UNIT - I

SEMICONDUCTOR PHYSICS

1. BAND THEORY

a. Introduction

Based on the electrical conductivity, materials can be divided into three types. They are

(i) conductors (ii) insulators and (iii) semiconductors.

Those material which have plenty of free electrons and in which these electrons can move easily from one atom to another are called **conductors**. All metals are good conductors (ex) silver, copper etc. The resistivity of a good conductor at room temperature will be of the order of 1.7×10^{-8} ohm-m.

Those material which have few free electrons and in which electrons are tightly bound to the nucleus are called **insulators**. Examples are: glass, mica, plastic, rubber, air etc. The resistivity of an insulator will be of the order of 7.5×10^{11} ohm-m.

Those material whose resistivity lies between that of a good conductor and insulators are called **semiconductors**. Examples are germanium, silicon etc. The resistivity of a semiconductor will be of the order of 10×10^{-2} to 10.0×10^{-3} ohm-m.

There is some difference in the electrical properties of a semiconductor and a conductor. When the temperature of a conductor increases, its resistivity also increases. But in a semiconductor, the resistivity decreases as the temperature increases. Hence conductors have a positive temperature coefficient of resistance, but semiconductors have negative temperature of coefficient of resistance. Moreover, the electrical conductivity of a semiconductor is very much affected by even a very minute amount of impurity added.

The existence of electron energy band provides an explanation of electrical conduction in the solid state. There are two ways to consider how energy bands arise. One way is based on the wave mechanical treatment of the motion of electrons in a periodic field of positive ion cores in the solid. There is also another way of developing the concept of energy band that uses the concepts of molecular orbital theory. We will discuss the second way.

b. Theory of Energy Bands in Crystals.

In an isolated atom, each electron is characterized by a set of four quantum numbers n, *l*, m and s. According to the Pauli's exclusion principle, no two electrons in any particular atom can have the same set of four quantum numbers. So an energy state defined by a set of three quantum numbers (n, *l* and m) can be occupied by only two electrons. These electrons differ in the fourth quantum number + or – spin.

[illegible] → தனித்தநிலைகளுடைய

In an isolated atom, these energy states are discrete. This idea of an electron possessing discrete energy holds good in the case of gas. In gases the different atoms are separated sufficiently from one another. So they have got very little influence upon one another.

In a crystalline solids, the atoms are situated close to one another. Therefore, the discrete energy possessed by an electron in a single free atom, may not be possessed by the electron in the same atom inside a crystal. But the inner electrons are not very much affected by the presence of neighbouring atom. So the energy level of the inner electrons will be changed only by a negligible amount. But the outer valence electrons take part in the chemical bonds. Hence the energy of these electrons are greatly affected.

Consider, for example, two hydrogen atoms. Each atom has a single electron in the 1s energy state. When the two atoms are brought together, according to the molecular orbital theory, a more stable molecular orbitals and a less stable molecular orbitals are formed. Thus when two atoms are brought together, for each energy state of the isolated atom, a more stable state and a less stable states are formed. If N atoms are brought together, N new energy states are formed for each energy state in an isolated atom.

In the case of crystals, N is very large of the order of 10^{23} per cm^3. Therefore, the energy levels are also very large and they are very close to each other. The large number of discrete, but very close energy sublevels is an energy band. The energy band occupied by the valence electrons is called the **valence band**. It may be either completely filled or partially filled with electrons but can never be empty.

When the temperature of the crystal is raised or an electric field is applied to it, a valence electron in the valence band may acquire sufficient energy. So it will leave the valence band. These electrons which have left the valence band are called conduction electrons. The band occupied by these electrons are called the **conduction band**. This band lies next to the valence band. It may either be empty or partially filled with electron. In fact, it may be defined as the lowest unfilled energy band in an atom. In conduction band electrons move freely and conduct electric current through the solid. That is why they are called conduction electrons.

The valence band and the conduction band are separated by a gap known as **forbidden energy gap**. In this region there is no allowed state for an electron. Therefore, electrons are never found in this forbidden gap. An electron may move from the valence band to conduction band and vice versa. The width of the forbidden energy gap is the energy difference between the bottom of the conduction band and the top of the valence band.

c. Distinction between conductors, Insulators and semiconductors

The difference between conductors, semiconductors and insulators may be explained on the basis of energy gap between the valence and conduction band. Fig 1-1 shows the energy gap diagram of (i) an insulator (ii) a semiconductor and (iii) a conductor.

(i) Insulators : The energy band structure of an insulator indicated schematically in Fig 1-1 (a). In this, the forbidden energy gap E_g is very large. In general, it is more than 3 ev. For a diamond (carbon), the energy gap E_g = 6 ev. This large forbidden band separates the filled valence band from the vacant conduction band. Hence a very large amount of energy must be supplied to a valence electron to enable it to move to the conduction band. An energy of this magnitude cannot be supplied to an electron in a crystal. Hence in diamond, conduction is impossible and it is an insulator. For the insulators to conduct electricity a large electric field of the order of 10^8 V-m^{-1} is necessary to shift the electrons from the valence band to the conduction band.

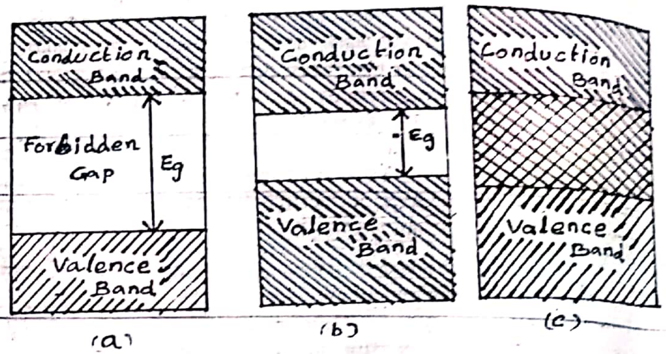


Fig 1-1

(ii) Semiconductor : A substance for which the width of the forbidden energy gap is relatively small (~ 1 ev) is called a semiconductors. The most important practical semiconductor material are germanium and silicon. At O K, the energy gap E_g for Ge is 0.785 ev and for Si is 1.21 ev. Energies of this magnitude cannot be acquired from an applied field. Hence the valence band remains full and the conduction band is empty. Therefore at low temperature the material are insulators. At higher temperature, the conductivity of semiconductors increases. These substances are known as **intrinsic semiconductor.**

When the temperature is increased, some of the valence electrons acquire thermal energy greater than E_g. So these electrons move from valence band to conduction band. These are now free electrons. So they can move about under the influence of even a small applied field. The insulator has now become slightly conducting; it is a semiconductor. The absence of an electron in the valence band is called a **hole.** The hole behaves as a positively charged carriers of electricity. Thus as electron move to the conduction band, an equal number of holes is created in the valence band. In the process of conduction, under the action of an electric field electrons move from one hole to the next. Thus Ge and Si, which are insulators at low temperature, becomes slightly conducting when the temperature is increased. The conductivity increases with the temperature.

(iii) Conductors : The band structure of a crystal may contain no forbidden energy gap. So the valence band merges into an empty band as shown in Fig 1-1 (c). Under the influence of an applied electric field, the electron may acquire additional energy and move into higher energy state. Since these mobile electrons constitute a current. Since these mobile electron constitute a current, this substance is a conductor. The empty region is the conduction band. A metal is characterized by a band structure containing overlapping valence and conduction bands.

2. SEMICONDUCTORS

Semiconductors can be classified into two types. They are,

i) Intrinsic or pure semiconductors. and

ii) Extrinsic or impure semiconductors.

Extrinsic semiconductors are of two types. They are

i) N-type semiconductor and

ii) P-type semiconductor.

a. Intrinsic Semiconductors

When the conductivity of a semiconductor is solely determined by the thermally generated carriers, the semiconductor is called an **intrinsic semiconductors.** When a pure semiconductor is kept at O K, the valence band is completely filled and the conduction band is completely empty. Therefore at O K, a pure semiconductor behaves as an insulator. On the otherhand, when a pure semiconductor is kept at room temperature, some of the valence electron gain sufficient energy, jump into the conduction band and becomes free. Thus free electrons are produced. The vacancies formed in the valance band when some of the valence electron jump into the conduction band are called holes. A hole carries a positive charge equal in magnitude to the electron charge. The number of thermally generated electron always equal to the number of holes.

When an external voltage is applied to the semiconductor, the electrons and holes move in the opposite directions. The electron move towards the positive electrode while the holes move towards the negative electrode.

Germanium and silicon are the two important semiconductors. These substances belong to the IV group of the periodic table and have four electrons in the outermost shell.

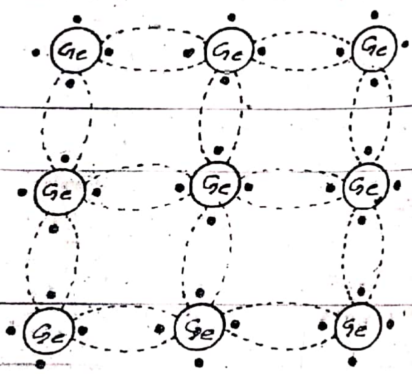


Fig 1-2

Each of the four outer electron forms a bond with one electron of the nearest neighbouring atoms. The covalent bond is represented in Fig1-2 by two dashed line which joins each atom to each of its neighbours. At absolute zero temperature intrinsic semiconductors behaves as a perfect insulator because at this temperature there is no free electron, to act as a charge carriers. This is so, because all the

bonds of the molecules are intact as shown in Fig1-2. So there is no free electron for the conduction. So the semiconductor behaves as an insulator.

As temperature increases the thermal energy is sufficient to break more and more covalent bonds and set free the electrons. The absence of the electron is called a hole. A hole has a positive charge and hence behave exactly opposite to that of an electron. So in a semiconductor for every covalent bond that is broken, we get two carriers of electricity. The electron and holes are as shown in Fig1-3.

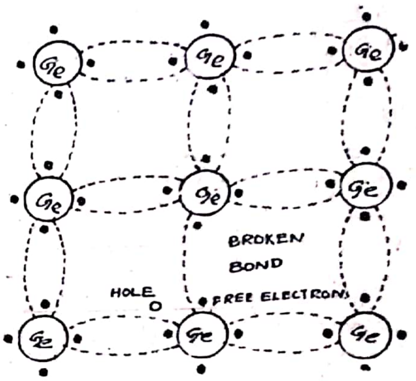


Fig1-3.

Thus at room temperature, a semiconductor has two types of charge carriers (i) the electron (–ve) and (ii) the hole (+ ve). If a d.c voltage is applied across a silicon or germanium at room temperature, the electrons will move to the positive terminal and the hole towards the negative terminal. In an intrinsic semiconductor, the number of free electron is equal to the number of holes. The conductivity increases with the increase in temperature. In other words, the resistivity decreases with the increase in temperature. The pure or intrinsic semiconductor is not of much use because its conductivity is a function of temperature and cannot be controlled otherwise.

b. Extrinsic Semiconductor

If a minute amount of impurity is added to a pure semiconductor, its current conducting property is altered appreciable. Semiconductors belong to the group IV of the periodic table. The impurities which are added to the semiconductor are elements belonging to the III and IV group of the periodic table. There are two types of extrinsic semiconductors. They are (i) **N-type semiconductor** and (ii)**P-type semiconductor**. If the electrons are excess, it is called N-type semiconductor and if the holes are in excess, it is called the P-type.

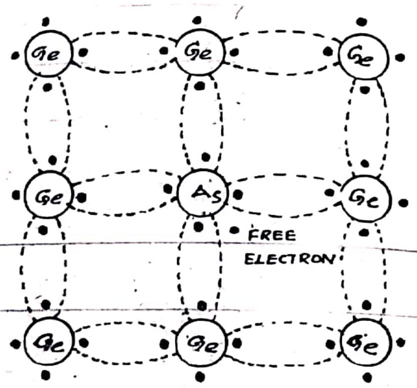


Fig 1-4

(i) N-type Semiconductors: Semiconductors generally belong to the IV group of the periodic table. Hence it will have 4 electrons in the outermost orbit. Let us consider a small amount of impurity such as arsenic or antimony which belong to the V group of the periodic table is doped in the germanium atom. The amount of impurity added is extremely small say 1 or 2 atoms of impurity for 10^6 germanium atom. The impurity has five valence electrons while germanium has four. Four of the five valence electrons of the impurity atom form covalent bonds with neighbouring four equidistance germanium atom. For the fifth electron of the impurity, there is no place is the covalent bond. Hence this extra electron is free to more. This situation is

represented in the Fig 1-4. Due to dopping of impurity, a hole is not created with the liberation of a free electron as in the case of intrinsic semiconductor. Also the overall crystal structure is not altered by the addition of the pentavalent impurity atom. For every pentavalent impurity atom that is added, one extra carrier is provided in the crystal. In this one free electron is donated by the impurity and hence the impurity is called **donor impurity**.

If we apply a voltage across this germanium crystal, the free electrons will move towards the +ve polarity and constitute an electrical current. Hence these material is called the **N-type semiconductor**. In this, the number of electrons will outnumber the hole and hence the current is mainly due to the electron. Hence in N-type semiconductor, the **electron** are the **majority carrier** of charge and the **hole** is called the **minority** carrier of charge.

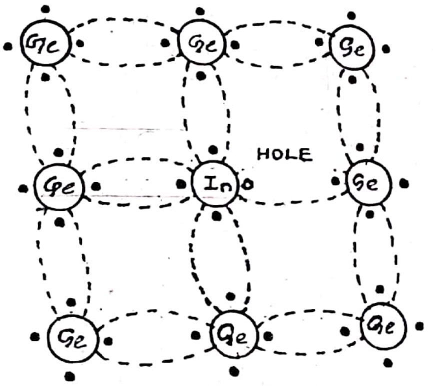


Fig 1-5

(ii) P-type Semiconductor: When the impurity element of group III like indium, gallium or aluminium is added to a semiconductor, there is an opposite effect upon the lattice structure. Now the impurity indium with three valence electron in each atom takes part in the covalent bond structure of germanium. The fourth bond is left incomplete or broken as shown in Fig 1- 5. The broken covalent bond constitutes the hole in the semiconductor sense. This hole will

be free to move about the crystal;. The impurity atom with three valence electron will readily accept an electron. Hence this impurity is called the **acceptor impurity**.

In P-type semiconductor, the current is due to hole (+ ve charge) and hence it is P-type semiconductor. In addition to the holes created by the impurity atoms, there will be electron-hole pairs due to thermal agitation. In the P-type semiconductor the holes outnumber the electron. Hence in this holes are called the majority carrier of charges and electrons are called the minority carrier of charge.

c. Hall Effect in Semiconductors

Hall, in 1897, discovered that when a magnetic field is applied at right angle to the direction of the current flow in a conductor, an electric field is developed in the direction perpendicular to the direction of the current and of the applied magnetic field. This phenomenon is known as **Hall effect**. Using the phenomenon of Hall effect, carrier concentration and carrier mobility can be experimentally found.

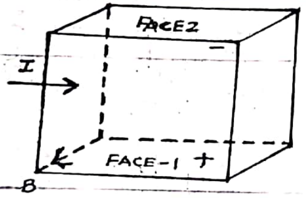


Fig 1-6

Consider a rectangular slab as shown in Fig 1-6. A current density J_x flows in the X- direction and the magnetic induction B_z is applied in the Z - direction and a potential difference E_y develops along the Y-direction. Using a sensitive voltmeter, the p.d developed between the faces 1 and 2 can be measured.

When a steady magnetic field B_Z is applied in the Z-direction, each charge in the current stream experiences the Lorentz force given by the equation $F = -e(B \times v)$. The component of the force along the Y-direction is

$$F_y = -e\,v_x B_z \qquad \text{......(1)}$$

This force is upward if the charge is + ve and downward if the charge is −ve. In our cases, the current is in the +ve direction and hence the electron motion must be in the −ve direction. So v_x is negative. The current is cased only by the electrons and so the charge is negative.

$$\therefore \text{ Lorentz force } F_y = -ev_x B_z \qquad \text{.......(2)}$$

Due to the Lorentz force, the electrons are pushed up. So the top edge acquires a −ve charge and the bottom edge acquire a +ve charge. Hence an electric field is setup in the −ve Y-direction. This voltage is called the **Hall Voltage** and the phenomenon is called **Hall effect**.

The Hall field acting upwards exerts a downward force on the electrons which are pushed up by the Lorentz force. In the equilibrium position, the force on the electron due to Hall field is exactly equal and opposite to the force on it due to the Lorentz force. If E_y is the Hall electric field set up in the −ve Y-direction, then Hall force acting on the electron is eE_y. At the equilibrium position.

Lorentz force = Hall force

$$\therefore \quad B_z e v_x = eE_y \qquad \text{.......(3)}$$

$$\therefore \quad \text{Hall field } E_y = B_Z v_x \qquad \text{.......(4)}$$

$$\text{But the current density } J_x = nev_x \qquad \text{.......(5)}$$

$$\therefore \quad v_x = \frac{J_x}{ne} \qquad \text{.......(6)}$$

$$\therefore \quad E_y = \frac{B_z J_x}{ne} \qquad \text{.......(7)}$$

The ratio $\frac{E_y}{B_z J_x}$ is called the **Hall constant** and is denoted by R_H.

$$\therefore \quad R_H = \frac{E_y}{B_z J_x} = \frac{1}{ne} \qquad \text{......(8)}$$

Since all the three quantities E, B and J can be accurately determined, the Hall coefficient R_H and the carrier concentration n can be calculated. In the case of metals, the carriers are electrons and so E_y will be negative.

In the case of semiconductors, there are two types of charge carriers. In the N-type semiconductors, the majority carriers are the electrons and so the Hall coefficient is –ve. The Hall coefficient R_n for an electron is defined by the relation.

$$R_n = \frac{E_y}{B_z J_x} \qquad \text{......(9)}$$

If n is the density of electron in the N-type material, then

$$R_n = -\frac{1}{ne} \qquad \text{......(10)}$$

In the case of P-type semiconductor, the majority carrier are the hole and so the Hall coefficient is + ve. The Hall coefficient of hole R_p is defined by the relation.

$$R_p = \frac{E_y}{B_z J_x} \qquad \text{......(11)}$$

if p is the density of holes in the P-type material, then $J_x = pev_x$

$$\therefore \quad R_p = \frac{1}{pe} \qquad \text{......(12)}$$

3. JUNCTION DIODES

a. P-N Junction Diode

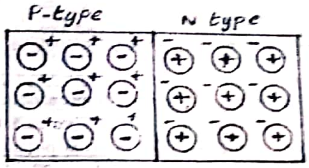


Fig 1-7

When a P-type semiconductor is in intimate contact with an N-type semiconductor, a P-N junction is formed. The circles represent the fixed impurity atoms in the two types with their charges marked on them. The sign represented outside the circle represent the free electrons and the holes. The P-type material contains a large number of holes and N-type contains large number of electrons.

When the P-N junction is formed, some of the electrons tend to diffuse towards the P-type material and holes will diffuse towards the N-type material. In this process, the electron crossing the junction from N-region into P-region recombine with the holes in the P-region. Similarly the holes crossing the junction, recombine with electrons in the N-region. Due to this recombination of electrons and holes near the junction, the +ve donor and –ve. acceptors are left uncompensated in the vicinity of the junction. This region does not contain mobile charges. This region is called the **depletion region**. Thus an internal potential difference is produced across this junction. This p.d is called the **internal potential barrier**. In circuits, the P-N junction diode is represented as shown in Fig 1-8.

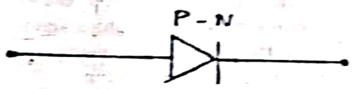


Fig. 1.8.

For a P-N junction diode, two types of connections are given. They are (i) Forward bias and (ii) Reverse bias.

(i) Forward Bias: A potential is applied across the P–N junction. The +ve pole is connected to the P-region and the –ve pole is connected to N-region. So the holes in the P-region are repelled towards the junction and the electron in the N–region are repelled towards the junction. Thus there is a great deal of concentration of charges at the junction crossing over from the +ve to –ve and vice versa. If the applied potential is greater than the internal potential barrier, there is a considerable flow of current due to this. This type of connection given to the P-N junction is said to be **Forward biased**. When it is forward biased, it offers less resistance for the flow of current.

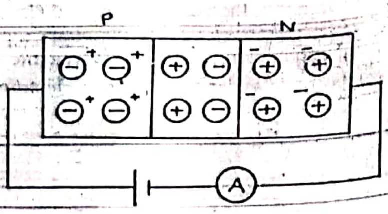


Fig. 1.9.

(ii) Reverse Bias: In this the negative terminal of the battery is connected to P-type and the positive terminal is connected to the N-type. This type of connection is called Reverse biased. In this the holes in P-region are attracted to the –ve of the battery terminal and moves away from the junction while the electrons in the N–region also move away from the junction because of the attraction of the +ve terminal. Since there are effectively no hole and electron carriers in the vicinity of the junction, current flow stops almost. There may be a very small current of the order of microampere. This is due to minority carrier of opposite charges present in the P and N- type areas. Hence when it is reverse biased, it offers a very high resistance to the flow of current.

