

# UNIT I

## PN DIODE AND ITS APPLICATIONS

### INTRODUCTION

The current-voltage characteristics is of prime concern in the study of semiconductor devices with light entering as a third variable in optoelectronics devices. The external characteristics of the device is determined by the interplay of the following internal variables:

1. Electron and hole currents
2. Potential
3. Electron and hole density
4. Doping
5. Temperature

### Semiconductor equations

The semiconductor equations relating these variables are given below:

#### Carrier density:

$$n = n_i \exp\left(\frac{E_{FN} - E_i}{kT}\right) \quad (1)$$

$$p = n_i \exp\left(\frac{E_i - E_{FP}}{kT}\right) \quad (2)$$

where  $E_{FN}$  is the electron quasi Fermi level and  $E_{FP}$  is the hole quasi Fermi level. These two equations lead to

$$np = n_i^2 \exp\left(\frac{E_{FN} - E_{FP}}{kT}\right) \quad (3)$$

In equilibrium  $E_{FN} = E_{FP} = \text{Constant}$

#### Current:

There are two components of current; electron current density  $J_N$  and hole current density  $J_P$ .

There are several mechanisms of current flow:

- (i) Drift
- (ii) Diffusion
- (iii) thermionic emission
- (iv) tunneling

The last two mechanisms are important often only at the interface of two different materials such as a metal-semiconductor junction or a semiconductor-semiconductor junction where the two semiconductors are of different materials. Tunneling is also important in the case of PN junctions where both sides are heavily doped.

In the bulk of semiconductor, the dominant conduction mechanisms involve drift and diffusion.

The current densities due to these two mechanisms can be written as

$$J_N = qn\mu_N\varepsilon + qD_N \frac{dn}{dx} \quad (4)$$

$$J_P = qn\mu_P\varepsilon + qD_P \frac{dp}{dx} \quad (5)$$

where  $\mu_N, \mu_P$  are electron and hole mobilities respectively and  $D_N, D_P$  are their diffusion constants.

### Potential:

The potential and electric field within a semiconductor can be defined in the following ways:

$$(i) \quad \Psi = -\frac{E_c}{q} + \text{constant}; \quad \varepsilon = \frac{1}{q} \frac{dE_c}{dX}$$

$$(ii) \quad \Psi = -\frac{E_v}{q} + \text{constant}; \quad \varepsilon = \frac{1}{q} \frac{dE_v}{dX}$$

$$(iii) \quad \Psi = -\frac{E_i}{q} + \text{constant}; \quad \varepsilon = \frac{1}{q} \frac{dE_i}{dX}$$

$$(iv) \quad \Psi = -\frac{E_o}{q} + \text{constant}; \quad \varepsilon = \frac{1}{q} \frac{dE_o}{dX}$$

All these definitions are equivalent and one or the other may be chosen on the basis of convenience. The potential is related to the carrier densities by the Poisson equation: -

$$\frac{\partial^2 \Psi}{\partial X^2} = -\frac{q}{\varepsilon} (p - n + N_D^+ - N_A^-) \quad (6)$$

where the last two terms represent the ionized donor and acceptor density.

### Continuity equations

These equations are basically particle conservation equations:

Electron continuity equation: 
$$\frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_N}{\partial x} + G_N - R_N \quad (7)$$

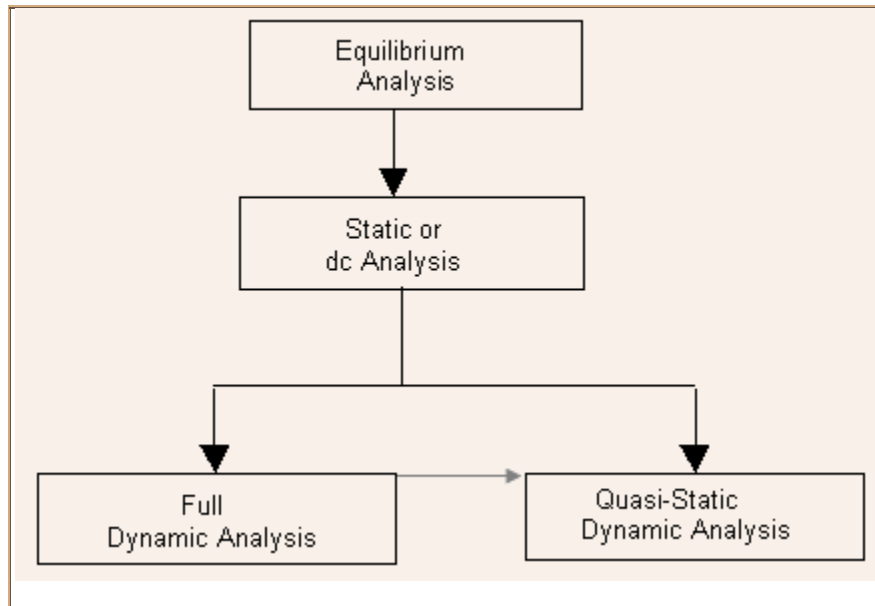
Hole continuity equation: 
$$\frac{\partial p}{\partial t} = -\frac{1}{q} \frac{\partial J_P}{\partial x} + G_P - R_P \quad (8)$$

Where G and R represent carrier generation and recombination rates. Equations **(1-8)** will form the basis of most of the device analysis that shall be discussed later on. These equations require models for mobility and recombination along with models of contacts and boundaries.

### Analysis Flow

Like most subjects, the analysis of semiconductor devices is also carried out by starting from simpler problems and gradually progressing to more complex ones as described below:

- (i) Analysis under zero excitation i.e. equilibrium.
- (ii) Analysis under constant excitation: in other words dc or static characteristics.
- (iii) Analysis under time varying excitation but with quasi-static approximation dynamic characteristics.
- (iv) Analysis under time varying excitation: non quasi-static dynamic characteristics.



Even though there is zero external current and voltage in equilibrium, the situation inside the device is not so trivial. In general, voltages, charges and drift-diffusion current components at any given point within the semiconductor may not be zero.

Equilibrium in semiconductors implies the following:

$$\frac{\partial Z}{\partial t} = 0$$

(i) steady state:

Where  $Z$  is any physical quantity such as charge, voltage electric field etc

(ii) no net electrical current and thermal currents:

Since current can be carried by both electrons and holes, equilibrium implies zero values for both net electron current and net hole current. The drift and diffusion components of electron and hole currents need not be zero.

$$\frac{dE_F}{dX} = 0$$

(iii) Constant Fermi energy:

The only equations that are relevant (others being zero!) for analysis in equilibrium are:

**Poisson Eq:**

$$\frac{\partial^2 \Psi_0}{\partial X^2} = -\frac{q}{\epsilon} (p_0 - n_0 + N_D^+ - N_A^-)$$

$$n_0 = n_i \exp\left(\frac{E_F - E_i}{kT}\right)$$

$$p_0 = n_i \exp\left(\frac{E_i - E_F}{kT}\right)$$

$$n_0 p_0 = n_i^2$$

In equilibrium, there is only one independent variable out of the three variables :  $n_0, p_0, \Psi_0$

If one of them is known, all the rest can be computed from the equations listed above. We shall take this independent variable to be potential.

The analysis problem in equilibrium is therefore determination of potential or equivalently, energy band diagram of the semiconductor device.

This is the reason why we begin discussions of all semiconductor devices with a sketch of its energy band diagram in equilibrium.

## Energy Band Diagram

This diagram in qualitative form is sketched by following the following procedure:

1. The semiconductor device is imagined to be formed by bringing together the various distinct semiconductor layers, metals or insulators of which it is composed. The starting point is therefore the energy band diagram of all the constituent layers.
2. The band diagram of the composite device is sketched using the fact that after equilibrium, the Fermi energy is the same everywhere in the system. The equalization of the Fermi energy is accompanied with transfer of electrons from regions of higher Fermi energy to region of lower Fermi energy and viceversa for holes.
3. The redistribution of charges results in electric field and creation of potential barriers in the system. These effects however are confined only close to the interface between the layers. The regions which are far from the interface remain as they were before the equilibrium

**Analysis in equilibrium:** Solution of Poisson's Equation with appropriate boundary conditions -

Non-equilibrium analysis:

- The electron and hole densities are no longer related together by the inverse relationship of Eq. (5) but through complex relationships involving all three variables  $n$ ,  $p$ ,  $\psi$
- The three variables are in general independent of each other in the sense that a knowledge of two of them does not lead automatically to a knowledge of the third.
- The concept of Fermi energy is no longer valid but new quantities called the quasi-Fermi levels are used and these are not in general constant.
- For static or dc analysis, the continuity equation becomes time independent so that only ordinary differential equations need to be solved.
- For dynamic analysis however, the partial differential equations have to be solved increasing the complexity of the analysis.

## Analysis of Semiconductor Devices

There are two complementary ways of studying semiconductor devices:

- (i) Through numerical simulation of the semiconductor equations.
- (ii) Through analytical solution of semiconductor equations.

- There are a variety of techniques used for device simulation with some of them starting from the drift diffusion formalism outlined earlier, while others take a more fundamental approach starting from the Boltzmann transport equation instead.
- In general, the numerical approach gives highly accurate results but requires heavy computational effort also.

- The output of device simulation in the form of numerical values for all internal variables requires relatively larger effort to understand and extract important relationships among the device characteristics.

The electrons in the valence band are not capable of gaining energy from external electric field and hence do not contribute to the current. This band is never empty but may be partially or completely with electrons. On the contrary in the conduction band, electrons are rarely present. But it is possible for electrons to gain energy from external field and so the electrons in these bands contribute to the electric current. The forbidden energy gap is devoid of any electrons and this much energy is required by electrons to jump from valence band to the conduction band.

In other words, in the case of conductors and semiconductors, as the temperature increases, the valence electrons in the valence energy move from the valence band to conductance band. As the electron (negatively charged) jumps from valence band to conductance band, in the valence band there is a left out deficiency of electron that is called Hole (positively charged).

Depending on the value of  $E_{\text{gap}}$ , i.e., energy gap solids can be classified as metals (conductors), insulators and semi conductors.

### **Semiconductors**

- Conductivity in between those of metals and insulators.
- Conductivity can be varied over orders of magnitude by changes in temperature, optical excitation, and impurity content (doping).
- Generally found in column IV and neighboring columns of the periodic table.
- Elemental semiconductors: Si, Ge.
- Compound semiconductors:

*Binary :*

GaAs, AlAs, GaP,  
etc. (III-V).

ZnS, ZnTe, CdSe (II-  
VI).

SiC, SiGe (IV  
compounds).

- *Ternary :* GaAsP.
- *Quaternary :* InGaAsP.
- Si widely used for rectifiers, transistors, and ICs.
- III-V compounds widely used in optoelectronic and high-speed applications.

## ❖ Applications

- Integrated circuits (ICs) SSI, MSI, LSI, and VLSI.
- Fluorescent materials used in TV screens II-VI (ZnS).
- Light detectors InSb, CdSe, PbTe, HgCdTe.
- Infrared and nuclear radiation detectors Si and Ge.
- Gunn diode (microwave device) GaAs, InP.
- Semiconductor LEDs GaAs, GaP.
- Semiconductor LASERS GaAs, AlGaAs.

## ❖ Energy Gap

- Distinguishing feature among metals, insulators, and semiconductors.
- Determines the absorption/emission spectra, the leakage current, and the intrinsic conductivity.
- Unique value for each semiconductor (e.g. 1.12 eV for Si, 1.42 eV for GaAs) function of temperature.

## ❖ Impurities

- Can be added in precisely controlled amounts.
- Can change the electronic and optical properties.
- Used to vary conductivity over wide ranges.
- Can even change conduction process from conduction by negative charge carriers to positive charge carriers and vice versa.
- Controlled addition of impurities doping.

### Energy Bands and Charge Carriers in Semiconductors

#### Bonding Forces and Energy Bands in Solids

- Electrons are restricted to sets of discrete energy levels within atoms, with large gaps among them where no energy state is available for the electron to occupy.
- Electrons in solids also are restricted to certain energies and are not allowed at other energies.
- Difference → in the solid, the electron has a *range (or band)* of available energies.
- The discrete energy levels of the isolated atom spread into bands of energies in the solid because
  - i) in the solid, the wave functions of electrons in neighboring atoms overlap, thus, it affects the potential energy term and the boundary conditions in the Schrödinger equation, and different energies are obtained in the solution, and
  - ii) an electron is not necessarily localized at a particular atom.
- The influence of neighboring atoms on the energy levels of a particular atom can be treated as a small perturbation, giving rise to shifting and splitting of energy states into energy bands.

## Bonding Forces in Solids

### ❖ Ionic Bonding

- Example: NaCl.
- Na ( $Z = 11$ ) gives up its outermost shell electron to Cl ( $Z=17$ ) atom, thus the crystal is made up of ions with the electronic structures of the inert atoms Ne and Ar.
- Note: the ions have net electric charges after the electron exchange  $\rightarrow$   $\text{Na}^+$  ion has a net positive charge, having lost an electron, and  $\text{Cl}^-$  ion has a net negative charge, having acquired an electron.
- Thus, an electrostatic attractive force is established, and the balance is reached when this equals the net repulsive force.
- Note: all the electrons are tightly bound to the atom.
- Since there are no loosely bound electrons to participate in current flow  $\Rightarrow$  NaCl is a good insulator.

### ❖ Metallic Bonding

- In metals, the outer shell is filled by no more than three electrons (loosely bound and given up easily)  $\Rightarrow$  great chemical activity and high electrical conductivity.
- Outer electron(s) contributed to the crystal as a whole  $\Rightarrow$  solid made up of ions with closed shells immersed in a sea of free electrons, which are free to move about the crystal under the influence of an electric field.
- Coulomb attraction force between the ions and the electrons hold the lattice together.

### ❖ Covalent Bonding

- Exhibited by the diamond lattice semiconductors.
- Each atom surrounded by four nearest neighbors, each having four electrons in the outermost orbit.
- Each atom shares its valence electrons with its four nearest neighbors.
- Bonding forces arise from a quantum mechanical interaction between the shared electrons.
- Both electrons belong to each bond, are indistinguishable, and have opposite spins.
- No free electrons available at 0 K, however, by thermal or optical excitation, electrons can be excited out of a covalent bond and can participate in current conduction  $\Rightarrow$  important feature of semiconductors.

### ❖ Mixed Bonding

- Shown by III-V compounds bonding partly ionic and partly covalent.
- Ionic character of bonding becomes more prominent as the constituent atoms move further away in the periodic table, e.g., II-VI compounds.

## ❖ Energy Bands

- As isolated atoms are brought together to form a solid, the electron wave functions begin to overlap.
- Various interactions occur, and, at the proper interatomic spacing for the crystal, the forces of attraction and repulsion find a balance.
- Due to Pauli exclusion principle, the discrete energy levels of individual atoms split into bands belonging to the pair instead of to individual atoms.
- In a solid, due to large number of atoms, the split energy levels for essentially *continuous bands of energy*.

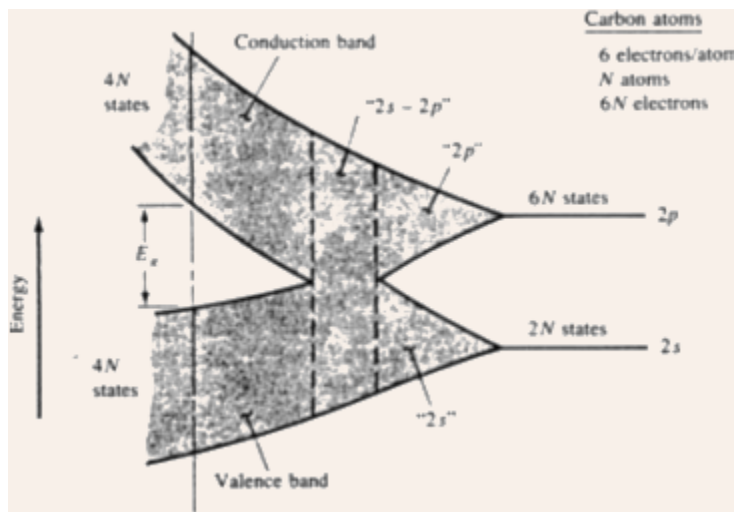


Fig.2.1 Splitting of individual energy levels to energy bands as atoms are brought closer together.

- Imaginary formation of a diamond crystal from isolated carbon atoms ( $1s^2 2s^2 2p^2$ ).
- Each atom has two 1s states, two 2s states, six 2p states, and higher states.
- For N atoms, the numbers of states are 2N, 2N, and 6N of type 1s, 2s, and 2p respectively.
- With a reduction in the interatomic spacing, these energy levels split into bands, and the 2s and 2p bands merge into a single band having 8N available states.
- As the interatomic spacing approaches the equilibrium spacing of diamond crystal, this band splits into two bands separated by an energy gap  $E_g$ , where no allowed energy states for electrons exist  $\Rightarrow$  *forbidden gap*.
- The upper band (called the *conduction band*) and the lower band (called the *valence band*) contain 4N states each.
- The lower 1s band is filled with 2N electrons, however, the 4N electrons residing in the original  $n = 2$  state will now occupy states either in the *valence band* or in the *conduction band*.
- At 0 K, the electrons will occupy the lowest energy states available to them  $\Rightarrow$  thus, the 4N states in the valence band will be completely filled, and the 4N states in the conduction band will be completely empty.

## ❖ Metals, Semiconductors, and Insulators

- For electrons to move under an applied electric field, there must be states available to them.
- A completely filled band cannot contribute to current transport; neither can a completely empty band.
- Thus, semiconductors at 0 K are perfect insulators.
- With thermal or optical excitation, some of these electrons can be excited from the valence band to the conduction band, and then they can contribute to the current transport process.
- At temperatures other than 0 K, the magnitude of the band gap separates an insulator from a semiconductor, e.g., at 300 K,  $E_g(\text{diamond}) = 5 \text{ eV}$  (insulator), and  $E_g(\text{Silicon}) = 1.12 \text{ eV}$  (semiconductor).
- Number of electrons available for conduction can be increased greatly in semiconductors by reasonable amount of thermal or optical energy.
- In metals, the bands are either partially filled or they overlap => thus, electrons and empty states coexist => great electrical conductivity.

## ❖ Direct and Indirect Semiconductors

- In a typical quantitative calculation of band structures, the wave function of a single electron traveling through a perfectly periodic lattice is assumed to be in the form of a plane wave moving in the x-direction (say) with propagation constant k, also called a *wave vector*.
- In quantum mechanics, the electron momentum can be given by  $\mathbf{p} = \hbar\mathbf{k}$ .
- The space dependent wave function for the electron is 
$$\psi_{\mathbf{k}}(\mathbf{x}) = U(\mathbf{k}, \mathbf{x}) e^{i\mathbf{k} \cdot \mathbf{x}} \quad (2.1)$$
 where the function  $U(\mathbf{k}, \mathbf{x})$  modulates the wave function according to the periodicity of the lattice.
- Allowed values of energy, while plotted as a function of k, gives the E-k diagram.
- Since the periodicity of most lattices is different in various directions, the E-k diagram is a complex surface, which is to be visualized in three dimensions.

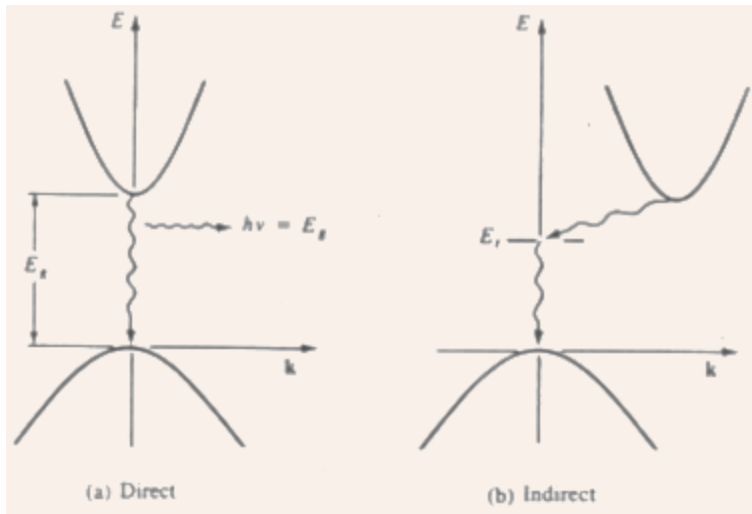


Fig.2.2 Direct and indirect transition of electrons from the conduction band to the valence band: (a) direct - with accompanying photon emission, (b) indirect via defect level.

- Direct band gap semiconductor: the minima of the conduction band and the maxima of the valence band occur at the same value of  $k$   $\Rightarrow$  an electron making the smallest energy transition from the conduction band to the valence band can do so without a change in  $k$  (and, the momentum).
- Indirect band gap semiconductor: the minima of the conduction band and the maxima of the valence band occur for different values of  $k$ , thus, the smallest energy transition for an electron requires a change in momentum.
- Electron falling from conduction band to an empty state in valence band  $\Rightarrow$  recombination.
- Recombination probability for direct band gap semiconductors is much higher than that for indirect band gap semiconductors.
- Direct band gap semiconductors give up the energy released during this transition ( $= E_g$ ) in the form of light  $\Rightarrow$  used for optoelectronic applications (e.g., LEDs and LASERS).
- Recombination in indirect band gap semiconductors occurs through some defect states within the band gap, and the energy is released in the form of heat given to the lattice.

### ❖ Variation of Energy Bands with Alloy Composition

- The band structures of III-V ternary and quaternary compounds change as their composition is varied.
- There are three valleys in the conduction band:  $\Gamma$  (at  $k = 0$ ), L, and X.
- In GaAs, the  $\Gamma$  valley has the minimum energy (direct with  $E_g = 1.43$  eV) with very few electrons residing in L and X valleys (except for high field excitations).
- In AlAs, the X valley has minimum energy (indirect with  $E_g = 2.16$  eV).

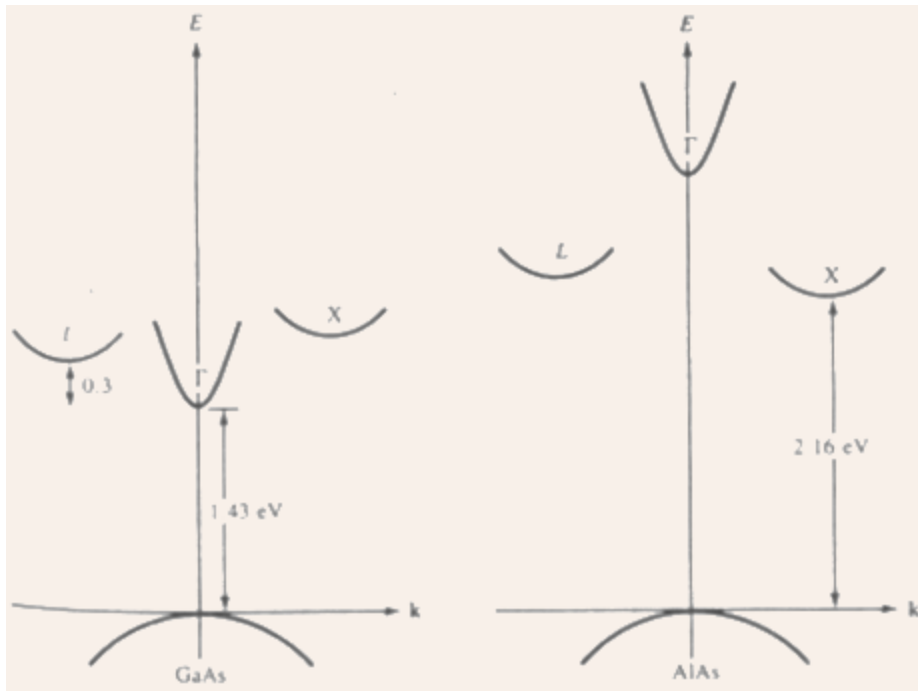


Fig.2.3 The E-k diagram of (a) GaAs and (b) AlAs, showing the three valleys (L,  $\Gamma$ , and X) in the conduction band.

### ❖ Charge Carriers in Semiconductors

- In a metal, the atoms are imbedded in a "sea" of free electrons, and these electrons can move as a group under the influence of an applied electric field.
- In semiconductors at 0 K, all states in the valence band are full, and all states in the conduction band are empty.
- At  $T > 0$  K, electrons get thermally excited from the valence band to the conduction band, and contribute to the conduction process in the conduction band.
- The empty states left in the valence band can also contribute to current conduction.
- Also, introduction of impurities has an important effect on the availability of the charge carriers.
- Considerable flexibility in controlling the electrical properties of semiconductors.

### ❖ Electrons and Holes

- For  $T > 0$  K, there would be some electrons in the otherwise empty conduction band, and some empty states in the otherwise filled valence band.
- The empty states in the valence band are referred to as holes.
- If the conduction band electron and the valence band hole are created by thermal excitation of a valence band electron to the conduction band, then they are called electron-hole pair (EHP).

- After excitation to the conduction band, an electron is surrounded by a large number of empty states, e.g., the equilibrium number of EHPs at 300 K in Si is  $\sim 10^{10}/\text{cm}^3$ , whereas the Si atom density is  $\sim 10^{22}/\text{cm}^3$ .
- Thus, the electrons in the conduction band are free to move about via the many available empty states.
- Corresponding problem of charge transport in the valence band is slightly more complex.
- Current transport in the valence band can be accounted for by keeping track of the holes themselves.
- In a filled band, all available energy states are occupied.
- For every electron moving with a given velocity, there is an equal and opposite electron motion somewhere else in the band.
- Under an applied electric field, the net current is zero, since for every electron  $j$  moving with a velocity  $v_j$ , there is a corresponding electron  $j'$  moving with a velocity  $-v_j$ .
- In a unit volume, the current density  $J$  can be given by

$$J = (-q) \sum_i^N v_i = 0 \quad (\text{filled band}) \quad (2.2)$$

where  $N$  is the number of electrons/cm<sup>3</sup> in the band, and  $q$  is the electronic charge.

- Now, if the  $j^{\text{th}}$  electron is removed and a hole is created in the valence band, then the net current density

$$J = (-q) \sum_i^N v_i - (-q)v_j \quad (j^{\text{th}} \text{ electron missing})$$

$$= (+q)v_j \quad (\text{since the first term is zero})$$

- Thus, the current contribution of the empty state (hole), obtained by removing the  $j^{\text{th}}$  electron, is equivalent to that of a positively charged particle with velocity  $v_j$ .
- Note that actually this transport is accounted for by the motion of the uncompensated electron  $j'$  having a charge of  $q$  and moving with a velocity  $v_j$ .
- Its current contribution  $(-q)(-v_j)$  is equivalent to that of a positively charged particle with velocity  $+v_j$ .
- For simplicity, therefore, the empty states in the valence band are called holes, and they are assigned positive charge and positive mass.
- The electron energy increases as one moves up the conduction band, and electrons gravitate downward towards the bottom of the conduction band.
- On the other hand, hole energy increases as one moves down the valence band (since holes have positive charges), and holes gravitate upwards towards the top of the valence band.

## ❖ Effective Mass

- The "wave-particle" motion of electrons in a lattice is not the same as that for a free electron, because of the interaction with the periodic potential of the lattice.
- To still be able to treat these particles as "free", the rest mass has to be altered to take into account the influence of the lattice.
- The calculation of effective mass takes into account the shape of the energy bands in three-dimensional k-space, taking appropriate averages over the various energy bands.
- The effective mass of an electron in a band with a given (E,k) relation is given by

$$m^* = \frac{[\hbar/(2\pi)]^2}{d^2E/dk^2} \quad (2.4)$$

---

**EXAMPLE 2.1:** Find the dispersion relation for a free electron, and, thus, observe the relation between its rest mass and effective mass.

**SOLUTION:** For a free electron, the electron momentum is  $\mathbf{p} = m_0\mathbf{v} = \hbar\mathbf{k}$ . Thus,  $E = m_0v^2/2 = \mathbf{p}^2/(2m_0) = \hbar^2\mathbf{k}^2/(2m_0)$ . Therefore, the dispersion relation, i.e., the E-k relation is parabolic. Hence,  $m^* = \hbar^2/(\partial^2E/\partial\mathbf{k}^2) = m_0$ . This is a very interesting relation, which states that for a free electron, the rest mass and the effective mass are one and the same, which is due to the parabolic band structure. Most materials have non-parabolic E-k relation, and, thus, they have quite different rest mass and effective mass for electrons.

Note: for severely non-parabolic band structures, the effective mass may become a function of energy, however, near the minima of the conduction band and towards the maxima of the valence band, the band structure can be taken to be parabolic, and, thus, an effective mass, which is independent of energy, may be obtained.

