q_0 Q_1 \ldots Q_{n-1}$$

An example of a four-bit counter based on these expressions is given in Figure.



Instead of using AND gates of increased size for each stage, which may lead to fan-in problems, we use a factored arrangement, as shown in the figure. This arrangement does not slow down the response of the counter, because all flip-flops change their states after a propagation delay from the positive edge of the clock. Note that a change in the value of Q_0 may have to propagate through several AND gates to reach the flip-flops in the higher stages of the counter, which requires a certain amount of time. This time must not exceed the clock period. Actually, it must be less than the clock period minus the setup time for the flip-flops.

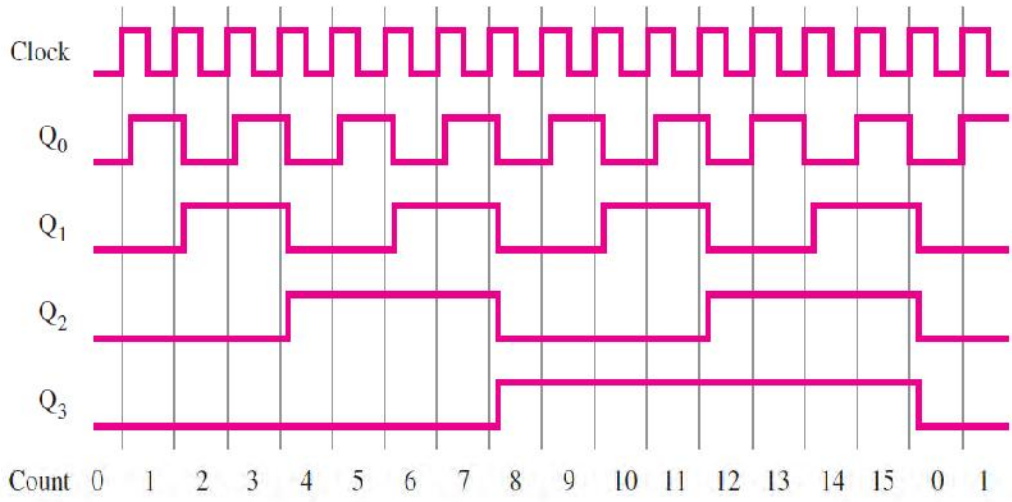


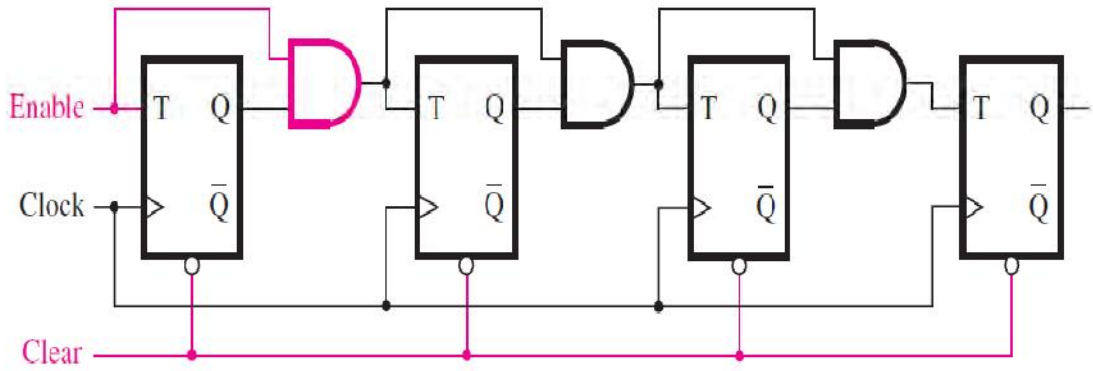
Figure gives a timing diagram. It shows that the circuit behaves as a modulo-16 up-

counter. Because all changes take place with the same delay after the active edge of the Clock signal, the circuit is called a synchronous counter.

Enable and Clear Capability

Asynchronous counter through synchronous counter change their contents in response to each clock pulse. Often it is desirable to be able to inhibit counting, so that the count remains in its present state. This may be accomplished by including an Enable control signal, as indicated in Figure. The circuit is the *synchronous counter*, where the Enable signal controls directly the T input of the first flip-flop. Connecting the Enable also to the AND gate chain means that

- if Enable = 0, then all T inputs will be equal to 0.
- If Enable = 1, then the counter operates as explained previously.



In many applications it is necessary to start with the count equal to zero. The clear inputs on all flip-flops can be tied together and driven by a Clear control input.

Synchronous Counter with D Flip-Flops:

While the toggle feature makes T flip-flops a natural choice for the implementation of counters, it is also possible to build counters using other types of flip-flops. The JK flip-flops can be used in exactly the same way as the T flip-flops because if the J and K inputs are tied together, a JK flip-flop becomes a T flip-flop. We will now consider using D flip-flops for this purpose.

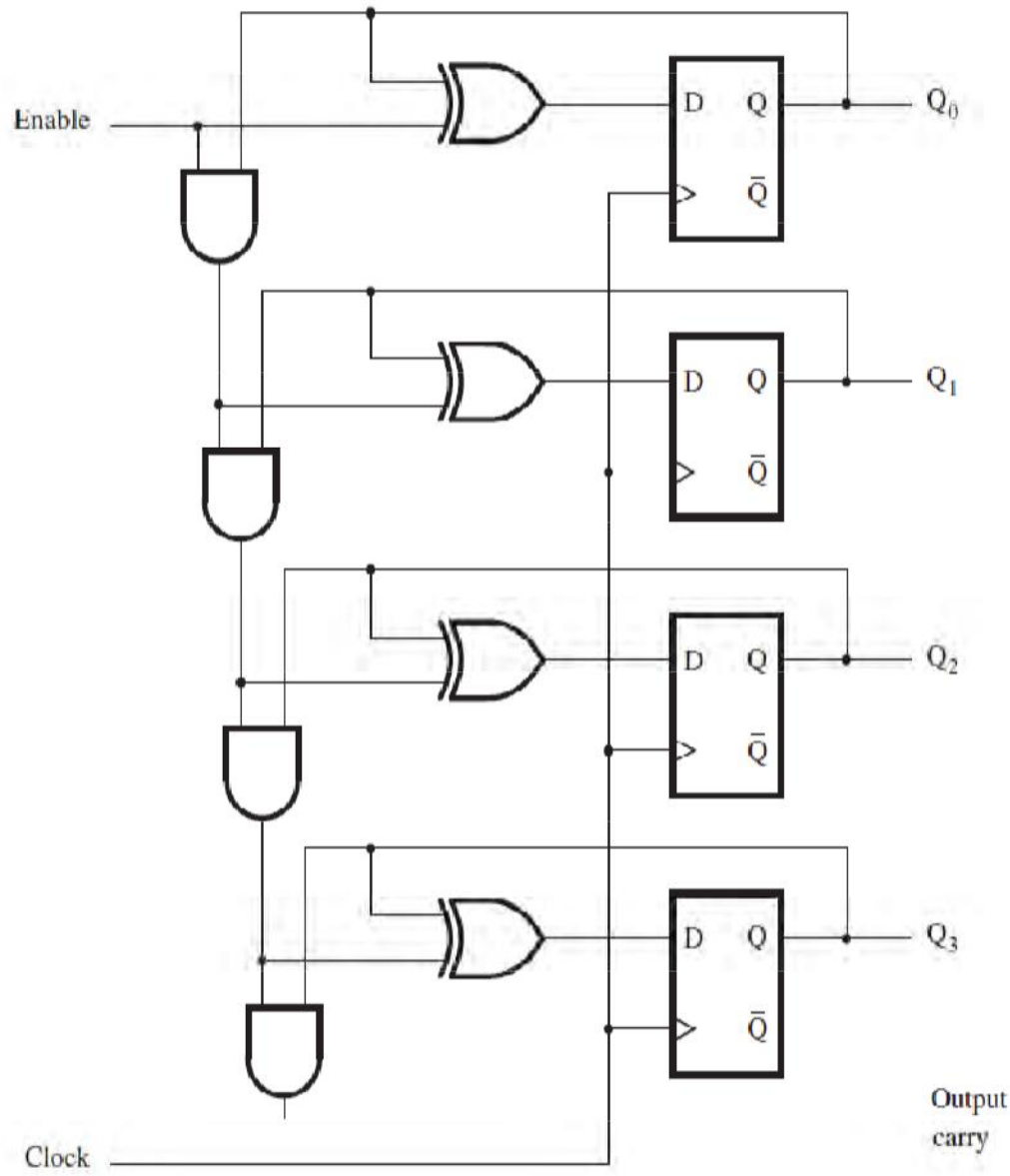


Figure gives a four-bit up-counter that counts in the sequence 0, 1, 2, . . . , 14, 15, 0, 1, and so on. The count is indicated by the flip-flop outputs $Q_3Q_2Q_1Q_0$. If we assume that Enable = 1, then the D inputs of the flip-flops are defined by the expressions

$$D_0 = \overline{Q}_0 = 1 \oplus Q_0$$

$$D_1 = Q_1 \oplus Q_0$$

$$D_2 = Q_2 \oplus Q_1Q_0$$

$$D_3 = Q_3 \oplus Q_2 Q_1 Q_0$$
$$\vdots$$
$$D_i = Q_i \oplus Q_{i-1....} Q_1 Q_0$$

We have included the Enable control signal so that the counter counts the clock pulses only if Enable = 1. In effect, the above equations are modified to implement the circuit in the figure as follows

$$D_0 = Q_0 \oplus Enable$$

$$D_1 = Q_1 \oplus Q_0 . Enable$$

$$D_2 = Q_2 \oplus Q_1 . Q_0 . Enable$$

$$D_3 = Q_3 \oplus Q_{2.} Q_1 . Q_0 . Enable$$

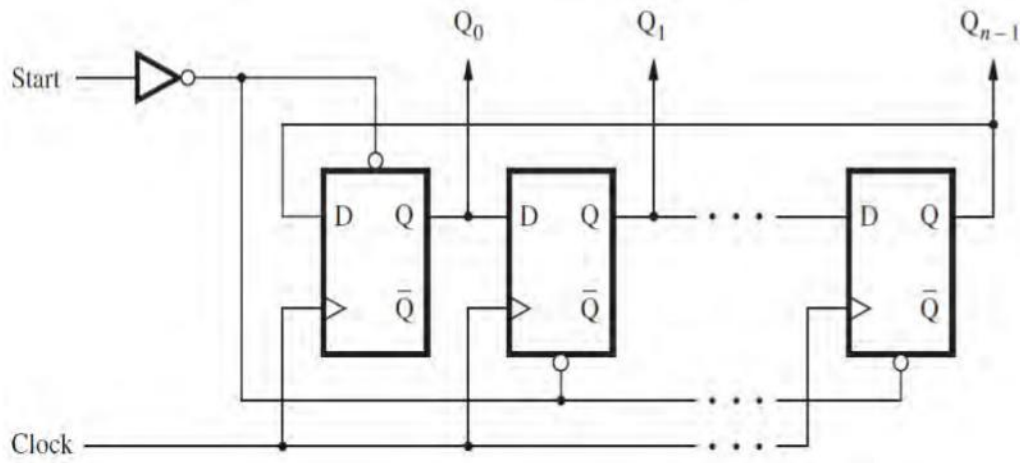
The operation of the counter is based on our observation for Table that the state of the flip-flop in stage i changes only if all preceding flip-flops are in the state Q = 1. This makes the output of the AND gate that feeds stage i equal to 1, which causes the output of the XOR gate connected to D_i to be equal to $\overline{Q}_i$. Otherwise, the output of the XOR gate provides $D_i = Q_i$, and the flip-flop remains in the same state. This resembles the carry propagation in a carry-lookahead adder circuit; hence the AND-gate chain can be thought of as the carry chain. Even though the circuit is only a four-bit counter, we have included an extra AND that produces the "output carry." This signal makes it easy to concatenate two such four-bit counters to create an eight-bit counter.

22.9 Ring Counter

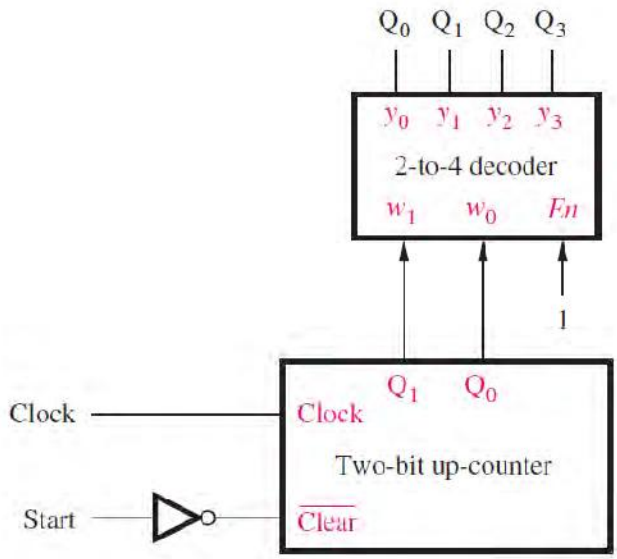
In the preceding counters the count is indicated by the state of the flip-flops in the counter. In all cases the count is a binary number. Using such counters, if an action is to be taken as a result of a particular count, then it is necessary to detect the occurrence of this count. This may be done using AND gates.

It is possible to devise a counterlike circuit in which each flip-flop reaches the state $Q_i = 1$ for exactly one count, while for all other counts $Q_i = 0$. Then Q_i indicates directly an occurrence of the corresponding count. Actually, since this does not

represent binary numbers, it is better to say that the outputs of the flips-flops represent a code. Such a circuit can be constructed from a simple shift register, as indicated in Figure.



The Q output of the last stage in the shift register is fed back as the input to the first stage, which creates a ringlike structure. If a single 1 is injected into the ring, this 1 will be shifted through the ring at successive clock cycles. For example, in a four-bit structure, the possible codes $Q_0Q_1Q_2Q_3$ will be 1000, 0100, 0010, and 0001. As we said in section 6.2, such encoding, where there is a single 1 and the rest of the code variables are 0, is called a one-hot code.



The circuit in Figure is referred to as a ring counter. Its operation has to be initialized by injecting a 1 into the first stage. This is achieved by using the Start control signal, which presets the left-most flip-flop to 1 and clears the others to 0. We assume that all changes in the value of the Start signal occur shortly after an active clock edge so that the flip-flop timing parameters are not violated.