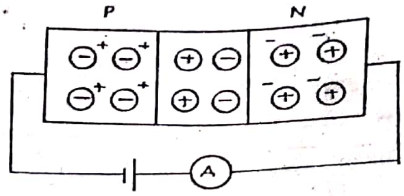


Fig. 1.10.

b. Breakdown in P–N junction

When a reverse-bias is given to a junction diode, a very small current flows through the diode. When the reverse bias voltage is increased, the current increases slowly. At a particular reverse voltage, the diode current begins to increase sharply. This particular voltage is called the **break-down Voltage.** At this voltage, if the current through the diode is not controlled, the diode may burn out. When the product of d.c voltage and the current exceeds the maximum power rating of the diode, the diode is also burn out. The breakdown phenomenon in a P-N junction is caused by either of the two effect **Avalanche** or **Zener.**

Avalanche Breakdown: When a reverse-bias is given to a diode, a very small current flows through it. This current is due to the movement of the minority carrier of charge. When the reverse bias voltage is increased, the kinetic energy of the electron also increases. When these electrons collide with a crystal atom, they may knock out electron from the covalent bonds of the semi-conductor material. In this process the covalent bond is broken and a pair of a electron and a hole is generated. Thus one electron, on collision with a crystal atom, generates a pair of an electron and a hole. In this process, the number of free electrons and holes goes on increasing. This cumulative effect is known as **avalanche multiplication.** Due to this a large reverse current flows through the junction. Now the diode is said to be in the avalanche breakdown region. If the current is not controlled by some external resistance, the diode may be permanently damaged.

Zener Breakdown: When the reverse bias Voltage across the P-N junction is increased, the electric field across the junction also increases. Hence the reverse current also increases. At a certain reverse voltage below 6V, the reverse current increases very rapidly and the junction breaks down. This breakdown at such a low range voltage cannot be due avalanche multiplication of electron-hole pair. In this, there is no collision of minority carrier with the semiconductor atoms. This breakdown can be explained on the basis of zener effect.

When the electric field is high, a large force is exerted on the valance electron of atoms in the depletion region. So a large number of electron-hole pairs are produced. These carriers are accelerated away from the junction by the applied voltage. Hence the reverse current increases rapidly. This process by which covalent bond in the depletion regions are directly broken by a strong electric field is zener breakdown and the reverse voltage at which the breakdown takes place is called the zener breakdown voltage. This effect is called zener effect.

c. Zener Diode

When a P-N junction is reverse biased, it offers a high resistance for the current flow. If the reverse-bias is increased at some particular voltage, the reverse current across the junction suddenly increases. This particular potential is called the breakdown voltage or zener breakdown voltage. The zener voltage varies from as low as a few volt to several hundred depending on the dopant density and the depletion layer. There are two distinct process by which the breakdown may occur. They are (i) zener breakdown and (ii) avalanche breakdown. A zener diode is a P-N junction diode which make use of either of the two breakdown. A zener diode is usually operated at a reverse bias voltage a little more than the breakdown voltage. Under these conditions, the voltage drop across the diode is practically independent of the current through it. This means that the diode acts as a voltage regulator. The zener diode in circuits is represented as shown in Fig 1-11.

Fig. 1-11

Zener Diode Characteristic

The Zener diode characteristic can be studied using the circuit shown in Fig 1-12. The zener diode is connected to a battery through a commutator as shown in fig. By changing the commutator position, the diode can be forward biased or reverse biased.

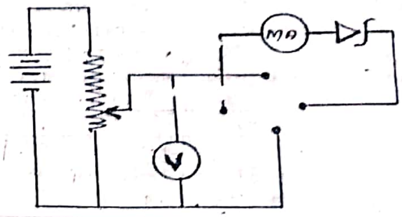


Fig. 1-12

First forward bias is given. Using the rheostat different voltage is given to the diode and for each voltage, the milliammeter reading is noted. Now a reverse bias is given to the diode and the milliammeter is replaced by a microammeter. The same procedure is repeated as for the forward bias.

Now we can draw a graph taking the voltmeter reading in X- axis and the current in Y-axis. The graph will be as shown in Fig. 1-13. Under forward bias condition, zener diode acts just like an ordinary junction diode.

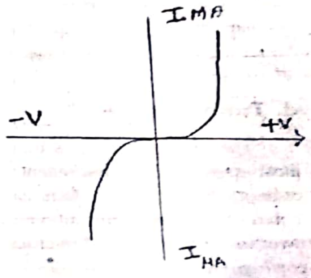


Fig 1-13

Under reverse bias, a small reverse current flows through it. This current almost remains constant until a certain critical voltage is reached. Beyond this voltage, the reverse current increases rapidly. This voltage is called **turnover voltage**. All zener diodes are silicon P-N junction diodes which have a sharp reverse voltage knee. Zener diodes are assorted according to their breakdown voltage. For a diode with low breakdown voltage, the knee on the curve is more obtuse. For high voltage zener diode, the knee is sharp.

The zener voltage $-V_z$ is defined as the voltage in the breakdown region for the average value of the zener current. When an a.c signal is superimposed, the slope of the curve can be measured at the average of the zener current $-I_z$. The zener resistance is defined as the ratio between the incremental voltage and the current.

$$r_Z = \frac{\delta V_Z}{\delta I_Z}$$

The Zener resistance has got a minimum value of approximately 10 ohm at about 6 volts.

Uses: Zener diodes are used

1) as voltage regulator
2) as a fixed reference voltage in a network for biasing and comparison purposses and for calibrating voltmeter.
3) as peak clippers
4) for meter protection against damage from accidental application of excessive voltage.

d. Tunnel Diode

The tunnel diode was invented by Leo Esaki and is one of the most significant development in semiconductor electronics. An ordinary P-N junction diode has an impurity concentration of about 1 part in 10^8. With this amount of doping, the width of the depletion layer is of the order of 5 microns (5×10^{-4} cm). The potential barrier corresponding to this width of depletion layer restrains the flow of hole from P to N region and electrons from N to P region. If the concentration of impurity atom is 1 part in 10^3, the depletion layer will reduce to about 100 A°. The shape of the volt- ampere characteristic curve of this device is quite different from that of a normal P-N junction diode. This diode utilizes the phenomenon called tunnelling and hence this diode is referred to as **tunnel diode** or **Easki diode**. The circuit symbol for a tunnel diode is as shown in Fig 1-14.

Fig. 1-14

Tunnel diode cannot rectify alternating voltage and exhibits negative resistance properties. Though it is a diode, it can function as an amplifier, an oscillator or a switch. As a switch it can operate in a time less than 10^{-9} second. Tunnel diodes have been constructed from germanium, silicon, gallium arsenide and indium antimonide.

Tunnel Diode Characteristics

The Volt-ampere characteristic of a tunnel diode is as shown in Fig. 1-15. From the fig, it is clear that the tunnel diode is an excellent conductor in the reverse direction. For a small forward voltage upto 50 mv for Ge, the resistance remains small of the order of 5 ohm.

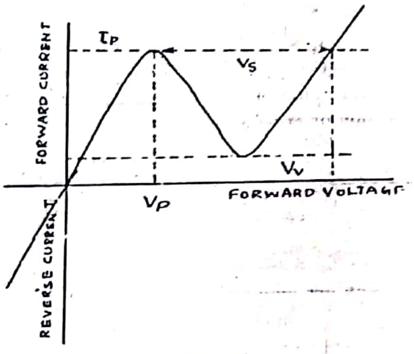


Fig. 1.15.

If the forward voltage is increased, the current increased and reaches a **peak current** I_p, corresponding to the voltage V_p. At this the slope dI/dV of the characteristic is zero. If the voltage is increased further, then the current decreases. In this region, the dynamic conductance is negative. Thus the tunnel diode exhibits negative resistance characteristic between the peak current I_p and the minimum value I_v called the **Valley current**. At the **valley voltage** at which $I = I_v$, the conductance (dI/dV) is again zero. Now if the forward voltage is increased, the current increases beyond the value V_v. Again the current reaches the value I_p at the voltage V_p. This is called **peak forward voltage**. The voltage swing V_s is defined as the voltage between two points on the tunnel diode characteristic current where the current is equal to I_p. The value of V_s depends on the material of the tunnel diode.

For currents whose value are between I_v and I_p, the curve is triple valued. It is because each current can be obtained at three different applied voltages. Because of this property, the tunnel diode can be used is pulse and digital circuits.

With the aid of energy-band and the concept of quantum mechanical tunnelling, the tunnel diode characteristic can be explained. In a tunnel diode the density of the impurity is of the order of 10^{19} per cubic centimeter. Hence, at room temperature, in the N-region the Fermi level shifts to conduction band. Similarly the Fermi level lies within the valence band of the P-region. When both P and N regions of the tunnel diode have the same temperature and the tunnel diode is unbiased, the Fermi level on both side of the junction will be the same. Thus there is a partial overlapping between conduction band of the N-type region and the valence band of the P-type region. At zero voltage bias condition, there will be tunneling of electron in either direction. The net current must be zero since the current in both directions cancel each other.

When the tunnel diode is forward biased, there are many filled states at the bottom of the conduction band of the N- region in line with unfilled state in the valence band of the P-type side. Hence electron will tunnel from N to the P-region. This gives rise to forward current. The current continues to rise with the applied voltage until the valence band of the P-type side and the conduction band of N-type side are maximally overlapped. If the forward bias voltage is increased further, some of the electrons in the conduction band on the N-side find themselves at the same energy as the forbidden energy band on the P-type. Hence the over lapping decreases. So the tunneling current also decreases. In this manner, the tunneling current continues to decrease as the forward bias voltage increases. When the amount of overlap becomes zero, the tunnel current becomes zero. If the forward bias is further increased, the conventional injected current of a normal P-N junction flows. This current occurs due to the injection of hole from P to N-region and electrons from N to P-region.

Uses: Tunnel diode is used for the following purposes

1) as a ultrahigh speed switch.
2) as a logic memory storage device
3) as microwave oscillator at a frequency of about 10 GH_Z
4) in relaxation oscillator circuit.

Advantage: The advantage of tunnel diodes are (1) low cost (2) low noise (3) high speed (4) environmental immunity and (5) low power.

Disadvantage: The disadvantages are (1) low output voltage swing and (2) it is a two terminal device. Because of this, there is no isolation between input and output and this leads to series circuit-design difficulties.

c. Backward Diode

Zener diodes normally have breakdown voltage greater than 2 volt. By increasing the doping level, we can get zener effect below 2 volt. In fact we can get breakdown near zero as shown in Fig 1-16. Forward conduction still occurs around 0.7 volt, but reverse conduction starts at about 0.1 volt or less. A diode with a curve like this in called a backward diode. The circuit symbol of a backward diode is as shown in Fig 1-17.

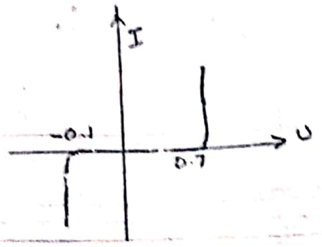


Fig. 1.16.

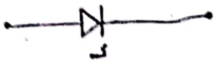


Fig. 1.17.

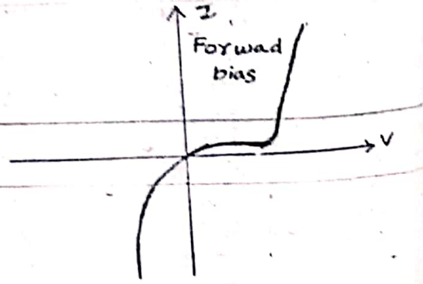


Fig. 1.18.

The volt-ampere characteristic of a backward diode is as shown in Fig 1-18. Because this device is a better conductor in the reverse than in the forward direction, it is called a backward diode or simple **back diode**. In the neighborhood of zero voltage, in response to either a forward biasing or reverse biasing voltage, the tunnel diode responds with a current which is large in comparison to the corresponding current in a conventional diode. These large currents are as a result of the tunneling effect. In the backward diode, the current due to tunneling is large only in the reverse direction. For this reason, the back diode is also called the **unitunnel diode**.

The temperature sensitivity of the back diode is appreciately less than the sensitivity of the ordinary diode. The backward diode has a sensitivity of about −0.1 mv/ C° compared with −2 mv/C° for the conventional diode. Moreover the conventional silicon diode has a break point, at room temperature between 0.6 and 0.7V, the back diode have a break point at 0V. The back diode is, therefore, very useful when the rectifying action is required in connection with small-amplitude wave form.

As an example, the Fig 1-19 shows a sine wave with a peak of 0.5V driving a backward diode. This 0.5 volt is not enough to forward bias the diode into conduction, but it is enough to break down the diode. For this reasons, the output wave has a peak of 0.4 volt (0.1 volt is lost across the diode)

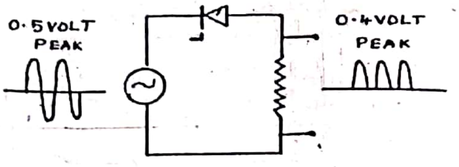


Fig. 1.19

MODEL QUESTION AND ANSWERS

I. Choose the correct Answer

1. Major part of the current in an intrinsic semiconductor is due to

(a) Conduction band electron

(b) holes in the valence band

(c) valence band electron

(d) thermally generated electron.

TRANSISTORS

1. Transistors

a. P-N-P and N-P-N Transistor

When a thin layer of N-type semiconductor is sandwitched between two P-type semiconductor, the result is in the formation of a P-N-P transistor. Similarly when a P-type area is sandwitched between two N-type semiconductor, we have a N-P-N transistor. P-N-P and N-P-N transistors and their circuit symbols are as shown in Fig 2-1.

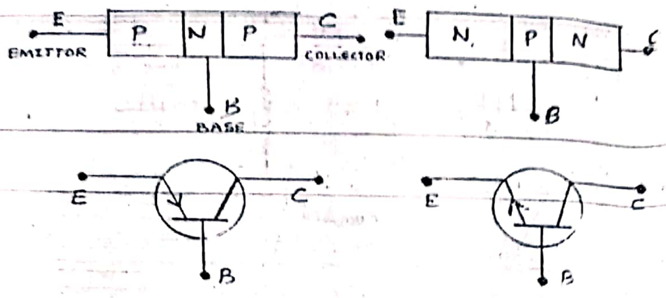


Fig 2-1

There are three regions in the transistor. They are (i) emitter (ii) collector and (c) base. The middle semiconductor in a transistor is called as base. The thickness of the base will be of the order 0.025 m.m. The two end regions are called the emitter and the collector. The emitter-base junction is always forward biased and hence this junction offers a low resistance for the flow of current. The collector-base junction is always reverse biased and hence this junction offers a high resistance for the current flow. The emitter,

base and collector of a transistor can be compared to cathode, grid and plate of a triode.

b. Transistor Action

The transistor action can be understood by analysing the current flow through it under the influence of externally applied voltage. The emitter-base junction will be forward biased and the collector-base junction will be reverse biased. We will study the action of a P-N-P transistor and the circuit is as shown in Fig 2-2.

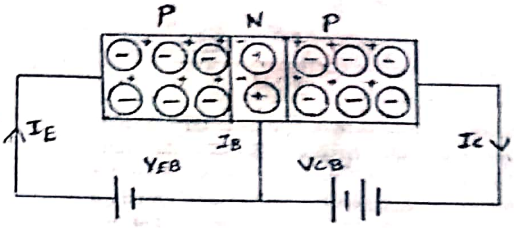


Fig. 2-2

As the emitter is forward biased, the holes in P-regions are injected to the base and the electrons are injected from N- region to P-region. But as the base is lightly doped and the emitter is heavily doped, the number of electron from base to the emitter is very small compared to the number of holes. In the base region, some of the holes are neutralised by the electron. Since the base region is very thin, most of holes will cross this region and will reach the collector region.

Since the collector-base region is reverse biased, the holes arriving at the junction attracted and this constitute a collector current. The emitter current is denoted by I_E, the collector current by I_c and the base current by I_B. It will be found that $I_c = \alpha I_E$ where α is called current gain whose value will be between 0.95 to 0.98. The emitter current is equal to the sum of the collector current and base current.

$$\therefore \quad I_E = I_C + I_b.$$

If the emitter-base voltage is increased, the collector will also increase. If the emitter current is decreased, the collector current will also decrease. If an alternating voltage is applied to the emitter as input, we will get amplified output in the collector circuit. Hence the transistor can be used as an amplifier.

In a similar way, we can explain the action of a P-N-P transistor.

c. Transistor circuit configuration

There are three basic circuit configuration in which transistor can be connected. They are

i) Common base configuration
ii) common emitter configuration and
iii) Common collector configuration.

The term common is used to denote the electrode that is common to the input and output circuits. Because the common electrode is generally grounded, these modes of operation are frequently referred to as grounded base, grounded emitter and grounded collector. The three type of configurations are as shown in Fig 2-3.

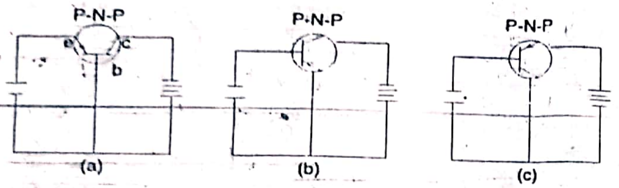


Fig 2-3.

The common base connection is as shown in Fig 2-3(a). In this, the input is applied between the emitter and the base and the output

is taken out between the collector and base. Thus the base is common to both input and output. So this connection is called common base connection.

For this connection.

1. Current gain is less than unity
2. Voltage gain is more than 100
3. Power gain is medium
4. input impedence is very low
5. Output impedence is very low
6. There is no phase change.

The common emitter connection is as shown in Fig. 2-3(b). In this the input is given between the base and the emitter and the output is taken across the collector and emitter. Thus the emitter is common to both the input and output. So this type of connection is called the common emitter connection. For this connection.

1. the current gain is medium about 50
2. the voltage gain is several hundred
3. Power gain is high
4. Input impedence is low
5. Output impedence is high and
6. there is a change of phase between input and output.

The common collector connection is as shown in Fig. 2-3(c). In this the collector is common to both input and output. Hence this configuration is called common collector connection. For this connection.

1. the current gain is medium
2. the voltage gain is less than unity
3. power gain is low
4. input impedence is very high
5. output impedence is very low and
6. there is no phase change between the input and output.

d. D.C Characteristic of a Transistor

The relation between different D.C. current and voltages of a transistor can be represented in a graph. These curves are called the static characteristic curves of a transistor. These curves are useful to determine the performance of a transistor. We will discuss these characteristic curves in detail for common base and common emitter configuration.

i. Common Base Characteristics

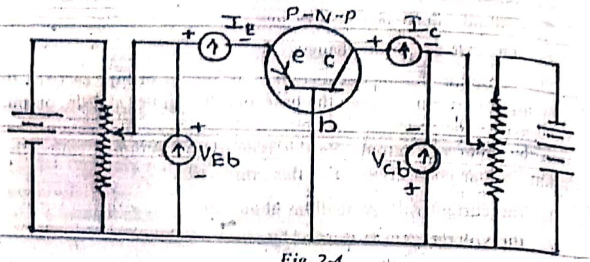


Fig. 2-4

The static characteristics of an P-N-P transistor connected in common base can be studied using the circuit arrangement shown in Fig 2-4. Using a potential divider arrangement, a forward bias is given between the emitter and base. The voltage between the emitter and base is measured using the voltmeter V_{EB} and the emitter current is measured using the milliammeter I_E. For the collector-base a reverse bias is given. The voltage between the collector and the base is measured using the voltmeter V_{CB} and the collector current is measured using the milliammeter I_C.

Input Characteristic: The Collector-base voltage V_{CB} is kept at a fixed Voltage. The emitter-base Voltage V_{EB} is varied step by step. For each voltage V_{EB}, the emitter current I_E and the collector current I_C is noted. The experiment can be repeated for various V_{CB}.

Now we can draw a curve taking the emitter current I_E is Y-axis and the emitter voltage V_{EB} is X-axis. The graph will be as shown

in Fig. 2- 5. This characteristic curve may be used to determine the input impedence.

Input impedence =

$$\left(\frac{\Delta V_{EB}}{\Delta I_E}\right)_{V_{CB}} \text{ ohm.}$$

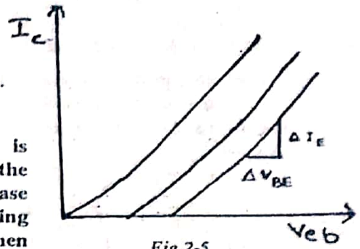


Fig 2-5

The input impedence is defined as the ratio between the small change in emitter-base voltage to the corresponding change in emitter current, when the collector-base voltage remains constant.

The reciprocal of the slope of the curve at a particular point gives the input impedence. It is measured in the unit of ohm.

Transfer Characteristic: When the collector to base voltage remains constant, we can show the relation between the input current I_E and the output current I_C. Taking I_E in X-axis and I_C in Y-axis, we can draw a graph and it will be as shown in Fig. 2-6. From the graph, we can determine the current amplification factor

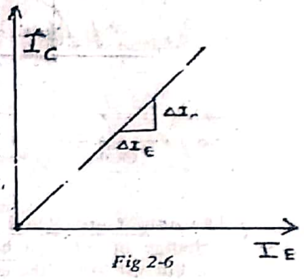


Fig 2-6

Current amplification Factor

$$\alpha = \left(\frac{\Delta I_C}{\Delta I_E}\right)_{V_{CB}}$$

Current amplification factor is defined as the ratio between the small change in collector current to the small change in emitter current when the collector-base voltage V_{CB} remains constant.

The slope of the curve gives the current amplification factor. It has no unit.

Output Characteristic: The emitter current I_E is kept at a fixed value. The collector-base voltage V_{CB} is varied step by step. For each V_{CB}, the collector current I_C is noted. The experiment may be repeated for different values of I_E

Now we can draw a graph taking the collector current I_C in Y-axis and the collector-base voltage is X-axis. The graph will be as shown in Fig. 2-7. From the graph, the output impedence can be determined.

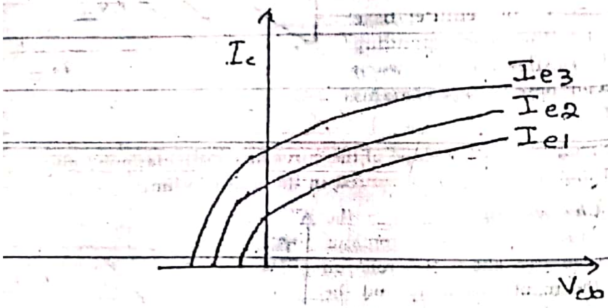


Fig. 2-7.

$$\text{Output impedence} = \left(\frac{\Delta V_{CB}}{\Delta I_C}\right)_{I_E} \text{ ohm.}$$

The output impedence is defined as the ratio between the small change in collector-base voltage to the small change in collector current when the emitter current remains constant.