- Thus, the effective mass is an inverse function of the curvature of the E-k diagram: weak curvature gives large mass, and strong curvature gives small mass.
- Note that in general, the effective mass is a tensor quantity, however, for parabolic bands, it is a constant.
- Another interesting feature is that the curvature  $d^2E/dk^2$  is positive at the conduction band minima, however, it is negative at the valence band maxima.
- Thus, the electrons near the top of the valence band have negative effective mass.
- Valence band electrons with negative charge and negative mass move in an electric field in the same direction as holes with positive charge and positive mass.
- Thus, the charge transport in the valence band can be fully accounted for by considering hole motion alone.
- The electron and hole effective masses are denoted by  $m_n^*$  and  $m_p^*$  respectively.

Table 2.1  
The effective mass for electrons and holes.  
( $\times m_0$ , where  $m_0$  is the rest mass for electrons)

	Ge	Si	GaAs
$m_n^*$	0.55	1.1	0.067
$m_p^*$	0.37	0.56	0.48

## ❖ Intrinsic Material

- A perfect semiconductor crystal with no impurities or lattice defects.
- No carriers at 0 K, since the valence band is completely full and the conduction band is completely empty.
- For  $T > 0$  K, electrons are thermally excited from the valence band to the conduction band (EHP generation).
- EHP generation takes place due to breaking of covalent bonds  $\Rightarrow$  required energy =  $E_g$ .
- The excited electron becomes free and leaves behind an empty state (hole).
- Since these carriers are created in pairs, the electron concentration ( $n/\text{cm}^3$ ) is always equal to the hole concentration ( $p/\text{cm}^3$ ), and each of these is commonly referred to as the *intrinsic carrier concentration* ( $n_i$ ).
- Thus, for intrinsic material  $n = p = n_i$ .
- These carriers are not localized in the lattice; instead they spread out over several lattice spacings, and are given by quantum mechanical probability distributions.
- Note:  $n_i = f(T)$ .
- To maintain a steady-state carrier concentration, the carriers must also recombine at the same rate at which they are generated.
- Recombination occurs when an electron from the conduction band makes a transition (direct or indirect) to an empty state in the valence band, thus annihilating the pair.
- At equilibrium,  $r_i = g_i$ , where  $g_i$  and  $r_i$  are the generation and recombination rates respectively, and both of these are temperature dependent.
- $g_i(T)$  increases with temperature, and a new carrier concentration  $n_i$  is established, such that the higher recombination rate  $r_i(T)$  just balances generation.
- At any temperature, the rate of recombination is proportional to the equilibrium concentration of electrons and holes, and can be given by  $r_i = \alpha_r n_i p_i = \alpha_r n_i^2 = g_i$  (2.5) where  $\alpha_r$  is a constant of proportionality (depends on the mechanism by which recombination takes place).

## ❖ Extrinsic Material

- In addition to thermally generated carriers, it is possible to create carriers in the semiconductor by purposely introducing impurities into the crystal  $\Rightarrow$  *doping*.
- Most common technique for varying the conductivity of semiconductors.
- By doping, the crystal can be made to have predominantly electrons (n-type) or holes (p-type).

- When a crystal is doped such that the equilibrium concentrations of electrons ( $n_0$ ) and holes ( $p_0$ ) are different from the intrinsic carrier concentration ( $n_i$ ), the material is said to be *extrinsic*.
- Doping creates additional levels within the band gap.
- In Si, column V elements of the periodic table (e.g., P, As, Sb) introduce energy levels very near (typically 0.03-0.06 eV) the conduction band.
- At 0 K, these levels are filled with electrons, and very little thermal energy (50 K to 100 K) is required for these electrons to get excited to the conduction band.
- Since these levels donate electrons to the conduction band, they are referred to as the *donor* levels.
- Thus, Si doped with donor impurities can have a significant number of electrons in the conduction band even when the temperature is not sufficiently high enough for the intrinsic carriers to dominate, i.e.,  $n_0 \gg n_i$ ,  $p_0 \approx 0 \Rightarrow$  n-type material, with electrons as *majority carriers* and holes as *minority carriers*.
- In Si, column III elements of the periodic table (e.g., B, Al, Ga, In) introduce energy levels very near (typically 0.03-0.06 eV) the valence band.
- At 0 K, these levels are empty, and very little thermal energy (50 K to 100 K) is required for electrons in the valence band to get excited to these levels, and leave behind holes in the valence band.
- Since these levels accept electrons from the valence band, they are referred to as the *acceptor* levels.
- Thus, Si doped with acceptor impurities can have a significant number of holes in the valence band even at a very low temperature, i.e.,  $p_0 \gg n_i$ ,  $n_0 \approx 0 \Rightarrow$  p-type material, with holes as majority carriers and electrons as minority carriers.
- The extra electron for column V elements is loosely bound and it can be liberated very easily  $\Rightarrow$  ionization; thus, it is free to participate in current conduction.
- Similarly, column III elements create holes in the valence band, and they can also participate in current conduction.
- Rough calculation of the ionization energy can be made based on the Bohr's model for  $H_2$  atoms, considering the loosely bound electron orbiting around the tightly bound core electrons. Thus,

$$E = \frac{m_n^* q^4}{2(4\pi\epsilon_0\epsilon_r)^2 [h/(2\pi)]^2} \quad (2.6) \text{ where } \epsilon_r \text{ is the relative permittivity of Si.}$$

**EXAMPLE 2.2:** Calculate the approximate donor binding energy for Si ( $r = 11.7$ ,  $m_n^* = 1.18 m_0$ ).

**SOLUTION:** From Eq.(2.6), we have

$$E = \frac{m_n^* q^4}{8(\epsilon_0\epsilon_r)^2 h^2} = \frac{1.18 \times 9.11 \times 10^{-31} \times (1.6 \times 10^{-19})^4}{8 \times (8.854 \times 10^{-12} \times 11.7)^2 \times (6.63 \times 10^{-34})^2} = 1.867 \times 10^{-20} \text{ J} = 0.117 \text{ eV.}$$

Note: The effective mass used here is an average of the effective mass in different

crystallographic directions, and is called the "conductivity effective mass" with values of  $1.28 m_0$  (at 600 K),  $1.18 m_0$  (at 300 K),  $1.08 m_0$  (at 77 K), and  $1.026 m_0$  (at 4.2 K).

- In III-V compounds, column VI impurities (e.g., S, Se, Te) occupying column V sites act as donors. Similarly, column II impurities (e.g., Be, Zn, Cd) occupying column III sites act as acceptors.
- When a column IV material (e.g., Si, Ge) is used to dope III-V compounds, then they may substitute column III elements (and act as donors), or substitute column V elements (and act as acceptors) => *amphoteric dopants*.
- Doping creates a large change in the electrical conductivity, e.g., with a doping of  $10^{15}/\text{cm}^3$ , the resistivity of Si changes from  $2 \times 10^5 \Omega\text{-cm}$  to  $5 \Omega\text{-cm}$ .

### ❖ Carrier Concentrations

- For the calculation of semiconductor electrical properties and analyzing device behavior, it is necessary to know the number of charge carriers/cm<sup>3</sup> in the material.
- The majority carrier concentration in a heavily doped material is obvious, since for each impurity atom, one majority carrier is obtained.
- However, the minority carrier concentration and the dependence of carrier concentrations on temperature are not obvious.
- To obtain the carrier concentrations, their distribution over the available energy states is required.
- These distributions are calculated using statistical methods.

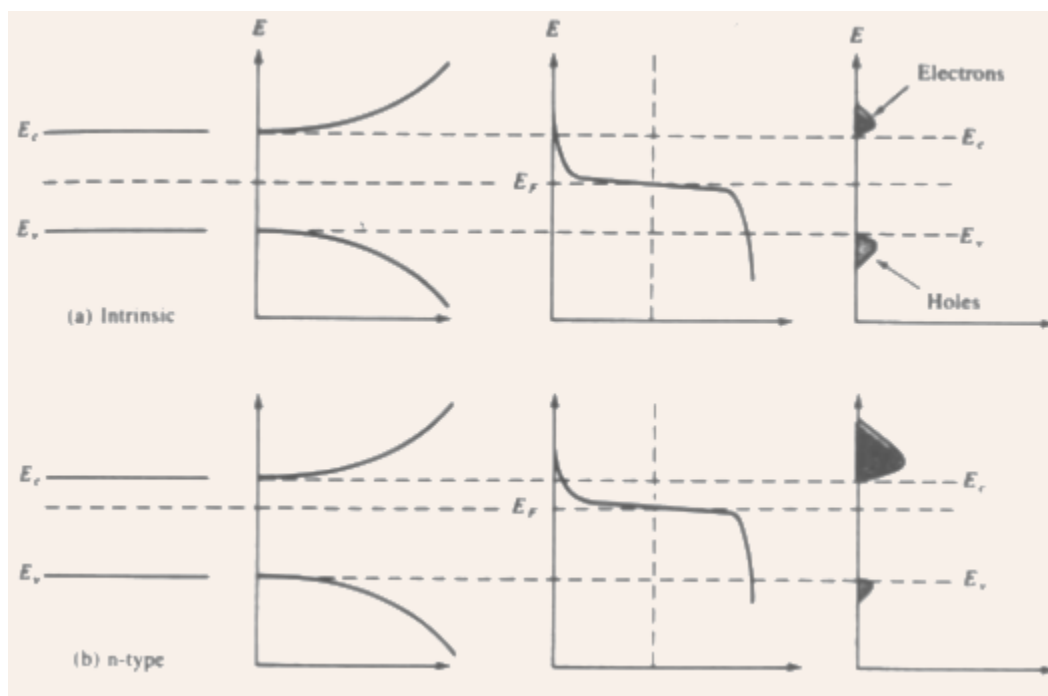
### ❖ The Fermi Level

- Electrons in solids obey *Fermi-Dirac* (FD) statistics.
- This statistics accounts for the indistinguishability of the electrons, their wave nature, and the Pauli exclusion principle.
- The Fermi-Dirac distribution function  $f(E)$  of electrons over a range of allowed energy levels at thermal equilibrium can be given by

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}} \quad (2.7) \text{ where } k \text{ is Boltzmann's constant } (= 8.62 \times 10^{-5} \text{ eV/K} = 1.38 \times 10^{-23} \text{ J/K}).$$

- This gives the probability that an available energy state at  $E$  will be occupied by an electron at an absolute temperature  $T$ .
- $E_F$  is called the Fermi level and is a measure of the average energy of the electrons in the lattice => an extremely important quantity for analysis of device behavior.
- Note: for  $(E - E_F) > 3kT$  (known as *Boltzmann approximation*),  $f(E) \approx \exp[-(E - E_F)/kT]$  => this is referred to as the Maxwell-Boltzmann (MB) distribution (followed by gas atoms).

- The probability that an energy state at  $E_F$  will be occupied by an electron is 1/2 at all temperatures.
- At 0 K, the distribution takes a simple rectangular form, with all states below  $E_F$  occupied, and all states above  $E_F$  empty.
- At  $T > 0$  K, there is a finite probability of states above  $E_F$  to be occupied and states below  $E_F$  to be empty.
- The F-D distribution function is highly symmetric, i.e., the probability  $f(E_F + \Delta E)$  that a state  $E$  above  $E_F$  is filled is the same as the probability  $[1 - f(E_F - \Delta E)]$  that a state  $E$  below  $E_F$  is empty.
- This symmetry about  $E_F$  makes the Fermi level a natural reference point for the calculation of electron and hole concentrations in the semiconductor.
- Note:  $f(E)$  is the probability of occupancy of an available state at energy  $E$ , thus, if there is no available state at  $E$  (e.g., within the band gap of a semiconductor), there is no possibility of finding an electron there.
- For intrinsic materials, the Fermi level lies close to the middle of the band gap (the difference between the effective masses of electrons and holes accounts for this small deviation from the mid gap).
- In n-type material, the electrons in the conduction band outnumber the holes in the valence band, thus, the Fermi level lies closer to the conduction band.
- Similarly, in p-type material, the holes in the valence band outnumber the electrons in the conduction band, thus, the Fermi level lies closer to the valence band.
- The probability of occupation  $f(E)$  in the conduction band and the probability of vacancy  $[1 - f(E)]$  in the valence band are quite small, however, the densities of available states in these bands are very large, thus a small change in  $f(E)$  can cause large changes in the carrier concentrations.



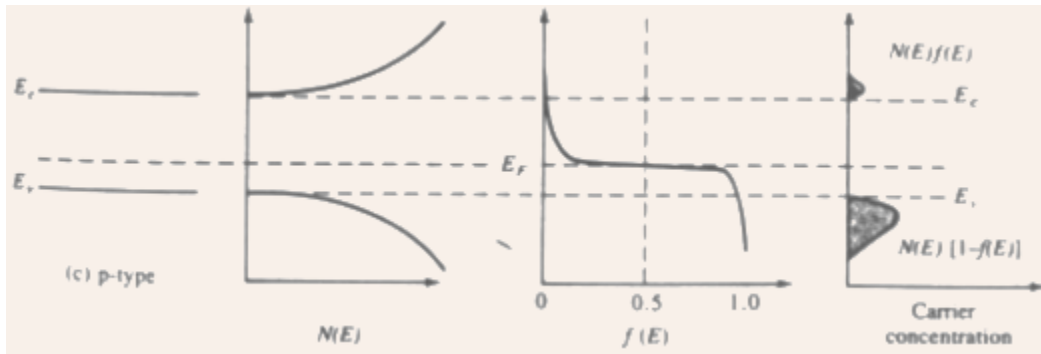


Fig.2.4 The density of states  $N(E)$ , the Fermi-Dirac distribution function  $f(E)$ , and the carrier concentration as functions of energy for (a) intrinsic, (b) n-type, and (c) p-type semiconductors at thermal equilibrium.

- Note: since the function  $f(E)$  is symmetrical about  $E_F$ , a large electron concentration implies a small hole concentration, and vice versa.
- In n-type material, the electron concentration in the conduction band increases as  $E_F$  moves closer to  $E_c$ ; thus,  $(E_c - E_F)$  gives a measure of  $n$ .
- Similarly, in p-type material, the hole concentration in the valence band increases as  $E_F$  moves closer to  $E_v$ ; thus,  $(E_v - E_F)$  gives a measure of  $p$ .

### Electron and Hole Concentrations at Equilibrium

- The F-D distribution function can be used to calculate the electron and hole concentrations in semiconductors, if the densities of available states in the conduction and valence bands are known.
- In equilibrium, the concentration of electrons in the conduction band can be given by

$$n_0 = \int_{E_c}^{\infty} f(E)N(E)dE \quad (2.8)$$

where  $N(E)dE$  is the density of available states/cm<sup>3</sup> in the energy range  $dE$ .

- Note: the upper limit of is theoretically not proper, since the conduction band does not extend to infinite energies; however, since  $f(E)$  decreases rapidly with increasing  $E$ , the contribution to this integral for higher energies is negligible.
- Using the solution of Schrödinger's wave equation under periodic boundary conditions, it can be shown that

$$N(E) = \frac{1}{2\pi^2} \left[ \frac{2m^*}{(\hbar/2\pi)^2} \right]^{3/2} E^{1/2} \quad (2.9)$$

- Thus,  $N(E)$  increases with  $E$ , however,  $f(E)$  decreases rapidly with  $E$ , thus, the product  $f(E)N(E)$  decreases rapidly with  $E$ , and very few electrons occupy states far above the conduction band edge, i.e., most electrons occupy a narrow energy band near the conduction band edge.
- Similarly, the probability of finding an empty state in the valence band  $[1 - f(E)]$  decreases rapidly below  $E_V$ , and most holes occupy states near the top of the valence band.
- Thus, a mathematical simplification can be made assuming that all available states in the conduction band can be represented by an effective density of states  $N_C$  located at the conduction band edge  $E_C$  and using Boltzmann approximation.

$$\text{Thus, } n_0 = N_C f(E_C) = N_C e^{-(E_C - E_F)/kT} \quad (2.10)$$

$$\text{where } N_C = 2 \left( \frac{2\pi m_n^* kT}{h^2} \right)^{3/2}$$

- Note: as  $(E_C - E_F)$  decreases, i.e., the Fermi level moves closer to the conduction band, the electron concentration increases.
- By similar arguments,

$$p_0 = N_V [1 - f(E_V)] = N_V e^{-(E_F - E_V)/kT} \quad (2.11)$$

where  $N_V$  is the effective density of states located at the valence band edge  $E_V$ .

- Note: the only terms separating the expressions for  $N_C$  and  $N_V$  are the effective masses of electrons ( $m_n^*$ ) and holes ( $m_p^*$ ) respectively, and since  $m_n^* \approx m_p^*$ , hence,  $N_C \approx N_V$ .
- Thus, as  $(E_F - E_V)$  decreases, i.e., the Fermi level moves closer to the valence band edge, and the hole concentration increases.
- These equations for  $n_0$  and  $p_0$  are valid in equilibrium, irrespective of the material being intrinsic or doped.
- For intrinsic material  $E_F$  lies at an intrinsic level  $E_i$  (very near the middle of the band gap), and the intrinsic electron and hole concentrations are given by  $n_i = N_C e^{-(E_C - E_i)/kT}$  and  $p_i = N_V e^{-(E_i - E_V)/kT}$  and (2.12)
- Note: At equilibrium, the product  $n_0 p_0$  is a constant for a particular material and temperature, even though the doping is varied,

$$\text{i.e., } n_0 p_0 = n_i^2 = N_C N_V e^{-E_g/kT} \quad (2.13)$$

- This equation gives an expression for the intrinsic carrier concentration  $n_i$  as a function of  $N_C$ ,  $N_V$ , and temperature:

$$n_i = \sqrt{N_C N_V} e^{-E_g/2kT} \quad (2.14)$$

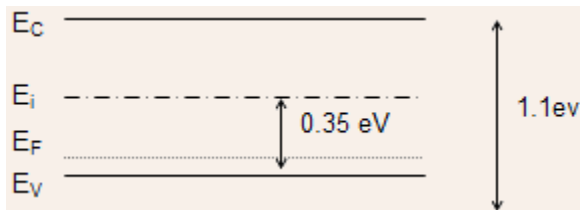
- These relations are extremely important, and are frequently used for calculations.
- Note: if  $N_C$  were to be equal to  $N_V$ , then  $E_i$  would have been exactly at mid gap (i.e.,  $E_C - E_i = E_i - E_V = E_g/2$ ).
- However, since  $N_C \neq N_V$ ,  $E_i$  is displaced slightly from mid gap (more for GaAs than that for Si).
- Alternate expressions for  $n_0$  and  $p_0$ :  

$$n_0 = n_i e^{(E_F - E_i)/kT} \quad \text{and} \quad p_0 = n_i e^{(E_i - E_F)/kT} \quad \text{and (2.15)}$$
- Note: the electron concentration is equal to  $n_i$  when  $E_F$  is at  $E_i$ , and  $n_0$  increases exponentially as  $E_F$  moves away from  $E_i$  towards the conduction band.
- Similarly, the hole concentration  $p_0$  varies from  $n_i$  to larger values as  $E_F$  moves from  $E_i$  towards the valence band.

**EXAMPLE 2.3:** A Si sample is doped with  $10^{16}$  B atoms/cm<sup>3</sup>. What is the equilibrium electron concentration  $n_0$  at 300 K? Where is  $E_F$  relative to  $E_i$ ? Assume  $n_i$  for Si at 300 K =  $1.5 \times 10^{10}$ /cm<sup>3</sup>.

**SOLUTION:** Since B (trivalent) is a p-type dopant in Si, hence, the material will be predominantly p-type, and since  $N_A \gg n_i$ , therefore,  $p_0$  will be approximately equal to  $N_A$ , and  

$$n_0 = \frac{n_i^2}{p_0} = \frac{2.25 \times 10^{20}}{10^{16}} = 2.25 \times 10^4 \text{ cm}^{-3}$$
Also,  
 $E_i - E_F = kT \ln(p_0/n_i) = 0.026 \ln[10^{16}/(1.5 \times 10^{10})] = 0.35 \text{ eV}$ . The resulting band diagram is:



### ❖ Temperature Dependence of Carrier Concentrations

- The intrinsic carrier concentration has a strong temperature dependence, given by  

$$n_i(T) = 2 \left( \frac{2\pi kT}{h^2} \right)^{3/2} (m_n^* m_p^*)^{3/4} e^{-E_g/2kT} \quad (2.16)$$
- Thus, explicitly,  $n_i$  is proportional to  $T^{3/2}$  and to  $e^{-1/T}$ , however,  $E_g$  also has a temperature dependence (decreasing with increasing temperature, since the interatomic spacing changes with temperature).

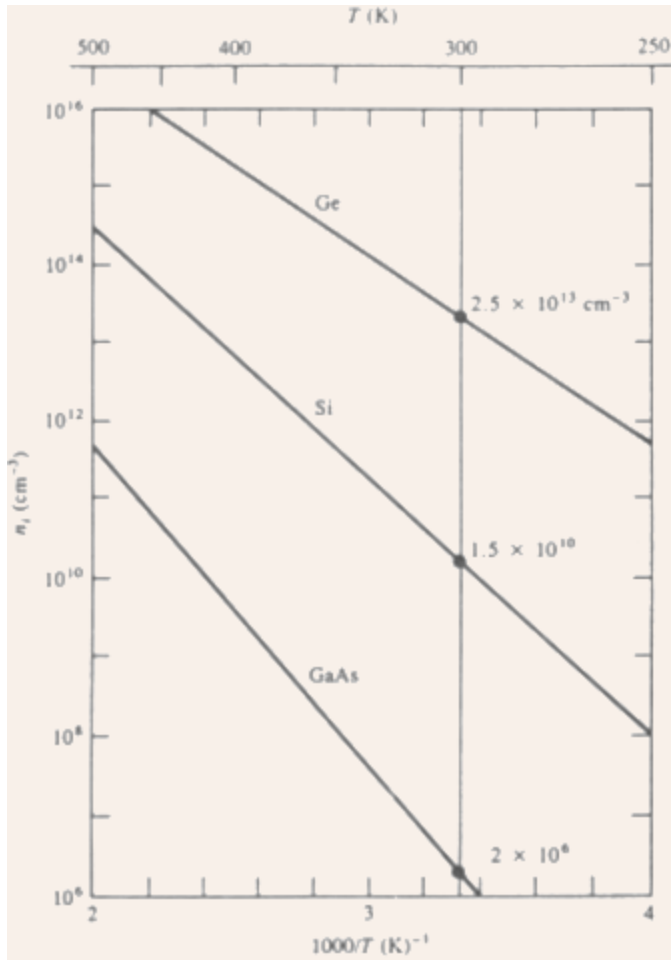


Fig.2.5 The intrinsic carrier concentration as a function of inverse temperature for Si, Ge, and GaAs.

- As  $n_i$  changes with temperature, so do  $n_0$  and  $p_0$ .
- With  $n_i$  and  $T$  given, the unknowns are the carrier concentrations and the Fermi level position with respect to  $E_i$ . One of these quantities must be given in order to calculate the other.
- Example: Si doped with  $10^{15}/\text{cm}^3$  donors ( $N_d$ ).
- At very low temperature, negligible intrinsic EHPs exist, and all the donor electrons are bound to the donor atoms.
- As temperature is raised, these electrons are gradually donated to the conduction band, and at about 100 K ( $1000/T = 10$ ), almost all these electrons are donated  $\Rightarrow$  this temperature range is called the ionization region.

Once all the donor atoms are ionized, the electron concentration  $n_0 \approx N_d$ , since for each donor atom, one electron is obtained.

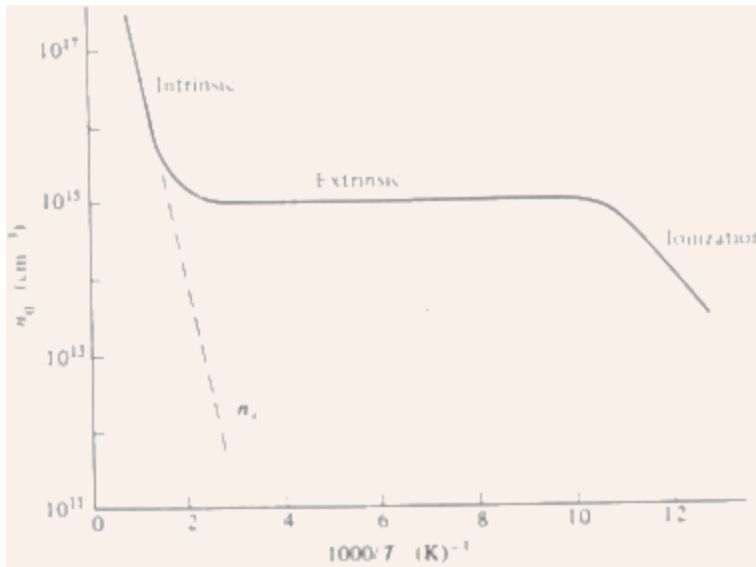


Fig.2.6 Variation of carrier concentration with inverse temperature clearly showing the three regions: ionization, extrinsic, and intrinsic.

- Thus,  $n_0$  remains virtually constant with temperature for a wide range of temperature (called the extrinsic region), until the intrinsic carrier concentration  $n_i$  starts to become comparable to  $N_d$ .
- For high temperatures,  $n_i \gg N_d$ , and the material loses its extrinsic property (called the intrinsic region).
- Note: in the intrinsic region, the device loses its usefulness => determines the maximum operable temperature range.

### ❖ Compensation and Space Charge Neutrality

- Semiconductors can be doped with both donors ( $N_d$ ) and acceptors ( $N_A$ ) simultaneously.
- Assume a material doped with  $N_d > N_A$  => predominantly n-type =>  $E_F$  lies above  $E_i$  => acceptor level  $E_a$  completely full, however, with  $E_F$  above  $E_i$ , the hole concentration cannot be equal to  $N_A$ .
- Mechanism:
  - Electrons are donated to the conduction band from the donor level  $E_d$ .
  - An acceptor state gets filled by a valence band electron, thus creating a hole in the valence band.
  - An electron from the conduction band recombines with this hole.
  - Extending this logic, it is expected that the resultant concentration of electrons in the conduction band would be  $N_d - N_A$  instead of  $N_d$ .
  - This process is called compensation.
- By compensation, an n-type material can be made intrinsic (by making  $N_A = N_d$ ) or even p-type (for  $N_A > N_d$ ).

**Note:** a semiconductor is neutral to start with, and, even after doping, it remains neutral

(since for all donated electrons, there are positively charged ions ( $N_d^+$ ); and for all accepted electrons (or holes in the valence band), there are negatively charged ions ( $N_a^-$ ).

- Therefore, the sum of positive charges must equal the sum of negative charges, and this governing relation, given by  $p_0 + N_d^+ = n_0 + N_a^-$  (2.17) is referred to as the equation for space charge neutrality.
  - This equation, solved simultaneously with the law of mass action (given by  $n_0 p_0 = n_i^2$ ) gives the information about the carrier concentrations.
- Note:** for  $N_d^+ - N_a^- \gg n_i^2 / n_0$ ,  $n_0 \approx N_d^+ - N_a^-$ , and, similarly, for  $N_a^- - N_d^+ \gg n_i^2 / p_0$ ,  $p_0 \approx N_a^- - N_d^+$

### Drift of Carriers in Electric and Magnetic Fields

- In addition to the knowledge of carrier concentrations, the collisions of the charge carriers with the lattice and with the impurity atoms (or ions) under electric and/or magnetic fields must be accounted for, in order to compute the current flow through the device.
- These processes will affect the ease (*mobility*) with which carriers move within a lattice.
- These collision and scattering processes depend on temperature, which affects the thermal motion of the lattice atoms and the velocity of the carriers.

### ❖ Conductivity and Mobility

- Even at thermal equilibrium, the carriers are in a constant motion within the lattice.
- At room temperature, the thermal motion of an individual electron may be visualized as random scattering from lattice atoms, impurities, other electrons, and defects.
- There is no net motion of the group of  $n$  electrons/cm<sup>3</sup> over any period of time, since the scattering is random, and there is no preferred direction of motion for the group of electrons and no net current flow.
- However, for an individual electron, this is not true the probability of an electron returning to its starting point after time  $t$  is negligibly small.
- Now, if an electric field  $E_x$  is applied in the  $x$ -direction, each electron experiences a net force  $qE_x$  from the field.
- This will create a net motion of group in the  $x$ -direction, even though the force may be insufficient to appreciably alter the random path of an individual electron.
- If  $p_x$  is the  $x$ -component of the total momentum of the group, then the force of the field on the  $n$  electrons/cm<sup>3</sup> is

$$-nqE_x = \left. \frac{dp_x}{dt} \right|_{\text{field}} \quad (2.18)$$

**Note:** this expression indicates a constant acceleration in the  $x$ -direction, which realistically cannot happen.

- In steady state, this acceleration is just balanced by the deceleration due to the collisions.
- Thus, while the steady field  $E_x$  does produce a net momentum  $p_x$ , for steady state current flow, the net rate of change of momentum must be zero when collisions are included.