The circuit in Figure can be used to build a ring counter with any number of bits, n. For the specific case of n = 4, part (b) of the figure shows how a ring counter can be constructed using a two-bit up-counter and a decoder. When Start is set to 1, the counter is reset to 00. After Start changes back to 0, the counter increments its value in the normal way. The 2-to-4 decoder, changes the counter output into a one-hot code. For the count values 00, 01, 10, 11, 00, and so on, the decoder produces Q0Q1Q2Q3 = 1000, 0100, 0010, 0001, 1000, and so on. This circuit structure can be used for larger ring counters, as long as the number of bits is a power of two. We will give an example of a larger circuit that uses the ring counter in Figure.

22.10 Self Learning Exercise- II

Q.1 What are the uses of Counter circuits.

Q.2 What is the difference between up and down counter?

Q.3 Draw Synchronous Counter with D Flip-Flops.

Q.4 Explain working of Ripple Counter.

22.11 Summary

The sequential circuits presented in this chapter, namely, registers and counters, have a regular structure that allows the circuits to be designed using an intuitive approach.

22.12 Glossary

Clear: Reset (synchronous or asynchronous).

Control section: The combinational logic portion of a synchronous circuit that

determines the next state of the circuit.

Count sequence: The specific series of output states through which a counter progresses.

Counter: A sequential digital circuit whose output progresses in a predictable repeating pattern, advancing by one state for each clock pulse.

DOWN counter: A counter with a descending sequence.

Left shift : A movement of data from the right to the left in a shift register. This is usually defined as moving from the LSB toward the MSB.

Memory section: A set of flip-flops in a synchronous circuit that holds its present state.

Parallel-load shift register: A shift register that can be preset to any value by directly loading a binary number into its internal flip-flops.

Present state: The current state of flip-flop outputs in a synchronous sequential circuit.

Right shift: A movement of data from the left to the right in a shift register. This is usually defined as data moving from the MSB toward the LSB.

Ring counter :A serial shift register with feedback from the output of the last flip-flop to the input of the first.

Serial shifting: Movement of data from one end of a shift register to the other at a rate of one bit per clock pulse.

Shift register: A synchronous sequential circuit that will store and move n-bit data, either serially or in parallel, in n flip-flops.

Synchronous counter: A counter whose flip-flops are all clocked by the same source and thus change in synchronization with each other.

22.13 Answer To Self Learning Exercises

Answer To Self Learning Exercise- I

Ans.1: A set of n flip-flops is used to store n bits of information, such as an n-bit number, we refer to these flip-flops as a register.

Ans.2: section 22.2

Ans.3: The control signal Shift/Load is used to select the mode of operation.

- If Shift/Load = 0, then the circuit operates as a shift register.
- If Shift/Load = 1, then the parallel input data are loaded into the register.

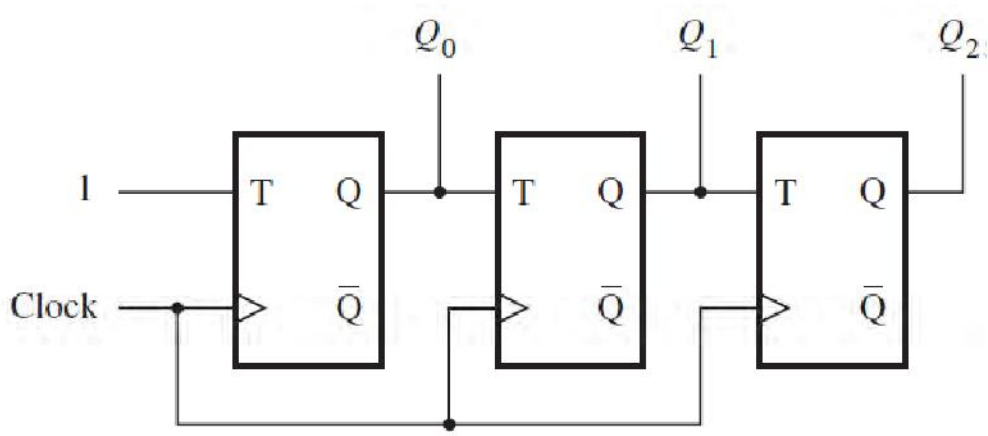
Answer To Self Learning Exercise- II

Ans.1: Counter circuits are used in digital systems for many purposes. They may count the number of occurrences of certain events, generate timing intervals for control of various tasks in a system, keep track of time elapsed between specific events, and so on.

Ans.2: The only difference between up and down counter is that in Figure the clock inputs of the second and third flip-flops are driven by the Q outputs of the preceding stages, rather than by the $\overline{Q}$ outputs.

22.14 Exercise

Q.1 The circuit in Figure looks like a counter. What is the sequence that this circuit counts in?



Q.2 A universal shift register can shift in both the left-to-right and right-to-left directions, and it has parallel-load capability. Draw a circuit for such a shift register.

Q.3 Short note on

1. Shift Register
2. Ring counter

Reference and Suggested Readings

1. Power Electronics by P.C.Sen, 1987.
2. Integrated circuits by K. R. Botkar, 1993.
3. Digital Principles and Applications by D. P. Leach and A. P. Malvino, 1975.
4. Digital Electronics By R. J.Tocci, 1980
5. Principles of Digital Design, by D. D. Gajski 1997.
6. Digital Logic Circuit Analysis and Design, by V. P. Nelson, H. T. Nagle, B. D. Carroll, and J. D. Irwin, 1995.
7. Contemporary Logic Design, by R. H. Katz,1994.
8. Introduction to Logic Design, by J. P. Hayes 1993.
9. Digital Design Principles and Practices, by J. F.Wakerly, 1990.

UNIT-23
Timing Circuits : Schmitt Trigger, Multivibrators

Structure of the Unit

23.0 Objectives

23.1 Introduction

23.2 Monostable multivibrator

23.3 Bistable multivibrator

23.4 Self Learning Exercise-I

23.5 Astable multivbrator

22.6 Waveform Generators

23.7 Self Learning Exercise-II

23.8 Summary

23.9 Glossary

23.10 Answers to Self Learning Exercises

23.11 Exercise

23.12 Answers to Exercise

References and Suggested Readings

23.0 Objectives

After interacting with the material presented here students will be able to describe the characteristic of the 3 basic types of multivibrator:

1. Astable (free running)
2. Monostable (one shot)

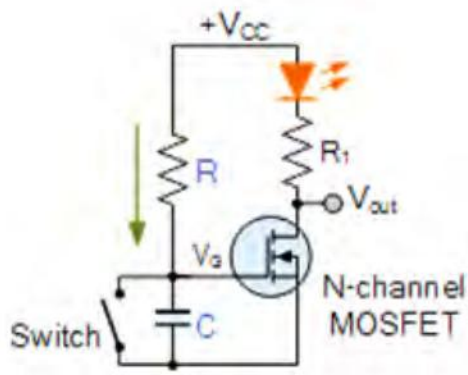
3. Bistable (flip flop)

23.1 Introduction

Multivibrators are electronic circuits where output takes some known states. These states may be *stable* or *unstable*. Depending on the status of the output state being *stable* or *unstable*, the circuit gets different names.

- When the output has only two stable states, it is called bistable multivibrator. Memory elements or flip-flops are typical bistable multivibrators. A bistable multivibrator with a hysteresis is called as Schmitt Trigger.
- If the output has two unstable states and output is switching from one state to the other continuously, it is an astable multivibrator.
- If the output has a single stable and the other state is unstable, then this circuit is called as monostable multivibrator and it is used to generate a single pulse of known period.

23.2 Monostable Multivibrator



Monostable Multivibrators have only ONE stable state (hence their name: "Mono"), and produce a single output pulse when it is triggered externally. Monostable Multivibrators only return back to their first original and stable state after a period of time determined by the time constant of the RC coupled circuit.

Consider the MOSFET circuit shown in figure. The resistor R and capacitor C form an RC timing circuit. The N-channel enhancement mode

MOSFET is switched "ON" due to the voltage across the capacitor with the drain connected LED also "ON".

When the switch is closed the capacitor is short circuited and therefore discharges while at the same time the gate of the MOSFET is shorted to ground. The MOSFET and therefore the LED are both switched "OFF". While the switch is closed the circuit will always be "OFF" and in its "unstable state".

When the switch is opened, the fully discharged capacitor starts to charge up through the resistor, Rat a rate determined by the RC time constant of the resistor-capacitor network. Once the capacitors charging voltage reaches the lower threshold voltage level of the MOSFETs gate, the MOSFET switches "ON" and illuminates the LED returning the circuit back to its stable state.

Then the application of the switch causes the circuit to enter its unstable state, while the time constant of the RC network returns it back to its stable state after a preset timing period thereby producing a very simple "one-shot" or Monostable Multivibrator MOSFET circuit.

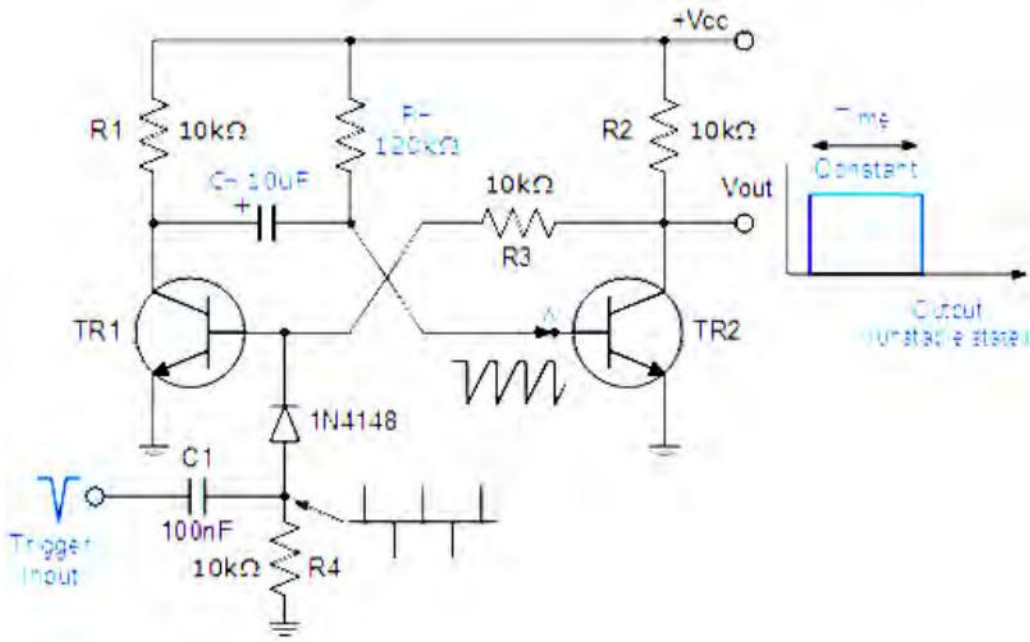
Monostable Multivibrators or "One-Shot Multivibrators" as they are also called, are used to generate a single output pulse of a specified width, either "HIGH" or "LOW" when a suitable external trigger signal or pulse T is applied. This trigger signal initiates a timing cycle which causes the output of the monostable to change its state at the start of the timing cycle and will remain in this second state.

The timing cycle of the monostable is determined by the time constant of the timing capacitor, C_T and the resistor, R_T until it resets or returns itself back to its original (stable) state. The monostable multivibrator will then remain in this original stable state indefinitely until another input pulse or trigger signal is received. Then, Monostable Multivibrators have only ONE stable state and go through a full cycle in response to a single triggering input pulse.

Monostable Multivibrator Circuit:

The basic collector-coupled transistor Monostable Multivibrator circuit and its associated waveforms are shown above. When power is firstly applied, the base of transistor TR_2 is connected to Vcc via the biasing resistor, R_T thereby turning the

transistor "fully-ON" and into saturation and at the same time turning TR_1 "OFF" in the process. This then represents the circuits "Stable State" with zero output. The current flowing into the saturated base terminal of TR_2 will therefore be equal to $I_b = (Vcc - 0.7)/R_T$.



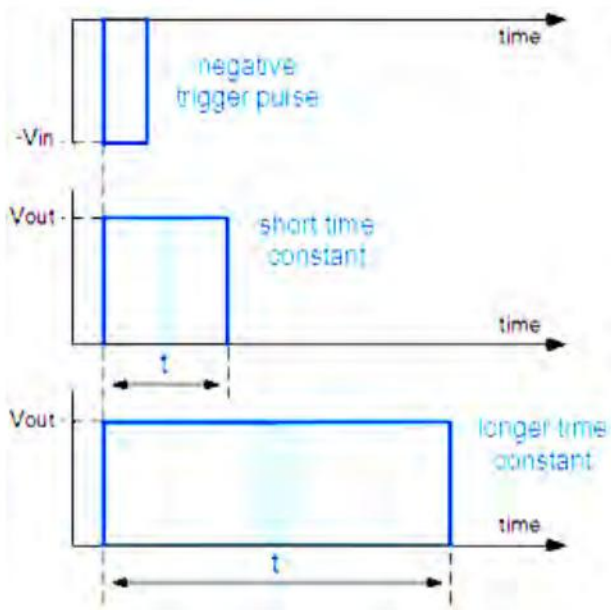
If a negative trigger pulse is now applied at the input, the fast decaying edge of the pulse will pass straight through capacitor, C1 to the base of transistor, TR_1 via the blocking diode turning it "ON". The collector of TR_1 which was previously at Vcc drops quickly to below zero volts effectively giving capacitor C_T a reverse charge of -0.7v across its plates. This action results in transistor TR2 now having a minus base voltage at point X holding the transistor fully "OFF". This then represents the circuits second state, the "Unstable State" with an output voltage equal to Vcc.

Timing capacitor, C_T begins to discharge this -0.7v through the timing resistor R_T, attempting to charge up to the supply voltage Vcc. This negative voltage at the base of transistor TR_2 begins to decrease gradually at a rate determined by the time constant of the R_T C_T combination. As the base voltage of TR_2 increases back up to Vcc, the transistor begins to conduct and doing so turns "OFF" again transistor TR1 which results in the monostable multivibrator automatically returning back to its original stable state awaiting a second negative trigger pulse to

restart the process once again.