The reciprocal of the output impedence is called the output admittance. It is measured in the unit of mho.

ii) Common Emitter Characteristic:

The static characteristic of a transistor can be studied using the circuit as shown in Fig. 2-8. Using a potential divider arrangement, a forward bias is given between the base and the emitter. The base-emitter voltage is measured using the voltmeter V_{BE} and the base current is measured using the microammeter I_B. For the collector a reverse bias is given. The voltage between the collector and emitter is measured using the voltmeter V_{CE} and the collector current is measured using the milliammeter I_C.

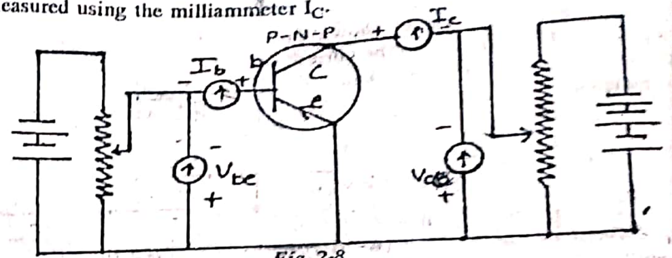


Fig. 2-8.

Input characteristic: The collector-emitter voltage V_{CE} is kept constant at a particular value. The base-emitter voltage V_{BE} is varied step by step. For each V_{BE}, the base current I_B and the collector current I_C is noted. The experiment may be repeated for different values of V_{CE}.

A graph may be drawn taking the V_{BE} along the X-axis and the base current I_B along the Y-axis. The graph will be as shown in Fig. 2-9. From the graph the input impedence can be calculated.

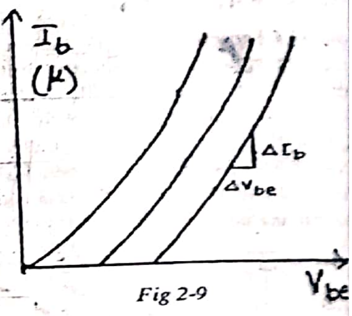


Fig 2-9

$$\text{Input impedence} = \left(\frac{\Delta V_{BE}}{\Delta I_B}\right)_{V_{CE}} \text{ ohm.}$$

The input impedence is defined as the ratio between the small change in base-emitter voltage to the corresponding change in base current when the collector-emitter voltage V_{CE} remains constant.

Transfer Characteristic: Transfer characteristic curve can be drawn taking the base current in X-axis and the collector current in Y-axis.

The graph will be as shown in Fig. 2-10. From the graph the current amplification factor β can be calculated.

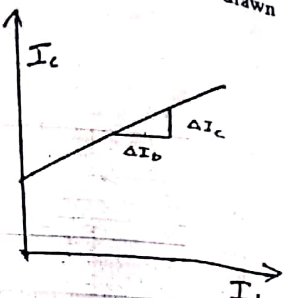


Fig 2-10

$$\beta = \left(\frac{\Delta I_C}{\Delta I_B}\right)_{V_{CE}}$$

The current amplification factor is defined as the ratio between the small change in the collector current to the small change in the base current when the collector-emitter voltage remains constant.

Output Characteristic: The base current I_B is kept at a particular constant value. The collector-emitter voltage V_{CE} is varied step by step. For each V_{CE}, the collector current I_C is noted. The experiment may be repeated by changing the value of I_B. Now we can draw a graph taking the collector-emitter voltage V_{CE} along the X-axis and the collector current I_C along the Y-axis. The graph will be as shown in Fig. 2-11. From the graph, the output impedence can be calculated.

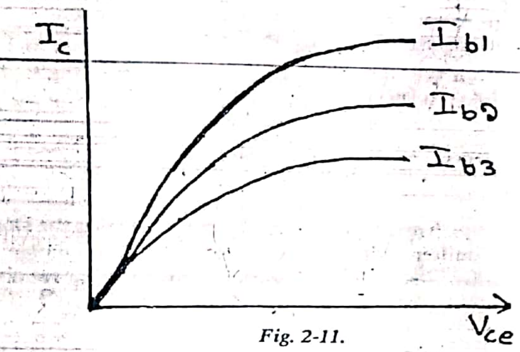


Fig. 2-11.

$$\text{Output impedence} = \left(\frac{\Delta V_{CE}}{\Delta I_C}\right)_{I_B} \text{ ohm.}$$

The output impedence is defined as the ratio between the small change in collector-emitter voltage (V_{CE}) to the corresponding change in the collector current when the base current (I_B) remains constant.

e. Hybrid Parameters

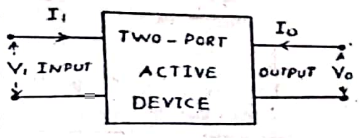


Fig 2-12

The terminal behaviour of a large class of two-port devices are specified by two voltages and two currents. The box in Fig. 2-12 represents such a two port network. We may select two of the four quantities as independent variable and the other two dependent variable. The type of parameters may be determined by the choice of independent variable. There are three types of transistor parameters.

i) Z-Parameters: In this the currents (I_1 and I_0) are taken as independent variable and the voltages (V_1, and V_0) are taken as dependent variable. In this the parameters will have the dimensions of impedence (ohm). Hence these parameters are known as Z-parameters.

ii) Y-Parameters: In this the voltages (V_1 and V_0) are taken as independent variable and the currents (I_1 and I_0) are taken as dependent variable. In this the parameters will have the dimensions of admittance (mho). Hence these parameters are known as y-parameters.

iii) H-Parameters: In this V_0 and I_1 are taken as independent variable and V_1 and I_0 are taken as dependent variable. So the different parameters will have different dimensions. Hence these parameters are known as Hybrid parameters.

In H-parameter V_1 and I_0 are taken as dependent variable.

$$\therefore \quad V_1 = f_1 (I_1, V_0) \qquad \cdots\cdots (1)$$

$$I_0 = f_2 (I_1, V_0) \qquad \cdots\cdots (2)$$

Taking the total differentials, we get

$$dV_1 = \Delta V_1 = \left(\frac{\partial V_1}{\partial I_1}\right)_{V_0} dI_1 + \left(\frac{\partial V_1}{\partial V_0}\right)_{I_1} dV_0 \qquad \cdots\cdots (3)$$

$$dI_0 = \Delta I_0 = \left(\frac{\partial I_0}{\partial I_1}\right)_{V_0} dI_1 + \left(\frac{\partial I_1}{\partial V_0}\right)_{I_1} dV_0 \qquad \cdots\cdots (4)$$

We have to develop a.c. equivalent circuit of a transistor. Hence the Δ-quantities may be replaced by the instantaneous values of these variable quantities.

$$\therefore V_1 = \left(\frac{\partial V_1}{\partial I_1}\right)_{V_0} i_L + \left(\frac{\partial V_1}{\partial V_0}\right)_{I_1} V_0 \qquad \cdots\cdots (5)$$

$$i_0 = \left(\frac{\partial I_0}{\partial I_1}\right)_{V_0} i_1 + \left(\frac{\partial I_0}{\partial V_0}\right)_{I_1} V_0 \qquad \cdots\cdots (6)$$

The partial derivatives in the above equation may be defined as follows.

$$h_i = \left(\frac{\partial V_1}{\partial I_1}\right)_{V_0} = \text{input impedence (ohm)}$$

$$h_r = \left(\frac{\partial V_1}{\partial V_0}\right)_{I_1} = \text{reverse voltage gain}$$

$$h_f = \left(\frac{\partial I_0}{\partial I_1}\right)_{V_0} = \text{forward current gain}$$

$$h_0 = \left(\frac{\partial I_0}{\partial V_0}\right)_{I_1} = \text{output admittance (mho)}$$

Substituting the above parameters in equation (5) and (6) we get

$$V_1 = h_i\, i_1 + h_r\, V_0 \qquad \cdots\cdots (7)$$

$$i_0 = h_f\, i_1 + h_0\, V_0 \qquad \cdots\cdots (8)$$

- From equation (7) it is found that the input voltage V_1 is the sum of two voltages. The first component is the voltage drop across the resistance h_i when an input a.c. i_1 flows through it. The second component is proportional to the output voltage. It can be represented by an active element which is a constant voltage source of magnitude $h_r V_0$

From equation (8), it is found that the output current i_0 is made up of two components. The first component is proportional to the input current i_1. It can be represented by an active element such as a constant current generator $h_f i_1$. The second component is represented by an admittance h_0 placed parallel to the current generator $h_f i_1$. The current through h_0 has the magnitude $h_0 V_0$. The complete equivalent circuit with hybrid parameter is as shown in Fig. 2-13.

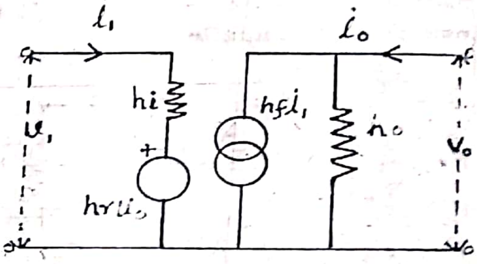


Fig. 2-13.

If V_0 is made zero by short circuiting the output terminal, equation (7) becomes,

$$V_1 = h_i i_1 \quad \text{or} \quad h_i = \frac{V_1}{i_1} \qquad \text{..... (9)}$$

Equation (8) becomes

$$i_0 = h_f i_1 \quad \text{or} \quad h_f = \frac{i_0}{i_1} \qquad \text{..... (10)}$$

If the input current i_1 is made zero by open circuiting the input terminal, equation (7) becomes

$$V_1 = h_r V_0 \quad \text{or} \quad h_r = \frac{V_1}{V_0} \qquad \text{.....(11)}$$

Equation (8) becomes

$$i_0 = h_0 V_0 \quad \text{or} \quad h_0 = \frac{i_0}{V_0} \qquad \text{.....(12)}$$

These hybrid parameters can be used for any type transistor connections such as common base, common emitter and common collector. But the value of these parameters will be different for different configurations. For the common base the parameters used are h_{ib}, h_{fb}, h_{rb} and h_{ob}. In a similar way these parameters are used for other type of configurations.

f. Transistor as an Amplifier

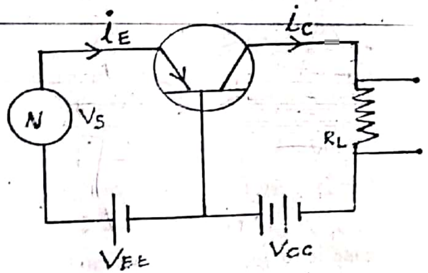


Fig. 2.14.

An amplifier is a circuit capable of magnifying the amplitude of electrical signal. In the input a small voltage is given and in the output we will get amplified output.

The basic common base amplifier-circuit is shown in Fig.2-14. The emitter-base junction is forward biased using the battery V_{EE} while the collector-base junction is reverse biased using the battery V_{CC}. The a.c. signal V_s which is to be amplified is included between the emitter and the base. The output is taken across the load resistance R_L.

When no input signal is applied, the emitter current is only the d.c. current I_E. Hence there will be no change in the collector current. It will remain constant I_C

$$\therefore I_E = I_C + I_B \qquad \text{.....(1)}$$

When a signal is applied, the voltage between the emitter and the base changes continuously. So there will be a change in the emitter current. Now the emitter current will be the sum of the d.c. current and the a.c. current due to the signal. Since there is a change in the emitter current, there will also be a change in the collector current and the base current. They will also consists of a.c. component.

If a.c. signal is not applied to the input terminal, a steady current flows through the load resistance. Hence the potential difference across the load will be a constant and it is $I_C R_L$. When the a.c. signal voltage is applied, there will be changes in V_{EB}, I_E, I_C and I_B. Let δV_{EB} be the small change in the input voltage. Hence there will be a small change in the collector current. Let δI_C be the change in the collector current. Therefore, the p.d. across the load resistance R_L is $R_L \delta I_C$. This is the a.c. output voltage across the load resistance.

$\therefore$ Voltage gain

$$\text{of the amplifier} = \frac{\text{a.c. output Voltage across } R_L}{\text{a.c. input voltage}}$$

$$A_U = -\frac{R_L\,\delta I_C}{\delta V_{EB}} \qquad \text{..... (2)}$$

But we know that $\delta I_C = \alpha \cdot \delta I_E$ where α is the current gain in the common base mode.

$$\therefore \quad A_V = \frac{R_L\,\alpha \cdot \delta I_C}{\delta V_{EB}} \qquad \text{..... (3)}$$

Since the emitter-base resistance is low, $R_L\,\delta I_E$ will be much larger than δV_{BE} and α is very nearly unity. Therefore the gain A_V will be greater than 1. Hence the transistor works as an amplifier.

$$\text{Power gain} = \text{Voltage gain} \times \text{current gain}$$
$$= A_V \times \alpha$$

As A_V is much greater than 1 and α is nearly unity, power gain is also achieved. Thus a transistor functions as an amplifier.

g. Transistor as an Oscillator

An oscillator may be defined as an electronic circuit which converts d.c. energy into an a.c. energy at a very high frequency. Electronic oscillator can be used to generate a.c. of the desired high frequency. In this there will be an excellent frequency stability, silent operation and ease in changing the frequencies. We can classify the oscillators according to the component which controls the frequency and wave form. Accordingly, we have L.C. oscillator, crystal oscillator, negative resistance oscillator and R.C. relaxation oscillator.

An L.C. circuit consist of a tank circuit. This consists of an inductance L in parallel to the capacitor C. In this oscillations are produced by the energy of the system being alteratively stored in the electric and magnetic field. Due to loss of energy; the amplitude of oscillation decay with time. By compensating the loss of energy, the oscillations can be maintained. The frequency of oscillation is $f = \frac{1}{2\pi\sqrt{LC}}$

In a transistor, there is a phase shift of 180° between the input and output. So by using a tank circuit and a feed back circuit, a transistor can be used in oscillator. The principle of transistor oscillator is as follows.

Oscillations produced by the tank circuit are fed to the transistor as input. The amplified output from the transistor has a phase shift of 180°. A small part of the output is feedback through the feedback circuit. It produces a phase shift of 180°. Thus a total phase shift of 0° or 360° is produced. This will produce a sustained oscillation.

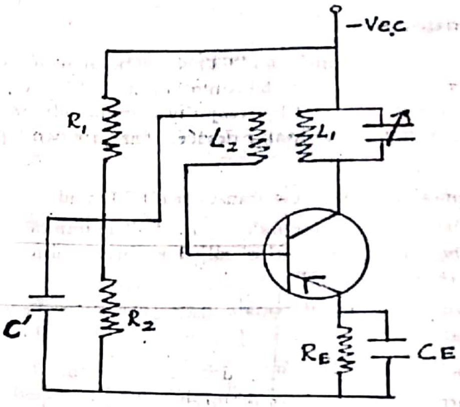


Fig. 2-15

A tuned collector oscillator is shown in Fig. 2-15. The tank circuit is connected between the emitter and the voltage source V_{CC}. The output voltage developed across the turned circuit is inductively coupled to the base circuit through the coil L_2. The winding direction of the two coils are so chosen that positive feedback takes place from the collector circuit to the base circuit.

When the collector supply voltage is switched on, a transient current is produced in the tank circuit. The varying current flowing through L_1 induces an emf across L_2 which is applied across the emitter and base. So the collector current increases. Part of this amplified energy is used to meet losses taking place in the oscillator circuit. The balance is radiated out in the form of electromagnetic waves. The frequency of oscillation is

$$f = \frac{1}{2\pi\sqrt{LC}} \text{Hz}$$

2. FIELD EFFECT TRANSISTOR

a. Introduction:

The field-effect transistor (FET) is a semiconductor device which depends for its operation on the control of current by an electric field. Since the current is carried by majority carriers only, the field effect transistor is said to be unipolar device. They are two types of FET. They are

1) Junction field-effect transistor (JFET) and
2) Metal-oxide semiconductor field effect transistor (MOSFET). The MOSFET is also called as insulated-gate field effect transistor (IGFET).

The following are the main difference between a FET and a conventional transistor.

i) The operation of FET depends upon the flow of majority carriers only. It is, therefore a unipolar device. But in a conventional transistor, the current flow is due to both electrons and holes. Hence it is called as bipolar junction transistor.

ii) A bipolar transistor are current controlled devices. In this the output current is controlled by the input current. But in a FET the flow of current is controlled by an electric field.

iii) The input impedence of a FET is high. The input impedence of JFET is of the order of 10^8 to 10^{12} ohm.

iv) A FET is less noisy than a bipolar transistor.

b. Junction Field Effect Transistor

Construction: In a JFET, the current flow is due to the majority carrier of charges. In a semiconductor, there are two types of carriers viz holes and electrons. Hence JFET are of two types.

i) n-channel FET-In this the current flow is due to electrons.
ii) p-channel FET-In this the current flow is due to holes.

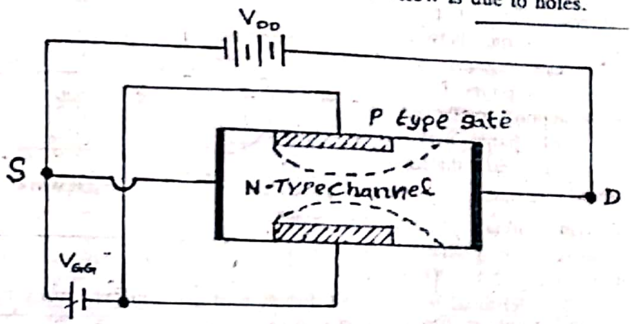


Fig. 2-16.

The structure of an n-channel FET is shown in Fig. 2-16. Ohmic contacts are made to the two ends of the semiconductor bar of n-type material. (if p-type silicon is used, the device is referred as p-channel FET). The junctions on both side of the bar are formed impurities opposite to that of the channel i.e. p-type impurities for n-type channel and vice versa. Current is allowed to flow along the length of the bar by applying a voltage between the end terminal of the bar. The current is carried by majority carriers which drift through the channel.

The following FET notation is standard.

Source: The source S is the terminal through which the majority carriers enter the bar. Conventional current entering the bar at S is designated by I_S.

Drain: The drain D is the terminal through which the majority carriers leave the bar. Conventional current entering the bar at D is designated by I_D. The drain-to-source voltage is called V_{DS}

Gate: On both side of the n-type bar in Fig. 2-16, heavily doped (P^+) regions of acceptor impurities have been formed by diffusion for creating p-n junction. These impurity regions are called the gate G. Between the gate and source a voltage V_{GS} is applied in the direction to reverse-bias the p-n junction. Conventional current entering the bar at G is designated by I_G

Channel: The regions of n-type material between the two gate regions is called channel (in p-type FET it will be p region): The majority carrier of charge move from source to drain through this region.

The circuit symbol of n-channel and p-channel are shown in Fig. 2-17. The arrow on the gate terminal refers to the direction of gate current, when the gate-source junction is forward biased.

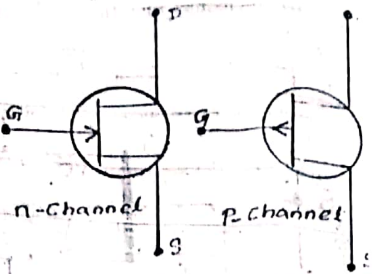


Fig 2-17

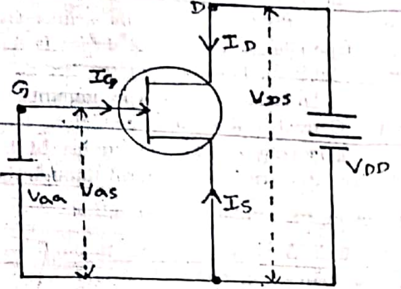


Fig. 2-18.

An n-channel FET with its terminal connected properly to voltage source is shown in Fig. 2-18. The source voltage V_{GG} and V_{DD} respectively supply the gate voltage and drain voltage. For a p-channel FET, the polarities of the voltage source should be reversed.

Operation of FET: To discuss the operation of a JFET, we have to give the following connections.

i) gates are always reverse-biased

ii) the source terminal is always connected to that end of the drain supply which provides the necessary charge carriers.

Let us consider an n-channel JFET (Fig. 2-16) and discuss its working when either V_{GS} or V_{DS} or both are changed.

i) When V_{GS} = 0 and V_{DS} =0: When no voltages are applied between D and S and G and S, the depletion region around the p-junction are of equal thickness and symmetrical.

ii) When V_{GS}=0 and V_{DS} is increased from Zero: For this the JFET is connected to the V_{DD} supply. The electrons flow from S to D whereas the conventional drain current I_D flows through the channel from D to S. When V_{DS} is applied, there is a gradual increase of positive potential along the channel as we go from S to D - i.e. as we go along the channel from S to D the reverse voltage across the p-n junction increases. Hence thickness of the depletion region also increases. Therefore the channel is wedge shaped.