- Note: the collision processes are totally random, thus, there is a constant probability of collision at any time for each electron.
- Consider a group of  $N_0$  electrons at time  $t = 0$ , and define  $N(t)$  as the number of electrons that have not undergone a collision by time  $t$

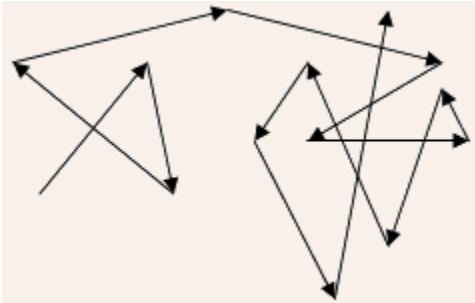


Fig.2.7 The random thermal motion of an individual electron, undergoing random scattering.

- The rate of decrease of  $N(t)$  at any time  $t$  is proportional to the number left unscattered at  $t$ , i.e.

$$-\frac{dN(t)}{dt} = \frac{1}{\bar{t}} N(t) \quad (2.19)$$

where  $\bar{t}^{-1}$  is the constant of proportionality.

- The solution is an exponential function

$$N(t) = N_0 e^{-t/\bar{t}} \quad (2.20)$$

and  $\bar{t}$  represents the mean time between scattering events, called the mean free time.

- The probability that any electron has a collision in time interval  $dt$  is  $dt/\bar{t}$ , thus, the differential change in  $p_x$  due to collisions in time  $dt$  is

$$dp_x = -p_x \frac{dt}{\bar{t}} \quad (2.21)$$

- Thus, the rate of change of  $p_x$  due to the decelerating effect of collisions is

$$\left. \frac{dp_x}{dt} \right|_{\text{collisions}} = -\frac{p_x}{\bar{t}} \quad (2.22)$$

- For steady state, the sum of acceleration and deceleration effects must be zero, thus,

$$-\frac{p_x}{\bar{t}} - nq E_x = 0 \quad (2.23)$$

- The average momentum per electron (averaged over the entire group of electrons) is

$$\langle p_x \rangle = \frac{p_x}{n} = -q\bar{t} E_x \quad (2.24)$$

- Thus, as expected for steady state, the electrons would have on the average a constant net velocity in the -x-direction

$$\langle v_x \rangle = \frac{\langle p_x \rangle}{m_n^*} = -\frac{q\bar{t}}{m_n^*} E_x \quad (2.25)$$

- This speed is referred to as the drift speed, and, in general, it is usually much smaller than the random speed due to thermal motion  $v_{th} (= \sqrt{3kT/m_n^*})$ .
- The current density resulting from this drift

$$J_x = -qn \langle v_x \rangle = \frac{nq^2\bar{t}}{m_n^*} E_x = \sigma E_x \quad (2.26)$$

- This is the familiar Ohm's law with  $\sigma \equiv \frac{nq^2\bar{t}}{m_n^*}$  being the conductivity of the sample, which can also be written as  $\sigma = qn\mu_n$ , with  $\mu_n = \frac{q\bar{t}}{m_n^*}$  is defined as the *electron mobility* (in  $\text{cm}^2/\text{V-sec}$ ), and it describes the ease with which electrons drift in the material.
- The mobility can also be expressed as the average drift velocity per unit electric field, thus  $\mu_n = -\langle v_x \rangle / E_x$ , with the negative sign denoting a positive value for mobility since electrons drift opposite to the direction of the electric field.
- The total current density can be given by  $J_x = q(n\mu_n + p\mu_p)E_x$  (2.27) when both electrons and holes contribute to the current conduction; on the other hand, for predominantly n-type or p-type samples, respectively the first or the second term of the above equation dominates.

**Note:** both electron and hole drift currents are in the same direction, since holes (with positive charges) move along the direction of the electric field, and electrons (with negative charges) drift opposite to the direction of the electric field.

- Since GaAs has a strong curvature of the E-k diagram at the bottom of the conduction band, the electron effective mass in GaAs is very small => the electron mobility in GaAs is very high since  $\mu_n$  is inversely proportional to  $m_n^*$ .
- The other parameter in the mobility expression, i.e.,  $\bar{t}$  (the mean free time between collisions) is a function of temperature and the impurity concentration in the semiconductor.
- For a uniformly doped semiconductor bar of length L, width w, and thickness t, the resistance R of the bar can be given by  $L/(wt\sigma) = \rho L/(wt)$  where  $\rho$  is the resistivity.

## ❖ Effects of Temperature and Doping on Mobility

- The two main scattering events that influence electron and hole motion (and, thus, mobility) are the lattice scattering and the *impurity scattering*.
- All lattice atoms vibrate due to temperature and can scatter carriers due to collisions.
- These collective vibrations are called phonons, thus *lattice scattering* is also known as *phonon scattering*.
- With increasing temperature, lattice vibrations increase, and the mean free time between collisions decreases => mobility decreases (typical dependence  $\sim T^{-3/2}$ ).
- Scattering from crystal defects and ionized impurities dominate at low temperatures.
- Since carriers moving with low velocity (at low temperature) can get scattered more easily by ionized impurities, this kind of scattering causes a decrease in carrier mobility with decreasing temperature (typical dependence  $\sim T^{3/2}$ ).
- Note: the scattering probability is inversely proportional to the mean free time (and to mobility), hence, the mobilities due to two or more scattering events add inversely:  
$$\frac{1}{\mu} = \frac{1}{\mu_1} + \frac{1}{\mu_2} + \dots \quad (2.28)$$
- Thus, the mechanism causing the lowest mobility value dominates.
- Mobility also decreases with increasing doping, since the ionized impurities scatter carriers more (e.g.,  $\mu_n$  for intrinsic Si is 1350  $\text{cm}^2/\text{V-sec}$  at 300 K, whereas with a donor doping of  $10^{17}/\text{cm}^3$ ,  $\mu_n$  drops to 700  $\text{cm}^2/\text{V-sec}$ ).

## ❖ High Field Effects

- For small electric fields, the drift current increases linearly with the electric field, since  $\sigma$  is a constant.
- However, for large electric fields (typically  $> 10^3 \text{ V/cm}$ ), the current starts to show a sublinear dependence on the electric field and eventually saturates for very high fields.
- Thus,  $\sigma$  becomes a function of the electric field, and this is known as the hot carrier effect, when the carrier drift velocity becomes comparable to its thermal velocity.
- The maximum carrier drift velocity is limited to its mean thermal velocity (typically  $10^7 \text{ cm/sec}$ ), beyond which the added energy imparted by the electric field is absorbed by the lattice (thus generating heat) instead of a corresponding increase in the drift velocity.

### 2.4.4 The Hall Effect

- An extremely important measurement procedure for determining the majority carrier concentration and mobility.

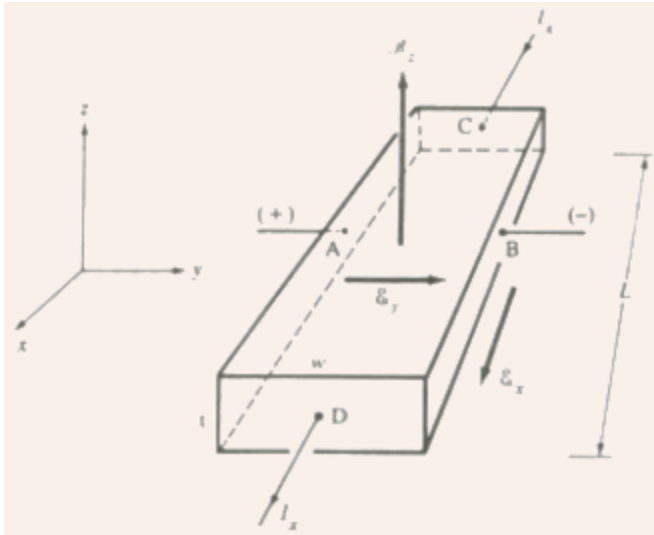


Fig.2.8 The experimental setup for the Hall Effect measurement.

- If a magnetic field is applied perpendicular to the direction of carrier flow, the path of the carriers get deflected due to the Lorentz force experienced by the carriers, which can be given by  

$$F = q(E + v \times B) \quad (2.29)$$
- Thus, the holes will get deflected towards the -y-direction, and establish an electric field along the y-direction, such that in steady state  $E_y = v_x B_z$ .
- The establishment of this electric field is known as the *Hall effect*, and the resulting voltage  $V_{AB} = E_y w$  is called the *Hall voltage*.
- Using the expression for the drift current,  $E_y = R_H J_x B_z$ , where  $R_H (= [q p_0]^{-1})$  is called the Hall coefficient.
- A measurement of the Hall voltage along with the information for magnetic field and current density gives the majority carrier concentration  $p_0 (= I_x B_z / [q t V_{AB}])$ .
- Also, the majority carrier mobility  $\mu_p (= R_H / \rho)$  can be obtained from a measurement of the resistivity  $\rho (= [V_{CD} / I_x] / [L / (wt)])$ .
- This experiment can be performed to obtain the variation of majority carrier concentration and mobility as a function of temperature.
- For n-type samples, the Hall voltage and the Hall coefficient are negative  $\Rightarrow$  a common diagnostic tool for obtaining the sample type.
- Note: caution should be exercised for near intrinsic samples.

---

**EXAMPLE 2.4:** A sample of Si is doped with  $10^{18}$  In atoms/cm<sup>3</sup>. What will be the measured value of its resistivity? What is the expected Hall voltage in a 150  $\mu\text{m}$  thick sample if  $I_x = 2$  mA and  $B_z = 5$  kG ( $1$  kG =  $10^{-5}$  Wb/cm<sup>2</sup>)?

### SOLUTION:

The sample is p-type with  $p_0 \gg n_0$ , and the mobility  $\mu_p \approx 400 \text{ cm}^2/\text{V}\cdot\text{sec}$ .

Thus, the resistivity  $\rho = (\sigma)^{-1} = (q\mu_p p_0)^{-1} = 1.5625 \text{ }\Omega\cdot\text{cm}$ .

The Hall coefficient  $R_H = +(qp_0)^{-1} = +625 \text{ cm}^3/\text{C}$  (note the unit of  $R_H$ ).

Note, alternately  $R_H = \rho\mu_p = 625 \text{ cm}^3/\text{C}$ .

The Hall voltage  $V_{AB} = (I_x B_z R_H)/t = (2 \times 10^{-3})(5 \times 10^{-5})(625)/(150 \times 10^{-4}) = 4.17 \text{ mV}$ .

### ❖ Equilibrium Condition

- In equilibrium, there is no external excitation except a constant temperature, no net transfer of energy, no net carrier motion, and no net current transport.
- An important condition for equilibrium is that no discontinuity or gradient can arise in the equilibrium Fermi level  $E_F$ .
- Assume two materials 1 and 2 (e.g., n- and p-type regions, dissimilar semiconductors, metal and semiconductor, two adjacent regions in a nonuniformly doped semiconductor) in intimate contact such that electron can move between them.
- Assume materials 1 and 2 have densities of state  $N_1(E)$  and  $N_2(E)$ , and F-D distribution functions  $f_1(E)$  and  $f_2(E)$  respectively at any energy  $E$ .

The sample is p-type with  $p_0 \gg n_0$ , and the mobility  $\mu_p \approx 400 \text{ cm}^2/\text{V}\cdot\text{sec}$ .

Thus, the resistivity  $\rho = (\sigma)^{-1} = (q\mu_p p_0)^{-1} = 1.5625 \text{ }\Omega\cdot\text{cm}$ .

The Hall coefficient  $R_H = +(qp_0)^{-1} = +625 \text{ cm}^3/\text{C}$  (note the unit of  $R_H$ ).

Note, alternately  $R_H = \rho\mu_p = 625 \text{ cm}^3/\text{C}$ .

The Hall voltage  $V_{AB} = (I_x B_z R_H)/t = (2 \times 10^{-3})(5 \times 10^{-5})(625)/(150 \times 10^{-4}) = 4.17 \text{ mV}$ .

- The rate of electron motion from 1 to 2 can be given by rate from 1 to 2  $N_1(E)f_1(E) \cdot N_2(E)[1 - f_2(E)]$  (2.30) and the rate of electron motion from 2 to 1 can be given by rate from 2 to 1  $N_2(E)f_2(E) \cdot N_1(E)[1 - f_1(E)]$  (2.31). At equilibrium, these two rates must be equal, which gives  $f_1(E) = f_2(E) \Rightarrow E_F1 = E_F2 \Rightarrow dE_F/dx = 0$ ; thus, the Fermi level is constant at equilibrium, or, in other words, there cannot be any discontinuity or gradient in the Fermi level at equilibrium.

### ❖ Practice Problems

**2.1** Electrons move in a crystal as wave packets with a group velocity  $v = d\omega/dk$ , where  $\omega$  is the angular frequency. Show that in a given electric field, these wave packets obey

Newton's second law of motion, i.e., the force  $F = m^*a$ , where  $m^*$  is the effective mass (given by  $m^* = \hbar^2/[\partial^2 E/\partial k^2]$ ) and  $a$  is the acceleration.

**2.2** Some semiconductors of interest have the dependence of its energy  $E$  with respect to the wave vector  $k$ , given by  $E(1 + \alpha E) = \hbar^2 k^2/[2m_n]$ , where  $m_n$  is the effective mass for  $E = 0$ ,  $k$  is the wave vector, and  $\alpha$  is a constant. Calculate the dependence of the effective mass  $m_n^*$  on energy.

**2.3** Determine the equilibrium recombination constant  $r$  for Si and GaAs, having equilibrium thermal generation rates of  $4.03 \times 10^5 \text{ (cm}^3\text{-sec)}^{-1}$  and  $3.18 \times 10^3 \text{ (cm}^3\text{-sec)}^{-1}$  respectively, and intrinsic carrier concentrations of  $1.5 \times 10^{10} \text{ cm}^{-3}$  and  $2.1 \times 10^6 \text{ cm}^{-3}$  respectively. Comment on the answers. Will  $\alpha_r$  change with doping at equilibrium?

**2.4** The relative dielectric constant for GaP is 10.2 and the electron effective mass is  $0.13m_0$ . Calculate the approximate ionization energy of a donor atom in GaP.

**2.5** Show that the probability that a state  $\Delta E$  above the Fermi level  $E_F$  is occupied is the same as the probability that a state  $\Delta E$  below  $E_F$  is empty.

**2.6** Derive an expression relating the intrinsic level  $E_i$  to the center of the band gap  $E_g/2$ , and compute the magnitude of this displacement for Si and GaAs at 300 K. Assume  $m_n^*/m_0 = 1.1$  and  $0.067$ , and  $m_p^*/m_0 = 0.56$  and  $0.48$  for Si and GaAs respectively.

**2.7** Show that in order to obtain maximum resistivity in a GaAs sample ( $E_g = 1.42 \text{ eV}$ ,  $n_i = 2.1 \times 10^6 \text{ cm}^{-3}$ ,  $\mu_n = 8500 \text{ cm}^2/\text{V-sec}$ , and  $\mu_p = 400 \text{ cm}^2/\text{V-sec}$ ), it has to be doped slightly p-type. Determine this doping concentration. Also, determine the ratio of the maximum resistivity to the intrinsic resistivity.

**2.8** A GaAs sample (use the data given in Problem 2.7) is doped uniformly with  $10^7 \text{ Si atoms/cm}^3$ , out of which 70% occupy Ga sites, and the rest 30% occupy As sites. Assume 100% ionization and  $T = 300 \text{ K}$ .  
a) Calculate the equilibrium electron and hole concentrations  $n_0$  and  $p_0$ .  
b) Clearly draw the equilibrium band diagram, showing the position of the Fermi level  $E_F$  with respect to the intrinsic level  $E_i$ , assuming that  $E_i$  lies exactly at midgap.  
c) Calculate the percentage change in conductivity after doping as compared to the intrinsic case.

**2.9** A Si sample is doped with  $10^{16} \text{ cm}^{-3}$  donor atoms. Determine the minimum temperature at which the sample becomes intrinsic. Assume that at this minimum temperature, the free electron concentration does not exceed by more than 1% of the donor concentration (beyond its extrinsic value). For Si,  $n_i(T) = 3.88 \times 10^{16} T^{3/2} e^{-7000/T} \text{ cm}^{-3}$  ( $T$  in K).

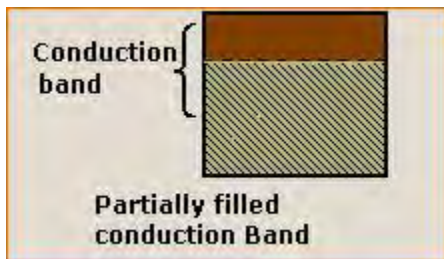
**2.10** Since the event of collision of an electron in a lattice is a truly random process, thus having a constant probability of collision at any given time, the number of particles left unscattered at time  $t$ ,  $N(t) = N_0 \exp(-t/\tau)$ , where  $N_0 = N(t)$  at  $t = 0$ , and  $\tau =$  mean free time. Hence, show that if there are a total of  $i$  number of scattering events, each with a mean

free time of  $\tau_i$ , then the net electron mobility  $\mu$  can be given by  $\mu^{-1} = \sum_i (\mu_i)^{-1}$ , where  $\mu_i$  is the mobility due to the  $i$ th scattering event.

**2.11** A Ge sample is oriented in a  $5 \text{ kG}$  ( $1 \text{ kG} = 10^{-5} \text{ Wb/cm}^2$ ) magnetic field (refer to Fig.2.8). The current is  $4 \text{ mA}$ , and the sample dimensions are  $w = 0.25 \text{ mm}$ ,  $t = 50 \text{ }\mu\text{m}$ , and  $L = 2.5 \text{ mm}$ . The following data are taken:  $V_{AB} = -2.5 \text{ mV}$ , and  $V_{CD} = 170 \text{ mV}$ . Find the type and concentration of the majority carrier, and its mobility. Hence, compute the net relaxation time for the various scattering events, assuming  $m_n^* = 0.55m_0$  and  $m_p^* = 0.37m_0$ .

**2.12** In the Hall effect experiment, there is a chance that the Hall Probes A and B (refer to Fig.2.8) are not perfectly aligned, which may give erroneous Hall voltage readings. Show that the true Hall voltage  $V_H$  can be obtained from two measurements of  $V_{AB}$ , with the magnetic field first in the  $+z$ -direction, and then in the  $-z$ -direction.

## Metals



or



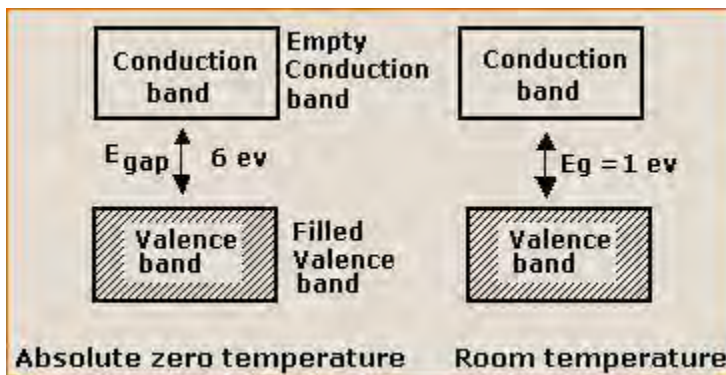
In metals, either the conduction band is partially filled or overlaps with valence band. There is no forbidden energy gap in between. Even if a small electric field is applied, free electrons start moving in a direction opposite to field and hence a good conductor of electricity.

### Insulators

Here the valency bands are completely filled and conduction band is empty and the forbidden gap is quite large. For example in diamond  $E_{\text{gap}}$  is 6eV. Even if an electric field is applied, no electron is able to go from valence band to conduction band.

### Semiconductors

The valence band is completely filled and conduction band is empty. The  $E_{\text{gap}}$  is also less i.e., of the order of few eV. At zero kelvin, electrons are not able to cross this forbidden gap and so behave like insulators. But as temperature is increased, electrons in valence band (VB) gain thermal energy and jump to conduction band (CB) and acquire small conductivity at room temperature and so behave like conductors. Hence they are called semiconductors.



### Charge carriers in semiconductors

At high temperature, electrons move from valence band to conduction band and as a result a vacancy is created in the valence band at a place where an electron was present before shifting to conduction band. The vacancy is a hole and is seat of positive charge having the same value of electron. Therefore the electrical conduction in semiconductors is due to motion of electrons in conduction band and also due to motion of holes in valence band.

## Semiconductor Basics

If **Resistors** are the most basic passive component in electrical or electronic circuits, then we have to consider the **Signal Diode** as being the most basic "Active" component. However, unlike a resistor, a diode does not behave linearly with respect to the applied voltage as it has an exponential I-V relationship and therefore can not be described simply by using Ohm's law as we do for resistors. Diodes are unidirectional semiconductor devices that will only allow current to flow through them in one direction only, acting more like a one way electrical valve, (Forward Biased Condition). But, before we have a look at how signal or power diodes work we first need to understand their basic construction and concept.

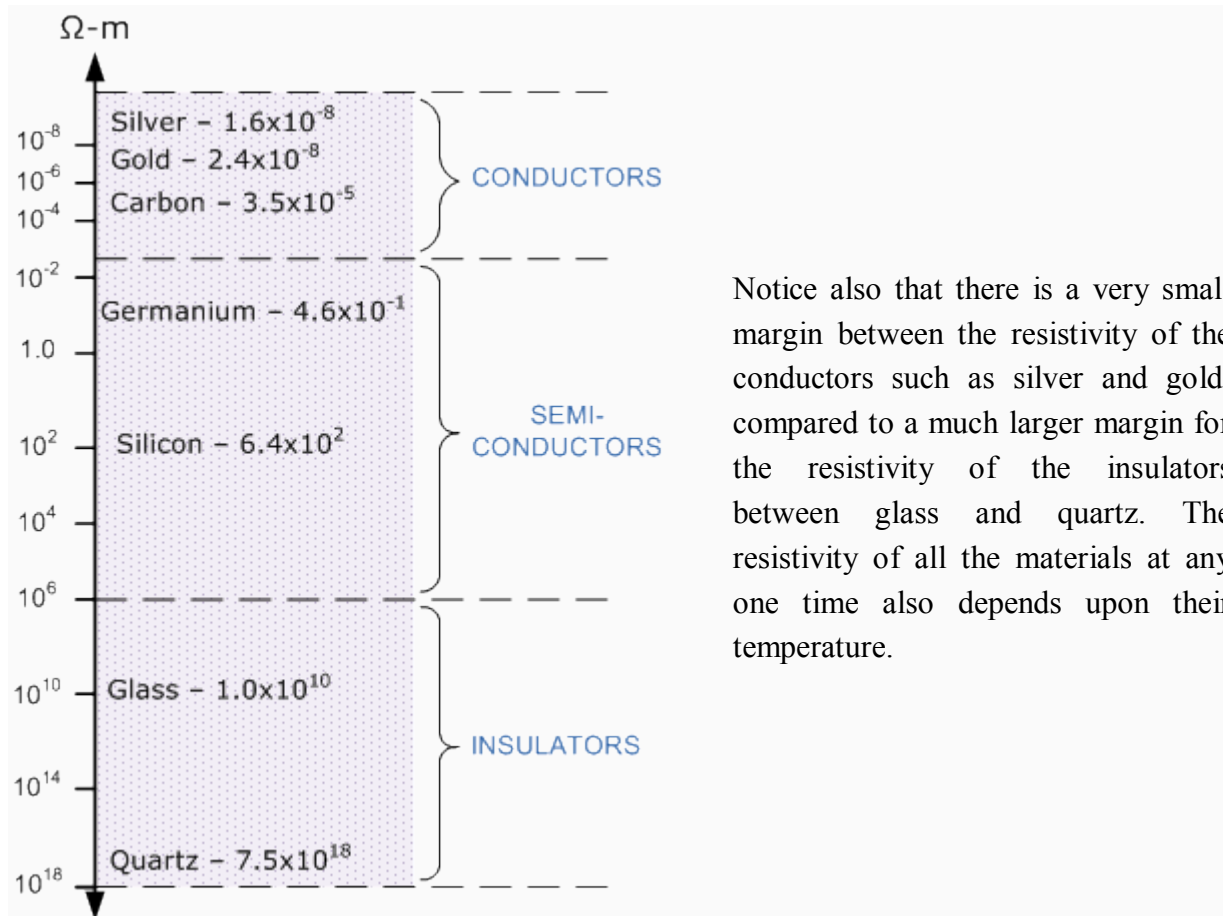
Diodes are made from a single piece of **Semiconductor** material which has a positive "P-region" at one end and a negative "N-region" at the other, and which has a resistivity value somewhere between that of a conductor and an insulator. But what is a "Semiconductor" material?, firstly let's look at what makes something either a **Conductor** or an **Insulator**.

### Resistivity

The electrical **Resistance** of an electrical or electronic component or device is generally defined as being the ratio of the voltage difference across it to the current flowing through it, basic **Ohm's Law** principals. The problem with using resistance as a measurement is that it depends very much on the physical size of the material being measured as well as the material out of which it is made. For example, If we were to increase the length of the material (making it longer) its resistance would also increase. Likewise, if we increased its diameter (making it fatter) its resistance would then decrease. So we want to be able to define the material in such a way as to indicate its ability to either conduct or oppose the flow of electrical current through it no matter what its size or shape happens to be. The quantity that is used to indicate this specific resistance is called **Resistivity** and is given the Greek symbol of  $\rho$ , (Rho). Resistivity is measured in Ohm-metres, ( $\Omega\text{-m}$ ) and is the inverse to conductivity.

If the resistivity of various materials is compared, they can be classified into three main groups, Conductors, Insulators and Semi-conductors as shown below.

## Resistivity Chart



### Conductors

From above we now know that **Conductors** are materials that have a low value of resistivity allowing them to easily pass an electrical current due to there being plenty of free electrons floating about within their basic atom structure. When a positive voltage potential is applied to the material these "free electrons" leave their parent atom and travel together through the material forming an electron drift. Examples of good conductors are generally metals such as Copper, Aluminium, Silver or non metals such as Carbon because these materials have very few electrons in their outer "Valence Shell" or ring, resulting in them being easily knocked out of the atom's orbit. This allows them to flow freely through the material until they join up with other atoms, producing a "Domino Effect" through the material thereby creating an electrical current.

Generally speaking, most metals are good conductors of electricity, as they have very small resistance values, usually in the region of micro-ohms per metre with the resistivity of

conductors increasing with temperature because metals are also generally good conductors of heat.

## **Insulators**

Insulators on the other hand are the exact opposite of conductors. They are made of materials, generally non-metals, that have very few or no "free electrons" floating about within their basic atom structure because the electrons in the outer valence shell are strongly attracted by the positively charged inner nucleus. So if a potential voltage is applied to the material no current will flow as there are no electrons to move and which gives these materials their insulating properties. Insulators also have very high resistances, millions of ohms per metre, and are generally not affected by normal temperature changes (although at very high temperatures wood becomes charcoal and changes from an insulator to a conductor). Examples of good insulators are marble, fused quartz, p.v.c. plastics, rubber etc.

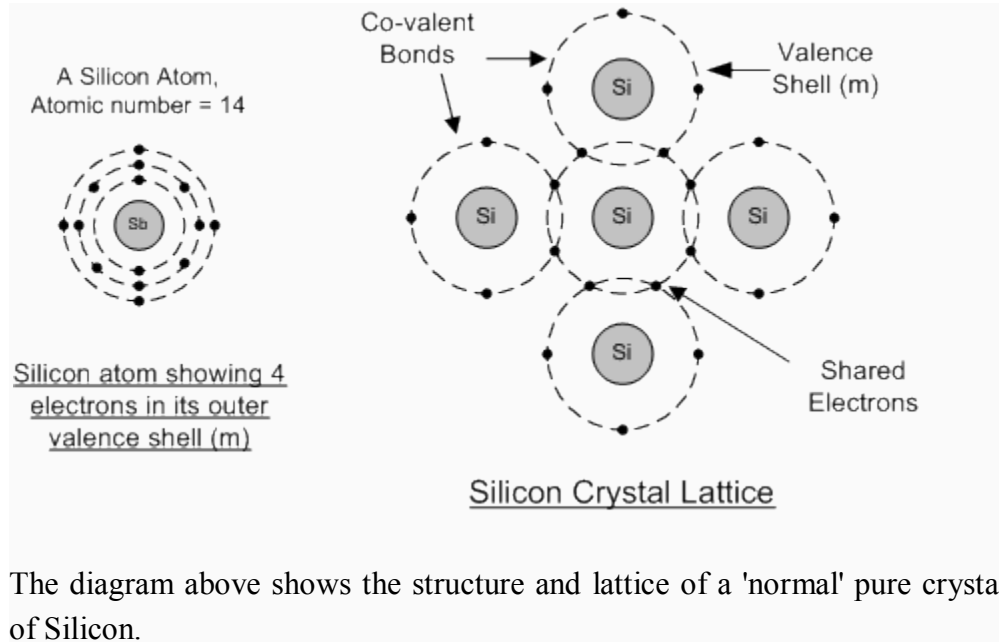
Insulators play a very important role within electrical and electronic circuits, because without them electrical circuits would short together and not work. For example, insulators made of glass or porcelain are used for insulating and supporting overhead transmission cables while epoxy-glass resin materials are used to make printed circuit boards, PCB's etc.