Monostable Multivibrators can produce a very short pulse or a much longer rectangular shaped waveform whose leading edge rises in time with the externally applied trigger pulse and whose trailing edge is dependent upon the RC time constant of the feedback components used. This RC time constant may be varied with time to produce a series of pulses which have a controlled fixed time delay in relation to the original trigger pulse as shown below.

Monostable Multivibrator Waveforms:



The time constant of Monostable Multivibrators can be changed by varying the values of the capacitor, C_T the resistor, R_T or both. Monostable multivibrators are generally used to increase the width of a pulse or to produce a time delay within a circuit as the frequency of the output signal is always the same as that for the trigger pulse input, the only difference is the pulse width.

23.3 Bistable Multivibrator

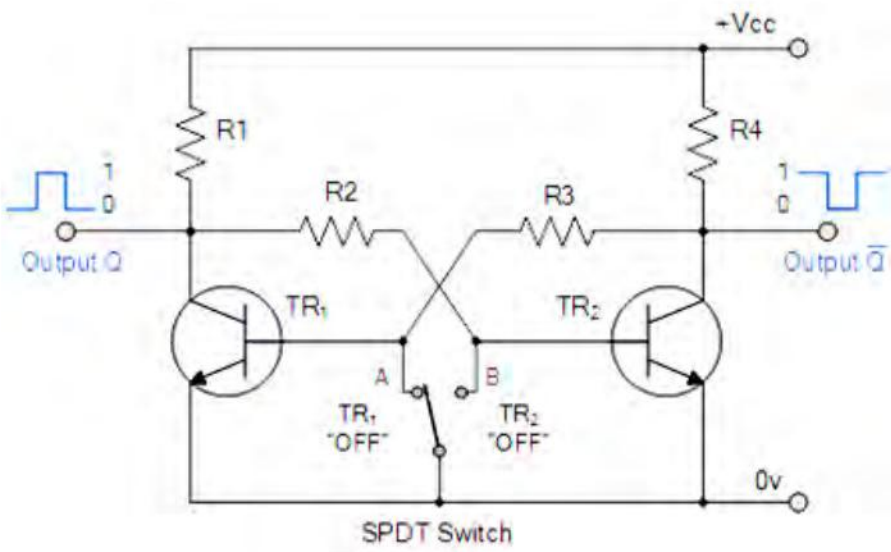
Bistable Multivibrators have TWO stable states (hence the name: "Bi" meaning two) and maintain a given output state indefinitely unless an external trigger is applied forcing it to change state.

The bistable multivibrator can be switched over from one stable state to the other by the application of an external trigger pulse thus, it requires two external trigger pulses before it returns back to its original state. As bistable multivibrators have two stable states they are more commonly known as Latches and Flip-flops for use in sequential type circuits.

The discrete Bistable Multivibrator is a two state non-regenerative device constructed from two cross-coupled transistors operating as "ON-OFF" Transistor Switches. In each of the two states, one of the transistors is cut-off while the other transistor is in saturation, this means that the bistable circuit is capable of remaining indefinitely in either stable state.

To change the bistable over from one state to the other, the bistable circuit requires a suitable trigger pulse and to go through a full cycle, two triggering pulses, one for each stage are required. Its more common name or term of "flip-flop" relates to the actual operation of the device, as it "flips" into one logic state, remains there and then changes or "flops" back into its first original state. Consider the circuit below.

Bistable Multivibrator Circuit:



The Bistable Multivibrator circuit above is stable in both states, either with one transistor "OFF" and the other "ON" or with the first transistor "ON" and the second "OFF". Lets suppose that the switch is in the left position, position "A". The base of transistor TR_1 will be grounded and in its cut-off region producing an

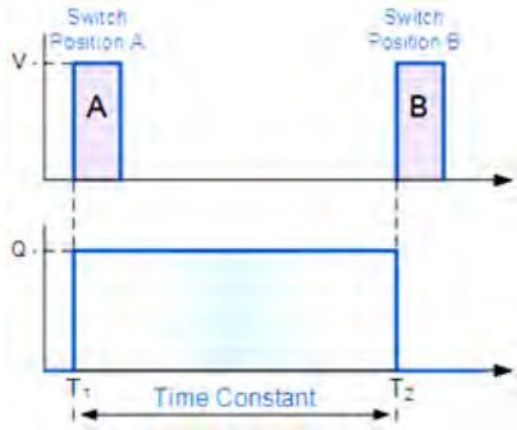
output at Q. That would mean that transistor TR_2 is "ON" as its base is connected to Vcc through the series combination of resistors R1 and R2. As transistor TR_2 is "ON" there will be zero output at Q, the opposite or inverse of Q.

If the switch is now move to the right, position "B", transistor TR_2 will switch "OFF" and transistor TR_1will switch "ON" through the combination of resistors R3 and R4 resulting in an output at Q and zero output at Q the reverse of above. Then we can say that one stable state exists when transistor TR_1 is "ON" and TR_2 is "OFF", switch position "A", and another stable state exists when transistor TR_1 is "OFF" and TR_2 is "ON", switch position "B".

Then unlike the monostable multivibrator whose output is dependent upon the RC time constant of the feedback components used, the bistable multivibrators output is dependent upon the application of two individual trigger pulses, switch position "A" or position "B".

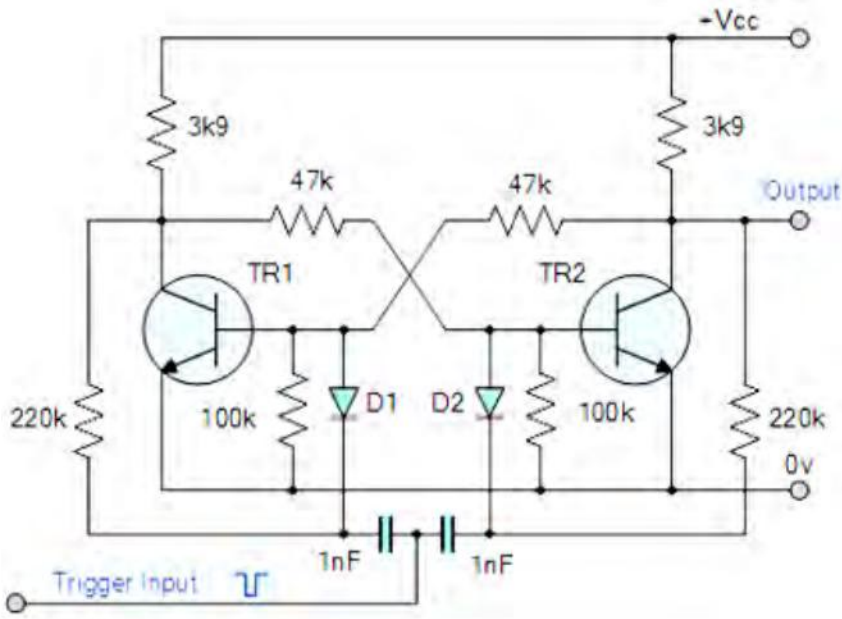
So Bistable Multivibrators can produce a very short output pulse or a much longer rectangular shaped output whose leading edge rises in time with the externally applied trigger pulse and whose trailing edge is dependent upon a second trigger pulse as shown below.

Bistable Multivibrator Waveform:



Manually switching between the two stable states may produce a bistable multivibrator circuit but is not very practical. One way of toggling between the two states using just one single trigger pulse is shown below.

Sequential Switching Bistable Multivibrator:



Switching between the two states is achieved by applying a single trigger pulse which in turn will cause the "ON" transistor to turn "OFF" and the "OFF" transistor to turn "ON" on the negative half of the trigger pulse. The circuit will switch sequentially by applying a pulse to each base in turn and this is achieved from a single input trigger pulse using a biased diodes as a steering circuit.

Then on the application of a first negative pulse switches the state of each transistor and the application of a second pulse negative pulse resets the transistors back to their original state acting as a divide-by-two counter. Equally, we could remove the diodes, capacitors and feedback resistors and apply individual negative trigger pulses directly to the transistor bases.

Bistable Multivibrators have many applications producing a set-reset, SR flip-flop circuit for use in counting circuits, or as a one-bit memory storage device in a computer. Other applications of bistable flip-flops include frequency dividers because the output pulses have a frequency that are exactly one half ($f/2$) that of the trigger input pulse frequency due to them changing state from a single input pulse. In other words the circuit produces Frequency Division as it now divides the input frequency by a factor of two (an octave).

23.4 Self Learning Exercise-I

Q.1 What is the other name of Monostable multivibrator.

Q.2 Classified multivibrator on the status of the output state.

Q.3 Draw waveform of Monostable and Bistable Multivibrator.

23.5 The Astable Multivibrator

Regenerative switching circuits such as Astable Multivibrators are the most commonly used type of relaxation oscillator because not only are they simple, reliable and ease of construction they also produce a constant square wave output waveform.

Unlike the Monostable Multivibrator or the Bistable Multivibrator we looked at in the previous sections that require an "external" trigger pulse for their operation, the Astable Multivibrator has automatic built in triggering which switches it continuously between its two unstable states both set and reset.

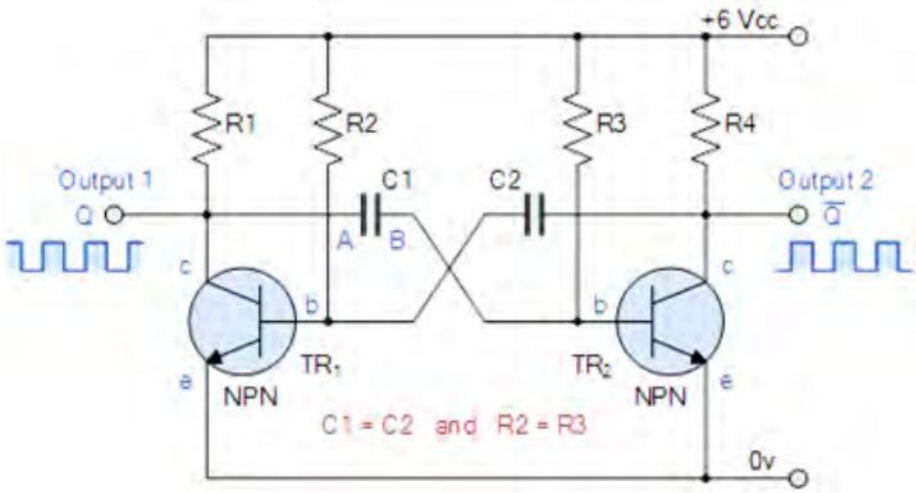
The Astable Multivibrator is another type of cross-coupled transistor switching circuit that has NO stable output states as it changes from one state to the other all the time. The astable circuit consists of two switching transistors, a cross-coupled feedback network, and two time delay capacitors which allows oscillation between the two states with no external trigger signal to produce the change in state.

In Electronic Circuits, astable multivibrators are also known as Free-running Multivibrator as they do not require any additional inputs or external assistance to oscillate. Astable oscillators produce a continuous square wave from its output or outputs, (two outputs no inputs) which can then be used to flash lights or produce a sound in a loudspeaker.

The basic transistor circuit for an Astable Multivibrator produces a square wave output from a pair of grounded emitter cross-coupled transistors. Both transistors either NPN or PNP, in the multivibrator are biased for linear operation and are operated as Common Emitter Amplifiers with 100% positive feedback.

This configuration satisfies the condition for oscillation when: ($\beta A = 1 \angle 0^{o}$). This results in one stage conducting "fully-ON" (Saturation) while the other is switched "fully-OFF" (cut-off) giving a very high level of mutual amplification between the two transistors. Conduction is transferred from one stage to the other by the discharging action of a capacitor through a resistor as shown below.

Astable Multivibrator Circuit:



Assume that transistor, TR_1 has just switched "OFF" and its collector voltage is rising towards Vcc, meanwhile transistor TR_2 has just turned "ON". Plate "A" of capacitor C1 is also rising towards the +6 volts supply rail of Vcc as it is connected to the collector of TR_1. The other side of capacitor, C1, plate "B", is connected to the base terminal of transistor TR_2 and is at 0.6v because transistor TR_2 is conducting therefore, capacitor C1 has a potential difference of 5.4 volts across it, 6.0 – 0.6v, (its high value of charge).

The instant that transistor, TR_1 switches "ON", plate "A" of the capacitor immediately falls to 0.6 volts. This fall of voltage on plate "A" causes an equal and instantaneous fall in voltage on plate "B" therefore plate "B" of the capacitor C1 is pulled down to -5.4V (a reverse charge) and this negative voltage turns transistor TR_2 hard "OFF". One unstable state.

Capacitor C_1 now begins to charge in the opposite direction via resistor R_3 which is also connected to the +6 volts supply rail, Vcc, thus the case of transistor TR_2 is

moving upwards in a positive direction towards Vcc with a time constant equal to the C_1 x R_3 combination.

However, it never reaches the value of Vcc because as soon as it gets to 0.6 volts positive, transistorTR_2 turns fully "ON" into saturation starting the whole process over again but now with capacitor C_2 taking the base of transistor TR_1 to -5.4v while charging up via resistor R_2 and entering the second unstable state. This process will repeat itself over and over again as long as the supply voltage is present.

The amplitude of the output waveform is approximately the same as the supply voltage, Vcc with the time period of each switching state determined by the time constant of the RC networks connected across the base terminals of the transistors. As the transistors are switching both "ON" and "OFF", the output at either collector will be a square wave with slightly rounded corners because of the current which charges the capacitors. This could be corrected by using more components as we will discuss later.

If the two time constants produced by C_2 x R_2 and C_1 x R_3 in the base circuits are the same, the mark-to-space ratio (t_1/t_2) will be equal to one-to-one making the output waveform symmetrical in shape. By varying the capacitors, C_1, C_2 or the resistors, R_2, R_3 the mark-to-space ratio and therefore the frequency can be altered.

We saw in the RC Discharging tutorial that the time taken for the voltage across a capacitor to fall to half the supply voltage, 0.5Vcc is equal to 0.69 time constants of the capacitor and resistor combination. Then taking one side of the astable multivibrator, the length of time that transistor $TR_2$is "OFF" will be equal to 0.69T or 0.69 times the time constant of C_1 x R_3. Likewise, the length of time that transistor TR_1 is "OFF" will be equal to 0.69T or 0.69 times the time constant of C_2 x R_2 and this is defined as.