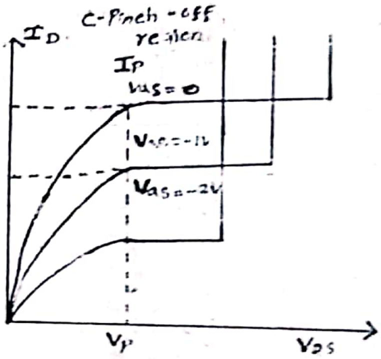


Fig. 2-19

As V_{DS} is gradually increased from zero, I_D increases proportionally as per ohm's law. This ohmic relationship between V_{DS} and I_D continues till V_{DS} reaches a certain critical value called **pinch of voltage** V_P. When $V_{DS} = V_P$, the current I_D is maximum. When V_{DS} is increased beyond V_P, the length of the pinch-off region increases. Hence there is no further increase of I_D. At a certain value of V_{DS}, I_D suddenly increases. This effect is due to the avalanche multiplication of electron caused by breaking of covalent bonds of silicon in the depletion region between the gate and the drain. The variation of I_D with V_{DS} when $V_{GS}=0$ is shown in Fig. 2-19.

iii) When V_{DS} =0 and V_{GS} is decreased from zero. When V_{GS} is made more and more negative the gate reverse bias increases. Hence the thickness of the depletion region also increases. At a particular voltage, the two depletion regions make contact with each other. In this condition, the channel is said to be cut-off. The value of V_{GS} which is required to cut- off the channel is called the **cut-off voltage**.

iv) When V_{GS} is negative and V_{DS} is increased. As V_{GS} is made more and more negative, value of V_p as well as break-down voltage are decreased. It is shown in Fig. 2-19.

Since gate voltage controls the drain current JFET is called a voltage controlled device. A p-channel JFET operates exactly in the same manner as n-channel JFET except that the current carriers are holes and polarities of both V_{DS} and V_{GS} are reversed.

c. Characteristics of JFET

The relation between the drain current I_D and the voltage between drain-source V_{DS} can be represented in a graph. This curve is called the static characteristic curve of a JFET. The circuit arrangement to study the characteristic of n-channel JFET is shown in Fig. 2-20.

i) Output Characteristic: The gate voltage V_{GS} is kept fixed at zero. The drain voltage V_{DS} is increased from zero in equal suitable steps. The drain current I_D is noted for each value of V_{DS}. The experiment is repeated for different values of V_{GS} such as - 1V - 2V, - 3V etc. A graph is plotted taking drain voltage V_{DS} along the X-axis and the drain current I_D on the Y-axis for each value of the gate voltage V_{GS}. The output characteristic curve is as shown in Fig. 2-19.

From the output characteristic curve the FET parameters can be calculated.

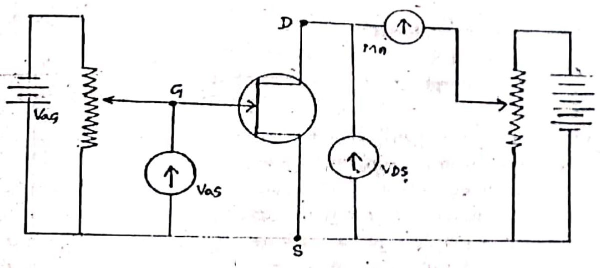


Fig. 2-20.

1) Pinch of Voltage: On the curve for V_{GS}=0, the point at which I_D becomes constant is noted. The corresponding value of V_{DS} gives the pinch-off voltage.

2) Drain Resistance r_d: It is the ratio of small change in the drain voltage to the corresponding small change in the drain current at a constant gate voltage.

$$r_d = \left(\frac{\partial V_{DS}}{\partial I_D}\right)_{V_{GS}} \text{ ohm}$$

3) Mutual conductance g_m : It is the ratio of a small change in the drain current to the corresponding small change in the gate voltage at a constant drain voltage.

$$g_m = \left[\frac{\partial I_D}{\partial V_{GS}}\right]_{V_{DS}}$$

4) Amplification Factor μ: It is the ratio of a small change in the drain voltage to the corresponding small change in the gate voltage at a constant drain current.

$$\mu = \left(\frac{\partial V_{DS}}{\partial V_{GS}}\right)_{I_D}$$

ii) Transfer characteristic: The same circuit can be used to study the transfer characteristic. The voltage V_{DS} is kept constant at a suitable value greater than the pinch-off voltage V_P. The gate voltage V_{GS} is decreased from zero in equal steps till I_D is reduced to zero. The drain current I_D is noted for each value of V_{GS}. Now a graph is plotted taking I_D along Y-axis and V_{GS} in X-axis. The graph will be as shown in Fig. 2-21.

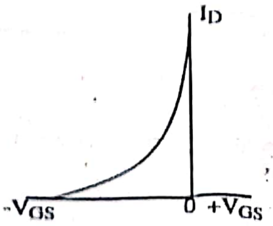


Fig. 2-21

It is seen that when $V_{GS} = 0$,

I_D is maximum and when $I_D = 0$, $V_{GS} = V_P$

d. FET Amplifier

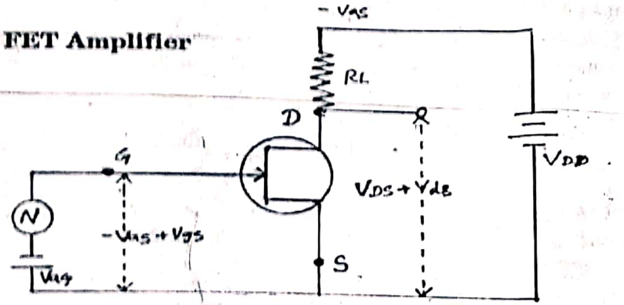


Fig. 2.22.

The common source JFET amplifier circuit is shown in Fig. 2-22. The a.c. signal which is to be amplified is connected in series with the gate bias battery V_{GG}. Due to this, there will be a variation in the total gate-to source voltage. The load resistance R_L is connected to the drain terminal. The output voltage is measured across the load resistance or across the JFET. For an input signal v_{gs}, let the output a.c. voltage be v_{ds}.

When the signal voltage is positive, the gate voltage becomes less negative with respect to source. Hence the drain current increases and hence the p.d. across the load R_L increases. Due to this, the drain terminal will be less positive with respect to source. A small voltage variation at the gate produces a large voltage variation across the load resistance. Hence this circuit functions as an amplifier. The increase in the gate voltage produces a decrease in the drain voltage. There will be a phase of 180° between the input and output.

Calculation of Gain: Applying Kirchhoff's law to the output circuit, we get for the d.c. voltage,

$$V_{DD} = V_{DS} + I_D R_L \qquad \text{..... (1)}$$

For a.c. voltage,

$$v_{ds} + i_d R_L = 0 \qquad \text{..... (2)}$$

For small signal voltage,

$$i_g = g_m v_{gs} + v_{ds}/r_d \qquad \text{..... (3)}$$

From equation (2) and (3),

$$v_{ds} + R_L(g_m v_{gs} + v_{ds}/r_d) = 0$$

$$\therefore v_{ds}(1 + R_L/r_d) = -R_L g_m v_{gs}$$

$$v_{ds}\left[\frac{R_L + r_d}{r_d}\right] = -R_L \frac{\mu}{r_d} v_{gs} \qquad \text{..... (4)}$$

$$[\text{ since } \mu = r_d g_m]$$

$$\therefore \text{Voltage gain } A_v = \frac{v_{ds}}{v_{gs}} = \frac{-\mu R_L}{r_d + R_L} \qquad \text{.....(5)}$$

The negative sign indicates that there is a phase shift of 180° between the input voltage and the output voltage.

Problems

1) **A transistor has a base current of 0.08 mA and the emitter current is 10 mA. Calculate (i) the collector current (ii) α and (iii) β**

Sol :

i) Collector current

We know that $I_E = I_C + I_B$

$\therefore I_C = I_E - I_B$

$= 10 - 0.08$

$\therefore I_C = 9.92$ mA.

ii) Current gains $\alpha = \dfrac{I_C}{I_E} = \dfrac{9.92}{10} = 0.992$

iii) Current gain $\beta = \dfrac{I_C}{I_B} = \dfrac{9.92}{0.08} = 124.$

2) **The collector current of a transistor is 6.6 mA and α is 0.95. Calculate (i) base current and (ii) β.**

Sol:

i) Base current

We know that $\alpha = \dfrac{I_C}{I_E}$

$\therefore \quad 0.95 = \dfrac{6.6}{I_E}$

$\therefore \quad I_E = \dfrac{6.6}{0.95} = 6.95$

We know that $I_E = I_C + I_B$

$\therefore I_B = I_E - I_C$

$= 6.95 - 6.6 = 0.35$ mA.

ii) Current gain β

$$\beta = \frac{I_C}{I_B} = \frac{6.6}{0.35} = 18.86$$

3) **In a common base transistor I_C=0.97 mA and I_B=30 mA. Find the value of α**

Sol:

We know that $I_E = I_C + I_B$

$\therefore \quad I_E = 0.97 + 30 \times 10^{-3}$

$= 0.97 + 0.03 = 1$ mA

Current gain $\alpha = \dfrac{I_C}{I_E} = \dfrac{0.97}{1} = 0.97$

4) **In a common base mode, if the collector-base voltage is changed by 0.5V, Collector current changes by 0.05 mA. Find the output impedence.**

Sol : Output impedence $= \left(\dfrac{\partial V_{CB}}{\partial I_C}\right)_{I_E}$

$\partial V_{CB} = 0.5$ V, $\partial I_C = 0.05$ mA $= 0.05 \times 10^{-3}$A.

$\therefore \quad$ Impedence $= \dfrac{0.5}{0.05 \times 10^{-3}} = 10^4$ ohm.

5) **For a constant drain-source voltage if the gate- source voltage is changed from 0 to –2V the corresponding change in drain current becomes 2 mA. Calculate the transconductance of the FET. If the a.c. drain resistance is 100 Kilo–ohm calculate the amplification factor of the FET.**

Sol: Transconductance $g_m = \left(\dfrac{\partial I_D}{\partial V_{GS}}\right)_{V_{DS}}$

Here $\partial I_D = 2$ mA $= 2 \times 10^{-3}$A

$\partial V_{GS} = 2$V

$g_m = \dfrac{2 \times 10^{-3}}{2} = 10^{-3}$ ohm.

Amplification Factor $\mu = g_m \times r_d$

$$\therefore \mu = 10^{-3} \times 100 \times 10^3 = 100$$

6) **The following readings were obtained experimentally from a FET**

V_{GS} =	0V	0V	-0.2V
V_{DS} =	7V	15V	15V
I_D =	10 mA	10.25 mA	9.65 mA

Determine (i) a.c. drain resistance (ii) transconductance and (iii) amplification factor.

Sol:

i) a.c. drain resistance

$$r_d = \left(\frac{\partial V_{DS}}{\partial I_D}\right)_{V_{GS}}$$

When $V_{GS} = 0$, the change in V_{DS} is from 7 to 15 V

$$\therefore \partial v_{DS} = 15 - 7 = 811$$

When $V_{GS} = 0$, change in drain current is from 10 mA to 10.25 mA

$$\therefore \partial I_D = 10.25 - 10.0 = 0.25 \text{ mA} = 0.25 \times 10^{-3} \text{A}$$

$$\therefore r_d = \frac{8}{0.25 \times 10^{-3}} = 32 \times 10^3$$

$\therefore r_d$ = 32 kilo-ohm.

ii) Transconductance

$$g_m = \left(\frac{\partial I_D}{\partial V_{GS}}\right)_{V_{DS}}$$

When $V_{DS} = 15$ V, the change in drain current is from 10.25 mA to 9.65 mA.

$$\therefore \partial I_D = 10.25 - 9.65 = 0.6 \text{ mA}$$

$$= 0.6 \times 10^{-3} \text{A}$$

When $V_{DS} = 15$V, the change in V_{GS}, is from 0V to −0.2V

$$\therefore \partial V_{GS} = 0.2 - 0 = 0.2 \text{ V}$$

$$\therefore g_m = \frac{0.6 \times 10^{-3}}{0.2} = 3 \times 10^{-3} \text{ mho}$$

$g_m = 3\ \mu$ mho.

iii) Amplification Factor:

$$\mu = g_m \times r_d$$

$$\therefore = 3 \times 10^{-3} \times 32 \times 10^3$$

$$\mu = 96$$

7) **A common-source FET amplifier has a load resistance R_L=500 k-ohm. If the a.c. drain resistance and amplification factor of the FET are 100 k ohm and 24 respectively, calculate the voltage gain of the amplifier.**

Sol:

Voltage gain $A_V = \dfrac{\mu R_L}{r_d + R_L}$

Here $\mu = 24$; $R_L = 500 \times 10^3$ ohm.

$r_d = 100 \times 10^3$ ohm.

$$\therefore A_V = \frac{24 \times 500 \times 10^3}{100 \times 10^3 + 500 \times 10^3} = \frac{24 \times 500 \times 10^3}{600 \times 10^3}$$

$\therefore A_V = 20.$

UNIT III

LASERS AND MASERS

1.Basic concepts of Lasers and Masers

a. Laser and Maser action

The word LASER is an acronym for "**Light amplification by stimulated emission of Radiation.**" Laser produces a highly intense, concentrated and parallel beam of monochromatic and coherent light. The word MASER is an acronym for "**Microwave Amplification by stimulated Emission of Radiation.**" The major difference between the Maser and Laser is their working frequency. The typical operating frequencies of Masers are nearly 10^{10} Hz while the frequencies in which Lasers operate are of the order of 10^{15} Hz. **Albert Einstein** was the first to introduce the idea of stimulated emission. It was A.L. Shawlow and C.H.Townes, the two American scientists who investigated the possibility of this method for developing optical generator. In July 1960, the first optical quantum generator was constructed.

To understand, the Laser action one should know about spontaneous and stimulated emission. In every non-radiating atomic system, there are discrete energy levels. Whenever an electron or an atom occupying a higher energy level E_2, jumps into a lower energy level E_1, the excess of energy of the particle is radiated. The frequency of the radiation is given by the relation.

$$E_2 - E_1 = h\gamma \qquad \text{.............. (1)}$$

where h is the Planck's constant whose value is 6.6256×10^{-34} J-S.

Suppose a radiation of energy hγ exactly equal to the energy difference between two levels E_1 and E_2 falls on the system. An atom in the lower energy level may absorb that energy and go to the higher energy level. This process is called **absorption.** Now the atom is in the excited state. In the excited state, the atom is unstable. Within a short time of the order of 10^{-8} to 10^{-3} second, the atom will return back to the original state without external stimulus. This process is called **spontaneous emission.** This spontaneous emission is a statistical process.

An interesting phenomenon, however, occurs if a photon of energy hγ exactly equal to the energy difference between the two states falls on an atom while it is still in the excited state. The atom gives off a photon hγ in addition to the one which triggered its emission. This process is called the **stimulated emission.** It is this process that is important in Maser and Laser action. The different process is shown in Fig 3-1.

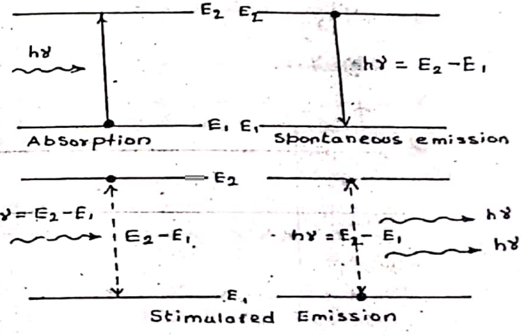


Fig 3-1

The population inversion is the other term which plays an important role in the Laser action. When the number of particle occupying in the upper energy level are more, stimulated emission will be dominating process than absorption. If the population in the lower state is more, absorption will be dominating. So for the stimulated emission to dominate it is necessary to increase the population of the upper energy level so that it is greater than that o the bottom. This is known as **population inversion.** Under norma thermal equilibrium, atoms or molecules try to occupy their lowes energy level. In order to obtain Laser action, it is necessary to find a technique by which the upper level is more populated. This was achieved by pumping the atom at the lower level to the higher energy level.

So for the Laser action the following conditions are necessary:

(1) There should be at least a pair of energy levels ($E_2 > E_1$) separated by the desired radiations which is to be stimulated.

(2) There must be a means of inverting the population N_2 at the higher energy level E_2 is always greater than the population N_1 at the lower energy level E_1.

(3) There must be an electromagnetic system which can retain the photon of energy $E_2 - E_1$ with low losses so that they can be used for stimulation.

b. Stimulated Emission.

Let us consider an atomic system placed in a radiation field of energy density E(γ). Let E_1 and E_2 be the two energy levels occupied by N_1 and N_2 atoms respectively. There are two possible process taking in an atomic system.

(1) Spontaneous transition from the level E_2 to E_1 may take place. Let A_{21} be the spontaneous emission transition probability per unit time.

(2) Secondly absorption of energy hγ can take place. This results in the atom being raised from the level E_1 to E_2. Absorption transition per unit time is proportional to the energy density E(γ). So the induced absorption transition probability per unit time will become $B_{12}E(\gamma)$ where B_{12} is the absorption probability per unit time per unit energy density.

(3) Thirdly the radiation field may interact with an atom in the higher energy state E_2 and induce it to fall to the lower energy state E_1. The probability of induced emission is proportional to the energy density of radiation field. So the probability of induced emission per unit time can be written as B_{21} E(γ) where the constant B_{21} represents the induced emission probability per unit time per unit energy density.

If there are N_2 atoms in the level E_2, the number of atoms that fall from the level E_2 to E_1 per unit time is equal to $[A_{21} + B_{21}E(\gamma)]N_2$. The number of atoms that rise up from the level E_2 to E_1 per unit time is $B_{12}E(\gamma)N_1$. So the net rate of change of atoms dN_2/dt in the level E_2 per unit time is given by

$$\frac{dN_2}{dt} = B_{12}E(\gamma)N_1 - [A_{21} + B_{21}E(\gamma)]N_2 \quad \text{..........(1)}$$

Under equilibrium condition, the net rate of change of atoms in any level must be zero.

$$\therefore \frac{dN_2}{dt} = 0 \text{ or } B_{12}E(\gamma)N_1 = [A_{21} + B_{21}E(\gamma)]N_2 \quad \text{..........(2}$$

The equilibrium distribution of particles among energy level is given by Boltzmann.

$$N_i = N \exp[-E_i/KT] \quad \text{.....(3)}$$

where N_i is the number of particle in the state of energy E_i, N is the total number of particles and K is the Boltzmann's constant. Using this the ratio between the number of atoms in the two definite energy level is given by

$$\frac{N_2}{N_1} = \text{Exp}[-(E_2 - E_1)/KT]$$

$$\text{or} \quad N_2 = N_1 \exp[-(E_2 - E_1)/KT]$$

$$\therefore \quad N_2 = N_1 \exp[-h\gamma/KT] \quad \text{.....(3)}$$

Under normal conditions, the population in the lower energy level is greater than the population in the higher energy level. Substituting the value of N_2/N_1 from equation (3) in equation (2) we get

$$B_{12}E(\gamma)\exp(h\gamma/KT) = A_{21} + B_{21}E(\gamma)$$

$$\therefore \quad E(\gamma) = \frac{A_{21}/B_{21}}{\exp(h\gamma/KT) - B_{21}/B_{12}} \quad \text{.....(4)}$$

The Planck's radiation formula gives the energy density of radiation at a given temperature as

$$E(\gamma) = \frac{8\pi h\gamma^3}{c^3} \cdot \frac{1}{\exp(h\gamma / KT)-1} \qquad .....(5)$$

Comparing equation (4) and (5) we get

$$A_{21}/B_{21} = 8\pi h\gamma/C^3 \qquad .....(6)$$

$$\text{and} \quad B_{21}/B_{12} = 1 \text{ or } B_{21} = B_{12} \qquad .....(7)$$

These relations were first obtained by Einstein in 1917. The constants A_{21}, B_{21} and B_{12} are called Einstein's coefficient.

Substituting equation (7) in equation (4) we get

$$\frac{\text{spontaneous emission probability}}{\text{Induced emission probability}} = \frac{A_{21}}{B_{21}\,E(\gamma)}$$

$$= \exp(h\gamma / KT)-1 \qquad .....(8)$$

If $h\gamma >> KT$, spontaneous emission is much more probable tha induced emission. If $h\gamma < KT$, stimulated emission will becom important. This is possible in the case of atomic transition in th microwave as well as in the visible region.

Let us now find the conditions under which the radiative emissio can be induced or stimulated continuously. In the case when matte is being continuously bombarded by photon and non-equilibriu condition prevail, we can write the ratio of emission rate to th absorption rate as

$$\frac{\text{Emission rate}}{\text{Absorption rate}} = \frac{[A_{21} + B_{21}\,E(\gamma)]N_2}{B_{12}\,E(\gamma)N_1}$$

$$= \left[1 + \frac{A_{21}}{B_{21}\,E(\gamma)}\right] \frac{N_2}{N_1} \qquad .....(9)$$

If the energy difference between the two levels is small such tha the ratio $h\gamma / KT$ is small, the quantity $A_{21} / B_{21}\,E(\gamma)$ becomes ver small as compared to unity as seen from equations (8). So the equatio (9) becomes

$$\frac{\text{Emission rate}}{\text{Absorption rate}} = \frac{N_2}{N_1} \qquad .....(10)$$

At any fixed temperature, the atomic population N_2 in the higher energy level E_2 is always less than the atomic population N_1 in the lower energy level E_1. If by some means we can able to effect population inversion so that $N_2 > N_1$, we can maintain the emission rate higher than absorption rate. This means that when a radiation of energy density $E(\gamma)$ passes through a system in which $N_2 > N_1$, the radiation comes out with more photon of frequency γ than that are incident on the system. This situation is referred to as the wave amplification by stimulated emission of radiation. If the radiation is in the microwave region, the phenomenon is the Maser. If the radiation is in the visible region, the phenomenon is the Laser.

c. Population Inversion

To obtain amplification of the electromagnetic wave through interaction with the laser material, a state of population inversion has to be created between the levels which take part in the optical transition. In the thermal equilibrium, there are fever atoms in higher energy level E_2 than in a lower lying level E_1. The ratio of population densities at room temperature is $N_2 / N_1 = 10^{-32}$. For Laser action, the number of atoms in the higher energy state should be larger. So more atoms have to be pumped from lower energy state to higher energy state. This is called the **population inversion.** The operation leading to population inversion is called **pumping**.