## **Semiconductor Basics**

Semiconductors materials such as silicon (Si), germanium (Ge) and gallium arsenide (GaAs), have electrical properties somewhere in the middle, between those of a "conductor" and an "insulator". They are not good conductors nor good insulators (hence their name "semi"-conductors). They have very few "free electrons" because their atoms are closely grouped together in a crystalline pattern called a "crystal lattice". However, their ability to conduct electricity can be greatly improved by adding certain "impurities" to this crystalline structure thereby, producing more free electrons than holes or vice versa. By controlling the amount of impurities added to the semiconductor material it is possible to control its conductivity. These impurities are called donors or acceptors depending on whether they produce electrons or holes. This process of adding impurity atoms to semiconductor atoms (the order of 1 impurity atom per 10 million (or more) atoms of the semiconductor) is called **Doping**.

The most commonly used semiconductor material by far is **silicon**. It has four valence electrons in its outer most shell which it shares with its adjacent atoms in forming covalent bonds. The structure of the bond between two silicon atoms is such that each atom shares one electron with its neighbour making the bond very stable. As there are very few free electrons available to move from place to place producing an electrical current, crystals of pure silicon (or germanium) are

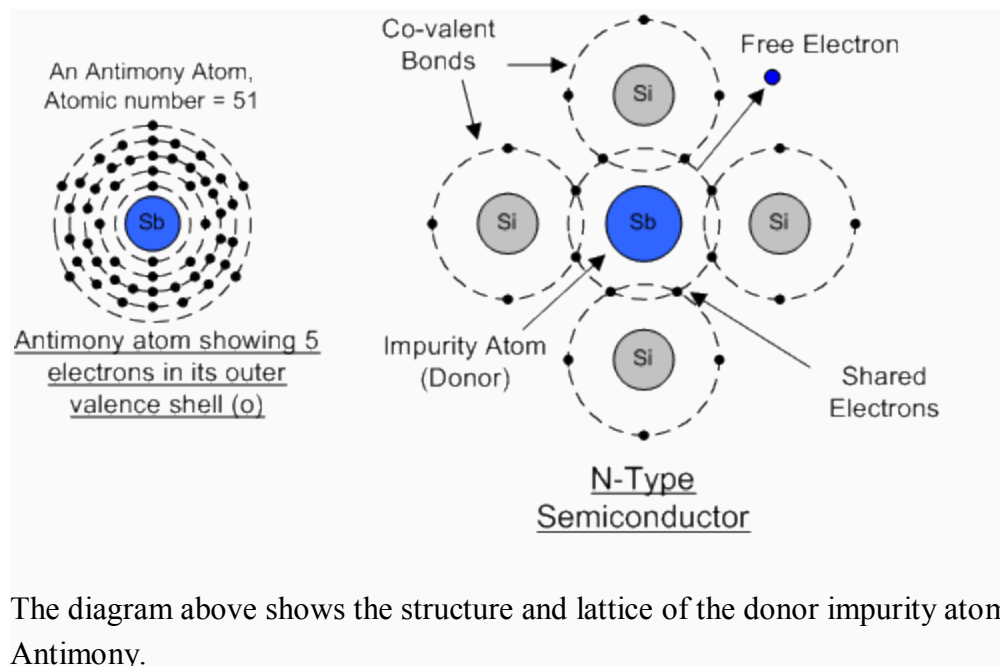
therefore good insulators, or at the very least very high value resistors. Silicon atoms are arranged in a definite symmetrical pattern making them a crystalline solid structure. A crystal of pure silicon (silicon dioxide or glass) is generally said to be an intrinsic crystal (it has no impurities).



### N-type Semiconductor Basics

In order for our silicon crystal to conduct electricity, we need to introduce an impurity atom such as Arsenic, Antimony or Phosphorus into the crystalline structure making it extrinsic (impurities are added). These atoms have five outer electrons in their outermost co-valent bond to share with other atoms and are commonly called "Pentavalent" impurities. This allows four of the five electrons to bond with its neighbouring silicon atoms leaving one "free electron" to move about when an electrical voltage is applied (electron flow). As each impurity atom "donates" one electron, pentavalent atoms are generally known as "donors".

**Antimony** (symbol Sb) is frequently used as a pentavalent additive as it has 51 electrons arranged in 5 shells around the nucleus. The resulting semiconductor material has an excess of current-carrying electrons, each with a negative charge, and is therefore referred to as "**N-type**" material with the electrons called "Majority Carriers" and the resultant holes "Minority Carriers". Then a semiconductor material is N-type when its donor density is greater than its acceptor density. Therefore, a N-type semiconductor has more electrons than holes.

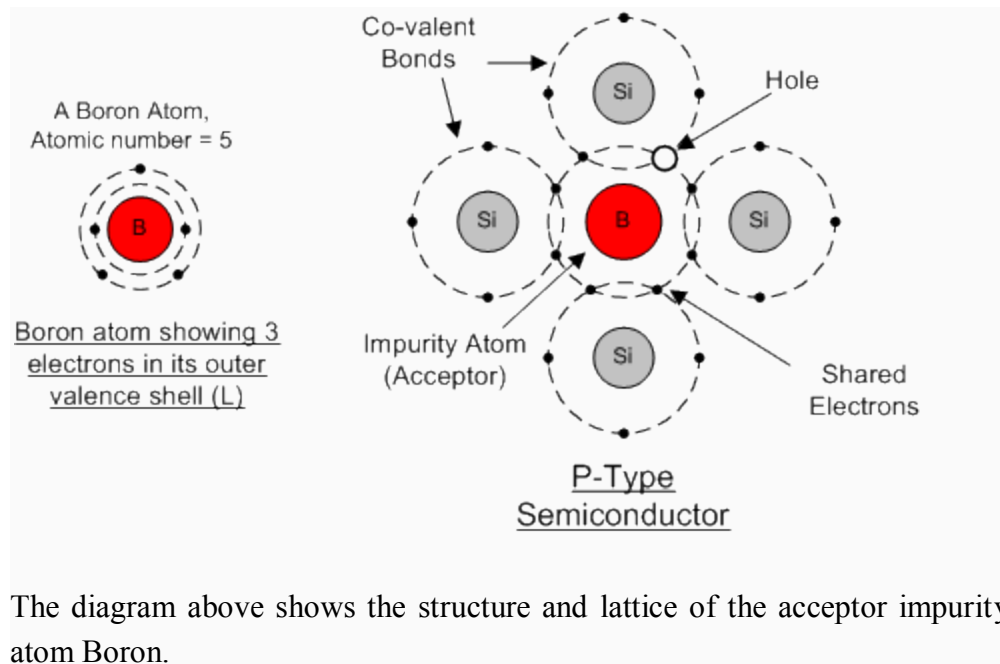


The diagram above shows the structure and lattice of the donor impurity atom Antimony.

### P-Type Semiconductor Basics

If we go the other way, and introduce a "Trivalent" (3-electron) impurity into the crystal structure, such as Aluminium, Boron or Indium, only three valence electrons are available in the outermost covalent bond meaning that the fourth bond cannot be formed. Therefore, a complete connection is not possible, giving the semiconductor material an abundance of positively charged carriers known as "holes" in the structure of the crystal. As there is a hole an adjoining free electron is attracted to it and will try to move into the hole to fill it. However, the electron filling the hole leaves another hole behind it as it moves. This in turn attracts another electron which in turn creates another hole behind, and so forth giving the appearance that the holes are moving as a positive charge through the crystal structure (conventional current flow). As each impurity atom generates a hole, trivalent impurities are generally known as "**Acceptors**" as they are continually "accepting" extra electrons.

**Boron** (symbol B) is frequently used as a trivalent additive as it has only 5 electrons arranged in 3 shells around the nucleus. Addition of Boron causes conduction to consist mainly of positive charge carriers results in a "**P-type**" material and the positive holes are called "Majority Carriers" while the free electrons are called "Minority Carriers". Then a semiconductor is P-type when its acceptor density is greater than its donor density. Therefore, a P-type semiconductor has more holes than electrons.



## Semiconductor Basics Summary

### N-type (e.g. add Antimony)

These are materials which have **Pentavalent** impurity atoms (Donors) added and conduct by "electron" movement and are called, **N-type Semiconductors**.

In these types of materials are:

- 1. The Donors are positively charged.
- 2. There are a large number of free electrons.
- 3. A small number of holes in relation to the number of free electrons.
- 4. Doping gives:
  - positively charged donors.
  - negatively charged free electrons.
- 5. Supply of energy gives:
  - negatively charged free electrons.
  - positively charged holes.

### P-type (e.g. add Boron)

These are materials which have **Trivalent** impurity atoms (Acceptors) added and conduct by "hole" movement and are called, **P-type Semiconductors**.

In these types of materials are:

- 1. The Acceptors are negatively charged.
- 2. There are a large number of holes.
- 3. A small number of free electrons in relation to the number of holes.
- 4. Doping gives:
  - negatively charged acceptors.
  - positively charged holes.
- 5. Supply of energy gives:
  - positively charged holes.
  - negatively charged free electrons.

and both P and N-types as a whole, are electrically neutral.

In the next tutorial about semiconductors and diodes, we will look at joining the two semiconductor materials, the P-type and the N-type materials to form a **PN Junction** which can be used to produce diodes.

## **Pn junction**

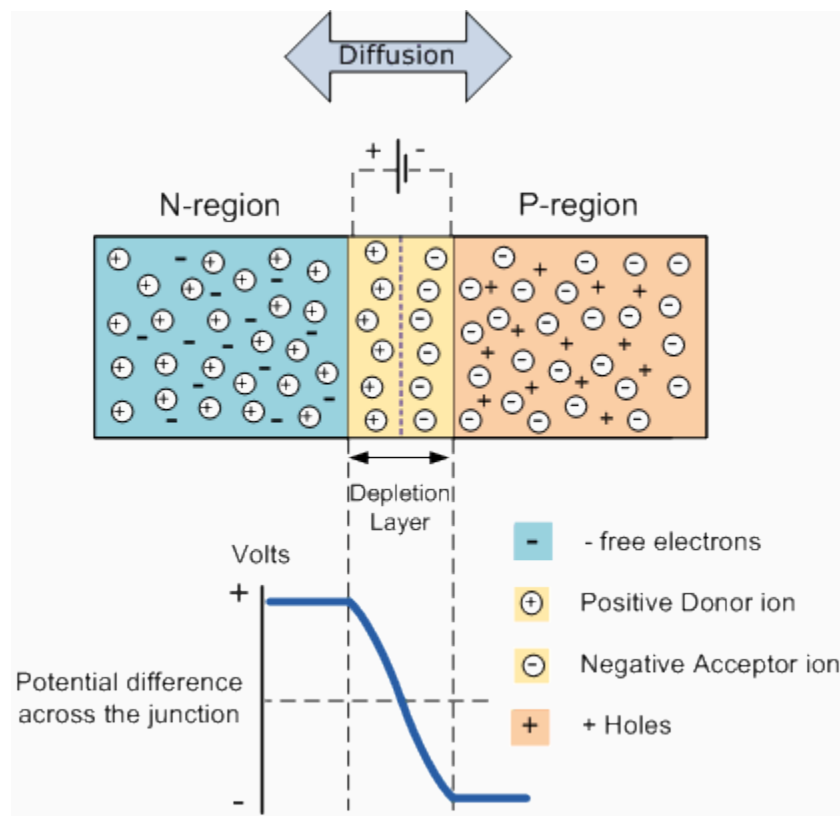
### **The PN junction**

In the previous tutorial we saw how to make an N-type semiconductor material by doping it with Antimony and also how to make a P-type semiconductor material by doping that with Boron. This is all well and good, but these semiconductor N and P-type materials do very little on their own as they are electrically neutral, but when we join (or fuse) them together these two materials behave in a very different way producing what is generally known as a **PN Junction**.

When the N and P-type semiconductor materials are first joined together a very large density gradient exists between both sides of the junction so some of the free electrons from the donor impurity atoms begin to migrate across this newly formed junction to fill up the holes in the P-type material producing negative ions. However, because the electrons have moved across the junction from the N-type silicon to the P-type silicon, they leave behind positively charged donor ions ( $N_D$ ) on the negative side and now the holes from the acceptor impurity migrate across the junction in the opposite direction into the region where there are large numbers of free electrons. As a result, the charge density of the P-type along the junction is filled with negatively charged acceptor ions ( $N_A$ ), and the charge density of the N-type along the junction becomes positive. This charge transfer of electrons and holes across the junction is known as **diffusion**.

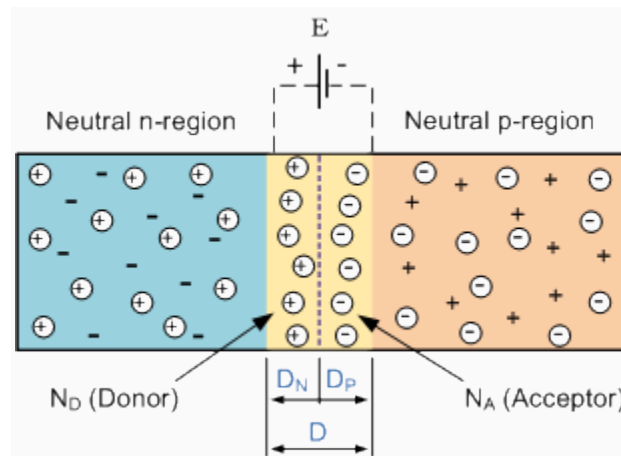
This process continues back and forth until the number of electrons which have crossed the junction have a large enough electrical charge to repel or prevent any more carriers from crossing the junction. The regions on both sides of the junction become depleted of any free carriers in comparison to the N and P type materials away from the junction. Eventually a state of equilibrium (electrically neutral situation) will occur producing a "**potential barrier**" zone around the area of the junction as the donor atoms repel the holes and the acceptor atoms repel the electrons. Since no free charge carriers can rest in a position where there is a potential barrier the regions on both sides of the junction become depleted of any more free carriers in comparison to the N and P type materials away from the junction. This area around the junction is now called the **Depletion Layer**.

### The PN junction



The total charge on each side of the junction must be equal and opposite to maintain a neutral charge condition around the junction. If the depletion layer region has a distance  $D$ , it therefore must therefore penetrate into the silicon by a distance of  $D_p$  for the positive side, and a distance of  $D_n$  for the negative side giving a relationship between the two of  $D_p \cdot N_A = D_n \cdot N_D$  in order to maintain charge neutrality also called equilibrium.

## PN junction Distance



As the N-type material has lost electrons and the P-type has lost holes, the N-type material has become positive with respect to the P-type. Then the presence of impurity ions on both sides of the junction cause an electric field to be established across this region with the N-side at a positive voltage relative to the P-side. The problem now is that a free charge requires some extra energy to overcome the barrier that now exists for it to be able to cross the depletion region junction.

This electric field created by the diffusion process has created a "built-in potential difference" across the junction with an open-circuit (zero bias) potential of:

$$E_0 = V_T \ln \left( \frac{N_D \cdot N_A}{n_i^2} \right)$$

Where:  $E_0$  is the zero bias junction voltage,  $V_T$  the thermal voltage of 26mV at room temperature,  $N_D$  and  $N_A$  are the impurity concentrations and  $n_i$  is the intrinsic concentration.

A suitable positive voltage (forward bias) applied between the two ends of the PN junction can supply the free electrons and holes with the extra energy. The external voltage required to overcome this potential barrier that now exists is very much dependent upon the type of semiconductor material used and its actual temperature. Typically at room temperature the voltage across the depletion layer for silicon is about 0.6 - 0.7 volts and for germanium is about 0.3 - 0.35 volts. This potential barrier will always exist even if the device is not connected to any external power source.

The significance of this built-in potential across the junction, is that it opposes both the flow of holes and electrons across the junction and is why it is called the potential barrier. In practice, a **PN junction** is formed within a single crystal of material rather than just simply joining or fusing together two separate pieces. Electrical contacts are also fused onto either side of the crystal to enable an electrical connection to be made to an external circuit. Then the resulting device that has been made is called a PN junction Diode or Signal Diode.

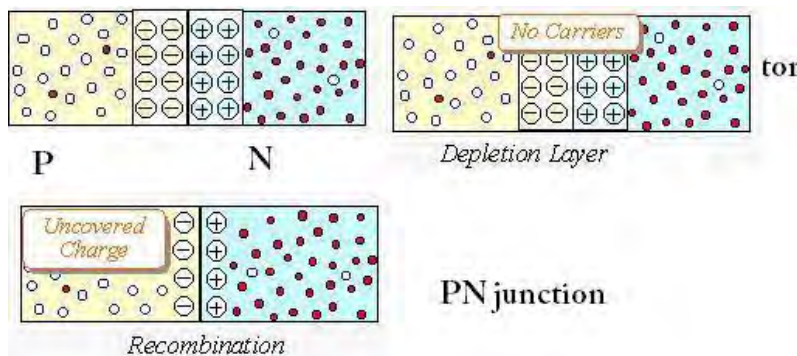
In the next tutorial about the PN junction, we will look at one of the most interesting aspects of the **PN junction** is its use in circuits as a diode. By adding connections to each end of the P-type and the N-type materials we can produce a two terminal device called a **PN Junction Diode** which can be biased by an external voltage to either block or allow the flow of current through it.

### Introduction to depletion layer pn junction:

If one side of crystal pure semiconductor Si(silicon) or Ge(Germanium) is doped with acceptor impurity atoms and the other side is doped with donor impurity atoms , a PN junction is formed as shown in figure.P region has high concentration of holes and N region contains large number of electrons.

### What is Depletion Layer of Pn Junction?

As soon as the junction is formed, free electrons and holes cross through the junction by the process of diffusion.During this process , the electrons crossing the junction from N- region into P-region , recombine with holes in the P-region very close to the junction.Similarly holes crossing the junction from the P-region into the N-region, recombine with electrons in the N- region very close to the junction. Thus a region is formed, which does not have any mobile charge very close to the junction. This region is called the depletion layer of pn junction.



In this region, on the left side of the junction, the acceptor atoms become negative ions and on the right side of the junction, the donor atoms become positive ions as shown in figure.

### **Function of Depletion Layer of Pn Junction :**

An electric field is set up, between the donor and acceptor ions in the depletion layer of the pn junction .The potential at the N-side is higher than the potential at P-side. Therefore electrons in the N- side are prevented to go to the lower potential of P-side. Similarly, holes in the P-side find themselves at a lower potential and are prevented to cross to the N-side. Thus, there is a barrier at the junction which opposes the movement of the majority charge carriers. The difference of potential from one side of the barrier to the other side of the barrier is called potential barrier. The potential barrier is approximately 0.7V for a silicon PN junction and 0.3V for germanium PN junction. The distance from one side of the barrier to the other side is called the width of the barrier, which depends on the nature of the material.

### **Introduction to Diodes and Transistors:**

Diodes are known as p-n junction in the physics or in most of the basic sciences but in the electronics and engineering sciences a p-n junction has the abbreviated name i.e. diode. Diodes are the electrical component of any electric circuit which prevents the circuit from the high electric current because diodes are get break i.e. they get burn when a large current is flowing through them and hence prevents the electrical circuits.

Transistors are the electrical device which mainly consists of two junctions thus they are called as the junction transistors. The transistors have three terminals instead of the two terminals and each terminal has its specific characteristics. In electronic circuits two transistors namely n-p-n and p-n-p are most preferably used.

### **More about Diodes and Transistors:**

When a p-type semiconductor is brought into a close contact with an n-type semiconductor, then the resultant arrangement is called a p-n junction or a junction diode or simply a diode.

A junction transistor is obtained by growing a thin layer of one semiconductor in between two thick layers of other similar type semiconductor. Thus a junction transistor is a semiconductor device having two junctions and three terminals.

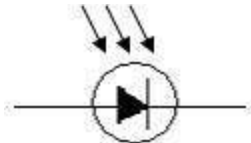
### **Example of Diodes and Transistors:**

Junction diodes are of many types. Important among them are:

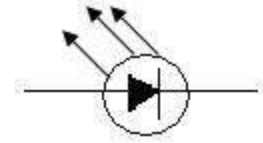
- (a) **Zener diode:** A zener diode is specially designed junction diode which can operate continuously without being damaged in the region of reverse break down voltage shown in **Fig.1**.
- (b) **Photo diode:** Its working is based on the electric conduction from light shown in **Fig.2**.
- (c) **Light emitting diode (LED):** Its working is based on the production of light from electric current shown in **Fig.3**.



**Fig.1** Zener Diode



**Fig.2** Photo Diode



**Fig.3** LED

The transistors are also of several types but most important among them are listed below:

- (a) **Junction Transistors:** These are of two kind p-n-p and n-p-n and they are basic transistors which are used in the electronic circuitry and other electronic equipment very rapidly because of there low cost and high reliability.
- (b) **Field effect transistors:** In present days most of the electronic integrated circuits are using these kinds of transistors because they are highly conductive and easy to prepare then the junction transistor.

## The Junction Diode

### The Junction Diode

The effect described in the previous tutorial is achieved without any external voltage being applied to the actual PN junction resulting in the junction being in a state of equilibrium. However, if we were to make electrical connections at the ends of both the N-type and the P-type materials and then connect them to a battery source, an additional energy source now exists to overcome the barrier resulting in free charges being able to cross the depletion region from one side to the other. The behaviour of the PN junction with regards to the potential barrier width produces an asymmetrical conducting two terminal device, better known as the **Junction Diode**.

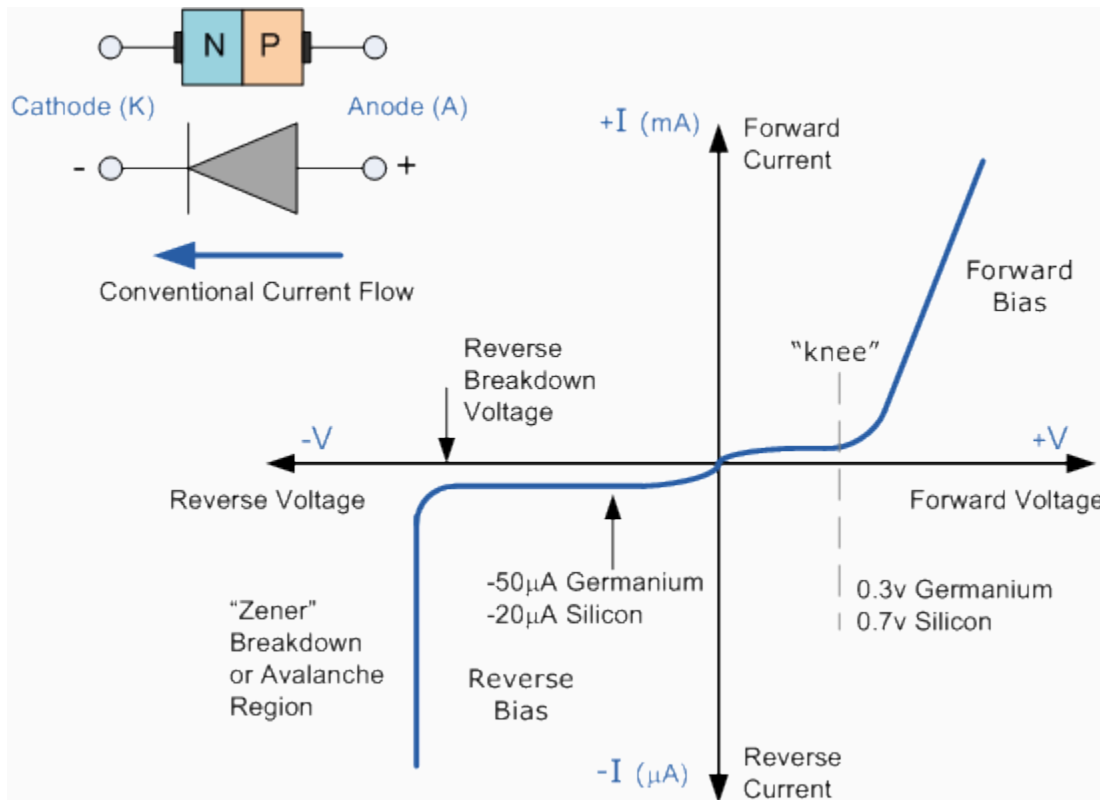
A diode is one of the simplest semiconductor devices, which has the characteristic of passing current in one direction only. However, unlike a resistor, a diode does not behave linearly with

respect to the applied voltage as the diode has an exponential I-V relationship and therefore we can not described its operation by simply using an equation such as Ohm's law.

If a suitable positive voltage (forward bias) is applied between the two ends of the PN junction, it can supply free electrons and holes with the extra energy they require to cross the junction as the width of the depletion layer around the PN junction is decreased. By applying a negative voltage (reverse bias) results in the free charges being pulled away from the junction resulting in the depletion layer width being increased. This has the effect of increasing or decreasing the effective resistance of the junction itself allowing or blocking current flow through the diode.

Then the depletion layer widens with an increase in the application of a reverse voltage and narrows with an increase in the application of a forward voltage. This is due to the differences in the electrical properties on the two sides of the PN junction resulting in physical changes taking place. One of the results produces rectification as seen in the PN junction diodes static I-V (current-voltage) characteristics. Rectification is shown by an asymmetrical current flow when the polarity of bias voltage is altered as shown below.

### Junction Diode Symbol and Static I-V Characteristics.



But before we can use the PN junction as a practical device or as a rectifying device we need to firstly **bias** the junction, ie connect a voltage potential across it. On the voltage axis above, "Reverse Bias" refers to an external voltage potential which increases the potential barrier. An external voltage which decreases the potential barrier is said to act in the "Forward Bias" direction.

There are two operating regions and three possible "biasing" conditions for the standard **Junction Diode** and these are:

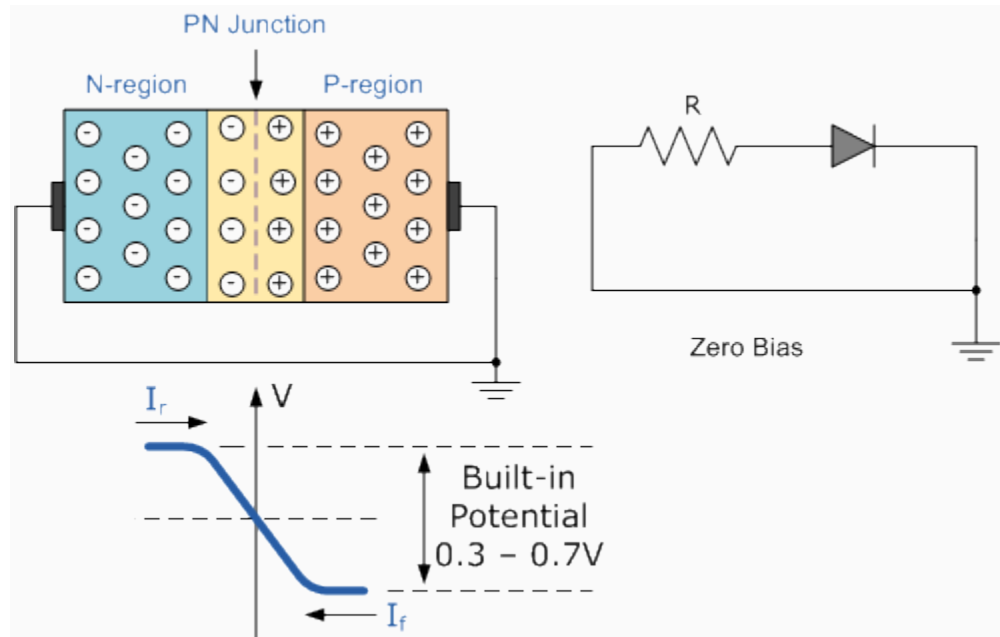
- 1. Zero Bias - No external voltage potential is applied to the PN-junction.
- 2. Reverse Bias - The voltage potential is connected negative, (-ve) to the P-type material and positive, (+ve) to the N-type material across the diode which has the effect of **Increasing** the PN-junction width.
- 3. Forward Bias - The voltage potential is connected positive, (+ve) to the P-type material and negative, (-ve) to the N-type material across the diode which has the effect of **Decreasing** the PN-junction width.