Astable Multivibrators Periodic Time

$$T = t_1 + t_2$$
$$t_1 = 0.69 C_1 R_3$$
$$t_2 = 0.69 C_2 R_2$$

where, R is in Ohms and C in Farads.

By altering the time constant of just one RC network the mark-to-space ratio and frequency of the output waveform can be changed but normally by changing both RC time constants together at the same time, the output frequency will be altered keeping the mark-to-space ratios the same at one-to-one.

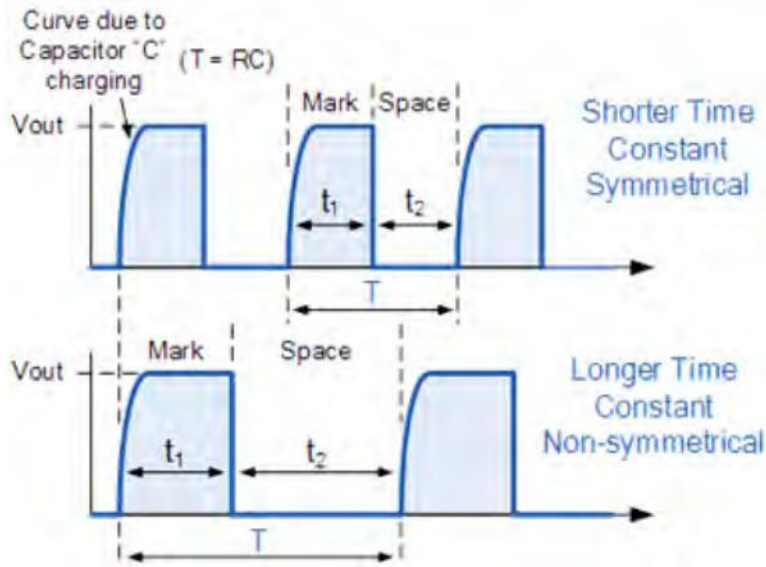
If the value of the capacitor C_1 equals the value of the capacitor, C_2, $C_1 = C_2$ and also the value of the base resistor R_2 equals the value of the base resistor, R_3, $R_2 = R_3$ then the total length of time of theMultivibrators cycle is given below for a symmetrical output waveform.

Frequency of Oscillation

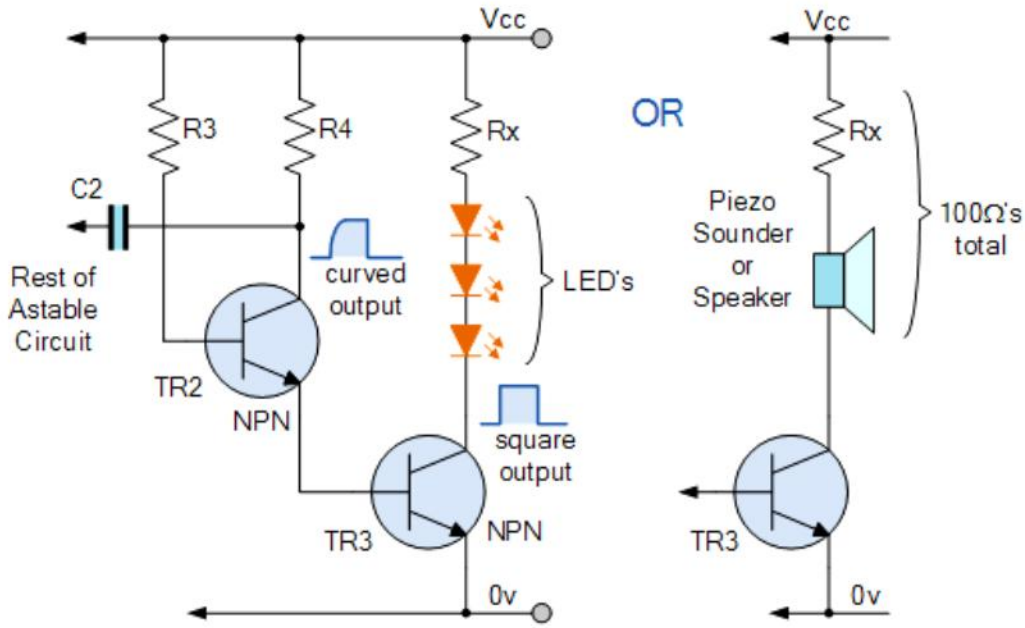
$$f=1/T=1/1.38RC$$

Where, R is in Ω's, C is in Farads, T is in seconds and *f* is in Hertz. and this is known as the "Pulse Repetition Frequency". So Astable Multivibrators can produce TWO very short square wave output waveforms from each transistor or a much longer rectangular shaped output either symmetrical or non-symmetrical depending upon the time constant of the RC network as shown below.

Astable Multivibrator Waveforms:



Astable Multivibrators Driving Circuit:



An output with a square leading edge is now produced from the third transistor, TR_3 connected to the emitter of transistor, TR_2. This third transistor switches "ON" and "OFF" in unison with transistorTR_2. We can use this additional transistor to switch Light Emitting Diodes, Relays or to produce a sound from a Sound Transducer such as a speaker or piezo sounder as shown above.

The load resistor, Rx needs to be suitably chosen to take into account the forward volt drops and to limit the maximum current to about 20mA for the LED circuit or to give a total load impedance of about 100Ω for the speaker circuit. The speaker can have any impedance less than 100Ω. By connecting an additional transistor, TR_4 to the emitter circuit of the other transistor, TR_1 in a similar fashion we can produce an astable multivibrator circuit that will flash two sets of lights or LED's from one to the other at a rate determined by the time constant of the RC timing network.

23.6 Waveform Generators

In the previous sections we have looked in detail at the three different types of basic transistor multivibrator circuits that can be used as relaxation oscillators to

produce either a square or rectangular wave at their outputs for use as clock and timing signals. But it is also possible to construct basic Waveform Generator circuits from simple integrated circuits or operational amplifiers connected to a resistor-capacitor (RC) tank circuit or to a quartz crystal to produce the required binary or square wave output waveform at the desired frequency.

This waveform generation would be incomplete without some examples of digital regenerative switching circuits, since it illustrates both the switching action and operation of waveform generators used for generating square waves for use as timing or sequential waveforms.

We know that regenerative switching circuits such as *Astable Multivibrators* are the most commonly used type of relaxation oscillator as they produce a constant square wave output, making them ideal as a digital Waveform Generator.

Astable multivibrators make excellent Oscillators because they switch continuously between their two unstable states at a constant repetition rate thereby producing a continuous square wave output with a 1:1 mark-space ratio ("ON" and "OFF" times the same) from its output and in this tutorial we will look at some of the different ways we can construct waveform generators using just standard TTL and CMOS logic circuits along with some additional discrete timing components.

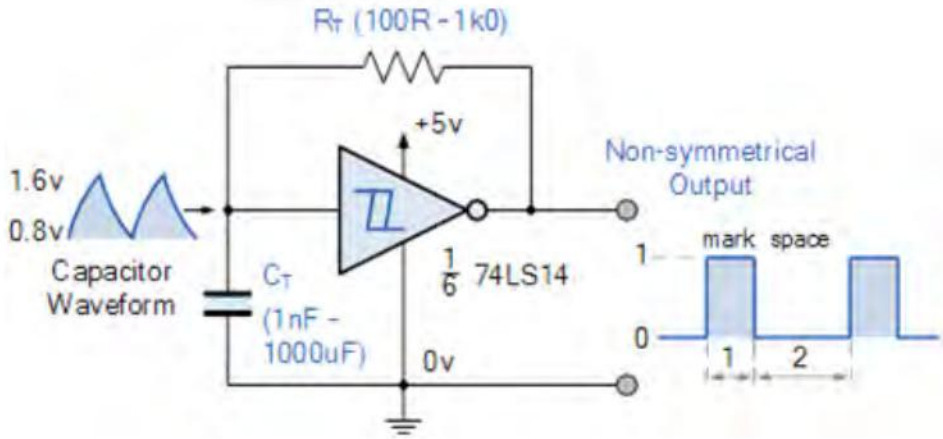
Schmitt Waveform Generators:

Simple Waveform Generators can be constructed using basic Schmitt trigger action inverters such as the TTL 74LS14. This method is by far the easiest way to make a basic astable waveform generator. When used to produce clock or timing signals, the astable multivibrator must produce a stable waveform that switches quickly between its "HIGH" and "LOW" states without any distortion or noise, and Schmitt inverters do just that.

We know that the output state of a Schmitt inverter is the opposite or inverse to that of its input state, (NOT gate principles) and that it can change state at different voltage levels giving it "hysteresis". Schmitt inverters use a Schmitt trigger action that changes state between an upper and a lower threshold level as the input voltage signal increases and decreases about the input terminal. This upper

threshold level “sets” the output and the lower threshold level “resets” the output which equates to a logic “0” and a logic “1” respectively for an inverter. Consider the circuit below.

Schmitt Inverter Waveform Generator:



This simple waveform generator circuit consists of a single TTL 74LS14 Schmitt inverter logic gate with a capacitor, C connected between its input terminal and ground, (0v) and the positive feedback required for the circuit to oscillate being provided by the feedback resistor, R.

Assume that the charge across the capacitors plates is below the Schmitt’s lower threshold level of 0.8 volt (Datasheet value). This therefore makes the input to the inverter at a logic “0” level resulting in a logic “1” output level (inverter principals).

One side of the resistor R is now connected to the logic “1” level (+5V) output while the other side of the resistor is connected to the capacitor, C which is at a logic “0” level (0.8V or below). The capacitor now starts to charge up in a positive direction through the resistor at a rate determined by the RCtime constant of the combination.

When the charge across the capacitor reaches the 1.6 volt upper threshold level of the Schmitt trigger (datasheet value) the output from the Schmitt inverter rapidly changes from a logic level “1” to a logic level “0” state and the current flowing through the resistor changes direction.

This change now causes the capacitor that was originally charging up through the resistor, R to begin to discharge itself back through the same resistor until the charge across the capacitors plates reaches the lower threshold level of 0.8 volts and the inverters output switches state again with the cycle repeating itself over and over again as long as the supply voltage is present.

So the capacitor, C is constantly charging and discharging itself during each cycle between the inputs upper and lower threshold levels of the Schmitt inverter producing a logic level "1" or a logic level "0" at the inverters output. However, the output waveform is not symmetrical producing a duty cycle of about 33% or 1/3 as the mark-to-space ratio between "HIGH" and "LOW" is 1:2 respectively due to the input gate characteristics of the TTL inverter.

The value of the feedback resistor, (R) MUST also be kept low to below 1kΩ for the circuit to oscillate correctly, 220R to 470R is good, and by varying the value of the capacitor, C to vary the frequency. Also at high frequency levels the output waveform changes shape from a square shaped waveform to a trapezoidal shaped waveform as the input characteristics of the TTL gate are affected by the rapid charging and discharging of the capacitor. The frequency of oscillation for Schmitt Waveform Generators is therefore given as:

$$f=1/1.2RC$$

With a resistor value between: 100R to 1kΩ, and a capacitor value of between: 1nF to 1000uF. This would give a frequency range of between 1Hz to 1MHz, (high frequencies produce waveform distortion).

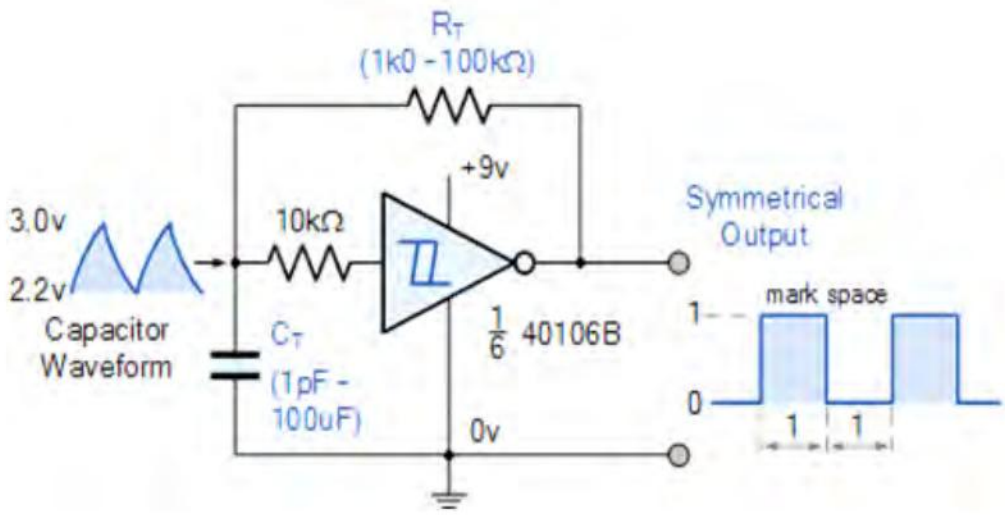
Generally, standard TTL logic gates do not work too well as waveform generators due to their average input and output characteristics, distortion of the output waveform and low value of feedback resistor required, resulting in a large high value capacitor for low frequency operation. Also TTL oscillators may not oscillate if the value of the feedback capacitor is too small. However, we can also make Astable Multivibrators using better CMOS logic technology that operate from a 3V to 15V supply.

The CMOS 40106 is a single input inverter with the same Schmitt-trigger action as

the TTL 74LS14 but with very good noise immunity, high bandwidth, high gain and excellent input/output characteristics to produce a more "squarer" output waveform as shown below.