With the aid of energy level diagram of a three level Laser (Fig 3-2), we may explain the condition necessary for the attainment of population inversion. In thermal equilibrium, most of the atoms will be in the

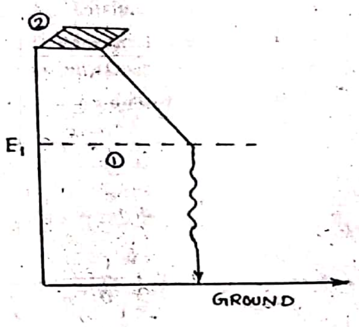


Fig. 3-2

ground state. Using a flash lamp, the atoms are raised from the ground state to broad energy band denoted by ② Once an atom exists in the broad energy level ②, it may relax to either the upper laser level ① or back to the ground state. But if the primary relaxation mechanism is a fast, due to non-radiative decay, most of the atoms will reach level ①. The upper Laser level is a metastable level with long relaxation time. Hence an accumulation of atom in that state is possible. If the rate with which atoms are pumped from the ground state to the broad energy level is sufficiently high, the number of atoms reaching and remaining in the metastable state exceeds the number of atoms in the ground state. This results in population inversion.

To achieve population inversion the rate of increase in the population of the broad excited level ② is given by

$$\frac{dN_2}{dt} = WN_o - WN_2 - \frac{N_2}{\tau_{21}} \qquad .....(1)$$

where N_o and N_1 are the population of the respective energy level, W is the pump-stimulated transition probability and τ_{21} is the fast non-radiative relaxation time. In equation (1) of the three terms on the right

(1) the first represents the rate of increase in the population of level ② due to absorption of energy from the incident pump.

(2) the second represents the rate of decrease of the population due to stimulated emission back to the ground state and

(3) the third represents the loss in population as a result of the primary relaxation mechanism to level ①

In the steady state,

$$\frac{dN_2}{dt} = 0 \qquad .....(2)$$

∴ The ratio of the population is given by

$$\frac{N_2}{N_0} = \frac{W\tau_{21}}{1 + W\tau_{21}} \approx W\tau_{21} \qquad .....(3)$$

In this we have assumed that $T_{21} < \frac{1}{W}$ Similarly, the rate of increase in the population of level ① is given by the equation

$$\frac{dN_1}{dt} = \frac{N_2}{\tau_{21}} - \frac{N_1}{\tau_{10}} \qquad .....(4)$$

where the first term on the right represents the rate of increase in the population due to relaxation from the higher broad energy level and the second term denotes the rate of decrease in the population of level ① as a result of atomic decay from level ① to the ground state.

In the steady state,

$$\frac{dN_1}{dt} = 0 \qquad .....(5)$$

$$\therefore N_2 = \frac{\tau_{21}}{\tau_{10}} N_1 \qquad .....(6)$$

Relating N_1 to N_o by means of the previously derived equation, we obtain

$$\frac{N_1}{N_o} = W\tau_{10} \qquad .....(7)$$

Hence population will be achieved when $W > \frac{1}{\tau_{10}}$ or when the pumping rate from the ground state to the broad energy level exceeding the relaxation rate from level ① back to the ground state. Thus if the pumping rate is greater than the relaxation rate from level ①, atoms will accumulate in that level which will eventually lead to a population inversion.

In a three level system, at least 50% of the atom have to be lifted out of the ground state if the population inversion is to be realised. So a four-level pumping is followed. In this to achieve a population inversion, only a small fraction of atoms have to be pumped from the ground state during Laser action. The simple model of such a system in shown in Fig 3-3

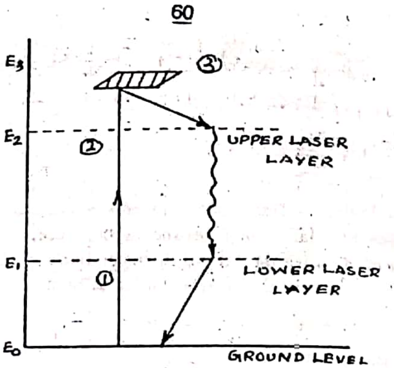


Fig 3.3

In this, first atoms are raised from the ground state into the broad energy level ③. A fast, non-radiative relaxation from this excited state ensure primary into level ② which serves as the upper Laser level and is usually a metastable level. By virtue of the long relaxation time of level ②, an accumulation of atom in that state occurs. Level ① which serves as the lower laser level possesses a fast relaxation time. These two levels will ensure the existence of population inversion. In this only a small percentage of the total atomic population has to be lifted out of the ground state in order to create the desired population inversion between level ① and ②

We can show that the population inversion between level ① and ② is independent of the pumping rate and will be achieved under the condition when $\tau_{21} > \tau_{10}$.

d. Metastable State

Initially most of the atoms of the Laser material are in the ground level. By pumping the atoms in the ground state are raised to the level ③. The level is usually made up of larger number of spectral levels. This makes pumping to be achieved over a wide spectral range. The excited atom from the broad energy band are transferred to sharp level ② by radiationless transition. The lifetime of the level ② is larger than other levels. So atoms accumulate in the state. So the population increases in this state. This state serves as upper Laser level. This level is called the metastable state.

The existence of metastable level is very important for the Laser action to occur. The relatively long life time provide mechanism helpful to achieve population inversion. Because of strong interatomic coupling, most of the excited atom decay through non- radiative way and hence short life time and broad line width. A few atoms of selected atoms give rise to radiative transitions.

The transition frequency from level ③ to level ②and from level ① to the ground fall within the frequency of lattice vibration of host crystal lattice. Hence all these atoms relax by providing phonons to lattice vibration. The order of τ_{32} and τ_{10} is 10^{-8} s to 10^{-11}s. But τ_{21} will be of the order of 10^{-5} s to 10^{-3} s. Hence the pumped atoms will have a long life time in this state and this will be the metastable state. The metastable level is one from which all dipole transition to lower energy state are forbidden.

Thus in the absence of metastable level, the excited atom will directly return to the ground state through spontaneous emission. Hence the existence of metastable level is a must. For population to build up relaxation out of lower level thus be fast i.e $\tau_{21} > \tau_{10}$

2. Types of Masers and Lasers

a. Ammonia Maser

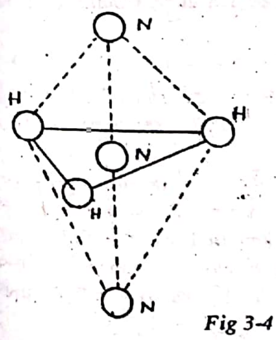


Fig 3-4

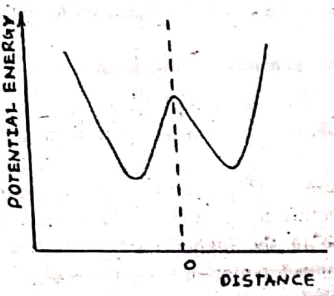


Fig 3-5

The first gas Maser was developed by **C.H.Townes.** In the ammonia molecule NH$_3$, the three hydrogen atoms lie in a plane at the corners of a triangle. The nitrogen atom is at the apex of this pyramid as shown in Fig 3-4. It has an energy minimum on either side of the plane of the hydrogen atoms and a potential barrier. It has a maximum value in the plane of the hydrogen atoms (fig 3-5). The nitrogen atom can tunnel through the plane of the hydrogen atoms and such a transition is called **inversion.** The transition between these inversion levels by the nitrogen atom gives rise to a strong emission line of frequency of 23,870 MHz or a wavelength of 1.26 cms. It is in the microwave region.

Classically, the nitrogen atom is pictured as flipping back and forth at the frequency 23,870 MHz. In the quantum picture, this is described as transition between inversion energy levels. At any temperature, the nitrogen population in the two energy level is given by the Boltzmann's formula

$$N_2 = N_1 \exp [-E_2 - E_1/KT]$$

$$= N_1 \exp [-h\gamma / KT]$$

where N$_2$ is the number of nitrogen atoms in the level E$_2$ and N$_1$ is the number of nitrogen atom in the level E$_1$. The above equation suggests that under normal condition $N_2 < N_1$. Population inversion is obtained by segregating the lower energy molecules from the higher energy molecules by an ingenious device. The difference in the dipole moments of the ammonia molecule in the two energy states enables us to separate the higher energy molecules by applying a non-homogeneous electric field.

The essential parts of ammonia Maser are shown in Fig 3-6. The ammonia gas from an oven kept at room temperature is allowed to flow at a uniform rate at a pressure of 1 m.m into an electrostatic focusser. The electric field is inhomogeneous. The molecules of ammonia which were in the lower energy state were attracted while those in the higher energy state were repelled by the field and continued in motion along the axis and into a cavity. The lower energy molecules are removed by a vacuum pump. Almost all of the molecules entering the cavity were in the higher energy state.

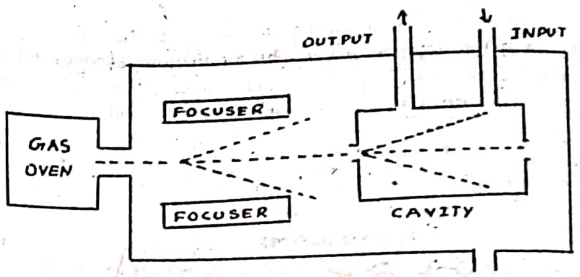


Fig 3-6

The upper level molecules were bombarded with a microwave of frequency 23, 870 MHz. The excited ammonia molecule gave up their energy in the form of additional photons at the same frequency. So the incoming waves were amplified. This produced a beam of microwave energy of a very high purity.

The excited molecules, then, entered a metal chamber. The inner walls of the chamber were highly reflective. The input signals were allowed to enter through a wave guide. This is reflected many times inside the chambers. So new photons were produced. These photons were also reflected in the chamber and additional photons were produced due to interaction with the molecules of the gases. When the number of ammonia molecules in the chamber was sufficient, the incoming signal could be stopped. But the production of new photon could be maintained. In this way the apparatus could also be used as a generator of 23,870 MHz frequency without any incoming signal. But the Maser first designed by Townes is not quiet suitable for practical applications since it can operate only at one frequency with no means of tunning. Basically it is a low noise amplifier, but with a suitable feedback, it can also be converted into a high precession generator.

b. Ruby Laser

The first successful operation of the Laser was achieved by **T.H Maiman** in 1960, using ruby crystal. Ruby is an aluminum oxide in which a few of the aluminium atoms have been replaced by chromium atom ($Al_2O_3 \cdot Cr^{3+}$) The Chromium atom take part in Laser action. Maiman used 0.05% Cr^{3+} to get a pink colour. It is because the chromium absorbs ultraviolet, green and yellow and transmits only red and blue.

Ruby Laser apparatus is shown in Fig 3-7. The Laser consists of three main parts.

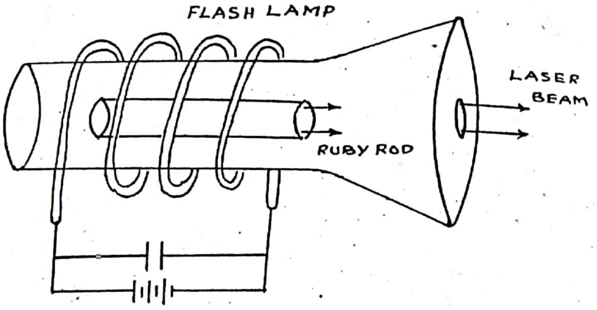


Fig 3-7.

(1) an active material

(2) a resonant system made as two parallel plates with reflecting coating applied on them and

(3) an exciting system usually made up of a helical Xenon flash tube and a power supply.

The working element of the ruby Laser is a cylinder of pink ruby, usually between 0.5 cm to 2 cm in diameter and 2 to 30 cm in length. The end faces are grounded and polished so that they are plane and parallel to a high degree of accuracy. One end of the faces is provided with a completely reflecting surface, the other end is partially reflecting. The ruby rod is arranged along the axis of the helical Xenon flash tube in such a manner that the coils of the helix encompass the rod. The flash of the tube lasts several millisecond. During this period of time, the tube consumes energy amounting to several thousand of joules and most of it is spent for heating the apparatus. The other smaller part of the energy in the form of blue and green radiation is absorbed by the ruby. This energy ensure the excitation of chromium ions.

The operation of ruby Laser can be explained with the help of energy level diagram shown in Fig 3-8. In this 1,2 and 3 correspond to energy levei of chromium ions.

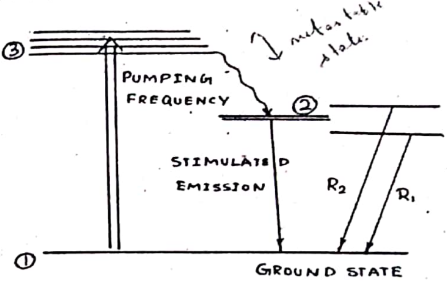


Fig 3.8

In the normal, non-excited state chromium ions are in the lower level 1. When the ruby crystal is irradiated with the light of Xenon flash tube, the chromium atoms are excited and pass to the upper level 3. In this the light absorption band is 5600A. The absorption band width of this level is about 800 A.

From level 3, part of the excited chromium atoms return to the ground level 1 and the other part to level 2. During this transition, chromium ions give off part of this energy to the crystal lattice in the form of heat. The probability of transition from level 3 to level 2 is 200 times greater and from level 2 to level 1, 300 times smaller than probability of transition from level 3 to level 1. The transition from level 3 to level 2 is non–radiative or raditionless. States such as the middle state are said to be metastable state. Population of the middle state builds up and population-inversion is achieved.

In such a system, the probability of spontaneous transition at any moment of time is very large. The very first photon appearing during such spontaneous transition will knock out a second photon from a neighbouring atom. The atom from which the first photon was emitted will be brought to its ground state. Now these two photons will knock out two more photons and their total numbers will be four and so on. The process grows practically instantaneously. The first wave of radiation on reaching the reflecting surface, will return and cause further increase in the number of induced transition and in the radiation intensity.

Such a process will repeat many times. The generation will rise and the power will increase till the majority of the excited particle of the active material give off the energy acquired at the moment of excitation. A very high intensity beam will emerge through the partially silvered end face of the ruby rod. The direction of this beam will be strictly parallel to the ruby axis.

In fact the level 2 consists of two close sub-level. In the case the exciting power is insufficient, two weak lines R_1 and R_2 are emitted. Their wave lengths are 6943 A and 6929 A respectively. This radiation is mainly due to spontaneous transition. With an increase in the exciting power, the intensity of radiation at the wavelength of 6929A is practically no more increasing and at the wavelength 6943 A generation is established.

The output has the enormous power output of more than 10^4 watts over a beam of cross-section 10^{-4} .square metre. The emitted band is within a wavelength interval of about 0.02 A. The beam from the end face of the ruby rod has an angular spread of less than 1°. The output of ruby laser is 10-20 killo watt in milli second pulses.

c. He-Ne Laser

The He-Ne was the first Laser to employ a gaseous medium for the lasing action. Its high temporal and spatial coherence and relative simplicity make it a highly suitable device for numerous measurement techniques and applications. The first He-Ne Laser was developed by **A.Javan in 1959**. A schematic representation of such a laser is shown in Fig 3-9.

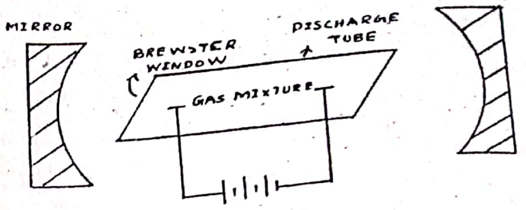


Fig 3-9

The gas Laser is a fused quartz tube with a diameter of about 1.5cm and 80 cm long. This tube is filled with a mixture of gases, neon (Ne) under a pressure of 0.1m.m of H_g and helium (He) under a pressure of 1 m.m of H_g. Atomic resonance lines in gas are determined primarly by Doppler width and are relatively narrow. Hence the most common pumping mechanism in a gas Laser involves excitation through collisions with the electrons and other atoms in a gaseous discharge. In a He-Ne Laser, the discharge may be achieved in a tube by passing a direct current through the gas mixture. When discharge occurs, some atoms are ionized, creating positive ions and free electrons. The free electrons are acted upon by the applied electric field which accelerate and increases their energy. This is the initial state in the excitation process.

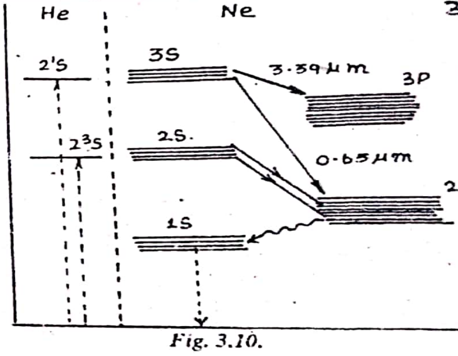


Fig. 3.10.

3s to 2p

2s to 2

3st

The attainment of population inversion between numerous electrons energy levels of Ne in the He-Ne Laser can be explained using the energy level diagram shown in Fig. 3-10.

When energetic electrons in the gas discharge strike He atoms, they excite these atoms, primarly into level 2^1S and 2^3S. When these excited He atom collide with unexcited Ne atoms, they may exchange energy raising the Ne atom from the ground state into the 3S and 2S energy levels respectively while they drop back to the ground state. This pumping mechanism will continuously act to populate the Ne 3S and 2S levels. Thus population inversion is obtained between these levels and lower Ne levels. The three most important He-Ne laser lines are due to the following transitions.

1) 3S to 2P. The wavelength emitted is red line at 632.8 nm.
2) 2S to 2P. The wavelength emitted is infra red at 1.15×10^3 nm (1.5 μm) and
3) 3S to 3P. The wavelength emitted is infra red at 3.39×10^3 nm (3.39μm)

If a single line oscillation such as the 632.8 nm line is desired, special measures have to be employed to suppress all the lines except the desired line. One such technique is the use of Littrow prims. This allows only the desired wavelength and others are reflected from the axis of the plasma tube.

Advantage :

The emission of a gas Laser is highly monochromatic and coherent. In a solid laser there may be crystalline imperfection, thermal distortion and scattering. But these are absent in gas lasers. So the emitted light is more directional and more monochromatic. Gas lasers are capable of operating continuously without need for cooling. Gas lasers are used in optical testing and optical alignment, scientific measurement of many types and accurate length and distance measurement both in surveying and at small distance.

UNIT - IV

OPTOELECTRONIC DEVICES

1. Light Emitting Diodes (LED)

a. Introduction

The devices in which photon plays a major role are called the **photonic devices or photo devices**. These devices operate on the principle of interaction of light with semiconductor. The absorption of light by semiconductors is associated with the transfer of radiant energy to the crystalline lattice. If the radiant energy is greater than the work function, the electron in the atom comes out of the atom. This effect is photoelectric emission. This is the so called extrinsic photoelectric effect. Photocells function on this principle. When light incident on a semiconductor, electron-hole pairs are generated. This increases the electrical conductivity. This is called intrinsic photoelectric effect. This is used in photo-resistors, photodiodes etc.

Photonic devices can be divided into three groups.

(i) devices that convert electrical energy into optical radiation - the LED and diode Laser.
(ii) devices that detect optical signals through electronic process- photodetectors and
(iii) devices which convert optical radiation into electrical energy - the photo voltaic device or solar cell.

The electroluminescence phenomenon was discovered in 1907. Electrolumeniscence is the generation of light by an electric current passing through a material under an applied electric field. Electroluminescent light differs from thermal radiation. It contains a narrow range of wavelength. In LED, the spectral line width is 100 to 500 °A. The light may even be nearly perfectly monochromatic as in the diode Laser (0.1 to 1 Å)

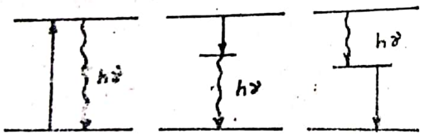
b. Radiative Transitions

The LED and semiconductor laser belong to the luminescent device family. Luminescence is observed in gaseous, liquid and solid substances. Luminescence is the emission of optical radiation (ultroviolet, visible or infrared) as a result of electronic excitation of a material. It is not associated with heat. Thus luminescent radiation is a cold radiation.

The luminescence effect is based on excitation. Under the action of certain external energy source, an electron is excited to a higher empty energy levels. This state is not stable. Upon termination of the excitation, the electrons return to the lower unoccupied levels. During this process a part of the energy is expended in the form of heat. The other portion of the electron can give off the energy in the form of quantum of electromagnetic radiation (photon). It is the flow of photons that is observed as luminescent radiations.

Luminescence may be classified by the method of electron excitation.

(i) Photoluminescence - excitation by light
(ii) electroluminescence - excitation by electric field.
(iii) Cathodoluminescence - excitation by an electron beam and
(iv) radioluminescence - excitation by high energy radiation.



(a) Direct (b) via imparity Centre (c) lattice defect

Fig 4-1

The luminescence properties exhibited by semiconductors are due to the transition of electrons from the valence band or from the impurity levels to the conduction band. These excess nonequilibrium electrons perform reverse quantum transition to the valence band or to the impurity levels and recombine with holes. The quantum transitions are divided into radiative and non-radiative. The emitted radiations are called recombination emission. In some of the semiconductor, radiative transition take place directly from the conduction band to valence band. In others, the transitions are excited only through the intermediary of one or two impurity levels or those of lattice defecates located in the band gap. These transitions are shown in Fig 4-1.