### **Zero Biased Junction Diode**

When a diode is connected in a **Zero Bias** condition, no external potential energy is applied to the PN junction. However if the diodes terminals are shorted together, a few holes (majority carriers) in the P-type material with enough energy to overcome the potential barrier will move across the junction against this barrier potential. This is known as the "**Forward Current**" and is referenced as  $I_F$

Likewise, holes generated in the N-type material (minority carriers), find this situation favourable and move across the junction in the opposite direction. This is known as the "**Reverse Current**" and is referenced as  $I_R$ . This transfer of electrons and holes back and forth across the PN junction is known as diffusion, as shown below.

## Zero Biased Junction Diode



The potential barrier that now exists discourages the diffusion of any more majority carriers across the junction. However, the potential barrier helps minority carriers (few free electrons in the P-region and few holes in the N-region) to drift across the junction. Then an "Equilibrium" or balance will be established when the majority carriers are equal and both moving in opposite directions, so that the net result is zero current flowing in the circuit. When this occurs the junction is said to be in a state of "**Dynamic Equilibrium**".

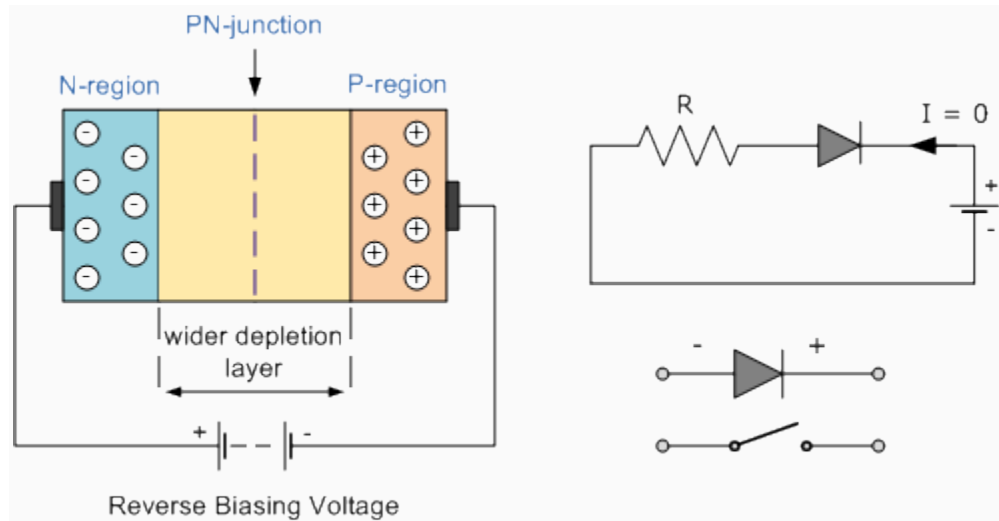
The minority carriers are constantly generated due to thermal energy so this state of equilibrium can be broken by raising the temperature of the PN junction causing an increase in the generation of minority carriers, thereby resulting in an increase in leakage current but an electric current cannot flow since no circuit has been connected to the PN junction.

## Reverse Biased Junction Diode

When a diode is connected in a **Reverse Bias** condition, a positive voltage is applied to the N-type material and a negative voltage is applied to the P-type material. The positive voltage applied to the N-type material attracts electrons towards the positive electrode and away from the junction, while the holes in the P-type end are also attracted away from the junction towards the negative electrode. The net result is that the depletion layer grows wider due to a lack of electrons and holes and presents a high impedance path, almost an insulator. The result is that a

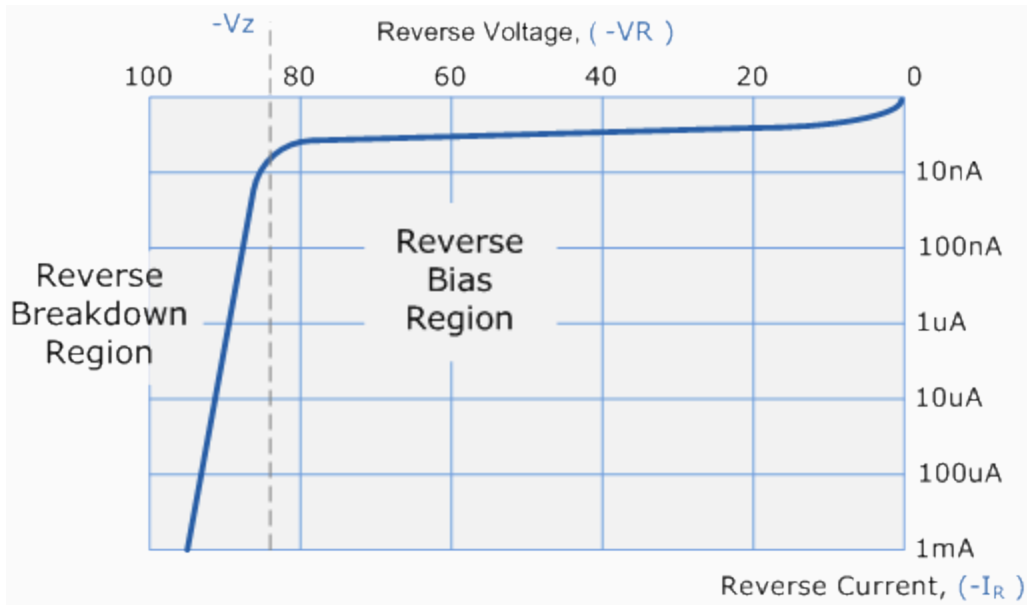
high potential barrier is created thus preventing current from flowing through the semiconductor material.

### Reverse Biased Junction Diode showing an Increase in the Depletion Layer



This condition represents a high resistance value to the PN junction and practically zero current flows through the junction diode with an increase in bias voltage. However, a very small **leakage current** does flow through the junction which can be measured in microamperes, ( $\mu\text{A}$ ). One final point, if the reverse bias voltage  $V_r$  applied to the diode is increased to a sufficiently high enough value, it will cause the PN junction to overheat and fail due to the avalanche effect around the junction. This may cause the diode to become shorted and will result in the flow of maximum circuit current, and this shown as a step downward slope in the reverse static characteristics curve below.

### Reverse Characteristics Curve for a Junction Diode

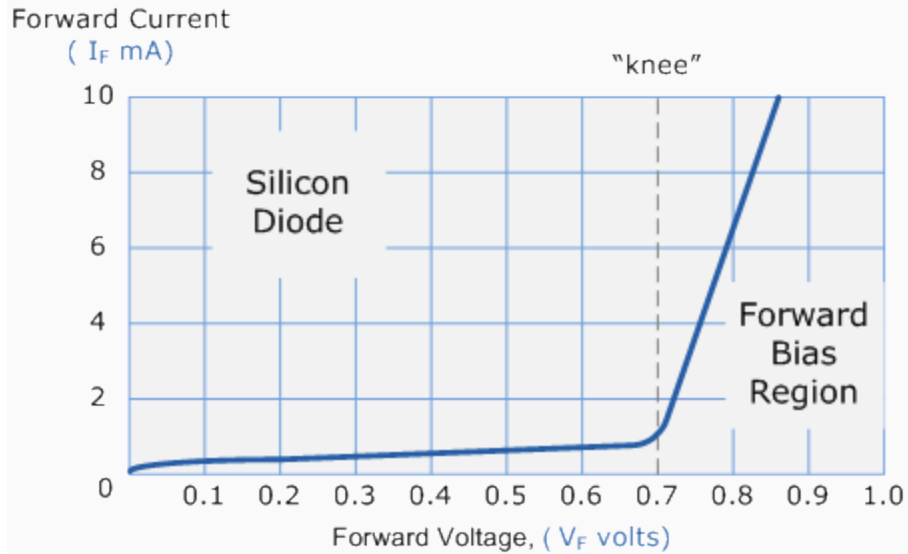


Sometimes this avalanche effect has practical applications in voltage stabilising circuits where a series limiting resistor is used with the diode to limit this reverse breakdown current to a preset maximum value thereby producing a fixed voltage output across the diode. These types of diodes are commonly known as **Zener Diodes** and are discussed in a later tutorial.

### Forward Biased Junction Diode

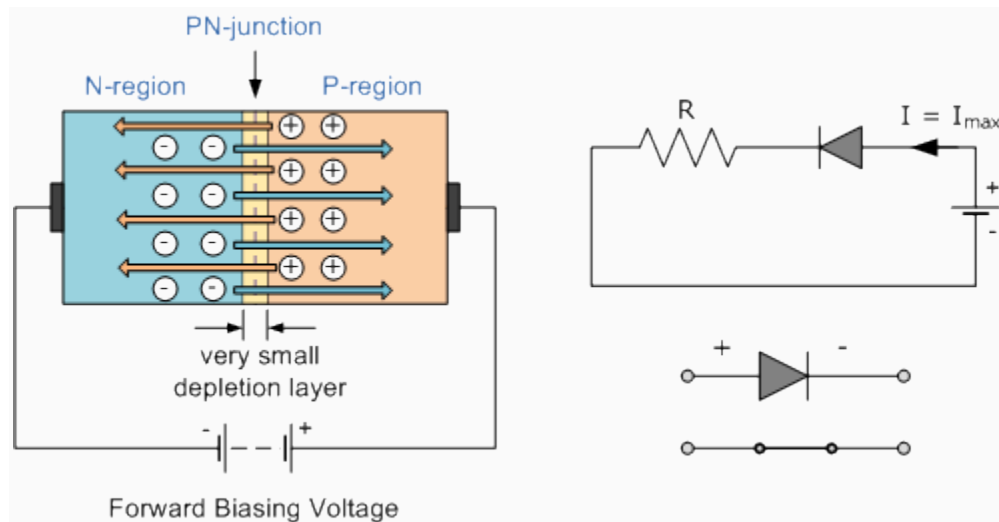
When a diode is connected in a **Forward Bias** condition, a negative voltage is applied to the N-type material and a positive voltage is applied to the P-type material. If this external voltage becomes greater than the value of the potential barrier, approx. 0.7 volts for silicon and 0.3 volts for germanium, the potential barriers opposition will be overcome and current will start to flow. This is because the negative voltage pushes or repels electrons towards the junction giving them the energy to cross over and combine with the holes being pushed in the opposite direction towards the junction by the positive voltage. This results in a characteristics curve of zero current flowing up to this voltage point, called the "knee" on the static curves and then a high current flow through the diode with little increase in the external voltage as shown below.

### Forward Characteristics Curve for a Junction Diode



The application of a forward biasing voltage on the junction diode results in the depletion layer becoming very thin and narrow which represents a low impedance path through the junction thereby allowing high currents to flow. The point at which this sudden increase in current takes place is represented on the static I-V characteristics curve above as the "knee" point.

### Forward Biased Junction Diode showing a Reduction in the Depletion Layer



This condition represents the low resistance path through the PN junction allowing very large currents to flow through the diode with only a small increase in bias voltage. The actual potential difference across the junction or diode is kept constant by the action of the depletion layer at approximately 0.3v for germanium and approximately 0.7v for silicon junction diodes. Since the diode can conduct "infinite" current above this knee point as it effectively becomes a short circuit, therefore resistors are used in series with the diode to limit its current flow. Exceeding its

maximum forward current specification causes the device to dissipate more power in the form of heat than it was designed for resulting in a very quick failure of the device.

### **Junction Diode Summary**

The PN junction region of a **Junction Diode** has the following important characteristics:

- 1). Semiconductors contain two types of mobile charge carriers, **Holes** and **Electrons**.
- 
- 2). The holes are positively charged while the electrons negatively charged.
- 
- 3). A semiconductor may be doped with donor impurities such as Antimony (N-type doping), so that it contains mobile charges which are primarily electrons.
- 
- 4). A semiconductor may be doped with acceptor impurities such as Boron (P-type doping), so that it contains mobile charges which are mainly holes.
- 
- 5). The junction region itself has no charge carriers and is known as the depletion region.
- 
- 6). The junction (depletion) region has a physical thickness that varies with the applied voltage.
- 
- 7). When a diode is **Zero Biased** no external energy source is applied and a natural **Potential Barrier** is developed across a depletion layer which is approximately 0.5 to 0.7v for silicon diodes and approximately 0.3 of a volt for germanium diodes.
- 
- 8). When a junction diode is **Forward Biased** the thickness of the depletion region reduces and the diode acts like a short circuit allowing full current to flow.
- 
- 9). When a junction diode is **Reverse Biased** the thickness of the depletion region increases and the diode acts like an open circuit blocking any current flow, (only a very small leakage current).

In the next tutorial about diodes, we will look at the small signal diode sometimes called a switching diode that are used in general electronic circuits. A signal diode is designed for low-voltage or high frequency signal applications such as in radio or digital switching circuits as opposed to the high-current mains rectification diodes in which silicon diodes are usually used, and examine the **Signal Diode** static current-voltage characteristics curve and parameters.

## Signal Diode

### The Signal Diode

The semiconductor **Signal Diode** is a small non-linear semiconductor devices generally used in electronic circuits, where small currents or high frequencies are involved such as in radio, television and digital logic circuits. The signal diode which is also sometimes known by its older name of the **Point Contact Diode** or the **Glass Passivated Diode**, are physically very small in size compared to their larger **Power Diode** cousins.

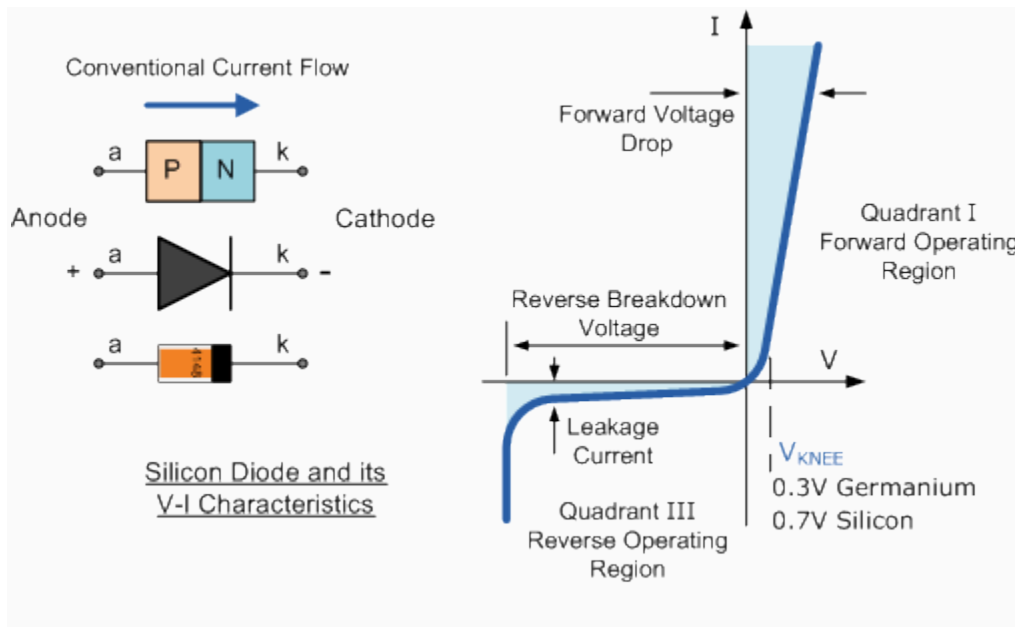
Generally, the PN junction of a small signal diode is encapsulated in glass to protect the PN junction, and usually have a red or black band at one end of their body to help identify which end is the cathode terminal. The most widely used of all the glass encapsulated signal diodes is the very common *1N4148* and its equivalent *1N914* signal diode. Small signal and switching diodes have much lower power and current ratings, around 150mA, 500mW maximum compared to rectifier diodes, but they can function better in high frequency applications or in clipping and switching applications that deal with short-duration pulse waveforms.

The characteristics of a signal point contact diode are different for both germanium and silicon types and are given as:

- Germanium Signal Diodes - These have a low reverse resistance value giving a lower forward volt drop across the junction, typically only about 0.2-0.3v, but have a higher forward resistance value because of their small junction area.
- 
- Silicon Signal Diodes - These have a very high value of reverse resistance and give a forward volt drop of about 0.6-0.7v across the junction. They have fairly low values of forward resistance giving them high peak values of forward current and reverse voltage.

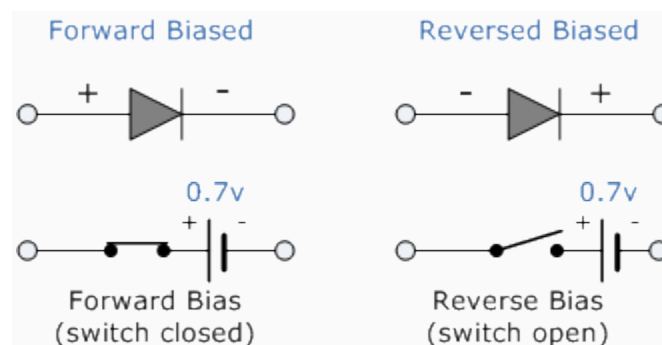
The electronic symbol given for any type of diode is that of an arrow with a bar or line at its end and this is illustrated below along with the Steady State V-I Characteristics Curve.

## Silicon Diode V-I Characteristic Curve



The arrow points in the direction of conventional current flow through the diode meaning that the diode will only conduct if a positive supply is connected to the Anode (a) terminal and a negative supply is connected to the Cathode (k) terminal thus only allowing current to flow through it in one direction only, acting more like a one way electrical valve, (Forward Biased Condition). However, we know from the previous tutorial that if we connect the external energy source in the other direction the diode will block any current flowing through it and instead will act like an open switch, (Reversed Biased Condition) as shown below.

### Forward and Reversed Biased Diode



Then we can say that an ideal small signal diode conducts current in one direction (forward-conducting) and blocks current in the other direction (reverse-blocking). Signal Diodes are used

in a wide variety of applications such as a switch in rectifiers, limiters, snubbers or in wave-shaping circuits.

## Signal Diode Parameters

**Signal Diodes** are manufactured in a range of voltage and current ratings and care must be taken when choosing a diode for a certain application. There are a bewildering array of static characteristics associated with the humble signal diode but the more important ones are.

### 1. Maximum Forward Current

The **Maximum Forward Current** ( $I_{F(max)}$ ) is as its name implies the *maximum forward current* allowed to flow through the device. When the diode is conducting in the forward bias condition, it has a very small "ON" resistance across the PN junction and therefore, power is dissipated across this junction (**Ohm's Law**) in the form of heat. Then, exceeding its ( $I_{F(max)}$ ) value will cause more heat to be generated across the junction and the diode will fail due to thermal overload, usually with destructive consequences. When operating diodes around their maximum current ratings it is always best to provide additional cooling to dissipate the heat produced by the diode.

For example, our small 1N4148 signal diode has a maximum current rating of about 150mA with a power dissipation of 500mW at 25°C. Then a resistor must be used in series with the diode to limit the forward current, ( $I_{F(max)}$ ) through it to below this value.

### 2. Peak Inverse Voltage

The **Peak Inverse Voltage** (PIV) or *Maximum Reverse Voltage* ( $V_{R(max)}$ ), is the maximum allowable **Reverse** operating voltage that can be applied across the diode without reverse breakdown and damage occurring to the device. This rating therefore, is usually less than the "avalanche breakdown" level on the reverse bias characteristic curve. Typical values of  $V_{R(max)}$  range from a few volts to thousands of volts and must be considered when replacing a diode.

The peak inverse voltage is an important parameter and is mainly used for rectifying diodes in AC rectifier circuits with reference to the amplitude of the voltage were the sinusoidal waveform changes from a positive to a negative value on each and every cycle.

### 3. Forward Power Dissipation

Signal diodes have a **Forward Power Dissipation**, ( $P_{D(max)}$ ) rating. This rating is the maximum possible power dissipation of the diode when it is forward biased (conducting). When current

flows through the signal diode the biasing of the PN junction is not perfect and offers some resistance to the flow of current resulting in power being dissipated (lost) in the diode in the form of heat. As small signal diodes are nonlinear devices the resistance of the PN junction is not constant, it is a dynamic property then we cannot use Ohms Law to define the power in terms of current and resistance or voltage and resistance as we can for resistors. Then to find the power that will be dissipated by the diode we must multiply the voltage drop across it times the current flowing through it:  $P_D = V \times I$

#### 4. Maximum Operating Temperature

The **Maximum Operating Temperature** actually relates to the *Junction Temperature* ( $T_J$ ) of the diode and is related to maximum power dissipation. It is the maximum temperature allowable before the structure of the diode deteriorates and is expressed in units of degrees centigrade per Watt, ( $^{\circ}\text{C}/\text{W}$ ). This value is linked closely to the maximum forward current of the device so that at this value the temperature of the junction is not exceeded. However, the maximum forward current will also depend upon the ambient temperature in which the device is operating so the maximum forward current is usually quoted for two or more ambient temperature values such as  $25^{\circ}\text{C}$  or  $70^{\circ}\text{C}$ .

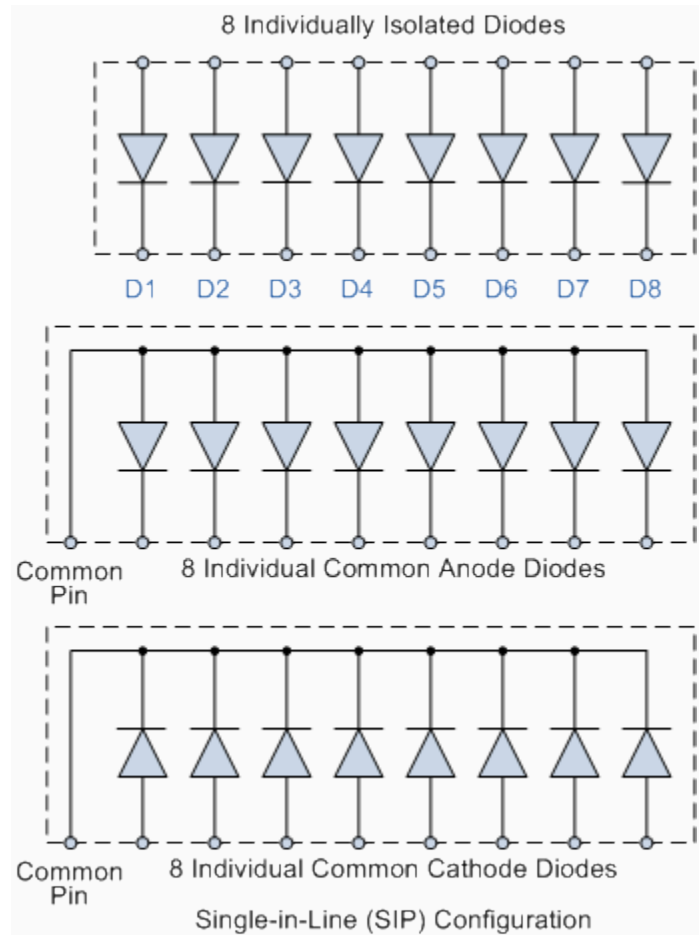
Then there are three main parameters that must be considered when either selecting or replacing a signal diode and these are:

- The Reverse Voltage Rating
- The Forward Current Rating
- The Forward Power Dissipation Rating

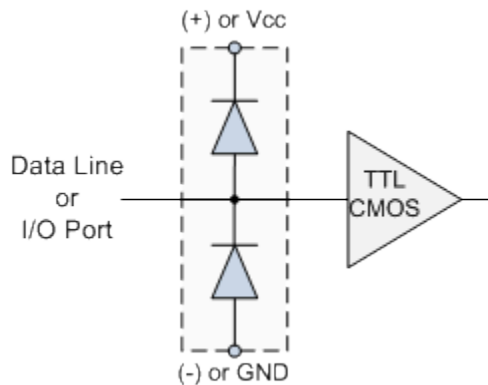
#### Signal Diode Arrays

When space is limited, or matching pairs of switching signal diodes are required, diode arrays can be very useful. They generally consist of low capacitance high speed silicon diodes such as the 1N4148 connected together in multiple diode packages called an array for use in switching and clamping in digital circuits. They are encased in single inline packages (SIP) containing 4 or more diodes connected internally to give either an individual isolated array, common cathode, (CC), or a common anode, (CA) configuration as shown.

## Signal Diode Arrays



Signal diode arrays can also be used in digital and computer circuits to protect high speed data lines or other input/output parallel ports against electrostatic discharge, (ESD) and voltage transients. By connecting two diodes in series across the supply rails with the data line connected to their junction as shown, any unwanted transients are quickly dissipated and as the signal diodes are available in 8-fold arrays they can protect eight data lines in a single package.



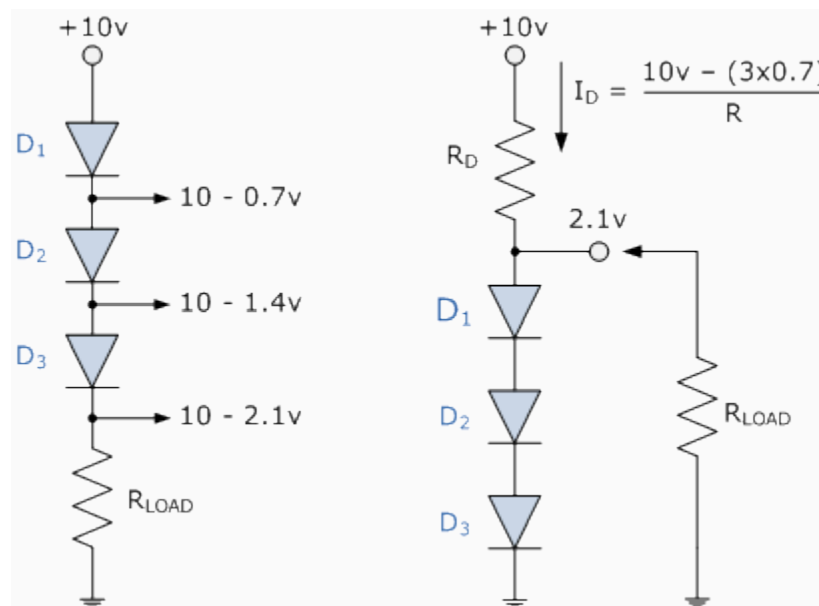
## CPU Data Line Protection

Signal diode arrays can also be used to connect together diodes in either series or parallel combinations to form voltage regulator or voltage reducing type circuits or to produce a known fixed voltage. We know that the forward volt drop across a silicon diode is about 0.7v and by connecting together a number of diodes in series the total voltage drop will be the sum of the individual voltage drops of each diode. However, when signal diodes are connected together in series, the current will be the same for each diode so the maximum forward current must not be exceeded.

### Connecting Signal Diodes in Series

Another application for the small signal diode is to create a regulated voltage supply. Diodes are connected together in series to provide a constant DC voltage across the diode combination. The output voltage across the diodes remains constant in spite of changes in the load current drawn from the series combination or changes in the DC power supply voltage that feeds them. Consider the circuit below.

### Signal Diodes in Series



As the forward voltage drop across a silicon diode is almost constant at about 0.7v, while the current through it varies by relatively large amounts, a forward-biased signal diode can make a simple voltage regulating circuit. The individual voltage drops across each diode are subtracted from the supply voltage to leave a certain voltage potential across the load resistor, and in our simple example above this is given as  $10v - (3 \times 0.7v) = 7.9v$ . This is because each diode has a

junction resistance relating to the small signal current flowing through it and the three signal diodes in series will have three times the value of this resistance, along with the load resistance  $R$ , forms a voltage divider across the supply.

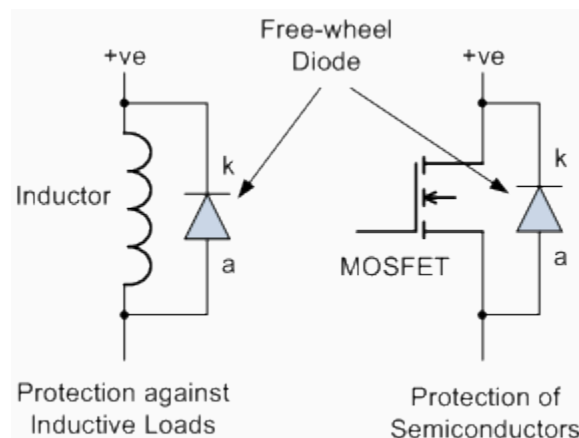
By adding more diodes in series a greater voltage reduction will occur. Also series connected diodes can be placed in parallel with the load resistor to act as a voltage regulating circuit. Here the voltage applied to the load resistor will be  $3 \times 0.7\text{v} = 2.1\text{v}$ . We can of course produce the same constant voltage source using a single **Zener Diode**. Resistor,  $R_D$  is used to prevent excessive current flowing through the diodes if the load is removed.

### Freewheel Diodes

Signal diodes can also be used in a variety of clamping, protection and wave shaping circuits with the most common form of clamping diode circuit being one which uses a diode connected in parallel with a coil or inductive load to prevent damage to the delicate switching circuit by suppressing the voltage spikes and/or transients that are generated when the load is suddenly turned "OFF". This type of diode is generally known as a "Free-wheeling Diode" or **Freewheel diode** as it is more commonly called.