CMOS Schmitt Waveform Generator:

The Schmitt waveform generators circuit for the CMOS 40106 is basically

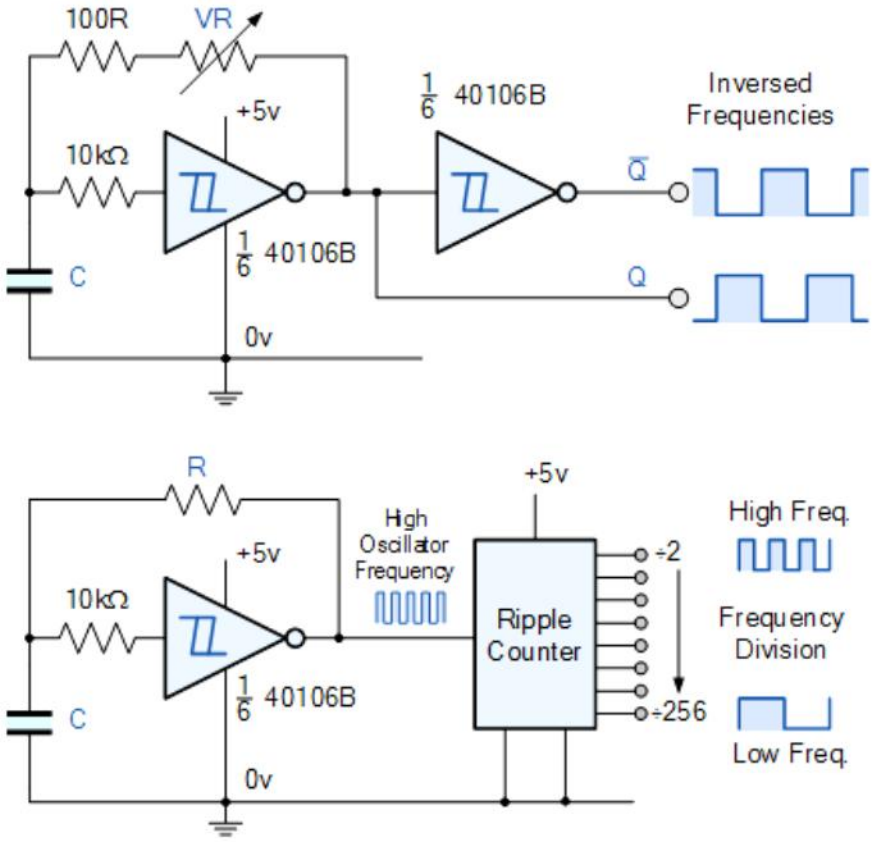


the same as that for the previous TTL 74LS14 inverter, except for the addition of the 10kΩ resistor which is used to prevent the capacitor from damaging the sensitive MOSFET input transistors as it discharges rapidly at higher frequencies.

The mark-space ratio is more evenly matched at about 1:1 with the feedback resistor value increased to below 100kΩ resulting in a smaller and cheaper timing capacitor, C. The frequency of oscillation is given the same: (1/1.2RC), with a resistor value between: 1kΩ and 100kΩ, and a capacitor value of between: 1pF to 100uF. This would give a frequency range of between 0.1Hz to 100kHz.

Schmitt Inverter Waveform Generators can also be made from a variety of different logic gates connected to form an inverter circuit. The basic Schmitt astable multivibrator circuit can be easily modified with some additional components to produce different outputs or frequencies. For example, two inverse waveforms or multiple frequencies and by changing the fixed feedback resistor to a potentiometer the output frequency can be varied as shown below.

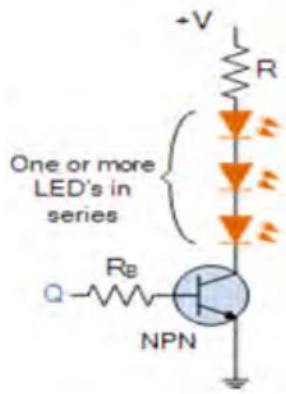
Clock Waveform Generators:



In the first circuit above, an additional Schmitt Inverter has been added to the output of the Schmitt waveform generator to produce a second waveform that is the inverse or mirror image of the first producing two complementary output waveforms, so when one output is "HIGH" the other is "LOW". This second Schmitt inverter also improves the shape of the inverse output waveform but adds a small "gate delay" to it so it is not exactly in synch with the first.

Also, the output frequency of the oscillator circuit can be varied by changing the fixed resistor, R into a potentiometer but a smaller feedback resistor is still required to prevent the potentiometer from shorting out the inverter when its at its minimum value, 0Ω's.

We can also use the two complementary outputs, Q and Q of the first circuit to alternatively flash two sets of lights or LED's by connecting their outputs directly to the bases of two switching transistors as shown.



In this way one or several LED's connected together in series with the collector of the switching transistors resulting in alternating flashes of each set of LED's as each transistor is switched "ON" in turn.

Also when using this type of circuit, remember to calculate a suitable series resistor, R to limit the LED current to below 20mA (red LED's) for the voltage you are using.

In order to generate a very low frequency output of a few Hertz to flash the LED's, Schmitt waveform generators use high value timing capacitors which themselves can be physically large and expensive.

One alternative solution is too use a smaller value capacitor to generate a much higher frequency, say 1kHz or 10kHz, and then divide this main clock frequency down into individual smaller ones until the required low frequency value is achieved, and the second circuit above does just that.

The lower circuit above shows the oscillator being used to drive the clock input of a ripple counter. Ripple counters are basically a number of divide-by-2, D-type flip-flops cascaded together to form a single divide-by-N counter, where N is equal to the counters bit-count such as the CMOS 4024 7-bit Ripple Counter or the CMOS 4040 12-bit Ripple Counter.

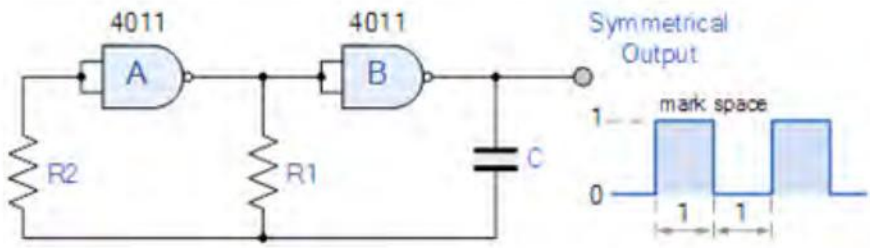
The fixed clock frequency produce by the Schmitt astable clock pulse circuit is divided into a number of different sub-frequencies such as, $f \div 2$, $f \div 4$, $f \div 8$, $f \div 256$, etc, up to the maximum "Divide-by-n" value of the ripple counter being used. This process of using either "Flip-flops", "Binary Counters" or "Ripple Counters" to divide a main fixed clock frequency into different sub-frequencies is known as

Frequency Division and we can use it to obtain a number of frequency values from a single waveform generator.

NAND Gate Waveform Generators:

Schmitt Waveform Generators can also be made using standard CMOS Logic NAND Gatesconnected to produce an inverter circuit. Here, two NAND gates are connected together to produce another type of RC relaxation oscillator circuit that will generate a square wave shaped output waveform as shown below.

NAND Gate Waveform Generator:



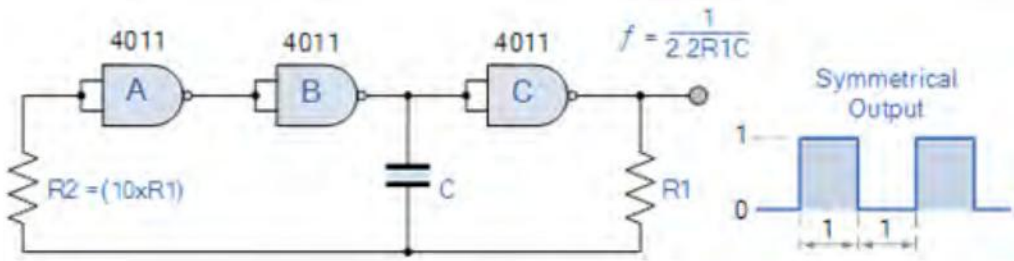
In this type of waveform generator circuit, the RC network is formed from resistor, R_1 and the capacitor, C with this RC network being controlled by the output of the first NAND gate. The output from this R_1C network is fed back to the input of the first NAND gate via resistor, R+ and when the charging voltage across the capacitor reaches the upper threshold level of the first NAND gate, the NAND gate changes state causing the second NAND gate to follow it, thereby change state and producing a change in the output level.

The voltage across the R_1C network is now reversed and the capacitor begins to discharge through the resistor until it reaches the lower threshold level of the first NAND gate causing the two gates to change state once again. Like the previous Schmitt waveform generators circuit above, the frequency of oscillation is determined by the R_1C time constant which is given as: $1/2.2R_1C$. Generally R_2 is given a value that is 10 times the value of resistor R_1.

When high stability or guaranteed startability is required, CMOS Waveform Generators can be made using three inverting NAND gates or any three logic inverters for that matter, connected together as shown below producing a circuit that is sometimes called “the ring of three” waveform generator. The frequency of

oscillation is determined again by the R_1C time constant, the same as for the two gate oscillator above, and which is given as: $1/2.2R_1C$ when R_2 has a value that is 10 times the value of resistor, R_1.

Stable NAND Gate Waveform Generator:



The addition of the extra NAND gate guarantees that the oscillator will start even with very low capacitor values. Also the stability of the waveform generator is greatly improved as it is less susceptible to power supply variations due to its threshold triggering level being nearly half of the supply voltage.

The amount of stability is mainly determined by the frequency of oscillation and generally speaking, the lower the frequency the more stable the oscillator becomes.

As this type of waveform generator operates at nearly half or 50% of the supply voltage the resultant output waveform has very nearly a 50% duty cycle, 1:1 mark-space ratio. The three gate waveform generator has many advantages over the previous two gate oscillator above but its one big disadvantage is that it uses an additional logic gate.

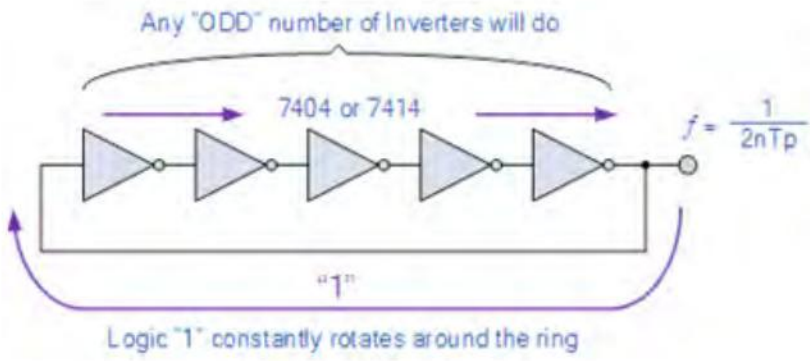
Ring Type Waveform Generator:

We have seen above that Waveform Generators can be made using both TTL and the better CMOS logic technology with an RC network producing a time delay within the circuit when connected across either one, two or even three logic gates to form a simple RC Relaxation Oscillator. But we can also make waveform generators using just Logic NOT Gates or in other wordsInverters without any additional passive components connected to them.

By connecting together any ODD number (3, 5, 7, 9 etc) of NOT gates to form a "ring" circuit, so that the output of the ring is connected straight back to the input

of the ring the circuit will continue to oscillate as a logic level "1" constantly rotates around the network producing an output frequency that is determined by the propagation delays of the inverters used.

Ring Waveform Generator:



The frequency of oscillation is determined by the total propagation delay of the Inverters used within the ring and which itself is determined by the type of gate technology, TTL, CMOS, BiCMOS that theinverter is made from. Propagation delay or propagation time, is the total time required (usually in Nanoseconds) for a signal to pass straight through the Inverter from a logic "0" arriving at the input to it producing a logic "1" at its output.

Also for this type of ring waveform generator circuit variations in the supply voltage, temperature and load capacitance all affect the propagation delay of logic gates. Generally an average propagation delay time will be given in the manufacturers data sheets for the type of digital logic gates being used with the frequency of oscillation given as:

f=1/2nTp Hz

Where: *f* is the frequency of Oscillation, n is the number of gates used and Tp is the propagation delay for each gate.

23.7 Self Learning Exercise- II

Q.1 What is the advantage of using Astable Multivibrator.

Q.2 Draw Astable Multivibrator Circuit.

Q.3 Explain working and principle of Schmitt Waveform Generators

23.8 Summary

In this chapter, we have learned that: Multivibrator circuits are used to produce free-running clock oscillator waveforms or to produce a timed digital level change triggered by an external source. We also discussed in detail various types of Multivibrator.

23.9 Glossary

Basic waveshape: The sinusoidal waveform can be viewed as a kind of a basic waveshape as all periodic waveforms of any waveshape can be expressed in terms of the sinusoidal waveshapes of different frequencies.

Bidirectional sampling gate: A bidirectional sampling gate can transmit an input signal that assumes both positive and negative polarities under the influence of an appropriate control input. This circuit is obtained by simply combining two unidirectional diode sampling gates. These two unidirectional gates take care of the two polarities of the input pulse

Clock: Clock is a periodic waveform that is formed by a train of pulses assuming high and low values alternately. It plays a very important role in the functioning of a sequential logic circuit.

Free running mode: If the voltage sweep waveform is generated in freerunning mode, the waveform is initiated immediately, and the circuit does not wait for any external trigger input to initiate the next cycle of the sweep waveform. Free-running mode operation is also described by terms such as astable mode and recurrent mode.

Relaxation oscillator: The term relaxation oscillator is employed to any circuit in which timing intervals are determined by the exponential charging (tensing) or discharging (relaxing) of a capacitor. A charged capacitor functions like a stiff spring that has stored potential energy in it. A discharging capacitor is analogous to a relaxing spring in which stored potential energy is being converted to kinetic energy.

Schmitt trigger circuit: The transistor Schmitt trigger circuit is a variation of the conventional bistable multivibrator. Transitions in stable states of an Schmitt trigger circuit can be induced by a waveform of an arbitrary shape and the only condition is that the maximum amplitude of the waveform has to exceed VUTP of the circuit. The Schmitt trigger circuit cannot function as a 1-bit memory element. Also see upper-tripping point and lowertripping point.

Stable state: If the multivibrator can remain indefinitely in a state until the circuit is triggered by an external signal, that state is known as a stable state.

Triggered mode: In triggered-mode operation, the sweep waveform generator does not generate a periodic output waveform and the output waveform is present at irregular intervals. The output of a triggered-mode circuit remains at a quiescent value and the circuit waits for the arrival of an external trigger pulse. The triggered-mode operation is also known as monostable operation.

Unsymmetrical triggering: In unsymmetrical triggering the trigger pulse is effective in inducing a transition in only one direction. In order to induce a reverse transition, the second trigger pulse from a second source must be applied at a different point.

23.10 Answer To Self Learning Exercises

Answer To Self Learning Exercise-I

Ans.1: One-Shot Multivibrators.