In non-radiative transition, the liberated energy is expended on heating the crystal lattice. This may occur either through impurity level or defect levels.

c. Luminescent Efficiency

For a given input excitation energy, the radiative recombination process is in direct competition with the non–radiative process. The quantum efficiency η is the fraction of the excited carriers that combine radiatively to the recombination. It may be written in terms of lifetime as

$$\eta = R_r/R = \tau_{nr}/(\tau_{nr} + \tau_r) \quad .....(1)$$

where τ_{nr} is the nonradiative lifetime and τ_r is the radiative lifetime and R_r and R are the radiative recombination rate and total recombination rate respectively. The recombination rate and lifetime are related for p-type layers by

$$R = (n - n_o)/\tau \quad .....(2)$$

For n - type layers by

$$R = (p - p_o)/\tau \quad .....(3)$$

where n_o and p_o are electron and hole concentration in thermal equilibrium and n and p are electron and hole concentration under optical excitation respectively. The minority carrier lifetime τ is given by

$$\tau = \tau_r \tau_{nr}/(\tau_{nr} + \tau_r) \quad ..... (4)$$

For high efficiency, the radiative lifetime τ_r should be very small.

d. Method of Excitation

Electroluminescence may be excited in different ways. They are (i) intrinsic (ii) avalanche (iii) tunneling and (iv) injection processes.

For the intrinsic excitation, a powder of semiconductor is suspended in a dielectric medium or in the form of thin homogeneous sublimated semicrystalline films. It will emit light only on applying an alternating voltage. Its efficiency is less than 1%. This mechanism is mainly caused by impact ionization of accelerated electron or field emission of electron from trapping centres.

For the avalanche excitation, a p-n junction is reverse biased. When the reverse voltage is in the avalanche breakdown, electron-hole pairs produced due to lionization. It may result in emission of light.

Electroluminescence may be excited by tunneling process. When a large reverse bias is given to a metal semiconductor barrier, holes at the metal Fermi level can tunnel into the valence band. It makes a radiative recombination with electrons that have tunneled from the valence band to the conduction band.

More promising are semiconductor emitters whose operation is based on the effect of injected electroluminescence. These are light - emitting diodes in which radiation originates through the recombination of electron injected in the p-region by the p-n junction biased in the forward direction. These injected electrons recombine in the narrow layer adjacent to the junction. The recombination luminescence so produced may be of sufficiently higher intensity. Light is emitted at small direct voltage of about several volts. The wavelength of recombination radiation is greater than the energy gap. The spectral distribution of recombination radiation lines within 200 to 600Å. The emission spectra obtained extends from far infrared to green and dark blue light region. The emission of light is easily modulated by the current of the light emitting diode.

e. Light Emitting Diodes

Recombination of charge carriers takes place at a p-n junction as electrons cross from the n-side and recombine with holes on the p-side. Free electrons are present in the conduction band of energy level while free holes are present in the valence band. Hence the electrons are at higher energy level than the holes. When recombination takes place, some of the energy is given up in the form of heat and light. If a semiconductor is translucent, the light will be emitted and the junction becomes a source of light i.e. a light Emitting Diode.

When a p-n junction diode is forward biased, carriers are injected across the junction to establish excess carriers above their thermal equilibrium values. These excess of carriers recombine and release electromagnetic energy. The electrons that are injected in the p-side make a downward transition from the conduction band to the valence band and recombine with valence band and recombine with holes. During this process photons are emitted whose energy is $h\gamma = E_g$. The wavelength of the emitted radiation is

$$\lambda = hC/E_g \qquad ......(1)$$

where h is Planck's constant ($h = 6.62 \times 10^{-34}$ J–S) C is the velocity of the electromagnetic wave ($C = 3 \times 10^8$ m/s) and E_g is the band gap Substituting the values of h and C and expressing E_g in the ev in the above expression we get

$$\lambda = \frac{1.24}{E_g} \mu m \qquad ......(2)$$

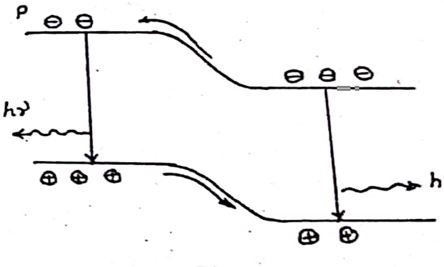


Fig 4-2

Light emission in the n-side of the junction takes place similarly except that the holes are the excess carrier. The downward transition of electron from the conduction band to valence band and subsequent emission of photons due to recombination with holes are shown in Fig 4-2.

The colour of the light emitted by the diode depends on the material used.

(i) Material for LED

Gallium arsenide (GaAs), Gallium Arsenede Phosphide (Ga As P) or Gallium Phosphide (GaP) are the semiconducting material which are used for the construction of LED. The radiation emitted by GaAs are in infrared region. GaAsP emitte either red or yellow colour while GaP emit red or green colour radiation.

The material which are used in LED should emit light that should be sensitive to human eye. Human eye is sensitive to light of energy $h\gamma \geq$ 1.8 ev. Hence the semiconductor should have a band gap greater than 1.8 ev. The most important semi conductor which is used in LED is $GaAs_{1-x}$m P_x For o 0.45, the energy gap is direct and increasing from $E_g = 1.427$ ev at $0 < c < 0.45$ $x = 0$ to $E_g = 1.977$ and $x = 0.45$ For $x > 0.45$, the energy gap is indirect. For direct band gap semiconductor, such as GaAs and $GaAs_{1-x}\,p_x (x \leq 0.45)$, interband transition may occur with high probability. The photon energy then approximately equal to the band gap energy of the semiconductor. The radiative transition is predominant in direct gap material. When $x > 0.45$, the probability for interband transition is extremely small. Therefore for indirect band gap semiconductor such as Gap, special recombination centres are incorporated to enhance the radiative process.

An efficient radiative recombination center in $GaAS_{1-x}\,P_x$ can be formed by adding an impurity such as nitrogen. The impurity replaces phosphorus atom in the lattice sites. Even though the outer electronic structure of the two atoms are the same, the core structures are different. This difference gives rise to an electron trap level close to the conduction band. A recombination center produced in this way is called an isoelectronic center. Normally these are neutral. In p-type material an injected electron is first trapped at the center. The negatively charged center then captures a hole from the valence band to form the bound excitation. Then the annihilation of electron-hole

pair yields a photon whose energy is equal to band gap minimum an energy approximately equal to the binding energy of the centers. Thus the probability of direct transition is greatly enhanced. This radiative recombination mechanism is predominant in indirect bandgap materials such as Gap. In the case of $GaAs_{1-x}\,P_x$, the efficiency without nitrogen drops sharply, but with nitrogen, the efficiency is considerable higher. The nitrogen doped alloy also show, a shift of the peak emission wavelength.

(ii) LED configuration

The basic LED structure is shown in fig4.3. A n-type epitaxial layer is grown upon a substrate and a p-region is created by diffusion. Charge carrier recombination occur in the p-region. Therefore it is kept uppermost. The p-region, therefore, becomes the surface of the device. The metal film anode connections are designed to allow most of the light to be emitted. This is done by making connections to the outside edges of the p-type layer. A gold film is applied to the bottom of the substrate to reflect as much of the incident light as possible, towards the surface of the device and to provide a cathode connection. This diode emits light only when it is forward biased.

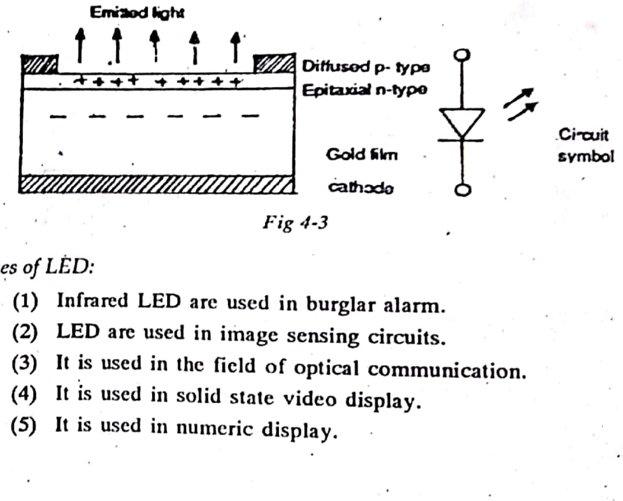


Fig 4-3

Uses of LED:

1. Infrared LED are used in burglar alarm.
2. LED are used in image sensing circuits.
3. It is used in the field of optical communication.
4. It is used in solid state video display.
5. It is used in numeric display.

2. PHOTODEVICES

a. Photoconduction

At any temperature, a semiconductor has a certain amount of free electrons and holes in equilibrium. The density of the free electrons depend upon the thermal generations of the carriers.

$$n_i = N_s e^{-\Delta E/2kT} \qquad \text{......(1)}$$

The absorption of light by a substance is associated with a release of free electron. If the energy of the incident radiation is greater than the energy of the band gap width of a semiconductor ($h\nu > \Delta E$), there will be a non- equilibrium of electron-hole pairs. These pairs add to the semiconductor conductivity by the value.

$$\Delta\sigma = q(\mu_n \Delta n + \mu_p \Delta p \qquad \text{.....(2)}$$

In a wide band gap semiconductor, the density of Δn and Δp will be greater than the density of electrons and holes in the equilibrium. Hence the photosensitivity will be higher than the semiconductor with narrow band gap width.

The mobilities of electrons and holes will be different in one and the same semiconductor. So we can consider the carries of only one sign in calculation. For n-type semiconductor,

$$\Delta\sigma = q\,\mu_n \Delta n \qquad \text{......(3)}$$

The densities of non-equilibrium electrons and holes are proportional to the radiant energy absorbed per unit time per unit volume. The energy absorbed by a unit area in unit time is

$$-\frac{dL}{dx} = KL \qquad \text{......(4)}$$

where K is the absorption coefficient and x is the layer thickness.

The densities of non-equilibrium electron and holes generated is proportional to the absorbed radiant energy.

$$\Delta p = \Delta n = \beta k L \qquad \text{......(5)}$$

where β is the quantum yield. The conductivity of the illuminated crystal increases with time. If the electron-hole pair density is increased, their recombination rate also increases. So for a given illumination intensity, a steady value of conductivity is set up in a certain time. At the steady value of the photoconductivity, the rate of carrier generation is equal to that of recombination.

The steady carrier density is produced in the process of illumination. This determines the steady photo conductivity. This must be proportional to the product of the densities of Δp and Δn generated under illumination in unit time by the life time of the carrier.

For electrons $\Delta n_{st} = \beta k L \tau_n$ ....(6)

For holes $\Delta p_{st} = \beta K L \tau_p$ ....(7)

τ_p and τ_n depend on the process of recombination. If greater the number of holes in a semiconductor, the higher the probability of electron running into a hole. Thus the less time, it stays in the conduction band until it recombine with a hole. Hence the life time of electron is,

$$\tau_n = \frac{1}{\gamma_n p} \qquad \text{..... (8)} \quad \text{and}$$

the life time of hole is

$$\tau_p = \frac{1}{\gamma_p n} \qquad \text{..... (9)}$$

where γ_n and γ_p are the recombination coefficient for electrons and holes respectively. The recombination coefficient depends on the thermal velocity $\bar{v}$ of an electron and capture cross-section $q_{n,p}$

$$\therefore \quad \gamma_n = \bar{v}_n q_n \qquad \text{..... (10)}$$

The cross section of electron capture by a trap centre is determined by the potential distribution near this centre.

Linear recombination occur when the concentration of recombination centres is high and does not depend on illumination. In this case,

$$\tau_n = \frac{1}{\gamma_n p} = \text{constant} \quad \ldots\ldots (11)$$

The recombination rate can be found on multiplying the recombination probability by the density of carriers generated by radiation.

$$\Delta n\, q_n\, \bar{v}\, p = \frac{\Delta n}{\tau_n} \quad \ldots\ldots (12)$$

The analogous expression for the hole recombination rate will be of the form

$$\Delta p\, q_p\, \bar{v}\, n = \frac{\Delta p}{\tau_p} \quad \ldots\ldots (13)$$

A change in the density on non-equilibrium carriers resulted from generation and recombination is described by the differential equation

$$\frac{d(\Delta n)}{dt} = \beta KL - \frac{\Delta n}{\tau_n} \quad \ldots\ldots (14)$$

The left-hand side refers to a change in the density of carriers due to generation and recombination. The first term of the right hand side describes the generation and the second term recombination.

The solution this equation will be of the form.

$$\Delta n = \tau \beta KL(1 - e^{-t/\tau_n}) \quad \ldots\ldots (15)$$

when t → ∞, the excess carrier density tend to an invariable value which determines a steady value of photoconductivity.

If a sample is not illuminated, the expression (14) can be written as

$$\frac{d(\Delta n)}{dt} = -\frac{\Delta n}{\tau_n} \quad \ldots\ldots (16)$$

For the initial condition $\Delta n = \Delta n_{st}$ and when $t = 0$, a fall in the photoconductivity can be found by integrating.

$$\Delta n = \Delta n_{st} e^{-t/\tau} \quad \ldots\ldots (17)$$

Relaxation curve describing the fall and rise of photo conductivity in the case of illumination by a square pulse is shown in Fig 4.4. By relaxation is meant the transient time required for the photo conductivity to reach a steady value after the light has been switched on or the time takes to a steady value of the dark conductivity after turning of the light. The curve shows that the rise time and the fall time of photoconductivity are equal.

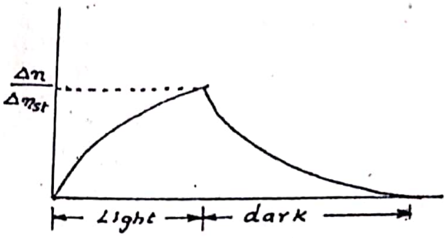


Fig 4-4

b. Photodiode

According to the quantum theory, light or any other radiation consists of elementary particles known as quanta or photon. Each photon possesses a definite amount of energy $h\gamma$. where γ is the frequency of radiation and h is the Planck's constant. When a radiation falls on the surface of a semiconductor material, a portion of the energy is absorbed by the semiconductor provided the energy of each photon $h\gamma$ is greater than the forbidden energy gap E_g of the semi conductor. As a result of the absorption, the valence electron of the semiconductor material increases. This property is known as the internal photoelectric effect which was discovered in 1873 at the time of studying the electrical conductivity of selenium in the presence of light. Devices like photodiode, phototransistor and solar cell are based on this effect. The internal photoeffect must not be confused with external photoelectric effect where free electrons are emitted from the surface of the semiconductor when radiation of appropriate frequency falls on it. Photoelectric cells are based on the external photoelectric effect.

If a reverse-biased p-n junction is illuminated, the current varies almost linearly with the light flux. This effect is exploited in the semi conductor photodiode. This device consists of a p-n junction embedded in a clear plastic. Radiation is allowed to fall upon one surface across the junction. The remaining sides of the plastic are either painted black or enclosed in a metallic case. The entire unit is extremely small compared to a photo tube. The semiconductor photodiode is of the order of tenth of an inch.

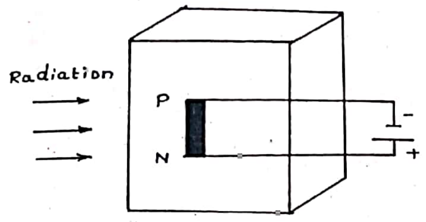


Fig 4.5

In order to operate a photodiode the battery E is connected in series with R_L and the diode is biased in the reverse direction as shown in fig 4.6. Without any radiation falling on the diode, it exhibits an extremely high resistance and a very small current of the order of a few micro- ampere flows in the circuit. This is the saturation current of a reverse biased p-n junction and is very sensitive to the variation of temperature. If the junction is illuminated by any radiation, an additional electron-hole pair will be created in the region depending on the intensity of the radiation. With a constant reverse voltage E, the reverse current increases proportionally with light intensity. Fig 4.7 shows the volt-ampere

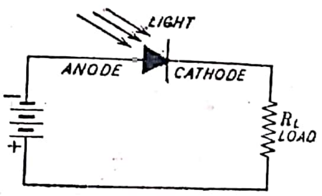


Fig 4-6

characteristic of photodiode with different value of light intensity as parameter.

If reverse voltage is excess of a few tenth of a volt are applied, an almost constant current is obtained. The dark current corresponds to the reverse saturation current due to thermally generated minority carriers. These minority carriers fall down to potential will at the junction whereas this barrier does not allow majority carrier to cross the junction. Now if light falls upon the surface, additional electron-hole pairs are formed. Since the concentration of majority carriers are greatly exceeds that of the minority carriers, the present increase in majority carriers is much smaller than the present increase in minority carriers. Hence it is justifiable to ignore the increase in majority density and to consider the radiation solely as a minority–carrier injector. These injected minority carriers diffuse to the junction, cross it and contribute to the current.

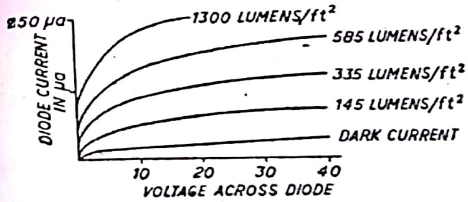


Fig 4-7

c. Phototransistor

The Phototransistor is a more sensitive device than the photodiode so far as the detection of radiation is concerned. The photo transistor is the same as the conventional transistor except that it uses radiation instead of electricity as the input signal. An important advantage feature of the conventional phototransistor over the photodiode is the current amplification produced in a transistor.

A p-n-p or n-p-n germanium transistor wafer is mounted so that the lens focuses the light ray on the area around the base- emitter junction. Leads are connected to the collector and the emitter region of the wafer. In some phototransistor a thread lead is connected to the base layer of the wafer as in conventional transistor. Fig 5-8 shows the symbols for p-n-p and n-p-n phototransistor.

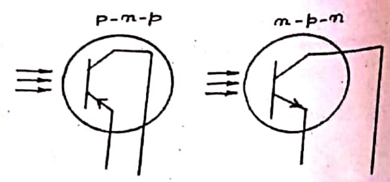


Fig 4.8

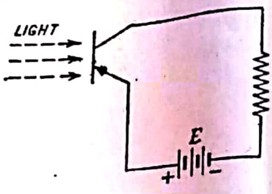


Fig 4-9

Fig 4.9 shows a common emitter type of phototransistor. The principle of operation of the device is as follows

When the device is biased as shown in Fig 4.9, the collector junction will be reverse biased as so most of the applied voltage will appear across this junction Before the phototransistor is exposed to any light, a small dark current flows and is equal to the saturation of the back biased n-p junction.

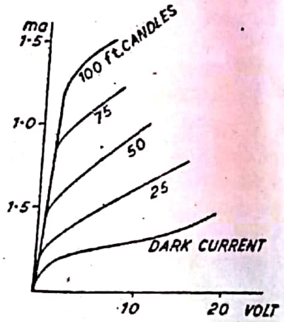


Fig 4-10

Since the base is open, $I_B = 0$, the collector current is given by

$$I_c = (\beta + 1)I_{co}$$

where β is the common emitter current gain and I_{co} is the reverse saturation collector current.

When light falls on the photo sensitive surface of the germanium wafer, additional electron - hole pairs are produced on absorption of light and a base current is set up. This base current is as usual amplified β times. Let I_p be the additional reverse saturation current component. The total collector current in the presence of radiation in given by

$$I_c = (\beta + 1)(I_{co} + I_p)$$

Fig 4.10 shows the characteristic of a phototransistor for different values of incident light power.

d. Digital Clock

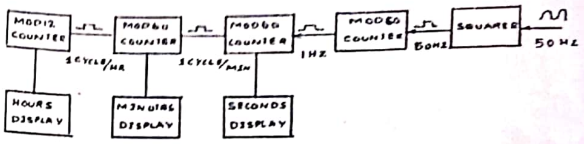


Fig 4.11

A digital clock is nothing but an electronic counter which counts clock pulses of frequency 1Hz. So the primary requirement of digital clock is to produce this clock pulse. The a.c. supply frequency of 50Hz which is maintained within ±0.1 per cent accuracy is made use of. The waveform of the supply voltage is sinusoidal but a clock pulse has to be a square wave. This sinusoidal wave is therefore transformed into a square wave by using a squaring circuit. To bring down the frequency to 1Hz a divider circuit is used. This divider is basically a Mod 50 counter which uses the 50Hz signal to produce 1Hz clock pulses. If this clock pulse train is applied to a mod 60 counter, the counter will count the number of clock pulses or in other words seconds. The output of the counter will be of frequency 1/60 Hz or 1 cycle per minute. It is fed to another mod 60 counter to obtain the minutes display. The output of the minute counter will have a frequency of 1 cycle per hour which is fed into a mod 12 counter to get the hours count. So the digital clock shows the hours, minutes and seconds which may be represented by the block diagram (Fig 4-11)

Using required number of flip-flops and different feed back connections, we get Mod 50, Mod 60 and Mod 12 counters. To obtain the digital display of hours, minutes and seconds, four input AND gates are used followed by the decoding procedure.

In the case of a digital wrist watch a battery is used as the primary source. This current is fed to a crystal oscillator to get an output of 10^6 Hz. A 10^{-6} divider converts the output of the oscillator to 1Hz. Then the process explained above is followed. If the supply frequency deviates too much from 50Hz, the clock will give wrong time. This difficulty is avoided in case of digital wrist watch which uses a d.c. source and is therefore more accurate.

e. Seven segment Displays (LED)

Panel indicators such as those used in calculators make use of a unit called Seven segment display Such a unit consists of LEDs arranged to form different segment of the figure of eight as shown in fig 4.12. Any number from 0 to 9 can be displayed by putting the proper LEDS on forward bias. When C and D are forward biased the number 1 will display, if b, c, g, f, e then 2, b, c, d, e, g then 3 and so on.

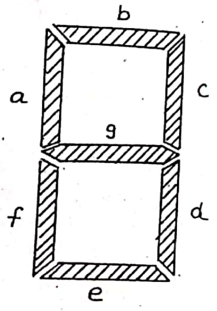


Fig 4-12

The anodes of all the segment may be connected together. The cathode of each segment is connected to an external input as shown in Fig 4.13. Such an arrangement is called a common anode seven segment display. Seven segment display are also available in which all the cathode are connected together and the anode to the external input.

The forward voltage to LED is 1.2V and forward current is 20mA. As compared to other display devices like LCD there is a larger current which the LED consumes. This is a major drawback or disadvantage of LEDs. Following example will make this point clear.