The **Freewheel diode** is used to protect solid state switches such as power transistors and MOSFET's from damage by reverse battery protection as well as protection from highly inductive loads such as relay coils or motors, and an example of its connection is shown below.

### Use of the Freewheel Diode



Modern fast switching, power semiconductor devices require fast switching diodes such as free wheeling diodes to protect them from inductive loads such as motor coils or relay windings. Every time the switching device above is turned "ON", the freewheel diode changes from a conducting state to a blocking state as it becomes reversed biased. However, when the device

rapidly turns "OFF", the diode becomes forward biased and the collapse of the energy stored in the coil causes a current to flow through the freewheel diode. Without the protection of the freewheel diode high di/dt currents would occur causing a high voltage spike or transient to flow around the circuit possibly damaging the switching device.

Previously, the operating speed of the semiconductor switching device, either transistor, MOSFET, IGBT or digital has been impaired by the addition of a freewheel diode across the inductive load with Schottky and Zener diodes being used instead in some applications. But during the past few years however, freewheel diodes had regained importance due mainly to their improved reverse-recovery characteristics and the use of super fast semiconductor materials capable at operating at high switching frequencies.

Other types of specialized diodes not included here are Photo-Diodes, PIN Diodes, Tunnel Diodes and Schottky Barrier Diodes. By adding more PN junctions to the basic two layer diode structure other types of semiconductor devices can be made. For example a three layer semiconductor device becomes a **Transistor**, a four layer semiconductor device becomes a Thyristor or Silicon Controlled Rectifier and five layer devices known as Triacs are also available.

In the next tutorial about diodes, we will look at the large signal diode sometimes called the **Power Diode**. Power diodes are silicon diodes designed for use in high-voltage, high-current mains rectification circuits.

## Power Diodes and Rectifiers

### The Power Diode

In the previous tutorials we saw that a semiconductor signal diode will only conduct current in one direction from its anode to its cathode (forward direction), but not in the reverse direction acting a bit like an electrical one way valve. A widely used application of this feature is in the conversion of an alternating voltage (AC) into a continuous voltage (DC). In other words, Rectification. Small signal diodes can be used as rectifiers in low-power, low current (less than 1-amp) rectifiers or applications, but were larger forward bias currents or higher reverse bias blocking voltages are involved the PN junction of a small signal diode would eventually overheat and melt so larger more robust **Power Diodes** are used instead.

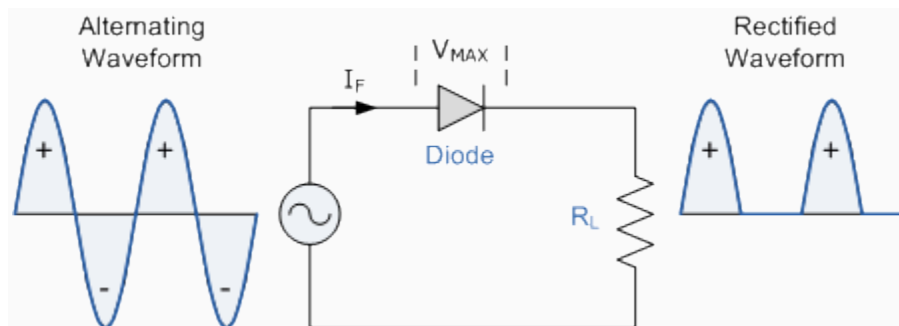
The power semiconductor diode, known simply as the **Power Diode**, has a much larger PN junction area compared to its smaller signal diode cousin, resulting in a high forward current

capability of up to several hundred amps (KA) and a reverse blocking voltage of up to several thousand volts (KV). Since the power diode has a large PN junction, it is not suitable for high frequency applications above 1MHz, but special and expensive high frequency, high current diodes are available. For high frequency rectifier applications **Schottky Diodes** are generally used because of their short reverse recovery time and low voltage drop in their forward bias condition.

Power diodes provide uncontrolled rectification of power and are used in applications such as battery charging and DC power supplies as well as AC rectifiers and inverters. Due to their high current and voltage characteristics they can also be used as freewheeling diodes and snubber networks. Power diodes are designed to have a forward "ON" resistance of fractions of an Ohm while their reverse blocking resistance is in the mega-Ohms range. Some of the larger value power diodes are designed to be "stud mounted" onto heatsinks reducing their thermal resistance to between 0.1 to 1°C/Watt.

If an alternating voltage is applied across a power diode, during the positive half cycle the diode will conduct passing current and during the negative half cycle the diode will not conduct blocking the flow of current. Then conduction through the power diode only occurs during the positive half cycle and is therefore unidirectional i.e. DC as shown.

### Power Diode Rectifier



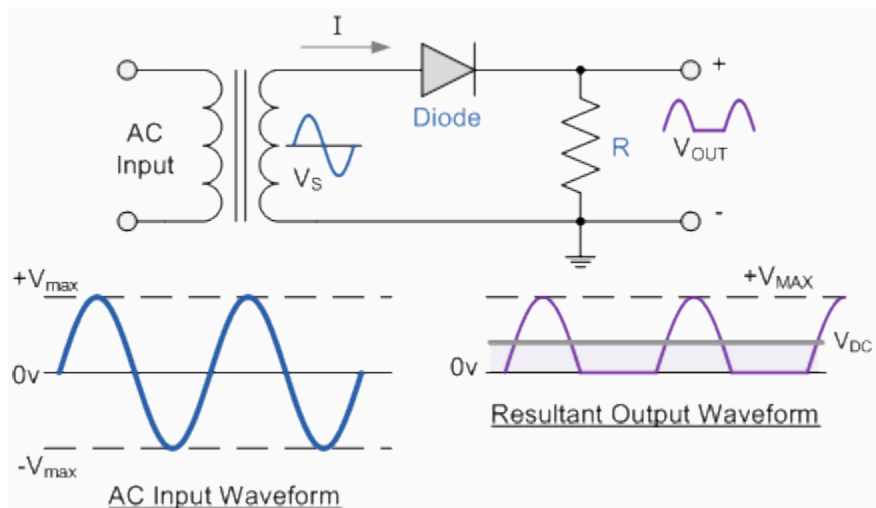
Power diodes can be used individually as above or connected together to produce a variety of rectifier circuits such as "Half-Wave", "Full-Wave" or as "Bridge Rectifiers". Each type of rectifier circuit can be classed as either uncontrolled, half-controlled or fully controlled were an uncontrolled rectifier uses only power diodes, a fully controlled rectifier uses thyristors (SCRs) and a half controlled rectifier is a mixture of both diodes and thyristors. The most commonly used individual power diode for basic electronics applications is the general purpose 1N400x Series Glass Passivated type rectifying diode with standard ratings of continuous forward rectified current of 1.0 amp and reverse blocking voltage ratings from 50v for the 1N4001 up to

1000v for the 1N4007, with the small 1N4007GP being the most popular for general purpose mains voltage rectification.

## Half Wave Rectification

A rectifier is a circuit which converts the *Alternating Current* (AC) input power into a *Direct Current* (DC) output power. The input power supply may be either a single-phase or a multi-phase supply with the simplest of all the rectifier circuits being that of the **Half Wave Rectifier**. The power diode in a half wave rectifier circuit passes just one half of each complete sine wave of the AC supply in order to convert it into a DC supply. Then this type of circuit is called a "half-wave" rectifier because it passes only half of the incoming AC power supply as shown below.

### Half Wave Rectifier Circuit



During each "positive" half cycle of the AC sine wave, the diode is *forward biased* as the anode is positive with respect to the cathode resulting in current flowing through the diode. Since the DC load is resistive (resistor,  $R$ ), the current flowing in the load resistor is therefore proportional to the voltage (**Ohm's Law**), and the voltage across the load resistor will therefore be the same as the supply voltage,  $V_s$  (minus  $V_f$ ), that is the "DC" voltage across the load is sinusoidal for the first half cycle only so  $V_{out} = V_s$ .

During each "negative" half cycle of the AC sine wave, the diode is *reverse biased* as the anode is negative with respect to the cathode therefore, No current flows through the diode or circuit. Then in the negative half cycle of the supply, no current flows in the load resistor as no voltage appears across it so  $V_{out} = 0$ .

The current on the DC side of the circuit flows in one direction only making the circuit **Unidirectional** and the value of the DC voltage  $V_{DC}$  across the load resistor is calculated as follows.

$$V_{d.c.} = \frac{V_{\max}}{\pi} = 0.318V_{\max} = 0.45V_S$$

Where  $V_{\max}$  is the maximum voltage value of the AC supply, and  $V_S$  is the r.m.s. value of the supply.

### Example No1.

Calculate the current, ( $I_{DC}$ ) flowing through a  $100\Omega$  resistor connected to a 240v single phase half-wave rectifier as shown above. Also calculate the power consumed by the load.

$$V_{d.c.} = 0.45V_S = 0.45 \times 240 = 108v$$

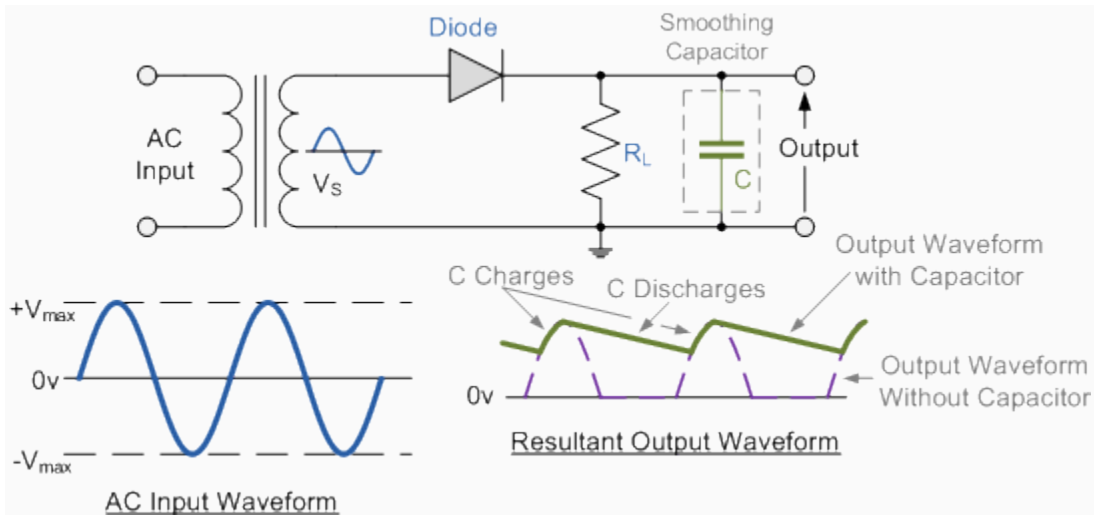
$$I_{d.c.} = \frac{V_{d.c.}}{R} = \frac{108}{100} = 1.08 \text{ amps}$$

$$\text{Power} = I^2 R = 1.08^2 \times 100 = 116 \text{ watts}$$

During the rectification process the resultant output DC voltage and current are therefore both "ON" and "OFF" during every cycle. As the voltage across the load resistor is only present during the positive half of the cycle (50% of the input waveform), this results in a low average DC value being supplied to the load. The variation of the rectified output waveform between this ON and OFF condition produces a waveform which has large amounts of "ripple" which is an undesirable feature. The resultant DC ripple has a frequency that is equal to that of the AC supply frequency.

Very often when rectifying an alternating voltage we wish to produce a "steady" and continuous DC voltage free from any voltage variations or ripple. One way of doing this is to connect a large value **Capacitor** across the output voltage terminals in parallel with the load resistor as shown below. This type of capacitor is known commonly as a "Reservoir" or *Smoothing Capacitor*.

## Half-wave Rectifier with Smoothing Capacitor



When rectification is used to provide a direct voltage power supply from an alternating source, the amount of ripple can be further reduced by using larger value capacitors but there are limits both on cost and size. For a given capacitor value, a greater load current (smaller load resistor) will discharge the capacitor more quickly (**RC Time Constant**) and so increases the ripple obtained. Then for single phase, half-wave rectifier circuits it is not very practical to try and reduce the ripple voltage by capacitor smoothing alone, it is more practical to use "Full-wave Rectification" instead.

In practice, the half-wave rectifier is used most often in low-power applications because of their major disadvantages being. The output amplitude is less than the input amplitude, there is no output during the negative half cycle so half the power is wasted and the output is pulsed DC resulting in excessive ripple. To overcome these disadvantages a number of **Power Diodes** are connected together to produce a **Full Wave Rectifier** as discussed in the next tutorial.

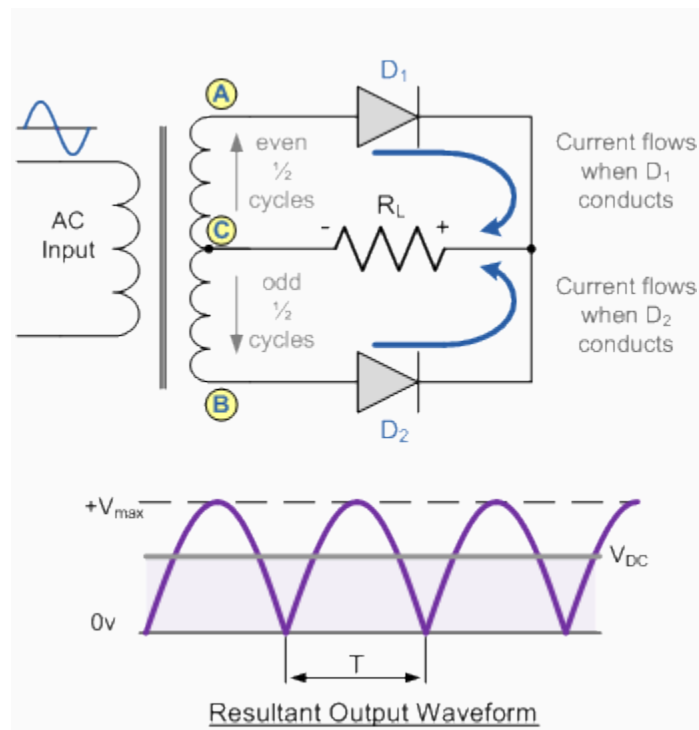
## Full Wave Rectifier

In the previous **Power Diodes** tutorial we discussed ways of reducing the ripple or voltage variations on a direct DC voltage by connecting capacitors across the load resistance. While this method may be suitable for low power applications it is unsuitable to applications which need a "steady and smooth" DC supply voltage. One method to improve on this is to use every half-cycle of the input voltage instead of every other half-cycle. The circuit which allows us to do this is called a **Full Wave Rectifier**.

Like the half wave circuit, a full wave rectifier circuit produces an output voltage or current which is purely DC or has some specified DC component. Full wave rectifiers have some fundamental advantages over their half wave rectifier counterparts. The average (DC) output voltage is higher than for half wave, the output of the full wave rectifier has much less ripple than that of the half wave rectifier producing a smoother output waveform.

In a **Full Wave Rectifier** circuit two diodes are now used, one for each half of the cycle. A transformer is used whose secondary winding is split equally into two halves with a common centre tapped connection, (C). This configuration results in each diode conducting in turn when its anode terminal is positive with respect to the transformer centre point C producing an output during both half-cycles, twice that for the half wave rectifier so it is 100% efficient as shown below.

### Full Wave Rectifier Circuit



The full wave rectifier circuit consists of two *power diodes* connected to a single load resistance ( $R_L$ ) with each diode taking it in turn to supply current to the load. When point A of the transformer is positive with respect to point B, diode  $D_1$  conducts in the forward direction as indicated by the arrows. When point B is positive (in the negative half of the cycle) with respect to point A, diode  $D_2$  conducts in the forward direction and the current flowing through resistor R is in the same direction for both circuits. As the output voltage across the resistor R is the phasor

sum of the two waveforms combined, this type of full wave rectifier circuit is also known as a "bi-phase" circuit.

As the spaces between each half-wave developed by each diode is now being filled in by the other diode the average DC output voltage across the load resistor is now double that of the single half-wave rectifier circuit and is about  $0.637V_{\max}$  of the peak voltage, assuming no losses. However in reality, during each half cycle the current flows through two diodes instead of just one so the amplitude of the output voltage is two voltage drops (  $2 \times 0.7 = 1.4V$  ) less than the input  $V_{\max}$  amplitude.

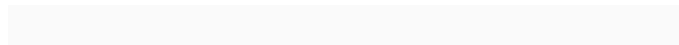
$$V_{d.c.} = \frac{2V_{\max}}{\pi} = 0.637V_{\max} = 0.9V_S$$

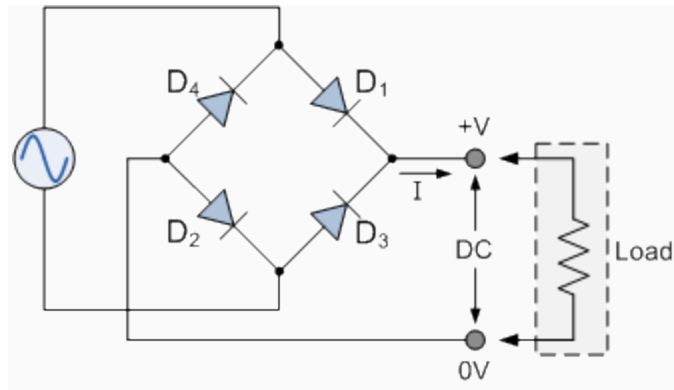
The peak voltage of the output waveform is the same as before for the half-wave rectifier provided each half of the transformer windings have the same rms voltage value. To obtain a different DC voltage output different transformer ratios can be used. The main disadvantage of this type of full wave rectifier circuit is that a larger transformer for a given power output is required with two separate but identical secondary windings making this type of full wave rectifying circuit costly compared to the "Full Wave Bridge Rectifier" circuit equivalent.

### **The Full Wave Bridge Rectifier**

Another type of circuit that produces the same output waveform as the full wave rectifier circuit above, is that of the **Full Wave Bridge Rectifier**. This type of single phase rectifier uses four individual rectifying diodes connected in a closed loop "bridge" configuration to produce the desired output. The main advantage of this bridge circuit is that it does not require a special centre tapped transformer, thereby reducing its size and cost. The single secondary winding is connected to one side of the diode bridge network and the load to the other side as shown below.

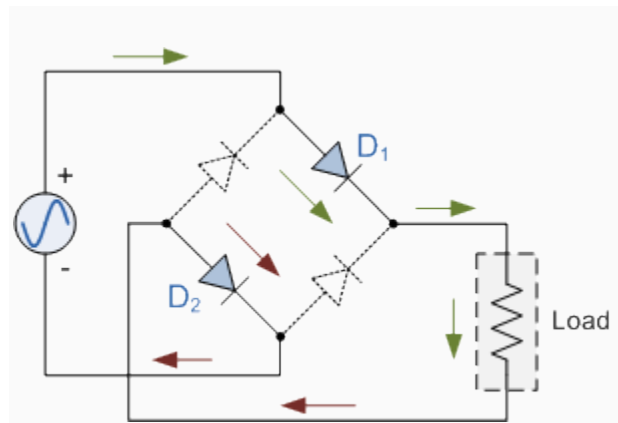
### **The Diode Bridge Rectifier**





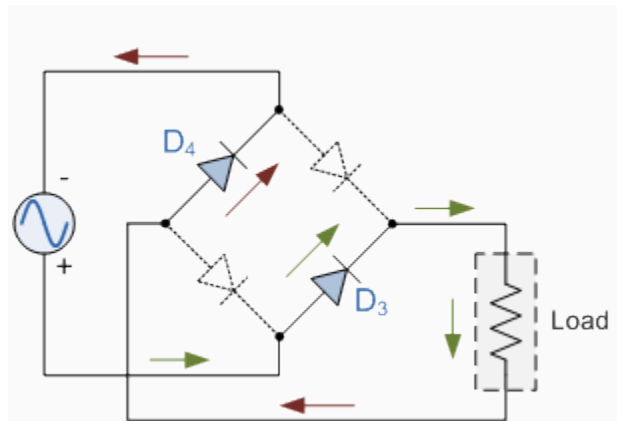
The four diodes labelled  $D_1$  to  $D_4$  are arranged in "series pairs" with only two diodes conducting current during each half cycle. During the positive half cycle of the supply, diodes  $D_1$  and  $D_2$  conduct in series while diodes  $D_3$  and  $D_4$  are reverse biased and the current flows through the load as shown below.

### The Positive Half-cycle



During the negative half cycle of the supply, diodes  $D_3$  and  $D_4$  conduct in series, but diodes  $D_1$  and  $D_2$  switch off as they are now reverse biased. The current flowing through the load is the same direction as before.

## The Negative Half-cycle



As the current flowing through the load is unidirectional, so the voltage developed across the load is also unidirectional the same as for the previous two diode full-wave rectifier, therefore the average DC voltage across the load is  $0.637V_{\max}$  and the ripple frequency is now twice the supply frequency (e.g. 100Hz for a 50Hz supply).



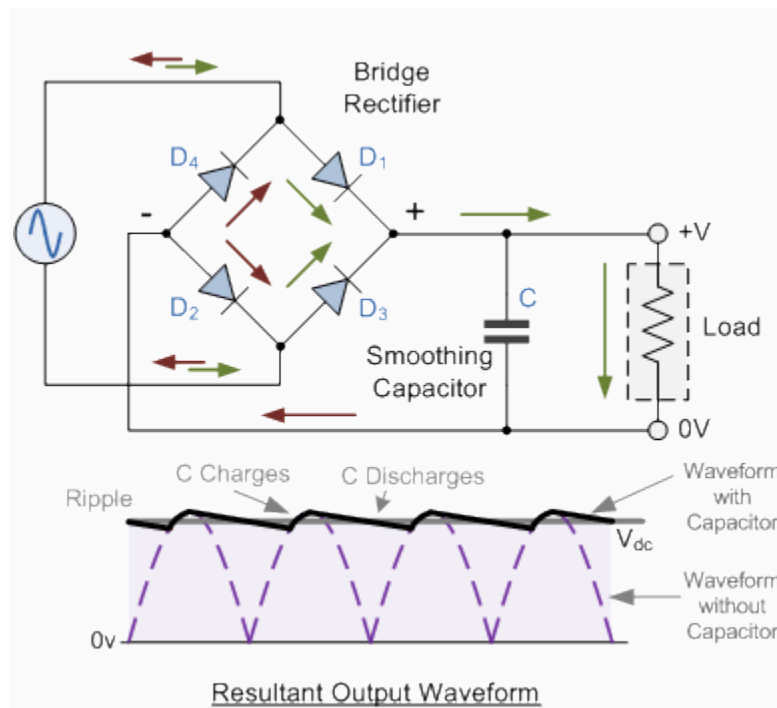
## Typical Bridge Rectifier

Although we can use four individual power diodes to make a full wave bridge rectifier, pre-made bridge rectifier components are available "off-the-shelf" in a range of different voltage and current sizes that can be soldered directly into a PCB circuit board or be connected by spade connectors. The image to the right shows a typical single phase bridge rectifier with one corner cut off. This cut-off corner indicates that the terminal nearest to the corner is the positive or +ve output terminal or lead with the opposite (diagonal) lead being the negative or -ve output lead. The other two connecting leads are for the input alternating voltage from a transformer secondary winding.

## The Smoothing Capacitor

We saw in the previous section that the single phase half-wave rectifier produces an output wave every half cycle and that it was not practical to use this type of circuit to produce a steady DC supply. The full-wave bridge rectifier however, gives us a greater mean DC value ( $0.637 V_{max}$ ) with less superimposed ripple while the output waveform is twice that of the frequency of the input supply frequency. We can therefore increase its average DC output level even higher by connecting a suitable smoothing capacitor across the output of the bridge circuit as shown below.

### Full-wave Rectifier with Smoothing Capacitor



The smoothing capacitor converts the full-wave rippled output of the rectifier into a smooth DC output voltage. Generally for DC power supply circuits the smoothing capacitor is an Aluminium Electrolytic type that has a capacitance value of 100 $\mu$ F or more with repeated DC voltage pulses from the rectifier charging up the capacitor to peak voltage. However, there are two important parameters to consider when choosing a suitable smoothing capacitor and these are its *Working Voltage*, which must be higher than the no-load output value of the rectifier and its *Capacitance Value*, which determines the amount of ripple that will appear superimposed on top of the DC voltage. Too low a value and the capacitor has little effect but if the smoothing capacitor is large enough (parallel capacitors can be used) and the load current is not too large, the output voltage will be almost as smooth as pure DC. As a general rule of thumb, we are looking to have a ripple voltage of less than 100mV peak to peak.

The maximum ripple voltage present for a **Full Wave Rectifier** circuit is not only determined by the value of the smoothing capacitor but by the frequency and load current, and is calculated as:

### Bridge Rectifier Ripple Voltage

$$V_{\text{RIPPLE}} = \frac{I_{\text{LOAD}}}{fC} \text{ volts}$$

Where:  $I$  is the DC load current in amps,  $f$  is the frequency of the ripple or twice the input frequency in Hertz, and  $C$  is the capacitance in Farads.

The main advantages of a full-wave bridge rectifier is that it has a smaller AC ripple value for a given load and a smaller reservoir or smoothing capacitor than an equivalent half-wave rectifier. Therefore, the fundamental frequency of the ripple voltage is twice that of the AC supply frequency (100Hz) where for the half-wave rectifier it is exactly equal to the supply frequency (50Hz).

The amount of ripple voltage that is superimposed on top of the DC supply voltage by the diodes can be virtually eliminated by adding a much improved  $\pi$ -filter (pi-filter) to the output terminals of the bridge rectifier. This type of low-pass filter consists of two smoothing capacitors, usually of the same value and a choke or inductance across them to introduce a high impedance path to the alternating ripple component. Another more practical and cheaper alternative is to use a 3-terminal voltage regulator IC, such as a LM78xx for a positive output voltage or the LM79xx for a negative output voltage which can reduce the ripple by more than 70dB (Datasheet) while delivering a constant output current of over 1 amp.

In the next tutorial about diodes, we will look at the **Zener Diode** which takes advantage of its reverse breakdown voltage characteristic to produce a constant and fixed output voltage across itself.

## Zener Diodes

### The Zener Diode

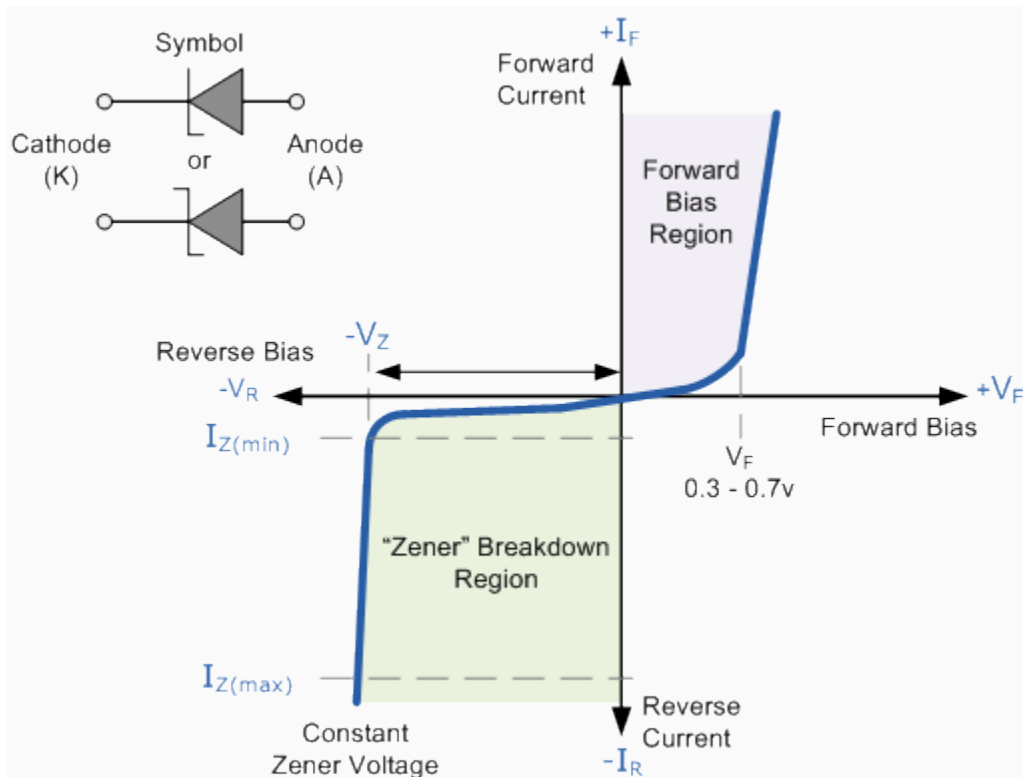
In the previous **Signal Diode** tutorial, we saw that a "reverse biased" diode blocks current in the reverse direction, but will suffer from premature breakdown or damage if the reverse voltage applied across it is too high. However, the **Zener Diode** or "Breakdown Diode" as they are sometimes called, are basically the same as the standard PN junction diode but are specially designed to have a low pre-determined **Reverse Breakdown Voltage** that takes advantage of

this high reverse voltage. The point at which a zener diode breaks down or conducts is called the "Zener Voltage" ( $V_Z$ ).