Ans.2: Astable (free running),Monostable (one shot),Bistable (flip flop)

Ans.3: section 23.2,23.3

Answer To Self Learning Exercise-II

Ans.1: Unlike the Monostable Multivibrator or the Bistable Multivibrator that require an "external" trigger pulse for their operation, the Astable Multivibrator has automatic built in triggering which switches it continuously between its two unstable states both set and reset.

23.11 Exercise

Q.1 What is the other name of Astable multivibrator.

Q.2 Write down the applications of Bistable Multivibrators.

Q.3 Draw any one type of a Schmitt trigger circuit and explain with waveforms.

Q.4 Draw the circuit of an astable multivibrator and explain its working.

Q.5 A monostable multivibrator is used as a voltage-to-time converter. Find the time period if R = 10 kΩ ,C = 0.01 μF, V_{BB}/V_{CC}= 5.0

23.12 Answers to Exercise

Ans.2: (i)Producing a set-reset, SR flip-flop circuit for use in counting circuits

(ii)Frequency dividers

Ans.5: Time period T is given by $T = RC \ln[1+ V_{CC}/ V_{BB}]$

After putting numbers T = 0.11 ms , F=1/T= 9.09 kHz.

Reference and Suggested Readings

10. Power Electronics by P.C.Sen, 1987.
11. Digital Principles and Applications by D. P. Leach and A. P. Malvino, 1975.
12. Digital Electronics By R. J.Tocci, 1980
13. Principles of Digital Design, by D. D. Gajski 1997.
14. Digital Logic Circuit Analysis and Design, by V. P. Nelson, H. T. Nagle, B. D. Carroll, and J. D. Irwin, 1995.
15. Contemporary Logic Design, by R. H. Katz,1994.
16. Introduction to Logic Design, by J. P. Hayes 1993.
17. Digital Design Principles and Practices, by J. F.Wakerly, 1990.

UNIT-24
D/A Conversion and A/D Conversion

Structure of the Unit

24.0 Objectives

24.1 Introduction

24.2 Digital to Analog (D/A) Converters

24.3 Variable Resistor Network For D/A Conversion

24.4 Binary ladder

24.5 Resolution D/A Converters

24.6 Accuracy D/A Converters

24.7 Self Learning Exercise-I

24.8 Analog to Digital (A/D) Converters

24.9 Accuracy A/D Converters

24.10 Resolution A/D Converters

24.11 Self Learning Exercise-II

24.12 Summary

24.13 Glossary

24.14 Answers to Self Learning Exercises

24.15 Exercise

References and Suggested Readings

24.0 Objectives

This unit introduces you to the concept of digital to analog converters. After studying this unit, you should be able to (i) know how to use a binary ladder to accomplish D/A Conversion and (ii) determine the accuracy and resolution of D/A converters. This unit gives you an understanding about analog-to-digital converters. After completing this unit, you should be able to explain the operation

of a counter type A/D converter. You should be able to determine the accuracy and the resolution of an A/D converter.

24.1 Introduction

In electronics, a digital to analog converter or (D-to-A) is a device which converts a digital (usually binary) code to an analog signal (current, voltage or electric charge).

The D-to-A converter fundamentally converts a finite- precision numbers (usually fixed- point binary numbers) into continuously varying physical quantity, usually an analogue electrical device. D/A conversion involve translating digital information into equivalent analog information. For ex: the output of a digital system might be changed to analog form for the purpose of driving a pen recorder. Similarly, an analog signal might be required for the servo motors which drive the cursor arms of a plotter. In this way, a D/A converter is sometimes considered as a decoding device.

D/A conversion is a straight forward process and is considerably easier than A/D conversion. A D/A converter is an integral part of any A/D converter. Therefore it is considered the digital -to- analog conversion process first.

The process of changing an analog signal to an equivalent digital signal is accomplished by the use of an A/D converter. For example, an A/D converter is used to change the analog output signals from transducers (measuring temperature, pressure, vibration, etc) into equivalent digital signals. These signals would then be in a form suitable for entry into a digital system. An A/D converter is often referred to as an encoding device since it is used to encode signals for entry into a digital system.

The operation of converting an analog voltage in an equivalent digital signal is some what more complicated than the converse operation of D/A Conversion. A number of different methods and techniques have been developed for this A/D conversion which is discussed in this unit.

24.2 Digital to Analog (D/A) Converters

Digital-to-Analog converters can be broadly classified in three categories: Current output, voltage output, and multiplying type. The current output DAC, as the name

suggests, provides current as the output signal. The voltage output DAC internally converts the current signal into the delay in converting the current signal into the voltage signal. However, in many applications, it is necessary to convert current into voltage by using an external operational amplifier. The multiplying DAC is similar to the other two types except its output represents the product of the input signal and the reference source and the product are linear over a broad range. Conceptually, there is not much difference between these three types; any DAC can be viewed as a multiplying DAC.

D/A converters are available as integrated circuits. Some are specially designed to be compatible with the microprocessor/ Microcontroller. Typical applications include digital voltmeters, peak detectors, panel meters, programmable gain and attenuation, and stepping motor drive.

24.3 Variable Resistor Network for D/A Conversion

The basic problem in converting a digital signal into an equivalent analog signal is to change the 'n' digital voltage levels into one equivalent analog voltage. This can be easily accomplished by designing a resistive network that will change each digital level into an equivalent binary weighted voltage or current.

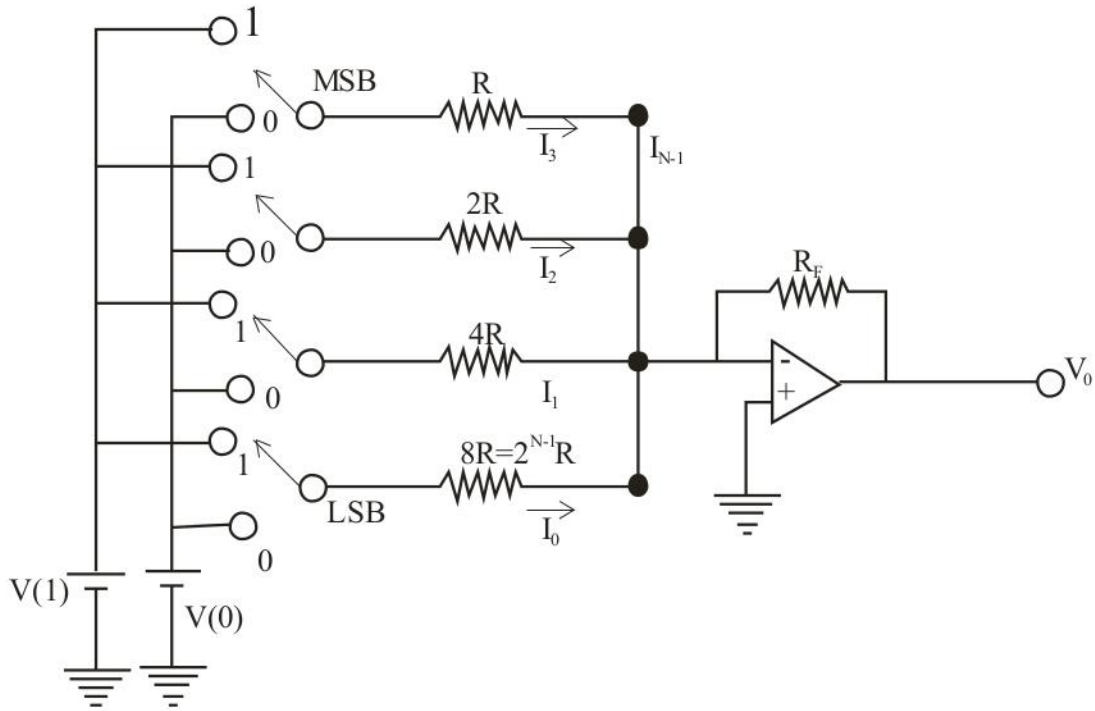


Fig. 24.1

Let us assume an N bit binary input to a resistor network, which produces a current I corresponding to logic 1 in most significant bit (MSB). A Logic 1 at the next lower bit causes a current I/2, still next lower bit having logic 1 produces a current I/4 and so on. For convertor of N bits, a logic 1 at its LSB (Least significant bit) causes a current of $\frac{1}{2^{N-1}}$. The total current thus produced will be proportional to the digital inputs. This current can be converted into voltage by using op-amplifier which will be proportional to the digital input.

Operations according to above principle. This current is called wergnea- resistor D/A converter since the resistances used in the converter are weighted in accordance with the binary weights.

Switches are used to connect digital 1 or 0 i.e. V (1) for 1 and V (0) for 0. Digital input from 0000 to 1111 can be applied through switches. The current I_i is given by

$$I_i = I_{N-1} + I_{N-2} + \ldots\ldots\ldots + I_2 + I_1 + I_0$$

From the Fig24.1 we have

$$I_{N-1} = V/R$$

$$I_{N-2} = V/2R$$

$$I_{N-3} = V/2^2R$$

$$I_{N-4} = V/2^3R$$

where V=1 if bit = 1; V = 0 if bit = 0.

$$I_3 = V/R$$

$$I_2 = \frac{V}{2R}$$

$$I_1 = V/2^2R$$

$$I_0 = V/2^3R$$

where V = V (1) if the bit = 1; V = V (0) if the bit = 0.

and $I_i = I_3 + I_2 + I_1 + I_0$.

For straight binary input V (0) = 0 and V (1) = - VR, The output voltage V_0 is given by:

$$V_0 = -(-V_R)\left[\frac{RF}{R}b_{N-1} + \frac{RF}{2^2R}b_{N-3} + \frac{RF}{2^3R}b_{N-4} + \ldots\ldots\frac{RF}{2^{N-1}R}b_0\right]$$

If $K = \frac{RF}{2^{N-1}R}\cdot V_R$ then

$$V_0 = K\,(2^{N-1} + b_{N-1} + 2^{N-2}\,b_{N-2} + 2_{N-3}\,b_{N-3} + \ldots\ldots\, 2^2\,b_2 + 2^1 b_1 + b_0)$$

K is a proportionality factor.

b_n = 1 if the nth bit of digital input is at 1.

b_n = 0 if the nth bit of digital input is at 0.

24.4 Binary Ladder

A Binary ladder also known as R-2R ladder network is a superior type of D/A converter, which eliminates the various drawbacks suffered by the resistine devider type of D/A converter. R-2R Ladder is a resistive network whose output is a property weighted sum of digital inputs. It is constructed of resistors that have only two values namely R and 2R. 4 bit D/A converter using R-2R Ladder is shown in figure 24.2.

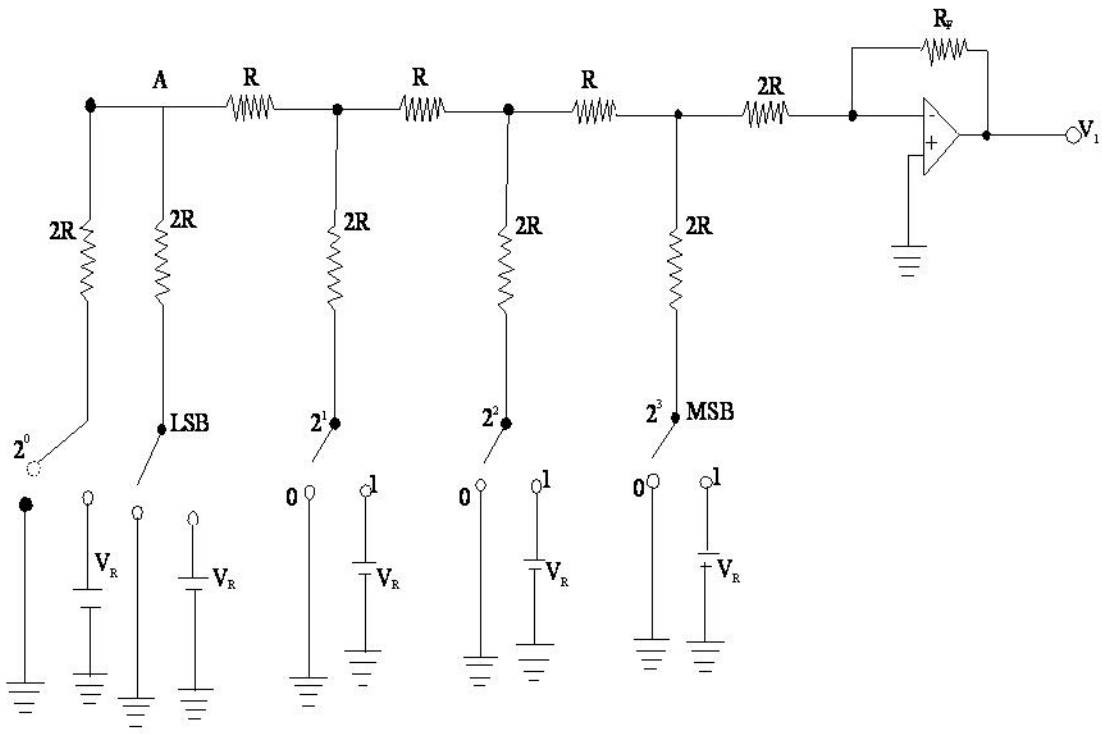


Fig. 24.2

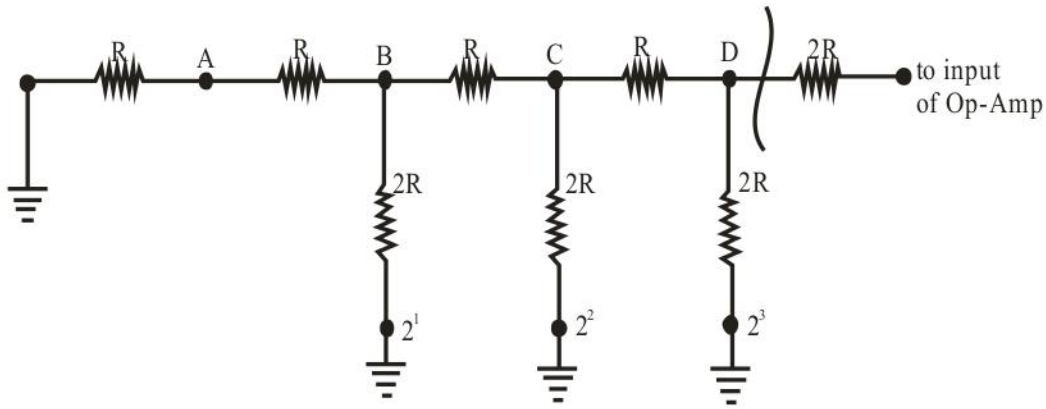


Fig. 24.3

Now looking from Node B towards branch A is 2R and also towards 2^1 input is also 2R. These two resistors can be combined to simplify the network as shown in fig. 24.4

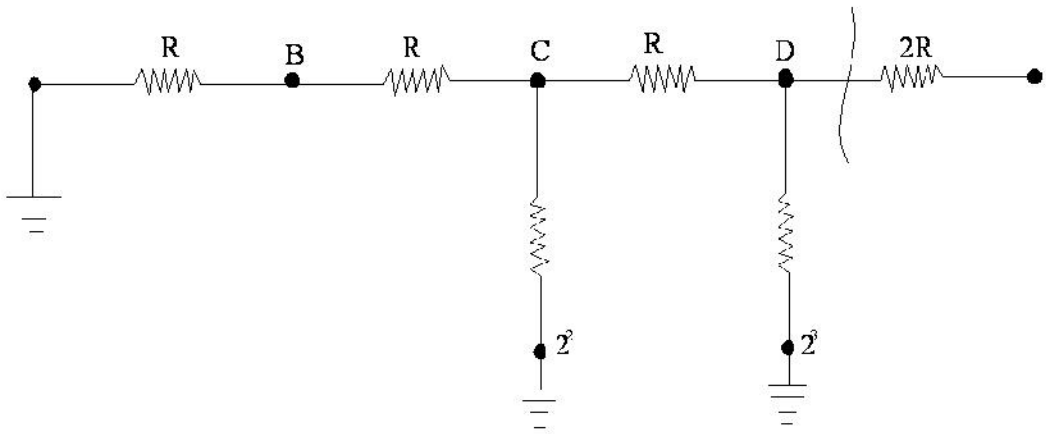


Fig. 24.4

From Figure 24.4, it can be seen that total resistance from node C towards branch B is 2R and towards digital input 2^2 is also 2R. This circuit is further reduced to Figure 24.5.