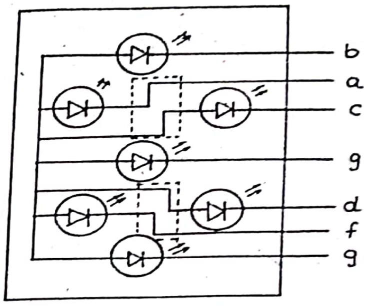


Fig 4.13

An electronic counter circuit needs about 100mA for its operation. If 4 seven segment display are added to the circuit, they will require about 500mA for operation. Hence for the whole unit more than 600 mA current will be needed which means a bulky power supply.

Advantage

1. LEDs are miniature in size and any number can be stacked together in a small space to form numerical display.
2. The light intensity of LEDs can be controlled easily by varying the current flow.
3. LEDs are rugged and can therefore withstand shocks and vibrations.
4. They can be operated over a wide range of temperatures from 0 to 70°C
5. They are very fast in operation. They can be switched on and off in a time less than 1ns
6. LEDs are available such that they emit light in different colours like red, green yellow and amber.
7. It has long life and has a high degree of reliability.

8. It has drive voltage and low noise.
9. Unlike LCD, the angle of viewing is not limited.
10. The brightness of these displays is sufficient for any environment.

Disadvantage :

1. These are not suited for large area display because of their high cost
2. Comparing to LCD, LEDs consume more energy.

f. Liquid Crystal Display (LCD)

Liquid crystals are used as displays in calculators and pocket computers etc. They are preferred over LEDs because they consume less power. They are two types of LCDs viz (i) Dynamic scattering type and (ii) Field effect type

(i) Dynamic scattering LCD : In normal liquid crystals the molecules are oriented in a definite crystal pattern. But on application of electric field to the liquid crystal disrupts the molecule alignment and cause a turbulance within the liquid.

The liquid is transparent when it is not activated by the flow of current. But when activated the molecular turbulance causes the light to be scattered in all directions and hence the activated areas appear bright. This phenomena is as Dynamic scattering. The actual liquid crystal material may be one of several organic compounds which exhibit the optical properties of a solid while retaining the fluidity of the liquid. Cholosteryl nonamate and p- azoxyanisole are such compounds.

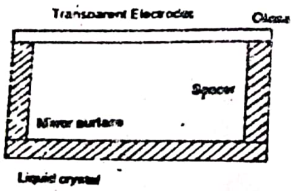
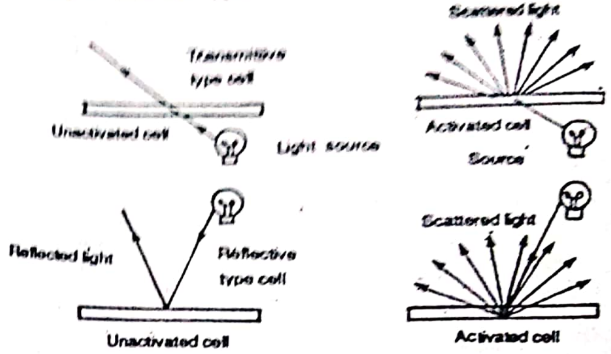


Fig 4-14

A liquid crystal consists of a layer of liquid crystal material sandwitched between glass sheets with transparent metal film electrodes deposited on the inside faces as shown in Fig 4.14.

When both glass sheets are transparent the cell is a transmitive type. On the other hand, when only one sheet is transparent and other has a reflective coating, the cell is known as reflective type cell. The working of these two types are clear form Fig 4.15



Transmitted type cell **Reflective type cell**

Fig 4.15

When transmitive type is not activated, the cell does not appear bright. When activated the incident light is diffused by scattering and cell appears bright even under high intensity light.

When reflected type is not activated, the light incident on it is simple reflected in usual manner and cell does not appear bright. When activated, the dynamic scattering occurs and cell appears quite bright.

(ii) Field Effect LCD : The construction of field effect LCD is similar to that of the dynamic scattering LCD. But in this two thin polarising optical filters are placed at the surface of each glass sheet. The liquid crystal material employed is known as twisted nematic type. When the cell is not energised, the material twists the light passing through it. The twisting allows the light to pass through the polarising filters. Thus in the case of transmittive type cell, the unenergised cell can appear dark against a bright background. When energised, the cell becomes transparent and disappears into the background.

As the LCD cells are reflectors or transmitted rather than generator of light, they consume very small amount of energy. The energy required by the cell is needed to activate the liquid crystal. The total current flow through a small seven segment display is about 25 μA for field effect cell. LCD require an a.c. Voltage supply either sinusoidal or square wave type. This is to avoid planting of cell electrode due the d.c. flow causing damage to the device. Dynamic scattering type requires 30V peak to peak square wave of 60Hz. A field effect cell requires 8V peak to peak square wave or sinusoidal type.

Fig 4.16 illustrates the square wave drive method for liquid crystal cells.

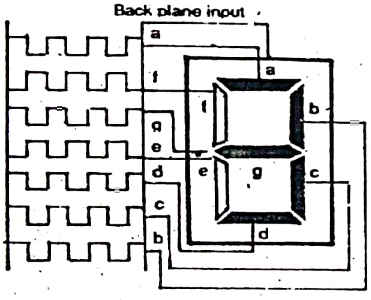


Fig 4-16

Unlike LED display which are usually small, LCD can fabricated of any convenient size. Power consumed in LCD is about 20 μw per segment or 140 μw per numeral when all segments are energised. In case of LEDs the power consumed is 400 mv per numeral including the series resistor. The major disadvantage of LCD is its decay time of 150 ms or more. This is very slow as compared to LED. A human eye can recognise fading of LCD segments switching off.

UNIT V

OPERATIONAL AMPLIFIER

1. Basic Operational Amplifier

a. Introduction:

The Operational Amplifier is a direct coupled high-gain differential input amplifier. Originally the operational amplifier had only one input and the output voltage was always inverted with respect to the input voltage. However common operational amplifier which are recently available are of differential type and these amplifiers have both inverting and non - inverting inputs.

An ideal operational amplifiers must have the following properties.

(i) Input resistance R_i must be infinite.
(ii) Output resistance R_o should be as small as zero.
(iii) Voltage gain A_v should be as high as infinite.
(iv) Band width should be as wide as infinite.
(v) Output voltage $V_0 = O$ when $V_1 = V_2$ and
(vi) The characteristics do not change with temperature.

Standard triangular symbol for an OP-AMP is shown in fig. 5-1. It has two input points and one output point. One is marked as plus and the other is marked as minus. The plus point may be compared as the emitter point of a transistor. So any signal applied at this point gets amplified without any phase inversion at output. So it is similar to common - base amplifier. Hence this input point is called the **non-inverting input point.**

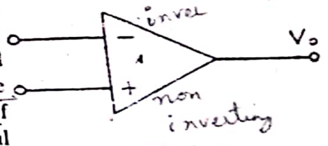


Fig. 5 - 1

A signal applied at the minus point is amplified with a phase inversion of 180°. So this point may be compared to the base terminal of a common emitter transistor amplifier. This input point is called **the inverting input point** of the OP-AMP

If two input signals V_1 and V_2 are applied the output voltage is related to the input by the equation.

$$V_0 = r\,(V_2 - V_1)$$

where A is the voltage gain

With $V_2 = 0$, $V_0 = -AV_1$, and with $V_1 = 0$, $V_0 = AV_2$. The sign (-) and (+) at the input terminal in the figure intended to serve as indication of their inversion and non- inversion. If $V_1 = V_2$ i.e if V_1 and V_2 make common voltage then $V_0 = 0$. Thus these plus and minus polarities indicate phase reversal only. It does not mean that voltages V_1 and V_2 are negative and positive respectively. It also does not mean that a positive input voltage has to be connected to the plus marked non-inverting point and negative input voltage to the negative-marked inverting point. All inputs and output voltages are referred to a common reference usually the ground.

b. Inverting Operational Amplifier

The basic inverting amplifier is shown in fig 5-2. The non inverting terminal has been grounded whereas R_1 connects the input signal V_1 to the inverting input. A feedback resistor R_f has been connected from the output to the inverting input.

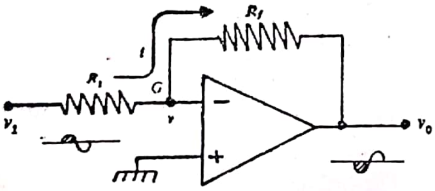


Fig. 5 - 2

The input voltage is V_1 and the output voltage is V_0. The gain of the operational amplifier is very large. So the voltage V at the inverting input terminal is very small. In fact , it will be close to the ground potential. It means that the point G is held virtually at ground potential irrespective of the magnitude of the potential V_1 and V_0. The current i flowing through R_1 is given by $i = (V_1 - V)/R_1$. The input impedance of OP-AMP is infinite. So the current i will flow through R_f and not into the OP-AMP. Applying Kirchhoff's current law at the point G, we can write,

$$\frac{V_1 - V}{R_1} = \frac{V - V_0}{R_f} \qquad \text{......(1)}$$

Since the point G is at a virtual ground, ie V = 0, we get from equation (1)

$$V_1/R_1 = -V_0/R_f \qquad \text{......(2)}$$

$$\therefore V_0/V_1 = -R_f/R_1 \qquad \text{......(3)}$$

$$\therefore \text{Gain } A_V = -R_f/R_1 \qquad \text{......(4)}$$

Thus the voltage gain is given by the ratio of the feedback resistance R_f to the input resistance R_1. The negative sign indicates that the output voltage is inverted with respect to the input voltage.

$$\text{Let} \quad R_f/R_1 = K.$$

$$\therefore V_0/V_1 = -K$$

$$\therefore V_0 = -K\,V_1 \qquad \text{......(5)}$$

Hence the closed -loop gain of the inverting amplifier depends on the ratio of the two external resistors R_f and R_1 and it is independent of amplifier parameters.

It is also seen that the OP-AMP works as a negative scaler. It scales the input i.e it multiplies the input by a minus constant factor K. It also works as a phase shifter and sign changer.

c. Non-inverting operational amplifier

The non - inverting amplifier is shown in fig. 5-3 . In this case, the input voltage V_2 is applied to the non-inverting terminal. So this circuit is called non-inverting amplifier.

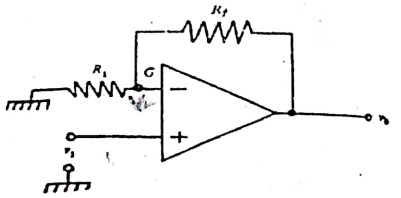


Fig 5-3

The gain of OP-AMP is infinite. So the potential of the point G is also V_2. The output voltage is V_0. The voltage across R_1 is V_2 and that across R_f is $(V_0 - V_2)$

$$\therefore i_1 = \frac{V_2}{R_1} \text{ and } i_2 = \frac{V_0 - V_2}{R_f} \qquad \text{......(1)}$$

Applying Kirchhoff's current law to the junction G, we get,

$$(-i_1) + i_2 = 0 \qquad \text{......(2)}$$

$$\therefore -\frac{V_2}{R_1} + \frac{V_0 - V_2}{R_f} = 0$$

$$\text{or } \frac{V_0 - V_2}{R_f} = \frac{V_2}{R_1}$$

$$\therefore \frac{V_0 - V_2}{V_2} = \frac{R_f}{R_1}$$

$$\frac{V_0}{V_2} - 1 = \frac{R_f}{R_1}$$

$$\therefore \frac{V_0}{V_2} = 1 + \frac{R_f}{R_1}$$

$$\therefore A_v = 1 + \frac{R_f}{R_1} \qquad \text{........(3)}$$

In this case, the gain is 1 plus the ratio of the two resistances R_f and R_1. Also the output voltage is in phase with the input voltage. This circuit offers a high input impedance and a low output impedance.

If $R_f = 0$ and $R_1 = \infty$, the gain of the amplifier is unity. Thus this circuit is referred as voltage follower or a **unit gain buffer.** This circuit can be used air an impedance matching device between a high-impedance source and a low impedance load.

d. Differential operational amplifier

A circuit that amplifies the difference between two signals is called a differential amplifier. This type of amplifiers are very useful in instrumentation circuit. A basic differential amplifier employing operational amplifier is shown in fig. 5-4

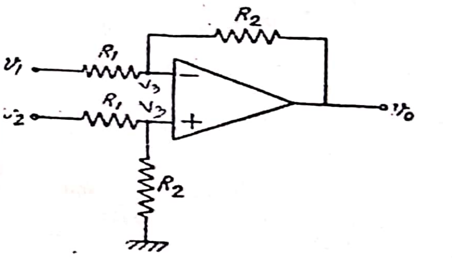


Fig. 5-4.

Here V_1 and V_2 are the input signal voltages and V_0 is the output voltage. Since the gain of the OP-AMP is infinite the potential of the points 1 and 2 will be the same say V_3. Application of Kirchhoff's current law at points 1 and 2 gives

$$\frac{V_1 - V_3}{R_1} = \frac{V_3 - V_0}{R_2} \qquad \text{.......... (1) and}$$

$$\frac{V_2 - V_3}{R_2} = \frac{V_3}{R_2} \qquad \text{.......... (2)}$$

R_1

Subtraction of (1) from (2) we get

$$V_0 = \frac{R_2}{R_1}(V_1 - V_2)$$

$$\therefore \frac{V_0}{(V_1 - V_2)} = \frac{R_2}{R_1}$$

$$\text{or} \qquad A_d = R_2/R_1 \qquad \text{......... (3)}$$

Thus its gain of DIFF-AMP is R_2/R_1.

Such a circuit is very useful in detecting very small difference is signals, since the gain R_2/R_1 can be chosen to be very large. For example if $R_2 = 100\ R_1$, then a small difference $(V_1 - V_2)$ is amplified 100 times.

e. Common - Mode Rejection Ratio (CMRR)

In an ideal DIFF-AMP, the output signal may be given by the equation,

$$V_0 = A_d (V_1 - V_2) \qquad \text{.......... (1)}$$

Where A_d is the gain of the differential amplifier. If $V_1 = V_2$, then $V_0 = 0$. That is the signals common to both inputs gets cancelled and produces no output voltage. This is true only for an ideal OP-AMP. But a practical OP-AMP exhibits some small response to the common mode component of the input voltage too. For example if one signal is $+25\ \mu V$ the second is $-25\mu V$, the output will not be exactly the same as if $V_0 = 900\ \mu V$ and $V_2 = 850\ \mu V$. even though the difference $V_d = 50\ \mu V$. is the same in the two cases. The output depends not only upon the difference signal V_d of the two signals, but also upon the average level called Common - mode signal V_c

The common mode signal

$$V_c = \frac{1}{2}(V_1 + V_2) \qquad \text{.....(2)}$$

For differential amplifier, the gain at the output with respect to the positive terminal is slightly different in magnitude to that of the negative terminal. So even with the same voltage applied to both inputs, the output is not zero. The output can be expressed as a linear combination of the two input voltages.

$$\therefore V_0 = A_1 V_1 + A_2 V_2 \qquad \text{.......(3)}$$

where A_1 is the voltage amplification from input 1 to the output under the condition that the input 2 is grounded and A_2 is the voltage amplification from input 2 to the output under the condition that the input 1 is grounded. We know that

$$V_c = \frac{1}{2}(V_1 + V_2)$$

$$V_d = (V_1 - V_2)$$

$$\therefore 2V_c = V_1 + V_2$$

$$V_d = V_1 - V_2 \qquad \text{----(4)}$$

From equation (4)

$$V_1 = V_c + V_d/2 \qquad \text{----(5)}$$

$$V_2 = V_2 - V_d/2 \qquad \text{-----(6)}$$

Substituting equation (5) and (6) in (3) we get

$$V_0 = A_d V_d + A_c V_c \qquad \text{---- (7)}$$

$$\text{Where } A_d = \frac{1}{2}(A_1 - A_2) \qquad \text{------ (8) and}$$

$$A_c = A_1 + A_2 \qquad \text{-----(9)}$$

The voltage gain for the difference signal is A_d and that for the common - mode signal is A_c.

The relative sensitivity of an OP-AMP to a difference signal as compared to a common - mode signal is called Common - Mode rejection ratio (CMRR). It gives the figure of merit ρ for the differential amplifier.

$$\therefore \text{CMRR} = \rho = \left|\frac{A_d}{A_c}\right| \qquad \text{--- (10)}$$

It is usually expressed in decibels (dB)

Problems

1 An inverting amplifier has $R_1 = 20$ kilo-ohm and $R_f = 100$ kilo-ohm. Find the output voltage, the input resistance and the input current for an input voltage 1V.

Sol:

$$\text{Output Voltage } V_0 = -\frac{R_f}{R_1} \times V_1$$

$$= -\frac{100}{20} \times 1 = -5V$$

Input resistance $R_{in} = R_1 = 20$ kilo-ohm.

Input current $= V_1/R_1 = 1/20 \times 10^{-3}$

$= 0.005$ mA.

2 In an inverting OP-AMP, the resistance R_1 is 10 kilo - ohm and the feedback resistance R_f is 100 kilo - ohm and $V_1 = 1V$. A load of 25 kilo-ohm is connected to the output terminal. Calculated the current through R_1 (2) the output voltage (3) the current through the load and (4) the total current in the output pin.

Sol :

(1) Current through R_1 is $I_1 = \dfrac{V_1}{R_1} = \dfrac{1}{10 \times 10^3} = 0.1$ mA

(2) Output voltage $V_0 = -\dfrac{R_f}{R_1} \times V_1 = -\dfrac{100}{10} \times 1 = -10V$

(3) Current through the load $I_L = \dfrac{V_0}{R_L} = \dfrac{10}{25 \times 10^3} = 0.4$ mA

(4) Total current $I = I_1 + I_L = 0.1 + 0.4 = 0.5$ mA.

3 Calculate the output Voltage of a non- inverting amplifier for a value of $V_1 = 2V$, $R_f = 500\ K\Omega$ and $R_1 = 100\ K\Omega$.

Sol :

For non-inverting amplifier $V_0 = \left(1 + \dfrac{R_f}{R_1}\right) V_1$

$$\therefore V_0 = \left(1 + \frac{500}{100}\right) 2 = 6 \times 2 = +12\ V.$$

2. Basic uses of Operational Amplifiers

a. OP-AMP as sign and scale changer and Phase Shifter

(i) ***Sign Changer :*** In an inverting OP-AMP, we know that,

$$A_V = -\frac{R_f}{R_1} = -\frac{Z'}{Z}$$

If $Z = Z'$, then $A_V = -1$ and the sign of the input signal has been changed. Hence such a circuits acts as a phase inverter. If two such amplifiers are connected in cascade, the output from the second stage equals the signal input without change of sign. Hence the output from the two stages are equal in magnitude but opposite in phase such a system is an excellent paraphase amplifier.

(ii) **Scale changer:** If the ratio $Z'/Z = K$, a real constant, then $A_V = -K$ and the scale has been multiplied by a factor $-K$ Usually in such a case of multiplication by a constant -1 or $-K$, Z and Z' are selected as resistances. If the inversion is not desired then a second amplifier with unit gain may be used in casecade with the scale changer.

(iii) Phase Shifter: Assume that Z and Z' are equal in magnitude but differ in angle. The operational amplifier shifts the phase of a sinusoidal input voltage while at the same time preserving its amplitude. Any phase shift from 0 to 360° (or ± 180°) may be obtained.

b. Integrating Amplifier

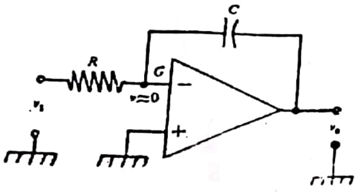


Fig. 5-5

An integrating amplifier is formed if a capacitor is connected between input and out as shown in fig. 5-5 . This circuit produces an output voltage that is proportional to the time integral of the input voltage. Hence this circuit is called an integrator. With the help of an integrating amplifier, it is possible to solve complex equations. Integrators are used in sweep generators, in filters and in simulation studies in analog computers.

If no current flows through the operational amplifier, we can write

$$I = I_c \quad \text{.......(1)}$$

$$\text{or} \quad \frac{V_1 - V_S}{R} = C.\frac{d(V_S - V_0)}{dt} \quad \text{......(2)}$$

V_S is very small so that

$$C\frac{dV_0}{dt} = -\frac{V_1}{R} \quad \text{......(3)}$$

$$\therefore dV_0 = -\frac{V_1}{CR}\,dt \quad \text{......(4)}$$

Taking integration on both sides

$$V_0 = -\frac{1}{CR}\int V_1\,dt \quad \text{...... (5)}$$

Thus the output voltage V_0 is equal to a constant $-1/CR$ times the integral of the input voltage V_1

c. Differentiators

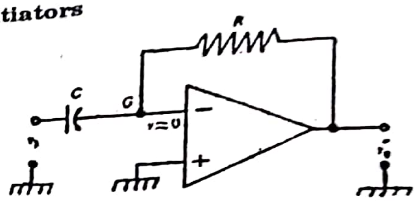


Fig. 5-6

Differentiation is the inverse of integration and may be obtained by interchanging R an C components of the integrator circuit. The differentiators is shown is Fig 5- 6.