The **Zener diode** is like a general-purpose signal diode consisting of a silicon PN junction. When biased in the forward direction it behaves just like a normal signal diode passing the rated current, but when a reverse voltage is applied to it the reverse saturation current remains fairly constant over a wide range of voltages. The reverse voltage increases until the diodes breakdown voltage  $V_B$  is reached at which point a process called *Avalanche Breakdown* occurs in the depletion layer and the current flowing through the zener diode increases dramatically to the maximum circuit value (which is usually limited by a series resistor). This breakdown voltage point is called the "zener voltage" for zener diodes.

The point at which current flows can be very accurately controlled (to less than 1% tolerance) in the doping stage of the diodes construction giving the diode a specific *zener breakdown voltage*, ( $V_Z$ ) ranging from a few volts up to a few hundred volts. This zener breakdown voltage on the I-V curve is almost a vertical straight line.

### Zener Diode I-V Characteristics



The **Zener Diode** is used in its "reverse bias" or reverse breakdown mode, i.e. the diodes anode

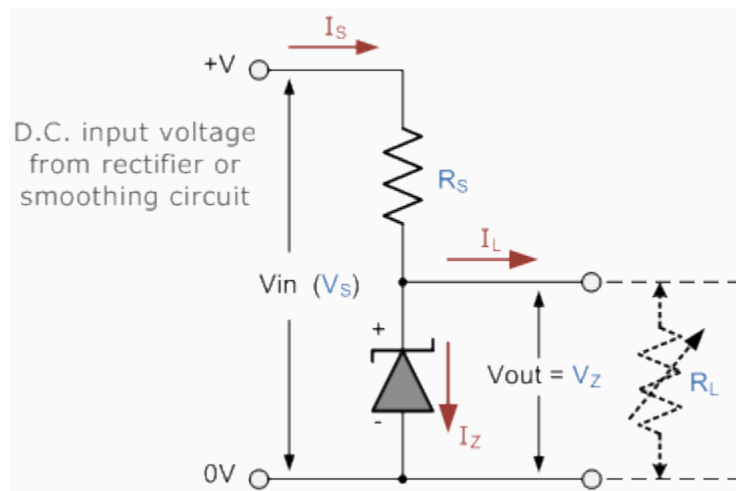
connects to the negative supply. From the I-V characteristics curve above, we can see that the zener diode has a region in its reverse bias characteristics of almost a constant negative voltage regardless of the value of the current flowing through the diode and remains nearly constant even with large changes in current as long as the zener diodes current remains between the breakdown current  $I_{Z(\min)}$  and the maximum current rating  $I_{Z(\max)}$ .

This ability to control itself can be used to great effect to regulate or stabilise a voltage source against supply or load variations. The fact that the voltage across the diode in the breakdown region is almost constant turns out to be an important application of the zener diode as a voltage regulator. The function of a regulator is to provide a constant output voltage to a load connected in parallel with it in spite of the ripples in the supply voltage or the variation in the load current and the zener diode will continue to regulate the voltage until the diodes current falls below the minimum  $I_{Z(\min)}$  value in the reverse breakdown region.

### The Zener Diode Regulator

**Zener Diodes** can be used to produce a stabilised voltage output with low ripple under varying load current conditions. By passing a small current through the diode from a voltage source, via a suitable current limiting resistor ( $R_S$ ), the zener diode will conduct sufficient current to maintain a voltage drop of  $V_{out}$ . We remember from the previous tutorials that the DC output voltage from the half or full-wave rectifiers contains ripple superimposed onto the DC voltage and that as the load value changes so to does the average output voltage. By connecting a simple zener stabiliser circuit as shown below across the output of the rectifier, a more stable output voltage can be produced.

### Zener Diode Regulator



The resistor,  $R_S$  is connected in series with the zener diode to limit the current flow through the diode with the voltage source,  $V_S$  being connected across the combination. The stabilised output voltage  $V_{out}$  is taken from across the zener diode. The zener diode is connected with its cathode terminal connected to the positive rail of the DC supply so it is reverse biased and will be operating in its breakdown condition. Resistor  $R_S$  is selected so to limit the maximum current flowing in the circuit.

With no load connected to the circuit, the load current will be zero, ( $I_L = 0$ ), and all the circuit current passes through the zener diode which in turn dissipates its maximum power. Also a small value of the series resistor  $R_S$  will result in a greater diode current when the load resistance  $R_L$  is connected and large as this will increase the power dissipation requirement of the diode so care must be taken when selecting the appropriate value of series resistance so that the zener's maximum power rating is not exceeded under this no-load or high-impedance condition.

The load is connected in parallel with the zener diode, so the voltage across  $R_L$  is always the same as the zener voltage, ( $V_R = V_Z$ ). There is a minimum zener current for which the stabilization of the voltage is effective and the zener current must stay above this value operating under load within its breakdown region at all times. The upper limit of current is of course dependant upon the power rating of the device. The supply voltage  $V_S$  must be greater than  $V_Z$ .

One small problem with zener diode stabiliser circuits is that the diode can sometimes generate electrical noise on top of the DC supply as it tries to stabilise the voltage. Normally this is not a problem for most applications but the addition of a large value decoupling capacitor across the zener's output may be required to give additional smoothing.

Then to summarise a little. A zener diode is always operated in its reverse biased condition. A voltage regulator circuit can be designed using a zener diode to maintain a constant DC output voltage across the load in spite of variations in the input voltage or changes in the load current. The zener voltage regulator consists of a current limiting resistor  $R_S$  connected in series with the input voltage  $V_S$  with the zener diode connected in parallel with the load  $R_L$  in this reverse biased condition. The stabilized output voltage is always selected to be the same as the breakdown voltage  $V_Z$  of the diode.

### **Example No1**

A 5.0V stabilised power supply is required to be produced from a 12V DC power supply input source. The maximum power rating  $P_Z$  of the zener diode is 2W. Using the zener regulator circuit above calculate:

a) The maximum current flowing through the zener diode.

$$\text{Maximum Current} = \frac{\text{Watts}}{\text{Voltage}} = \frac{2\text{w}}{5\text{v}} = 400\text{mA}$$

b) The value of the series resistor,  $R_S$

$$R_S = \frac{V_S - V_Z}{I_Z} = \frac{12 - 5}{400\text{mA}} = 17.5\Omega$$

c) The load current  $I_L$  if a load resistor of  $1\text{k}\Omega$  is connected across the Zener diode.

$$I_L = \frac{V_Z}{R_L} = \frac{5\text{v}}{1000\Omega} = 5\text{mA}$$

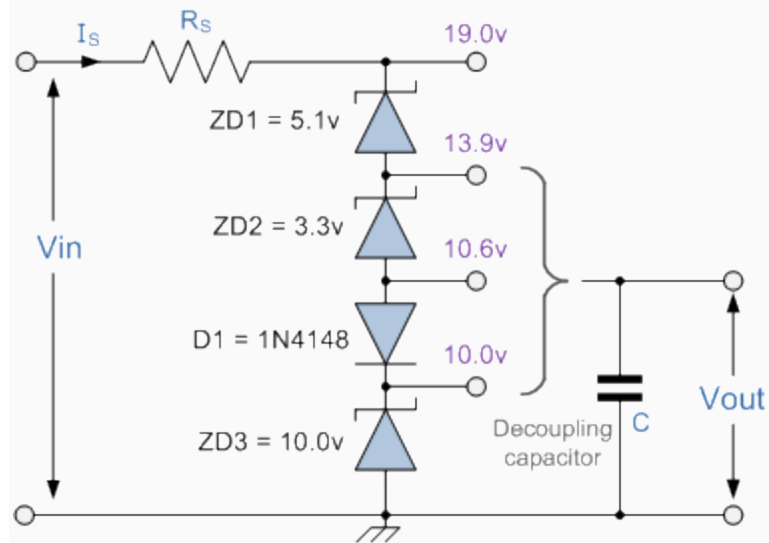
d) The total supply current  $I_S$

$$I_S = I_Z + I_L = 400\text{mA} + 5\text{mA} = 405\text{mA}$$

### **Zener Diode Voltages**

As well as producing a single stabilised voltage output, zener diodes can also be connected together in series along with normal silicon signal diodes to produce a variety of different reference voltage output values as shown below.

### **Zener Diodes Connected in Series**



The values of the individual Zener diodes can be chosen to suit the application while the silicon diode will always drop about 0.6 - 0.7V in the forward bias condition. The supply voltage,  $V_{in}$  must of course be higher than the largest output reference voltage and in our example above this is 19v.

A typical **zener diode** for general electronic circuits is the 500mW, *BZX55* series or the larger 1.3W, *BZX85* series were the zener voltage is given as, for example, *C7V5* for a 7.5V diode giving a diode reference number of *BZX55C7V5*. The 500mW series of zener diodes are available from about 2.4 up to about 100 volts and typically have the same sequence of values as used for the 5% (E24) resistor series with the individual voltage ratings for these small but very useful diodes are given in the table below.

### Zener Diode Standard Voltages

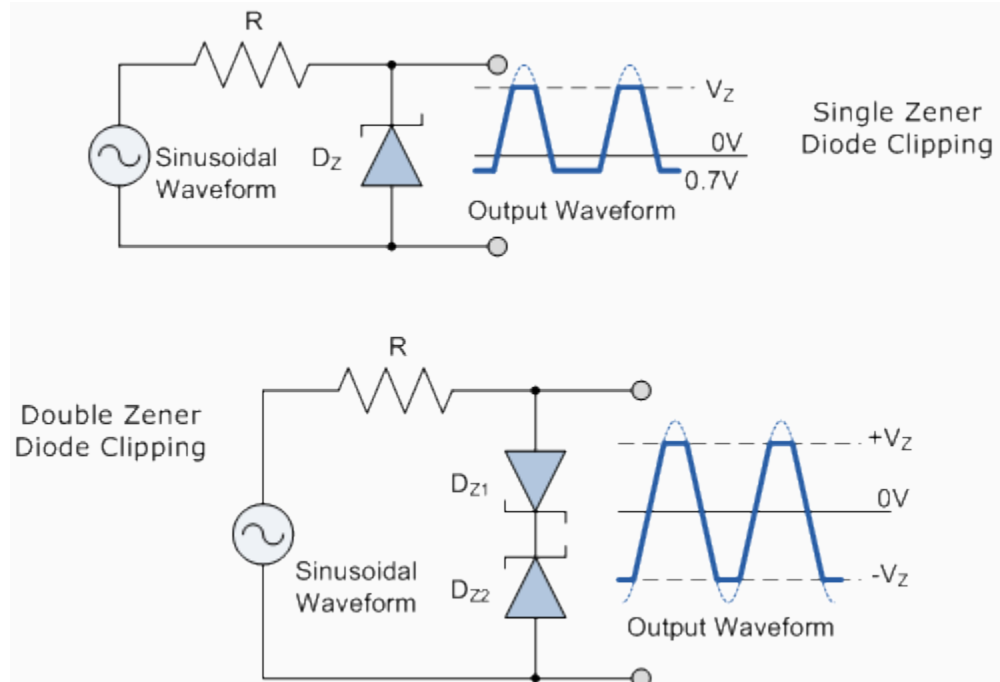
BZX55 Zener Diode Power Rating 500mW							
2.4V	2.7V	3.0V	3.3V	3.6V	3.9V	4.3V	4.7V
5.1V	5.6V	6.2V	6.8V	7.5V	8.2V	9.1V	10V
11V	12V	13V	15V	16V	18V	20V	22V
24V	27V	30V	33V	36V	39V	43V	47V
BZX85 Zener Diode Power Rating 1.3W							
3.3V	3.6V	3.9V	4.3V	4.7V	5.1V	5.6	6.2V
6.8V	7.5V	8.2V	9.1V	10V	11V	12V	13V
15V	16V	18V	20V	22V	24V	27V	30V
33V	36V	39V	43V	47V	51V	56V	62V

## Zener Diode Clipping Circuits

Thus far we have looked at how a zener diode can be used to regulate a constant DC source but what if the input signal was not steady state DC but an alternating AC waveform how would the zener diode react to a constantly changing signal.

Diode clipping and clamping circuits are circuits that are used to shape or modify an input AC waveform (or any sinusoid) producing a differently shape output waveform depending on the circuit arrangement. Diode clipper circuits are also called limiters because they limit or clip-off the positive (or negative) part of an input AC signal. As zener clipper circuits limit or cut-off part of the waveform across them, they are mainly used for circuit protection or in waveform shaping circuits. For example, if we wanted to clip an output waveform at  $+7.5\text{V}$ , we would use a  $7.5\text{V}$  zener diode. If the output waveform tries to exceed the  $7.5\text{V}$  limit, the zener diode will "clip-off" the excess voltage from the input producing a waveform with a flat top still keeping the output constant at  $+7.5\text{V}$ . Note that in the forward bias condition a zener diode is still a diode and when the AC waveform output goes negative below  $-0.7\text{V}$ , the zener diode turns "ON" like any normal silicon diode would and clips the output at  $-0.7\text{V}$  as shown below.

### Square Wave Signal



The back to back connected zener diodes can be used as an AC regulator producing what is jokingly called a "poor man's square wave generator". Using this arrangement we can clip the waveform between a positive value of  $+7.5\text{V}$  and a negative value of  $-7.5\text{V}$ . If we wanted to clip

an output waveform between different minimum and maximum values for example, +8V and -6V, use would simply use two differently rated zener diodes.

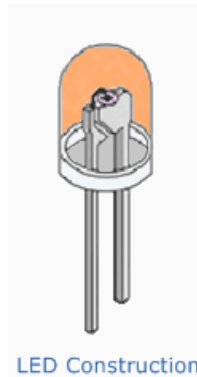
Note that the output will actually clip the AC waveform between +8.7V and -6.7V due to the addition of the forward biasing diode voltage, which adds another 0.7V voltage drop to it. This type of clipper configuration is fairly common for protecting an electronic circuit from over voltage. The two zeners are generally placed across the power supply input terminals and during normal operation, one of the zener diodes is "OFF" and the diodes have little or no affect. However, if the input voltage waveform exceeds its limit, then the zeners turn "ON" and clip the input to protect the circuit.

In the next tutorial about diodes, we will look at using the forward biased PN junction of a diode to produce light. We know from the previous tutorials that when charge carriers move across the junction, electrons combine with holes and energy is lost in the form of heat, but also some of this energy is dissipated as photons but we can not see them. If we place a translucent lens around the junction, visible light will be produced and the diode becomes a light source. This effect produces another type of diode known commonly as the **Light Emitting Diode** which takes advantage of this light producing characteristic to emit light (photons) in a variety of colours and wavelengths.

---

## Light Emitting Diodes

**Light Emitting Diodes** or **LED's**, are among the most widely used of all the different types of semiconductor diodes available today. They are the most visible type of diode, that emit a fairly narrow bandwidth of either visible light at different coloured wavelengths, invisible infra-red light for remote controls or laser type light when a forward current is passed through them. A "**Light Emitting Diode**" or **LED** as it is more commonly called, is basically just a specialised type of PN junction diode, made from a very thin layer of fairly heavily doped semiconductor material. When the diode is forward biased, electrons from the semiconductors conduction band recombine with holes from the valence band releasing sufficient energy to produce photons which emit a monochromatic (single colour) of light. Because of this thin layer a reasonable number of these photons can leave the junction and radiate away producing a coloured light output. Then we can say that when operated in a forward biased direction **Light Emitting Diodes** are semiconductor devices that convert electrical energy into light energy.



The construction of a light emitting diode is very different from that of a normal signal diode. The PN junction of an LED is surrounded by a transparent, hard plastic epoxy resin hemispherical shaped shell or body which protects the LED from both vibration and shock. Surprisingly, an LED junction does not actually emit that much light so the epoxy resin body is constructed in such a way that the photons of light emitted by the junction are reflected away from the surrounding substrate base to which the diode is attached and are focused upwards through the domed top of the LED, which itself acts like a lens concentrating the amount of light. This is why the emitted light appears to be brightest at the top of the LED.

However, not all LEDs are made with a hemispherical shaped dome for their epoxy shell. Some LEDs have a rectangular or cylindrical shaped construction that has a flat surface on top. Also, nearly all LEDs have their cathode, (K) terminal identified by either a notch or flat spot on the body, or by one of the leads being shorter than the other, (the Anode, A).

Unlike normal incandescent lamps and bulbs which generate large amounts of heat when illuminated, the light emitting diode produces a "cold" generation of light which leads to high efficiencies than the normal "light bulb" because most of the generated energy radiates away within the visible spectrum. Because LEDs are solid-state devices, they can be extremely small and durable and provide much longer lamp life than normal light sources.

### **Light Emitting Diode Colours**

So how does a light emitting diode get its colour. Unlike normal signal diodes which are made for detection or power rectification, and which are made from either Germanium or Silicon semiconductor materials, **Light Emitting Diodes** are made from exotic semiconductor compounds such as Gallium Arsenide (GaAs), Gallium Phosphide (GaP), Gallium Arsenide Phosphide (GaAsP), Silicon Carbide (SiC) or Gallium Indium Nitride (GaInN) all mixed together at different ratios to produce a distinct wavelength of colour. Different LED compounds emit light in specific regions of the visible light spectrum and therefore produce different

intensity levels. The exact choice of the semiconductor material used will determine the overall wavelength of the photon light emissions and therefore the resulting colour of the light emitted.

Typical LED Characteristics			
Semiconductor Material	Wavelength	Colour	V <sub>F</sub> @ 20mA
GaAs	850-940nm	Infra-Red	1.2v
GaAsP	630-660nm	Red	1.8v
GaAsP	605-620nm	Amber	2.0v
GaAsP:N	585-595nm	Yellow	2.2v
AlGaP	550-570nm	Green	3.5v
SiC	430-505nm	Blue	3.6v
GaInN	450nm	White	4.0v

Thus, the actual colour of a light emitting diode is determined by the wavelength of the light emitted, which in turn is determined by the actual semiconductor compound used in forming the PN junction during manufacture and NOT by the colouring of the LEDs plastic body although these are slightly coloured to both enhance the light and indicate its colour when its not be used. Light emitting diodes are

available in a wide range of colours with the most common being RED, AMBER, YELLOW and GREEN and are thus widely used as visual indicators and as moving light displays.

Recently developed blue and white coloured LEDs are also available but these tend to be much more expensive than the normal standard colours due to the production costs of mixing together two or more complementary colours at an exact ratio within the semiconductor compound and also by injecting nitrogen atoms into the crystal structure during the doping process.

From the table above we can see that the main P-type dopant used in the manufacture of **Light Emitting Diodes** is Gallium (Ga, atomic number 31) and that the main N-type dopant used is Arsenic (As, atomic number 31) giving the resulting compound of Gallium Arsenide (GaAs) crystal structure. The problem with using Gallium Arsenide on its own as the semiconductor compound is that it radiates large amounts of low brightness infra-red radiation (850nm-940nm approx.) from its junction when a forward current is flowing through it. This infra-red light is ok for television remote controls but not very useful if we want to use the LED as an indicating light. But by adding Phosphorus (P, atomic number 15), as a third dopant the overall wavelength of the emitted radiation is reduced to below 680nm giving visible red light to the human eye. Further refinements in the doping process of the PN junction have resulted in a range of colours spanning the spectrum of visible light as we have seen above as well as infra-red and ultra-violet wavelengths.

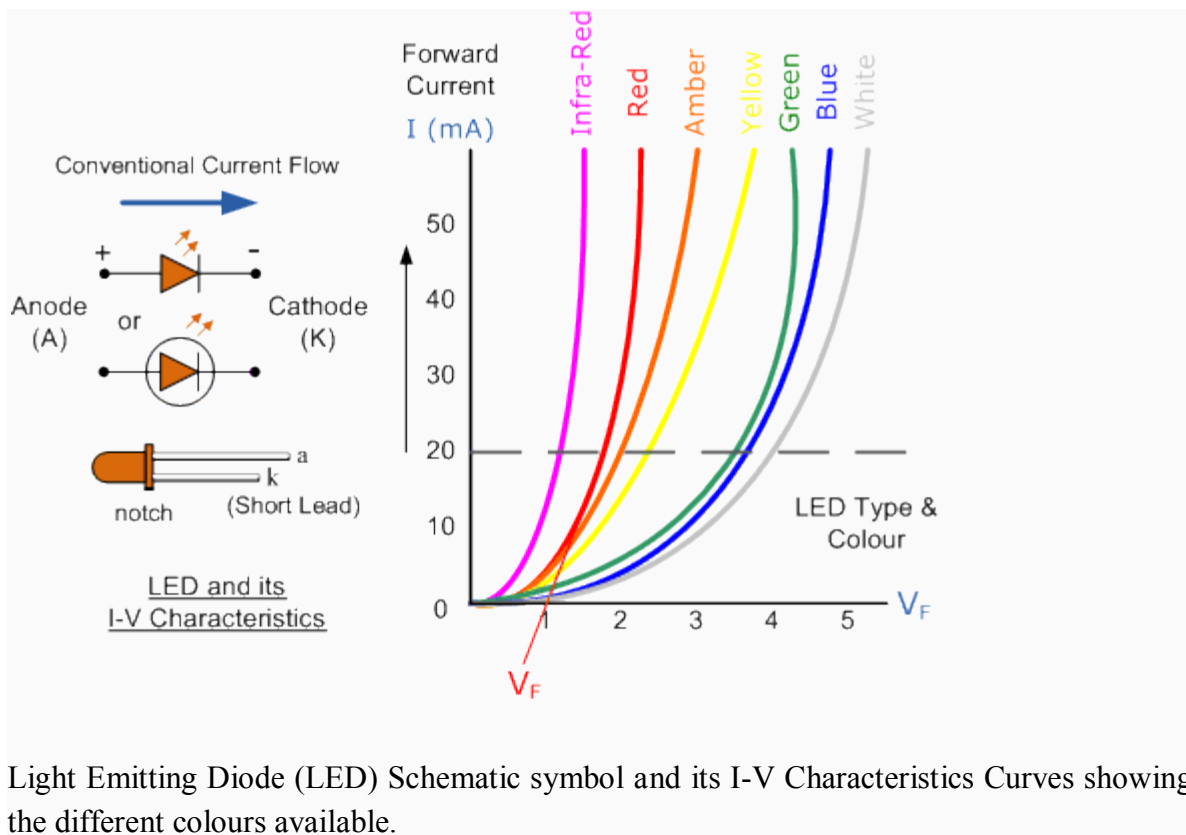
By mixing together a variety of semiconductor, metal and gas compounds the following list of LEDs can be produced.

- Gallium Arsenide (GaAs) - infra-red

- Gallium Arsenide Phosphide (GaAsP) - red to infra-red, orange
- Aluminium Gallium Arsenide Phosphide (AlGaAsP) - high-brightness red, orange-red, orange, and yellow
- Gallium Phosphide (GaP) - red, yellow and green
- Aluminium Gallium Phosphide (AlGaP) - green
- Gallium Nitride (GaN) - green, emerald green
- Gallium Indium Nitride (GaInN) - near ultraviolet, bluish-green and blue
- Silicon Carbide (SiC) - blue as a substrate
- Zinc Selenide (ZnSe) - blue
- Aluminium Gallium Nitride (AlGaN) - ultraviolet

Like conventional PN junction diodes, LEDs are current-dependent devices with its forward voltage drop  $V_F$ , depending on the semiconductor compound (its light colour) and on the forward biased LED current. The point where conduction begins and light is produced is about 1.2V for a standard red LED to about 3.6V for a blue LED. The exact voltage drop will of course depend on the manufacturer because of the different dopant materials and wavelengths used. The voltage drop across the LED at a particular current value, for example 20mA, will also depend on the initial conduction  $V_F$  point. As an LED is effectively a diode, its forward current to voltage characteristics curves can be plotted for each diode colour as shown below.

### **Light Emitting Diodes I-V Characteristics.**



Light Emitting Diode (LED) Schematic symbol and its I-V Characteristics Curves showing the different colours available.

Before a light emitting diode can "emit" any form of light it needs a current to flow through it, as it is a current dependant device with their light output intensity being directly proportional to the forward current flowing through the LED. As the LED is to be connected in a forward bias condition across a power supply it should be *current limited* using a series resistor to protect it from excessive current flow. Never connect an LED directly to a battery or power supply as it will be destroyed almost instantly because too much current will pass through and burn it out.

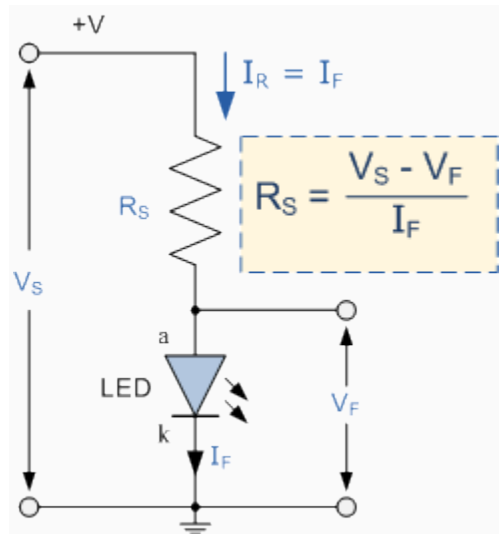
From the table above we can see that each LED has its own forward voltage drop across the PN junction and this parameter which is determined by the semiconductor material used, is the forward voltage drop for a specified amount of forward conduction current, typically for a forward current of 20mA. In most cases LEDs are operated from a low voltage DC supply, with a series resistor,  $R_S$  used to limit the forward current to a safe value from say 5mA for a simple LED indicator to 30mA or more where a high brightness light output is needed.

### LED Series Resistance.

The series resistor value  $R_S$  is calculated by simply using **Ohm's Law**, by knowing the required forward current  $I_F$  of the LED, the supply voltage  $V_S$  across the combination and the expected

forward voltage drop of the LED,  $V_F$  at the required current level, the current limiting resistor is calculated as:

### LED Series Resistor Circuit



### Example No1

An amber coloured LED with a forward volt drop of 2 volts is to be connected to a 5.0v stabilised DC power supply. Using the circuit above calculate the value of the series resistor required to limit the forward current to less than 10mA. Also calculate the current flowing through the diode if a 100Ω series resistor is used instead of the calculated first.

1). series resistor value at 10mA.

$$R_S = \frac{V_S - V_F}{I_F} = \frac{5\text{v} - 2\text{v}}{10\text{mA}} = \frac{3}{10 \times 10^{-3}} = 300\Omega$$

2). with a 100Ω series resistor.

$$R_S = \frac{V_S - V_F}{I_F}$$

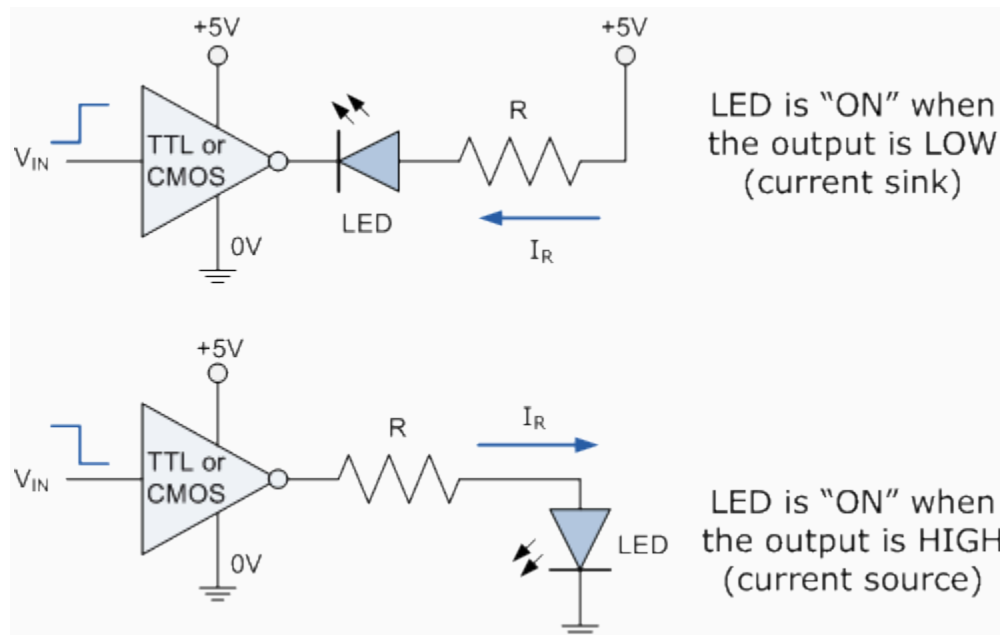
$$\therefore I_F = \frac{V_S - V_F}{R_S} = \frac{5 - 2}{100} = 30\text{mA}$$

We remember from the **Resistors** tutorials, that resistors come in standard preferred values. Our first calculation above shows to limit the current flowing through the LED to 10mA exactly, we would require a 300Ω resistor. In the E12 series of resistors there is no 300Ω resistor so we would need to choose the next highest value, which is 330Ω. A quick re-calculation shows the new forward current value is now 9.1mA, and this is ok.