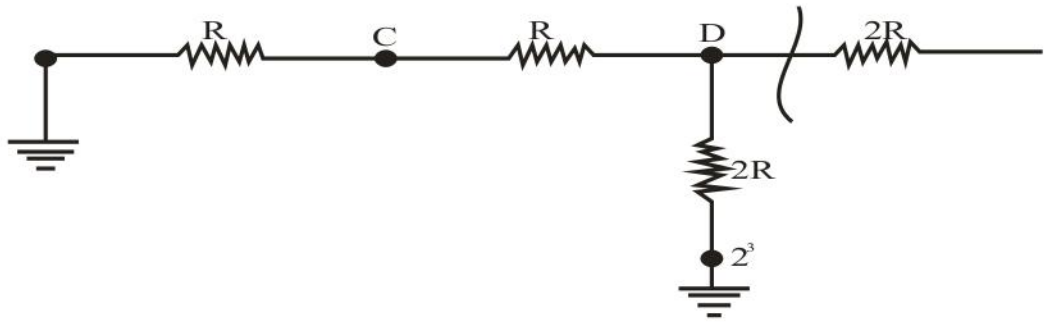


Fig. 24.5

From the figure 24.5 it is clear that looking towards branch C is 2R and towards digital input 2^3 is also 2R. From above discussion we can conclude that from any node looking towards the terminating resistor or out towards digital input is 2R. This is true irrespective of the inputs are grounded or at +V.

Using above discussion we try to determine output voltages for various digital inputs. Let us assume a digital input 1000. For this input Ladder Circuit is shown in Fig. 24.6.

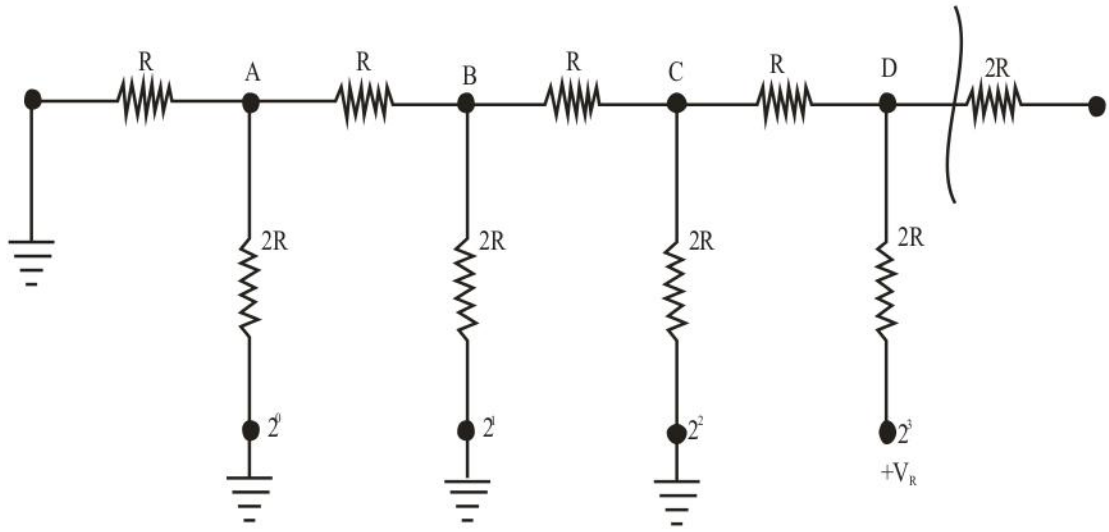


Fig. 24.6

As all inputs towards left of node D are grounded, entire network to the left of node D is replaced by resistance 2R to form a equivalent circuit shown fig. 24.7

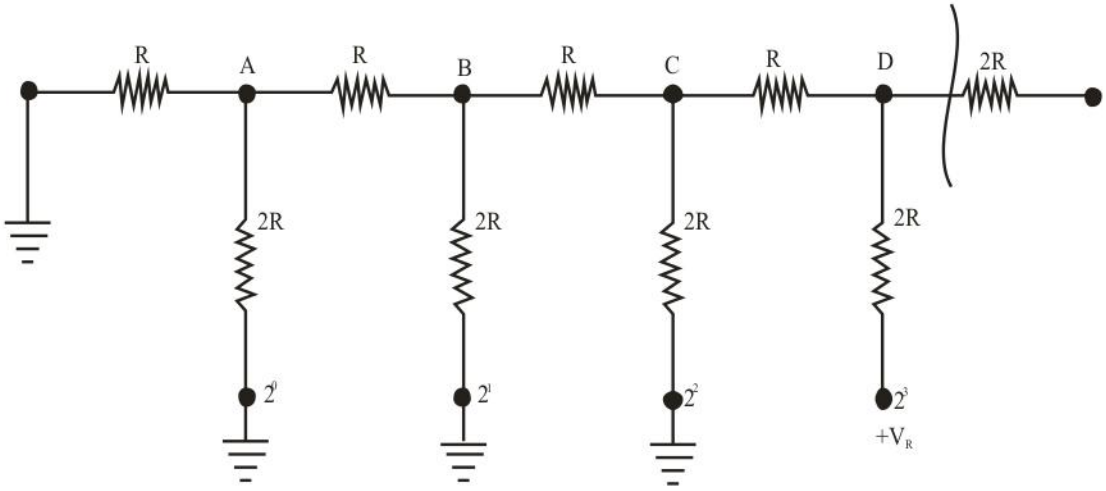


Fig. 24.7

From the circuit the output voltage is $V_A = V_R \times \frac{2R}{2R+2R} = \frac{+V_R}{2}$, thus 1 in MSB position produces an output $V_A = \frac{V_R}{2}$.

To determine the output voltage due to next MSB, consider a digital signal 0100. This is shown in figure 24.8

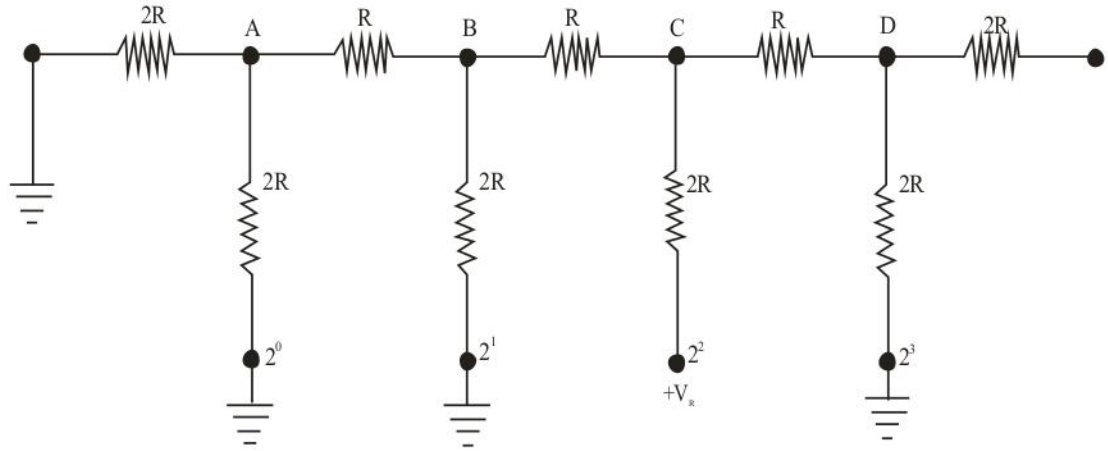


Fig. 24.8

As there is no voltage source towards left of node C, the entire network is replaced by a resistor 2R as shown in fig. 24.9

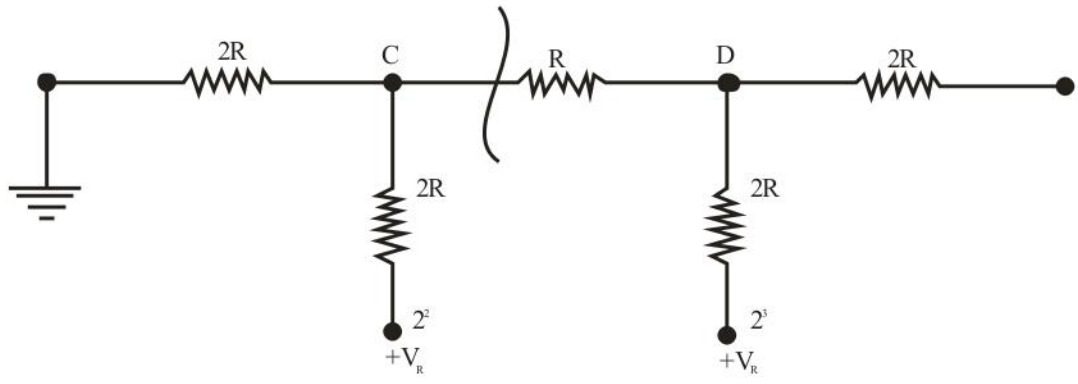


Fig. 24.9

Analyses the circuit towards left of the node C using thevenins Theorem. We get a thevenins voltage $\frac{+V_R}{2}$ in series with a resistance R final equivalent circuit with thevenin equivalent is included in figure 24.10.

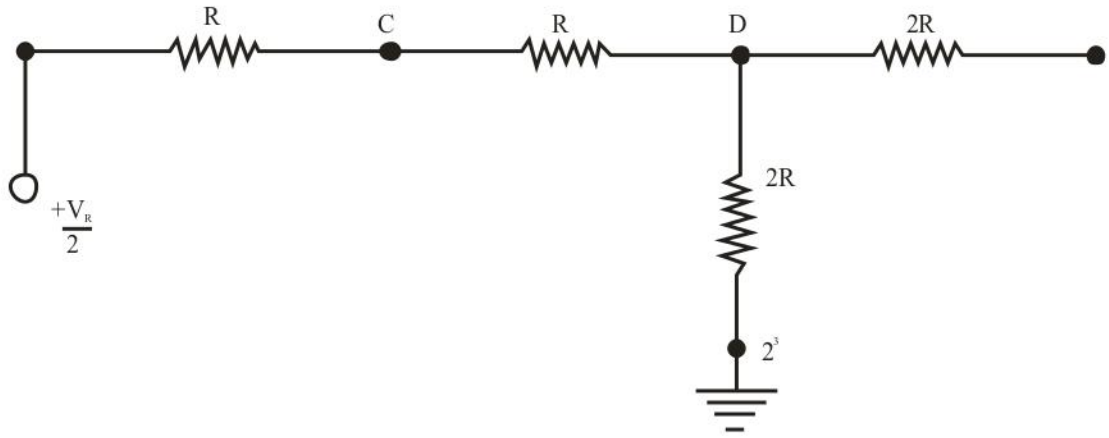


Fig. 24.10

From the above circuit shown in fig. 24.10 the output from node D is clearly $V_A = \frac{+V_R}{2} \times \frac{2R}{2R+2R} = \frac{+V_R}{4}$. Thus second MSB produces $V_A = \frac{+V_R}{4}$. This process can be continued for digital input 0010 and 0001 which provides an output voltages $V_A = \frac{+V_R}{8}$ and $V_A = \frac{+V_R}{16}$ respectively.

From preceeding discussion we noticed that MSB has binary weight of $\frac{+V_R}{2}$, Next MSB has $\frac{+V_R}{4}$ and so on and LSB has binary weight of $\frac{+V_R}{16}$. The value of equivalent resistance at node D is R which is in series with input resistance 2R of Op-Amp resulting in 3R in each case. Therefore we obtain an equivalent circuit of Figure 24.2 which is given in figure 24.11.