If we neglect the current through the amplifier and V_S is taken to be negligibly small, we get

$$I_c = I \quad \text{... (1)}$$

$$\therefore C\frac{d(V_1 - V_S)}{dt} = \frac{V_S - V_0}{R} \quad \text{.... (2)}$$

Neglecting V_S we get

$$C\frac{dV_1}{dt} = -\frac{V_0}{R}$$

$$\therefore V_0 = -CR\frac{dV_1}{dt} \qquad \text{.... (3)}$$

The output voltage V_0 is equal to a constant –CR times the time derivative of the input voltage V_1

d. Adder or Summing Amplifier

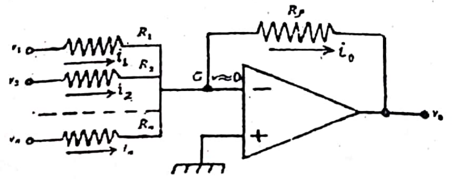


Fig. 5-7

An adder or a summing amplifier using an OP-AMP is shown in Fig. 5-7. Since the current flowing into the virtual ground is equal to that flowing out of it, we can write

$$i_1 + i_2 + \ldots i_n = i_0 \qquad \text{... (1)}$$

$$\text{or } \frac{V_1}{R_1} + \frac{V_2}{R_2} \ldots\ldots + \frac{V_n}{R_n} = -\frac{V_0}{R_f} \qquad \text{.... (2)}$$

$$\therefore V_0 = -\left(\frac{R_f}{R_1}V_1 + \frac{R_f}{R_2}V_2 + \ldots\ldots\ldots \frac{R_f}{R_n}V_n\right) \qquad \text{... (3)}$$

Let $R_1 = R_2 = \ldots\ldots\ldots\ldots = R_n = R.$

$$\therefore V_0 = -\frac{R_f}{R}(V_1 + V_2 + \ldots\ldots\ldots V_n) \qquad \text{... (4)}$$

If $R_f = R$, then

$$V_0 = -(V_1 + V_2 + \ldots\ldots V_n \qquad \text{... (5)}$$

i.e the output voltage V_0 is numerically equal to the algebraic sum of the input voltages.

e. D / A converters

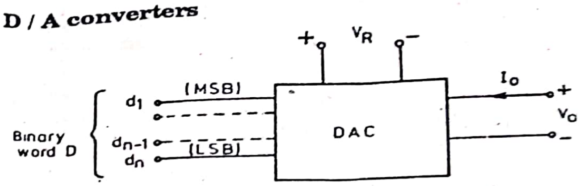


Fig. 5-8

The schematic D / A converter is shown in fig. 5-8. The input is an n-bit binary word and is combined with a reference voltage V_R to give an analog output signal. The output of a D/A convertor can be either a voltage or current. For a voltage output the D/A converter is mathematically described as

$$V_0 = K\,V_{FS}\,(d_1 2^{-1} + d_2 2^{-2} + \ldots\ldots\ldots + d_n 2^{-n})$$

Where

V_0	is the output voltage,
V_{FS}	is the full scale output voltage,
K	is the scaling factor usually adjusted to unity,
$d_1, d_2 \ldots\ldots d_n$	is n bit binary fractional word with the decimal point located at the left

d_1 is the most significant bit (MSB) with a weight of $V_{FS}/2$ and

d_n is the least significant bit (LSB) with a weight of $V_{FS}/2^n$

There are various ways to implement the above equation. In this we will discuss the following resistive methods.

(i) Binary weight method and

(ii) R -2R ladder method.

(i) D/A converter with Binary -weighted resistor:

$\frac{5}{10} =$ 0.5
5

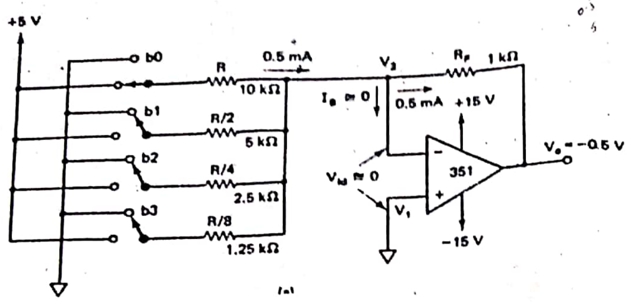


Fig. 5 - 9

Fig. 5-9 shows a D / A converter using an OP-AMP and binary weighted resistors. Here the OP-AMP is connected in the inverting mode. But it can also be connected in non- converting mode. d_0, d_1, d_2 and d_3 are four electronic switches which are controlled by binary input word. These switches are single pole double throw type. If the binary input to a particular switch is 1, it connects the resistance to the reference voltage. If the input bit is 0, the switch connects the resistor to the ground.

When the switch d_0 is closed, the voltage across R is 5V because $V_2 = V_1 = 0$ V. Therefore the current through R is 5V / 10KΩ = 0.5 mA. However, the input bias current is negligible. Hence the current through feedback resistor R_F is also 0.5 mA. This in turn produce an output voltage of is along closed, the current will be 1 mA. and the output voltage of −(1kΩ) (0.5mA) = −0.5V. Thus the OP–AMP work only as a current –to–voltage converter. When d1 is along closed, the current will be 1 mA and the output voltage V_0 is −1V. If both switches d_0 and d_1 are closed, the current through R_F will be 1.5 mA which will be converted to an output voltage of −1.5V.

Thus depending on whether switches d_0 to d_3 are open or closed, the binary weighted currents will be set-up in input resistors. The sum of these currents is equal to the currents through R_F which in turn is converted to a proportional output voltage. When all the switches are closed, the output will be maximum . The output voltage equation is given by

$$V_0 = R_F\left(\frac{d_0}{R} + \frac{d_1}{R/2} + \frac{d_2}{R/4} + \frac{d_3}{R/5}\right)$$

where each of the inputs d_3, d_2, d_1 and d_0 may either be high (+5V) or low (0V).

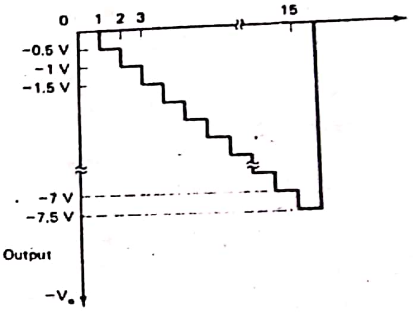


Fig 5-9(a)

Fig 5-9(a) shows analog outputs versus possible combination of inputs. The output is a negative going staircase waveform with 15 steps of −0.5V each. In practice, the steps may not all be the same size because of the variation in logic high voltage level. The size of the steps depend on the value of R_F. Therefore a desired step can be obtained by selecting a proper value of R_F.

The accuracy and stability depends upon the accuracy of the resistors and the tracking of each other with temperature. There are, however, a number of problems in this type. One of the disadvantages of binary weighted type is the wide range of resistors values required. For better resolution the input binary word length has to be increased.

Thus as the number of bit increases the range of resistance value increases. The fabrication of such a large resistance in IC is not practical. Also the voltage drop across such a large resistor due to the bias current would also affect the accuracy. The difficulty of achieving and maintaining accurate ratio restricts the use of weighted resistor type to below 8-bits.

(ii) *R-2R Ladder Method :*

The binary ladder is a resistive network whose output voltage is a properly weighted sum of the digital inputs. A ladder for 4 bits is shown in Fig 5-10. This ladder consists of resistances having only two valves. The left end of the ladder is terminated in a resistance of 2R.

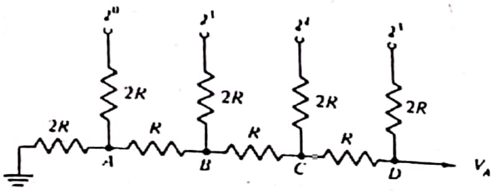


Fig 5-10

Let us consider all the inputs are at ground. Beginning at node A, the total resistance looking into the terminating resistor is 2R. The total resistance looking out towards the 2° input is also 2R. These two resistances can be combined to form an equivalent resistor of value R as shown in Fig. 5-11. Moving to node B, we can show that the total resistance looking into the branch towards node A is 2R, as is the total resistance looking out toward the 2^1 input. In a similar way we move to the node C and D and the resistance looking into the node is 2R. This is true regardless of whether the digital inputs are at gro

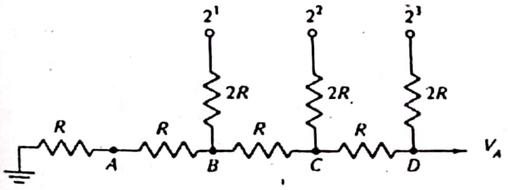


Fig 5-11

This resistance characteristic of ladder can be used to determine the output voltage for the various digital inputs. Let the digital input be 1000. Since there is no voltage source to the left of D, the equivalent circuit will be as shown in Fig. 5-12. In this case the output voltage is

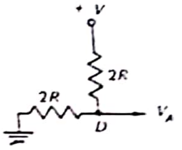


Fig. 5-12

$$V_A = +V\frac{2R}{2R + 2R} = +\frac{V}{2} \qquad \text{... (1)}$$

Thus a 1 in the MSB position will provide an output voltage of +V/2.

Now let the input be 0100. In this case the equivalent circuit will be as shown in Fig. 5-13. In this case we can show that the output will be

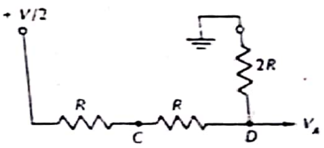


Fig. 5-13.

$$V_A = +\frac{V}{2} \cdot \frac{2R}{R + R + 2R}$$

$$= +\frac{V}{4} \qquad \text{... (2)}$$

Thus the second MSB provides an output voltage of +V/4. In a similar way we can show that the third will produce +V/8 and the fourth +V/16. Hence the total output voltage due to combination of input levels will be

$$V_A = \frac{V}{2} + \frac{V}{4} + \frac{V}{8} + \frac{V}{16} \qquad \text{... (3)}$$

This R-2R ladder can be used in the D/A converter. The D/A converter with R and 2R resistors is shown in Fig 5-14.

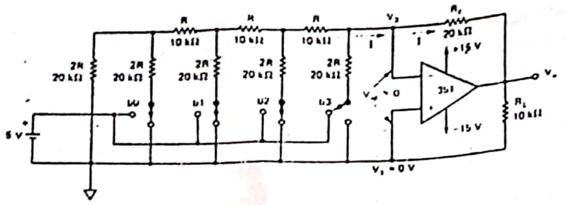


Fig. 5-14

The binary inputs are simulated by switches do through d_3. The output is proportional to the binary inputs. If the switches are closed then the input is high. If the switches are open, then the input is low. Assume that switch d_3 is closed. It is the most significant bit (MSB) so the input through d_3 is high (+5V). Other switches are connected to the ground. The resultant circuit will be as shown in Fig. 5-15. The resistance to the left of switch d_3 is 2R. In this figure, the (−) input is at virtual ground ($V_2 = 0$). Therefore the current through the equivalent resistance 2R in zero. However, the current through 2R connected to +5V is 5V/20KΩ = 0.25mA. The same current flows through R_F and in turn produces the output Voltage

$$V_0 = -(20K\Omega)(0.25mA) = -5V \qquad \text{..... (4)}$$

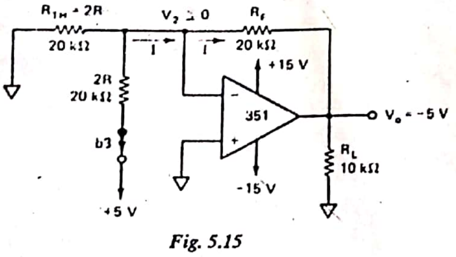


Fig. 5.15

Using the same analysis, the output Voltage corresponding to all possible combinations of binary inputs can be calculated. The output Voltage equation can be written as

$$V_0 = -R_F\left(\frac{d_3}{2R}+\frac{d_2}{4R}+\frac{d_1}{8R}+\frac{d_0}{16R}\right) \qquad \text{..... (3)}$$

where each of the inputs d_3, d_2, d_1 and do may be either high (+5V) or low (0V).

f. A/D converters

A/D converters convert an analog voltage to the digital output. As in the case of D/A converters analog converters are also specified as 8, 10, 12 or 16 bit. The block diagram of A/D converter is shown in Fig. 5-16. Fig. 5-16. Its function is just opposite to that of D/A converters. It accepts an analog input voltage V_A and produces an output binary word d_1, d_2 ..... d_n of functional value D so that

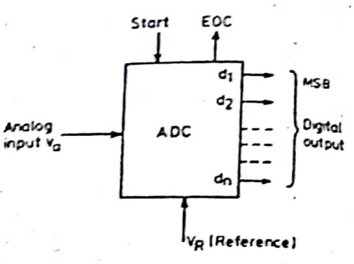


Fig. 5-16

$$D = d_1\,2^{-1} + d_2\,2^{-2} + \text{......}\ d_n\,2^{-n}$$

where d_1 is the most significant bit and d_n is the least significant bit. An A/D converters usually has two additional control lines. The START input to tell the A/D when to start the conversion and end of conversion (EOC) output to announce when the conversion is complete. A/D converters are classified into two groups according to their conversion technique. They are

(i) Direct type A/D converter and

(ii) Integrating type converter.

Direct type A/D converters compare a given analog signal with the internally generated equivalent signal. The integrating type of A/D converters perform conversion in an indirect manner by first changing the analog input signal to a linear function of time or frequency and then to a digital code. We will discuss successive approximation type converter and counter type converters. These two belong to direct type A/D converters.

(i) Successive Approximation Converter

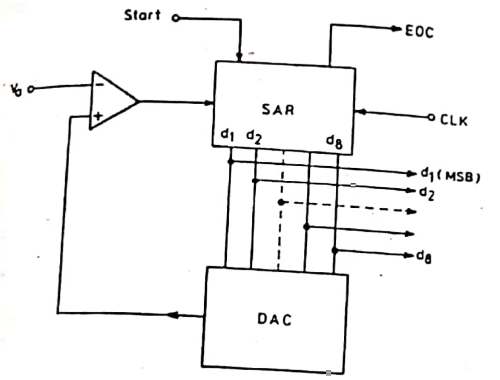


Fig. 5-17

Fig. 5-17 shows a successive approximation type of A/D converter. It uses a very efficient strategy to complete n-bit conversion in just n-clock periods. The heart of the circuit is an 8-bit successive-approximation register (SAR) Its output is applied to an 8-bit D/A converter. The analog output of the D/A converter is then compared to an analog input signal by the comparator. The output of the comparator is a serial data input to the SAR. The SAR then adjusts its digital output until it is equivalent to analog input. The 8-bit latch at the end of conversion holds onto the resultant digital data output. The circuit works as follows...

With the arrival of the START command, the SAR sets the MSB $d_1 = 1$, with other bits to Zero. So the trial code is 10000000. The D/A converter then generates an analog equivalent. The output V_d is now compared with the analog input V_A. If V_A is greater than the D/A converter output V_d, then 10000000 is less than the correct digital representation. The MSB is left at '1' and the next lower significant bit is made '1' and further tested.

However, if V_A is less than DAC output, then 10000000 is greater than the correct digital representation. So reset MSB to '0' and go on to the next lower significant bit. This procedure is repeated for all subsequent bits, one at a time until all bit positions have been tested. Whenever the DAC output crosses V_A, the comparator changes state and this can be taken as the end of conversion (EOC) command. The advantage of the successive approximation A/D converter is its high speed and excellent resolution. For example, the 8-bit successive approximation A/D converter requires only eight clock pulse.

(ii) Counter Type A/D Converter :

A higher resolution A/D converter using only one comparator could be constructed if a variable reference voltage were available. This reference voltage could then be applied to the comparator and when it became equal to the input analog voltage the conversion would be complete.

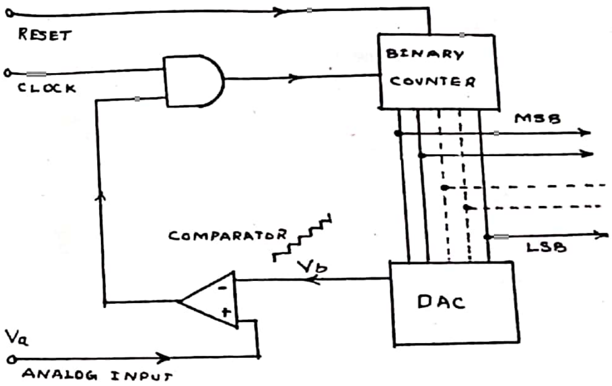


Fig. 5-18.

To construct such a converter, we begin with a simple binary counter. The digital output signal will be taken from this counter.

The output of this counter is given to a binary ladder to form a simple D/A converter. If a clock is applied to the input of the counter, the output of the binary ladder is the familiar staircase waveform. This waveform is exactly the reference voltage signal with a minimum of gating and control circuitry, this simple D/A converter, can be changed in the desired A/D converter.

Counter type A/D converter is shown in Fig. 5-18. The counter is reset to zero count by the reset pulse. Upon the release of reset, the clock pulses are counted by the binary counter. These pulses go through the AND gate which is enabled by the voltage comparator high output. The number of pulses counted increase with time. The binary word representing this count is used as the input of a D/A converter. The output of this is a staircase of the type shown in Fig. 5-18. The analog output V_d of DAC is compared to the analog input V_a by the comparator. IF $V_a > V_d$, the output of the comparator becomes high and the AND gate is enabled to allow the transmission of the clock pulses to the counter. When $V_a < V_d$ the output of the comparator becomes low and the AND gate is disabled. This stops the counting at the time $V_a \le V_d$. The digital output of the counter represents the analog input voltage V_a. For a new value of analog input V_a a second reset pulse is applied to clear the counter.

The counter-type A/D converter provides a very good method for digitizing to a high resolution. This method is much simple, but the conversion time required is longer. Since the counter always begins at zero and counts through its normal binary sequence, it may require as many as 2^n counts before conversion is complete. The average conversion time is 2^{n-1} counts. For example, a 12-bit system with 1 MHz clock frequency, the counter will take $(2^{12} - 1)\mu s = 4.095$ ms to convert a full scale input.

g. OP-AMP as a Comparator

A comparator circuit accepts input of linear voltage and provide a digital output that indicates when one input is less than or greater than the second. A basic comparator circuit can be represented as in Fig. 5-19. The output is a digital signal that stays at a high level when the non-inverting (+) input is greater than the inverting (−) input and switches to a lower voltage level, when the non-inverting input voltage goes below the inverting input reference voltage level.

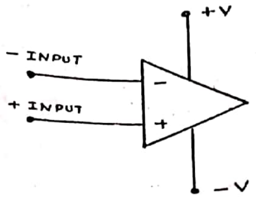


Fig. 5-19.

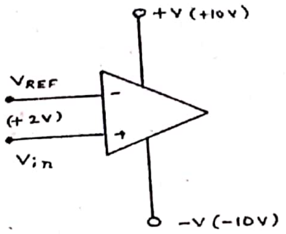


Fig. 5-20.

Fig. 5-20 shows a typical connection with one input connected to reference voltage, the other connected to the input signal voltage. As long as V_{in} is less than the reference voltage level of +2V, the output remains at a low-voltage level (near −10V). When the input rises just above +2V, the output quickly switches to a high voltage level (near +10V). Thus the high output indicates that the input signal is greater than +2V.

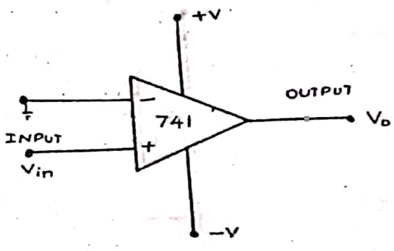


Fig. 5-21

We will examine the operation of a comparator using a 741 OP-AMP as shown in Fig. 5-21. With reference to input set to 0V, a sinusoidal signal is applied to the input terminal. It will cause the output to switch between its two output states as shown in Fig. 5-22. The input V_{in} going even a fraction of a millivolt above the 0V reference level will be amplified by the very high gain. So the output rises to its positive output saturation level and remains there while the input stays above $V_{ref} = 0V$. When the input drops just below the reference 0V level, the output is driven to its lower saturation level and stays there while the input remains below.

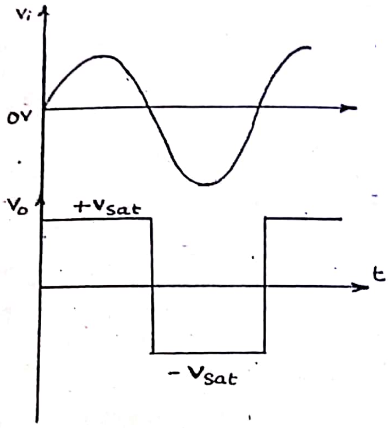


Fig. 5-22

$V_{ref} = 0V$. The Fig. 5-22 shows that the input signal is linea while the output is digital.

In actual circuits, the reference voltage is obtained from th supply voltage using a potential divider. The output of a comparato can also be made different from the saturation voltage. One such circuit is shown in Fig. 5-23 which is also a faster circuit. Dependin on the polarity of the input voltage, the amplifier gets large feedbac through one diode or the other and thus remains in the linear operatir

region. The output would switch its state when the input V_{in} would change its state compared to the trip point shown zero in figure.

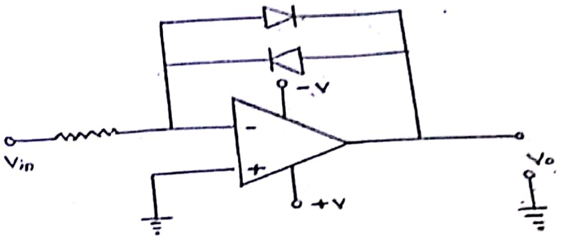


Fig. 5-23.

MODEL QUESTIONS AND ANSWER

I. Choose the Correct Answer

1) For an integrator using OP-AMP, the capacitor
   a) is connected between input and output terminal
   b) is connected in series with output
   c) is connected in series with input
   d) None of the above.
2) In an ideal OP-AMP, the input impedance is
   a) Zero b) 50 Ω c) infinite d) may be any
3) The CMRR of an OP-AMP is defined as
   a) output impedance/input impedance
   b) noise power at the output/noise power at input
   c) Differential gain/common mode gain
   d) Common mode gain/Differential gain
4) An OP-AMP with an ideal capacitor in the feedback pa works as
   a) a differentiator b) an integrator
   c) Wide band amplifier d) an adder