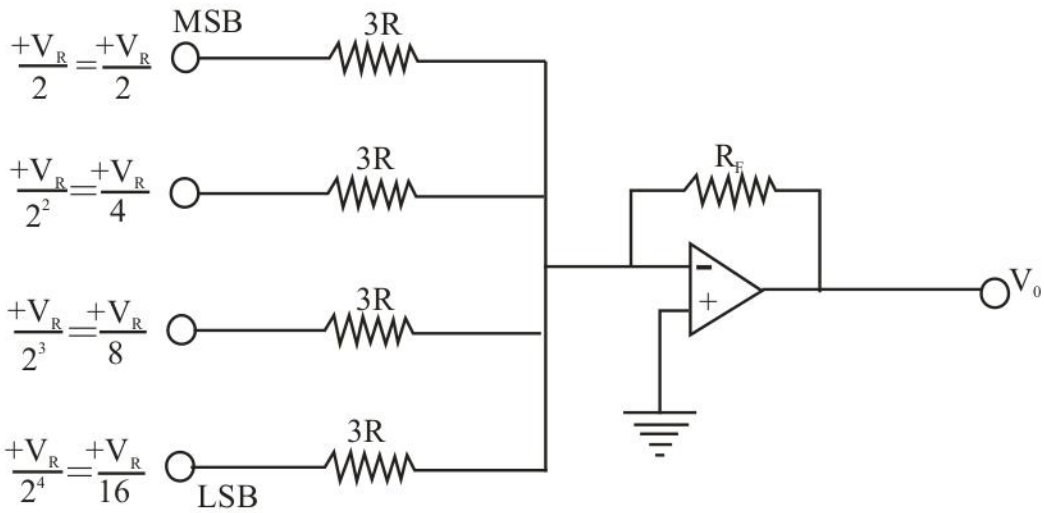


Fig. 24.11

The output analog voltage for the fig. 24.11 is given by,

$$V_0 = -\left(\frac{R_F}{3R}.\frac{V_R}{2^4}b_0 + \frac{R_F}{3R}.\frac{V_R}{2^3}b_1 + \frac{R_F}{3R}.\frac{V_R}{2^2}b_2 + \frac{R_F}{3R}.\frac{V_R}{2^1}b_3\right)$$

$$V_0 = \left(\frac{-R_F}{3R}\right)\left(\frac{V_R}{2^4}\right)\left[1b_0 + 2b_1 + 4b_2 + 8b_3\right]$$

In general for N bit D/A converter, the output voltage can be similarly determined and is given by

$$V_0 = (2^{N-1}b_{N-1} + 2^{N-2}b_{N-2} + 2^{N-3}b_{N-3} + \ldots.. 2^2b_2 + 2^1b_1 + 2^0b_0)$$

Where, $R_F = 3R$ and $V_R = -2^N V$.

The performance of a D/A converter is judged on the basis of the following:

1. Resolution
2. Accuracy

24.5 Resolution D/A Converters

The resolution capability determines the smallest change in the analog input, which can be discriminated by D/A converter. A converter which can accommodate large number of bits, will have a better resolution than one which accommodates a smaller number of bits. For instance a converter with 10-bits will have 2^{10} or 1024 output voltages. The smallest change which it can detect will be $\frac{1}{1024}$ of the full scale output. This converter has a resolution of 1 part in 1024. The resolution expressed as a percentage is $\frac{1}{1024}\times 100 = 0.1\%$. If the full scale output voltage is +10V, then the resolution in volts is $\frac{1}{1024}\times 10^V = 10\text{mV}$. A converter which can accommodate 8 bits will have 2^8 or 255 output voltages. It will have a resolution of 1 part in 255 or 0.4%. For full scale output voltage of +10V, the resolution in volts is $\frac{1}{255}\times 10^V \cong 40\text{mV}$. It is clear that 10 bit A/P converter has better resolution than 8 bit N_D converter.

24.6 Accuracy D/A Converters

The accuracy of the D/A converter is primarily a function of the accuracy of precision resistors used in the ladder and the precision of the reference voltage

supply used. Accuracy is a measure of how close the actual output voltage is to the theoretical output value. It is specified as a percentage of the maximum output voltage. If the converter has accuracy of $\pm 0.5\%$ it will cause an error of 0.5x10/100 or $\pm$ 50m for a converter having 10V full scale output voltage.

24.7 Self Learning Exercise-I

Q.1 Calculate the value of the LSB for an 8 bit digital to analog converter for 0 to 10V analog output range.

Q.2 Calculate the value of the MSB for an 8 bit D to A converter for 0 to 10V analog output range.

Q.3 Calculate the value of the full scale output for an 8-bit DAC for 0 to 10V analog output range.

Q.4 Find the output voltage of a 4-bit ladder, having the following digital inputs 0110 and $V_m = 10V$.

Q.5 What is the value of LSB?

Q.6 What is the formula of full scale output?

24.8 Analog to Digital (A/D) Converters

The A/D conversions a quantizing process whereby an analog signal is represented by equivalent binary states; this is opposite to the D/A conversion process. Analog-to-Digital converters can be classified into two general groups based on the conversion technique. One technique involves comparing a given analog signal with the internally generated equivalent signal. This group includes successive-approximation, counter, and flash-type converters. The second technique involves changing an analog signal into time or frequency and comparing these new parameters to known values. This group includes integrator converters and voltage-to-frequency converters. The trade-off between the two techniques is based on accuracy vs. speed. The successive approximation and the flash type are faster but generally less accurate than the integrator and the voltage-to-frequency type converters. Furthermore, the flash type is expensive and difficult to design for high accuracy.

The successive-approximation A/D converters are used in applications such as data loggers and instrumentation, where conversion speed is important. On the other

hand, integrating-type converters are used in applications such as digital meters, panel meters, and monitoring systems, where the conversion accuracy is critical. The most commonly used A/D converters the successive-approximation- is discussed in this section with several interfacing examples.

While all analog-to-digital converters are classified by their resolution or number of bits, how the A/D circuitry achieves this resolution varies from device to device. There are four primary types of A/D converters used for industrial and laboratory applications- successive approximation, flash /parallel, integrating, and ramp/counting. Some are optimized for speed, others for economy, and others for a compromise among competing priorities. Industrial and lab data acquisition tasks typically require 12 to 16 bits- 12 is the most common. As a rule, increasing resolution results in higher costs and slower conversion speed.

24.9 Accuracy A/D Converters

Accuracy is one of the most critical factors to consider when specifying an A/D converter for test and measurement applications. Unfortunately it is often confused with resolution, although related, they are distinctly different.

Since the A/D converter is a closed loop system involving both analog and digital systems, the overall accuracy must include errors from both the analog and digital positions. In determining the overall accuracy it is essential to separate the two sources of error. An analog to digital converter has several sources of errors. Quantization error and (assuming the A/D converter is intended to be linear) non-linearity is intrinsic to any analog-to-digital conversion. These error are measured in a unit called the LSB, (Least Significant Bit).

Quantization error is due to the finite resolution of the A/D converter, and is an unavoidable imperfection in all types of A/D converters. The magnitude of the quantization error at the sampling instant is between zero and half of one LSB. At lower levels the quantizing error becomes dependant of the input signal, resulting in distortion. This distortion is created after the anti-aliasing filter, and if these distortions are above 1/2 the sample rate they will alias back into the audio band. In order to make the quantizing error independent of the input signal, noise with amplitude of 1 quantization step is added to the signal. This slightly reduces signal to noise ratio, but completely eliminate the distortion.

24.10 Resolution A/D Converters

The resolution of the converter indicates the number of discrete values. It can produce over the range of analog values. The values are usually stored electronically in binary form, so the resolution is usually expressed in bits. In consequence, the number of discrete values available or "levels" is usually a power of two. For example, an A/D converter with a resolution of 8 bits can encode an analog input to one in 256 different levels, since $2^8 = 256$. The values can represent the ranges from 0 to 255 (i.e. unsigned integer) or from - 128 to 127 (i.e. signed integer), depending on the application.

Resolution can also be defined electrically, and expressed in volts. The voltage resolution of an A/D converter is equal to its overall voltage measurement range divided by the number of discrete intervals.

Example1 A 4- bit converter with V (1)= -1V, V (0) = 0V and $R_F = 8R$, obtain the analog output voltage for a digital input (i) 0010, (ii) 1010.

Sol. (i) $$V_0 = -(-1)\left[\frac{8R}{R}(0)+\frac{8R}{2R}(0)+\frac{8R}{4R}(1)+\frac{8R}{8R}(0)\right]$$

$$= 1\,[0+0+2+0] = 2V$$

(ii) $$V_o = -(-1)\left[\frac{8R}{R}(1)+\frac{8R}{2R}(0)+\frac{8R}{4R}(1)+\frac{8R}{8R}(0)\right]$$

$$= 1\,[8+0+2+0] = 10V$$

Example 2 Find the output voltage for a 4-bit ladder, having the digital inputs.

(a) 1010 (b) 0110 (c) 1001 (d) Full-Scale Output

Assume logic 0 = 0V and logic 1 = 12V

Sol. $V_A = V_m\,[V_D{}^0\,2^0 + V_D{}^2\,2^1 + V_D{}^2\,2^2 + V_{D3}\,2^3]$

(a) Digital input 1010

$$V_A = \frac{12\left[1\times 2^0+0\times 2^1+1\times 2^2+0\times 2^3\right]}{2^4}$$

$$= \frac{12\left[1+0+4+0\right]}{16} = \frac{60}{16} = 3.75$$

(b) Digital input 0110

$$V_A = \frac{12\left[0\times 2^0 + 1\times 2^1 + 1\times 2^2 + 0\times 2^3\right]}{16}$$

$$= \frac{12\left[0+2+4+0\right]}{16} = \frac{72}{16} = 4.5$$

(c) Digital input 1001

$$V_A = \frac{12\left[1\times 2^0 + 0\times 2^1 + 0\times 2^2 + 1\times 2^3\right]}{16}$$

$$V_A = \frac{12\left[1+0+0+8\right]}{16} = \frac{108}{16} = 6.75$$

(d) Digital input 1111

$$V_A = \frac{12\left[1\times 2^0 + 1\times 2^1 + 1\times 2^2 + 1\times 2^3\right]}{16}$$

$$V_A = \frac{12\left[1+2+4+8\right]}{16} = 11.22$$

Example 3

Calculate the values of LSB, MSB and full scale output for an 8 bit DAC for 0-12 V range.

Sol. Given data: n = 8, V_m = 12 V

$$\text{Value of LSB} = \frac{V_m}{2^n} = \frac{12}{2^8}$$

$$= \frac{12}{256} = 0.046\text{V}$$

$$\text{Value of MSB} = \frac{2^{n-1}.V_m}{2^n} = \frac{2^{8-1}.12}{2^8}$$

$$= \frac{2^7.12}{2^8} = 6\text{V}$$

Full Scale output = V_m - value of LSB

= 12-0.046

= 11.954 V

24.11 Self Learning Exercise- II

Q.1 What is Accuracy?

Q.2 What is a range of input voltage?

Q.3 Write the formula of resolution?

Q.4 Define Resolution.

Q.5 How many number of comparators in a parallel conversion type 8-bit A to D converters?

Q.6 How many number of comparisons carried out in a 4 bit flash type A to D converters?

24.12 Summary

D/A converters are readily obtainable commercial products. Converters are available with 6-, 8-, 10- and 12- bit resolution.

It is important to realize that resolution and accuracy in a system should be compatible.

Accuracy is a measure of how close the actual output voltage is to the theoretical output value. The accuracy of the D/A converter is primarily a function of the accuracy of the resistors used in the ladder and the precision of the reference voltage supply used resolution is defined as the smallest increments in voltage that can be discerned. Resolution is primarily a function of the number of bits in the digital input signal. That is, the smallest in output voltage is determined by the LSB.

The process of converting an analog input voltage to a number of equivalent digital output levels is known as A/D conversion. The counter type A/D converter is somewhat slower but represents a much more reasonable solution for digitizing high resolution signals. The successive approximation method is all variations of the basic counter type A/D converter which lead to a much faster conversion time. A/D converters are used virtually everywhere where an analog signal has to be processed, stored, or transported in digital form. Fast video A/D converters are used in TV tuner cards slow on-chip 8, 10, 12 or 16 bit A/D converters are common in micro controllers. Very fast A/D converters are needed in digital oscilloscopes, and are crucial for new applications like software defined ratio.

24.13 Glossary

Binary coding: The binary code ranges from all zeros to the positive full-scale value of all ones.

Resolution: ADC resolution is the number of bits used to represents the analog input signal.

Accuracy: Correctness of a single measurement.

24.14 Answers to Self Learning Exercises

Answers to Self Learning Exercise-I

Ans.1: 0.039 V

Ans.2: 5V

Ans.3: 9.961V

Ans.4: 3.75

Ans.5: $\frac{V_m}{2^n}$

Ans.6: Full scale output = V_m - value of LSB

Answers to Self Learning Exercise-II

Ans.1: It measures how closed the actual output voltage is to the theoretical output voltage. The accuracy depends on the quantization level. The accuracy of higher bits is more as compared to lower bit.

Ans.2: The range of the input voltage is defined by the maximum and minimum amplitude of an analog signal, which can be applied to the A to D converters for reliable operation.

Ans.3: Resolution = $V/ (2^n - 2)$

Ans.4: Resolution is defined as the voltage input change necessary for a 1-bit change in out put. It is also defined as voltage is the full-scale input voltage divided by the total number of levels.

Ans.5: 255

Ans.6: 15

24.15 Exercise

Section A: Very Short Answer Type Questions

Q.1 Explain the basic concepts of an A to D converters.

Q.2 Draw and explain the circuit of R-2R ladder D to A converters.

Q.3 Calculate the value of LSB, MSB and full-scale output for an 8-bit D to A converters for 0-12V range.

Q.4 Find the output voltage for a 4-bit ladder, having the digital inputs is 1111.

Assume logic 0 = 0v and logic 1 = 12V

Section-B :Short Answer Type Questions

Q.5 Describe how to find the output voltage of a resistive divider?

Q.6 Why is a D to A converter usually considered as a decoder?

Q.7 What are some of the advantages of the binary ladder over the resistive ladder?

Q.8 What is the difference between resolution and accuracy?

Section-C: Long Answer Type Questions

Q.9 How does flash A to D converter operates explain?

Q.10 Describe the operation of a counter-type A to D converter.

Q.11 Explain Binary ladder as R-2R ladder network.

Q.12 Explain Variable-Resistor Network for a D to A Conversion.

References and Suggested Readings

1. G. K. Kharate, Digital Electronics, Oxford University Press, 2013.
2. Gupta and Kumar, Hand Book of Electronics, Pragati Prakashan, 2008.
3. Y. N. Bapat, Electronic Circuits and system, Tata M.C. Graw hill publishing company, 2013.
4. R. S. Sedha. Applied Electronics, S. Chand and Company Ltd., 2009.