## LED Driver Circuits

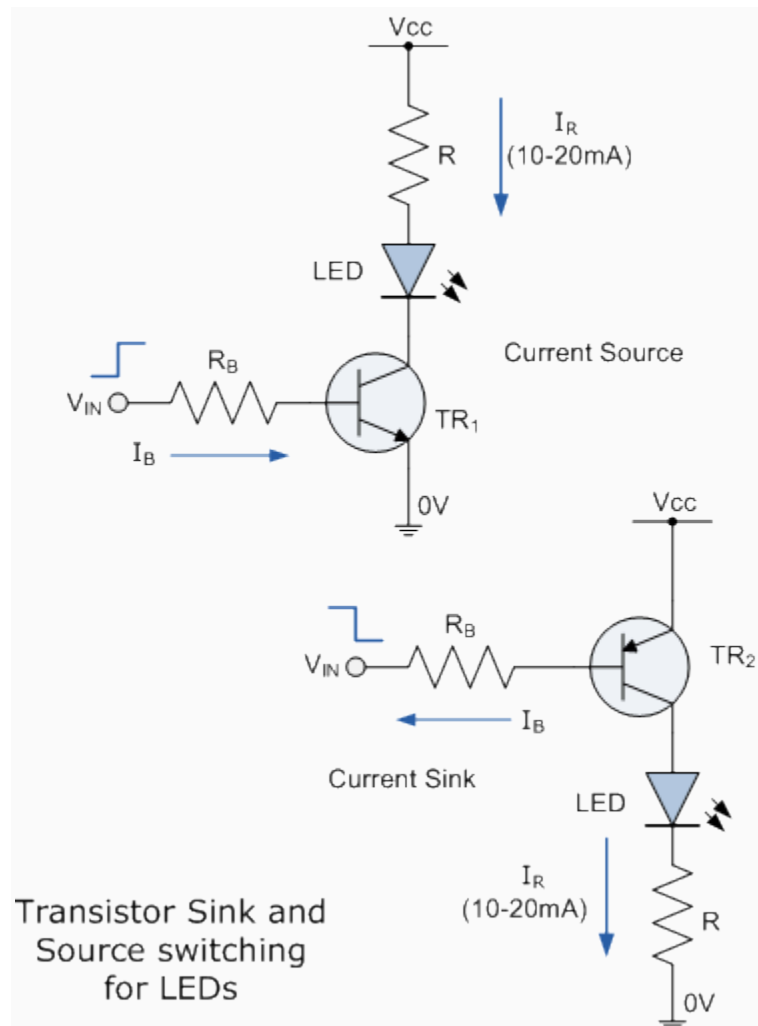
Now that we know what is an LED, we need some way of controlling it by switching it "ON" and "OFF". The output stages of both TTL and CMOS logic gates can both source and sink useful amounts of current therefore can be used to drive an LED. Normal integrated circuits (ICs) have an output drive current of up to 50mA in the sink mode configuration, but have an internally limited output current of about 30mA in the source mode configuration. Either way the LED current must be limited to a safe value using a series resistor as we have already seen. Below are some examples of driving light emitting diodes using inverting ICs but the idea is the same for any type of integrated circuit output whether combinational or sequential.

### IC Driver Circuit



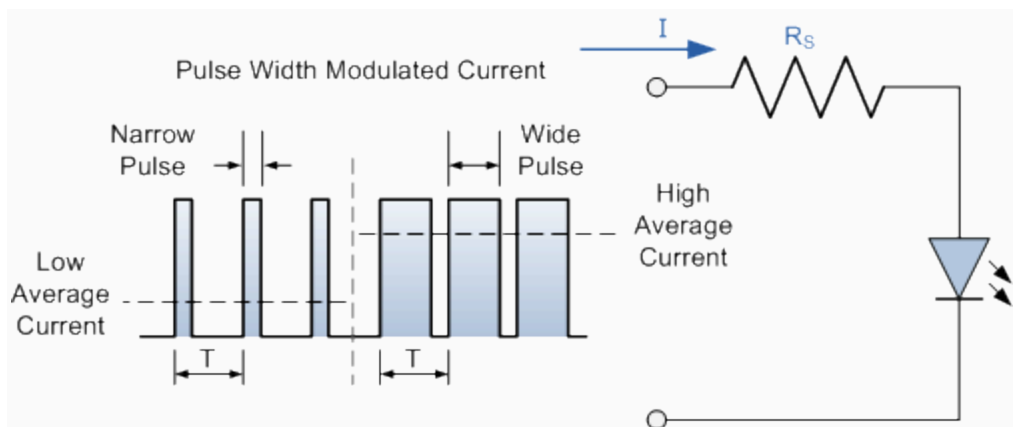
If more than one LED requires driving at the same time, such as in large LED arrays, or the load current is too high for the integrated circuit or we may just want to use discrete components instead of ICs, then an alternative way of driving the LEDs using either bipolar NPN or PNP **transistors as switches** is given below. Again as before, a series resistor,  $R_S$  is required to limit the LED current.

## Transistor Driver Circuit



The brightness of a light emitting diode cannot be controlled by simply varying the current flowing through it. Allowing more current to flow through the LED will make it glow brighter but will also cause it to dissipate more heat. LEDs are designed to produce a set amount of light operating at a specific forward current ranging from about 10 to 20mA. In situations where power savings are important, less current may be possible. However, reducing the current to below say 5mA may dim its light output too much or even turn the LED "OFF" completely. A much better way to control the brightness of LEDs is to use a control process known as "Pulse Width Modulation" or PWM, in which the LED is repeatedly turned "ON" and "OFF" at varying frequencies depending upon the required light intensity.

## LED Light Intensity using PWM



When higher light outputs are required, a pulse width modulated current with a fairly short duty cycle ("ON-OFF" Ratio) allows the diode current and therefore the output light intensity to be increased significantly during the actual pulses, while still keeping the LEDs "average current level" and power dissipation within safe limits. This "ON-OFF" flashing condition does not affect what is seen as the human eyes fills in the gaps between the "ON" and "OFF" light pulses, providing the pulse frequency is high enough, making it appear as a continuous light output. So pulses at a frequency of 100Hz or more actually appear brighter to the eye than a continuous light of the same average intensity.

### Multi-coloured Light Emitting Diode

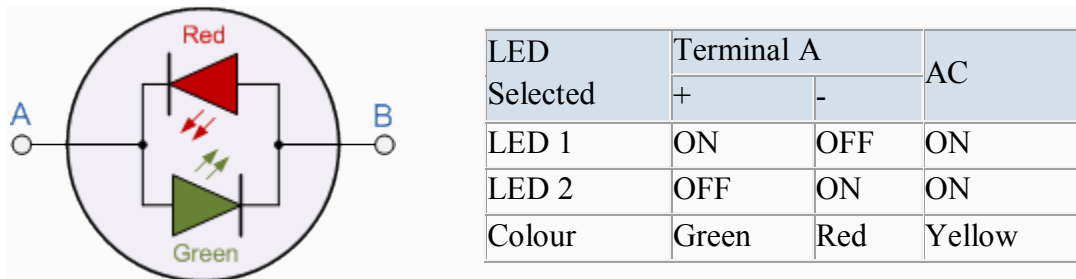
LEDs are available in a wide range of shapes, colours and various sizes with different light output intensities available, with the most common (and cheapest to produce) being the standard 5mm Red Gallium Arsenide Phosphide (GaAsP) LED. LED's are also available in various "packages" arranged to produce both letters and numbers with the most common being that of the "seven segment display" arrangement. Nowadays, full colour flat screen LED displays are available with a large number of dedicated ICs available for driving the displays directly. Most light emitting diodes produce just a single output of coloured light however, multi-coloured LEDs are now available that can produce a range of different colours from within a single device. Most of these are actually two or three LEDs fabricated within a single package.

### Bicolour Light Emitting Diodes

A bicolour light emitting diode has two LEDs chips connected together in "inverse parallel" (one forwards, one backwards) combined in one single package. Bicolour LEDs can produce any one of three colours for example, a red colour is emitted when the device is connected with current flowing in one direction and a green colour is emitted when it is biased in the other direction. This type of bi-directional arrangement is useful for giving polarity indication, for example, the

correct connection of batteries or power supplies etc. Also, a bi-directional current produces both colours mixed together as the two LEDs would take it in turn to illuminate if the device was connected (via a suitable resistor) to a low voltage, low frequency AC supply.

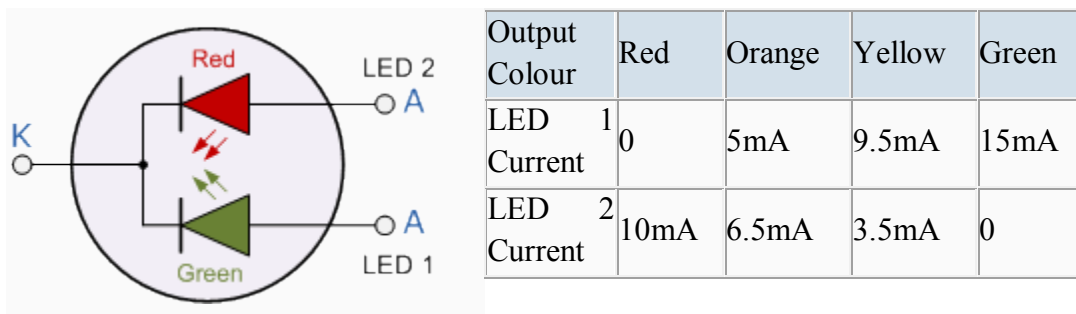
### A Bicolour LED



### Tricolour Light Emitting Diodes

The most popular type of tricolour LED comprises of a single Red and a Green LED combined in one package with their cathode terminals connected together producing a three terminal device. They are called tricolour LEDs because they can give out a single red or a green colour by turning "ON" only one LED at a time. They can also generate additional shades of colours (the third colour) such as Orange or Yellow by turning "ON" the two LEDs in different ratios of forward current as shown in the table thereby generating 4 different colours from just two diode junctions.

### A Multi or Tricolour LED



### LED Displays

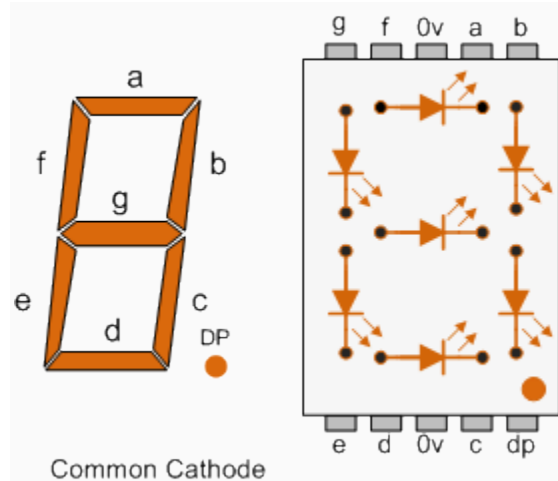
As well as individual colour or multi-colour LEDs, several light emitting diodes can be combined together within a single package to produce displays such as bargraphs, strips, arrays

and seven segment displays. A seven segment LED display provides a very convenient way when decoded properly of displaying information or digital data in the form of numbers, letters or even alpha-numerical characters and as their name suggests, they consist of seven individual LEDs (the segments), within one single display package.

In order to produce the required numbers or characters from 0 to 9 and A to F respectively, on the display the correct combination of LED segments need to be illuminated. A standard seven segment LED display generally has eight input connections, one for each LED segment and one that acts as a common terminal or connection for all the internal segments.

- The Common Cathode Display (CCD) - In the common cathode display, all the cathode connections of the LEDs are joined together and the individual segments are illuminated by application of a HIGH, logic "1" signal.
- 
- The Common Anode Display (CAD) - In the common anode display, all the anode connections of the LEDs are joined together and the individual segments are illuminated by connecting the terminals to a LOW, logic "0" signal.

### A Typical Seven Segment LED Display

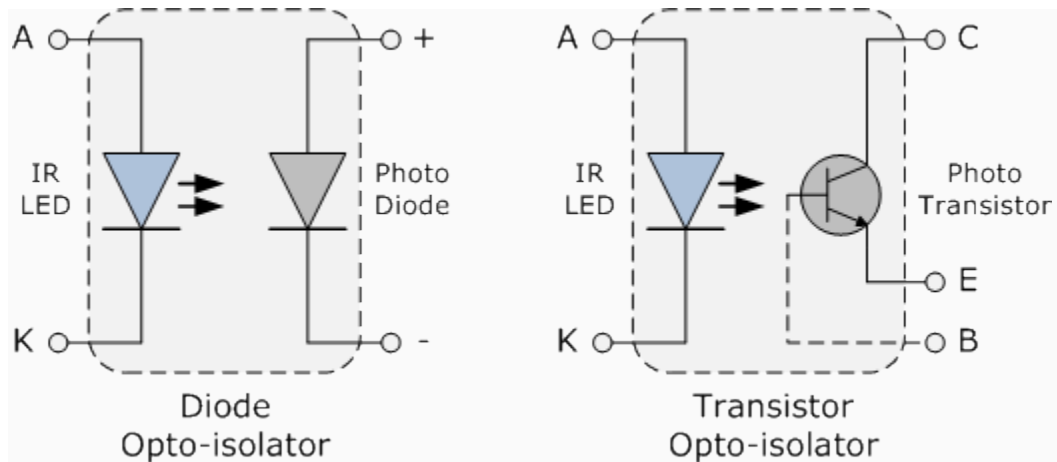


### Opto-coupler

Finally, another useful application of light emitting diodes is **Opto-coupling**. An opto-coupler or opto-isolator as it is also called, is a single electronic device that consists of a light emitting diode combined with either a photo-diode, photo-transistor or photo-triac to provide an optical signal path between an input connection and an output connection while maintaining electrical isolation between two circuits. An opto-isolator consists of a light proof plastic body that has a

typical breakdown voltages between the input (photo-diode) and the output (photo-transistor) circuit of up to 5000 volts. This electrical isolation is especially useful where the signal from a low voltage circuit such as a battery powered circuit, computer or microcontroller, is required to operate or control another external circuit operating at a potentially dangerous mains voltage.

### Photo-diode and Photo-transistor Opto-couplers



The two components used in an opto-isolator, an optical transmitter such as an infra-red emitting Gallium Arsenide LED and an optical receiver such as a photo-transistor are closely optically coupled and use light to send signals and/or information between its input and output. This allows information to be transferred between circuits without an electrical connection or common ground potential. Opto-isolators are digital or switching devices, so they transfer either "ON-OFF" control signals or digital data. Analogue signals can be transferred by means of frequency or pulse-width modulation.

### Diode

From Wikipedia, the free encyclopedia  
Jump to: navigation, search



Figure 1: Closeup of a diode, showing the square shaped semiconductor crystal (*black object on left*).

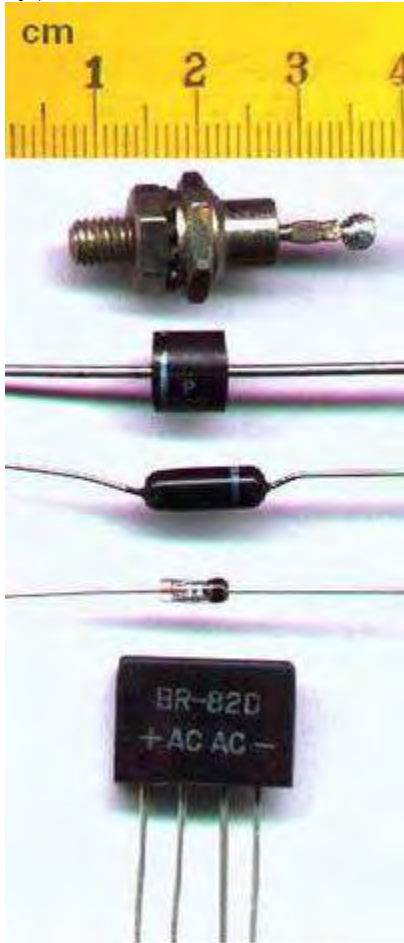


Figure 2: Various semiconductor diodes. Bottom: A bridge rectifier. In most diodes, a white or black painted band identifies the cathode terminal, that is, the terminal which conventional current flows out of when the diode is conducting.

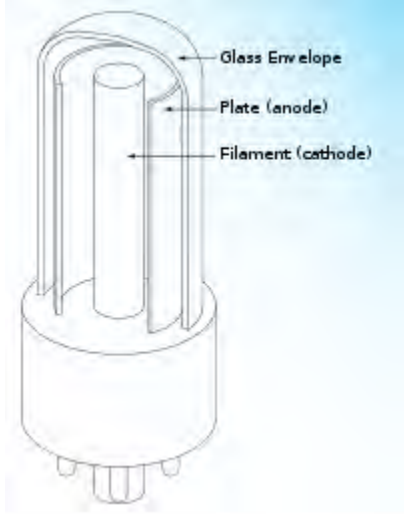


Figure 3: Structure of a vacuum tube diode. The filament may be bare, or more commonly (as shown here), embedded within and insulated from an enclosing cathode

In electronics, a **diode** is a two-terminal electronic component that conducts electric current in only one direction. The term usually refers to a **semiconductor diode**, the most common type today. This is a crystalline piece of semiconductor material connected to two electrical terminals.<sup>[1]</sup> A **vacuum tube diode** (now little used except in some high-power technologies) is a vacuum tube with two electrodes; a plate and a cathode.

The most common function of a diode is to allow an electric current to pass in one direction (called the diode's *forward* direction) while blocking current in the opposite direction (the *reverse* direction). Thus, the diode can be thought of as an electronic version of a check valve. This unidirectional behavior is called rectification, and is used to convert alternating current to direct current, and to extract modulation from radio signals in radio receivers.

However, diodes can have more complicated behavior than this simple on-off action, due to their complex non-linear electrical characteristics, which can be tailored by varying the construction of their P-N junction. These are exploited in special purpose diodes that perform many different functions. For example, specialized diodes are used to regulate voltage (Zener diodes), to electronically tune radio and TV receivers (varactor diodes), to generate radio frequency oscillations (tunnel diodes), and to produce light (light emitting diodes).

Diodes were the first semiconductor electronic devices. The discovery of crystals' rectifying abilities was made by German physicist Ferdinand Braun in 1874. The first semiconductor diodes, called cat's whisker diodes were made of crystals of minerals such as galena. Today most diodes are made of silicon, but other semiconductors such as germanium are sometimes used.<sup>[2]</sup>

o

### Thermionic and gaseous state diodes

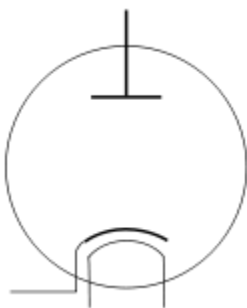


Figure 4: The symbol for an indirect heated vacuum tube diode. From top to bottom, the components are the anode, the cathode, and the heater filament.

Thermionic diodes are thermionic-valve devices (also known as vacuum tubes, tubes, or valves), which are arrangements of electrodes surrounded by a vacuum within a glass envelope. Early examples were fairly similar in appearance to incandescent light bulbs.

In thermionic valve diodes, a current through the heater filament indirectly heats the cathode, another internal electrode treated with a mixture of barium and strontium oxides, which are oxides of alkaline earth metals; these substances are chosen because they have a small work function. (Some valves use direct heating, in which a tungsten filament acts as both heater and cathode.) The heat causes thermionic emission of electrons into the vacuum. In forward operation, a surrounding metal electrode called the anode is positively charged so that it electrostatically attracts the emitted electrons. However, electrons are not easily released from the unheated anode surface when the voltage polarity is reversed. Hence, any reverse flow is negligible.

For much of the 20th century, thermionic valve diodes were used in analog signal applications, and as rectifiers in many power supplies. Today, valve diodes are only used in niche applications such as rectifiers in electric guitar and high-end audio amplifiers as well as specialized high-voltage equipment.

### **Semiconductor diodes**

A modern semiconductor diode is made of a crystal of semiconductor like silicon that has impurities added to it to create a region on one side that contains negative charge carriers (electrons), called n-type semiconductor, and a region on the other side that contains positive charge carriers (holes), called p-type semiconductor. The diode's terminals are attached to each of these regions. The boundary within the crystal between these two regions, called a PN junction, is where the action of the diode takes place. The crystal conducts conventional current in a direction from the p-type side (called the anode) to the n-type side (called the cathode), but not in the opposite direction.

Another type of semiconductor diode, the Schottky diode, is formed from the contact between a metal and a semiconductor rather than by a p-n junction.

### **Current–voltage characteristic**

A semiconductor diode's behavior in a circuit is given by its current–voltage characteristic, or I–V graph (see graph below). The shape of the curve is determined by the transport of charge carriers through the so-called *depletion layer* or *depletion region* that exists at the p-n junction between differing semiconductors. When a p-n junction is first created, conduction band (mobile) electrons from the N-doped region diffuse into the P-doped region where there is a large population of holes (vacant places for electrons) with which the electrons “recombine”. When a mobile electron recombines with a hole, both hole and electron vanish, leaving behind an immobile positively charged donor (dopant) on the N-side and negatively charged acceptor (dopant) on the P-side. The region around the p-n junction becomes depleted of charge carriers and thus behaves as an insulator.

However, the width of the depletion region (called the depletion width) cannot grow without limit. For each electron-hole pair that recombines, a positively charged dopant ion is left behind in the N-doped region, and a negatively charged dopant ion is left behind in the P-doped region. As recombination proceeds more ions are created, an increasing electric field develops through the depletion zone which acts to slow and then finally stop recombination. At this point, there is a “built-in” potential across the depletion zone.

If an external voltage is placed across the diode with the same polarity as the built-in potential, the depletion zone continues to act as an insulator, preventing any significant electric current flow (unless electron/hole pairs are actively being created in the junction by, for instance, light. see photodiode). This is the *reverse bias* phenomenon. However, if the polarity of the external voltage opposes the built-in potential, recombination can once again proceed, resulting in substantial electric current through the p-n junction (i.e. substantial numbers of electrons and holes recombine at the junction). For silicon diodes, the built-in potential is approximately 0.7 V (0.3 V for Germanium and 0.2 V for Schottky). Thus, if an external current is passed through the diode, about 0.7 V will be developed across the diode such that the P-doped region is positive with respect to the N-doped region and the diode is said to be “turned on” as it has a *forward bias*.

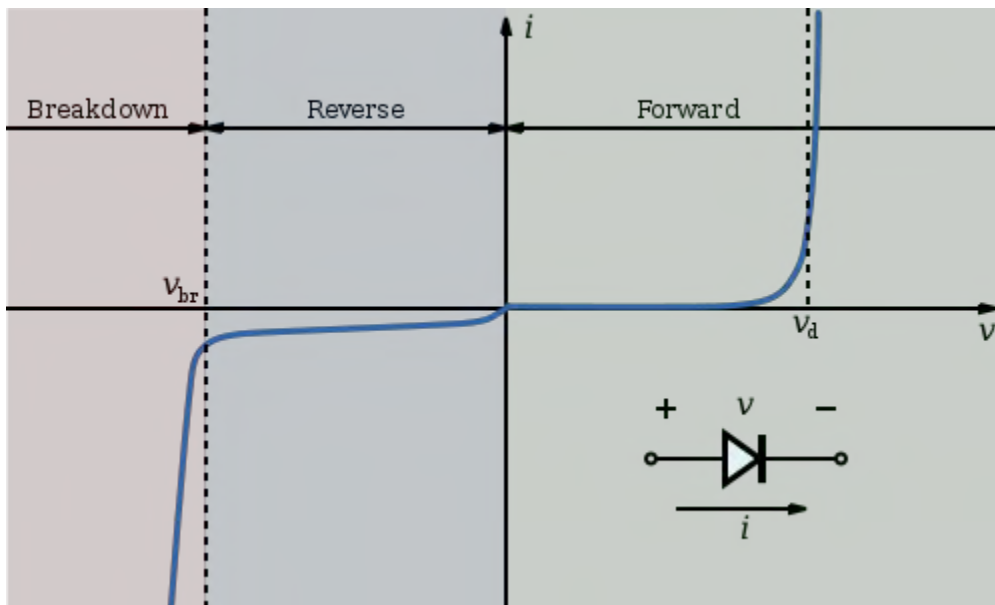


Figure 5: I–V characteristics of a P–N junction diode (not to scale).

A diode’s ‘*I–V characteristic*’ can be approximated by four regions of operation (see the figure at right).

At very large reverse bias, beyond the peak inverse voltage or PIV, a process called reverse breakdown occurs which causes a large increase in current (i.e. a large number of electrons and holes are created at, and move away from the pn junction) that usually damages the device permanently. The avalanche diode is deliberately designed for use in the avalanche region. In the zener diode, the concept of PIV is not applicable. A zener diode contains a heavily doped p-n junction allowing electrons to tunnel from the valence band of the p-type material to the

conduction band of the n-type material, such that the reverse voltage is “clamped” to a known value (called the *zener voltage*), and avalanche does not occur. Both devices, however, do have a limit to the maximum current and power in the clamped reverse voltage region. Also, following the end of forward conduction in any diode, there is reverse current for a short time. The device does not attain its full blocking capability until the reverse current ceases.

The second region, at reverse biases more positive than the PIV, has only a very small reverse saturation current. In the reverse bias region for a normal P-N rectifier diode, the current through the device is very low (in the  $\mu\text{A}$  range). However, this is temperature dependent, and at sufficiently high temperatures, a substantial amount of reverse current can be observed (mA or more).

The third region is forward but small bias, where only a small forward current is conducted.

As the potential difference is increased above an arbitrarily defined “cut-in voltage” or “on-voltage” or “diode forward voltage drop ( $V_d$ )”, the diode current becomes appreciable (the level of current considered “appreciable” and the value of cut-in voltage depends on the application), and the diode presents a very low resistance. The current–voltage curve is exponential. In a normal silicon diode at rated currents, the arbitrary “cut-in” voltage is defined as 0.6 to 0.7 volts. The value is different for other diode types — Schottky diodes can be rated as low as 0.2 V, Germanium diodes 0.25-0.3 V, and red or blue light-emitting diodes (LEDs) can have values of 1.4 V and 4.0 V respectively.

At higher currents the forward voltage drop of the diode increases. A drop of 1 V to 1.5 V is typical at full rated current for power diodes.

### Shockley diode equation

The *Shockley ideal diode equation* or the *diode law* (named after transistor co-inventor William Bradford Shockley, not to be confused with tetrode inventor Walter H. Schottky) gives the I–V characteristic of an ideal diode in either forward or reverse bias (or no bias). The equation is:

$$I = I_S \left( e^{V_D/(nV_T)} - 1 \right),$$

where

$I$  is the diode current,

$I_S$  is the reverse bias saturation current (or scale current),

$V_D$  is the voltage across the diode,

$V_T$  is the thermal voltage, and

$n$  is the *ideality factor*, also known as the *quality factor* or sometimes *emission coefficient*. The ideality factor  $n$  varies from 1 to 2 depending on the fabrication process and semiconductor material and in many cases is assumed to be approximately equal to 1 (thus the notation  $n$  is omitted).

The thermal voltage  $V_T$  is approximately 25.85 mV at 300 K, a temperature close to “room temperature” commonly used in device simulation software. At any temperature it is a known constant defined by:

$$V_T = \frac{kT}{q},$$

where  $k$  is the Boltzmann constant,  $T$  is the absolute temperature of the p-n junction, and  $q$  is the magnitude of charge on an electron (the elementary charge).

The *Shockley ideal diode equation* or the *diode law* is derived with the assumption that the only processes giving rise to the current in the diode are drift (due to electrical field), diffusion, and thermal recombination-generation. It also assumes that the recombination-generation (R-G) current in the depletion region is insignificant. This means that the Shockley equation doesn't account for the processes involved in reverse breakdown and photon-assisted R-G. Additionally, it doesn't describe the “leveling off” of the I-V curve at high forward bias due to internal resistance.

Under *reverse bias* voltages (see Figure 5) the exponential in the diode equation is negligible, and the current is a constant (negative) reverse current value of  $-I_S$ . The reverse *breakdown region* is not modeled by the Shockley diode equation.

For even rather small *forward bias* voltages (see Figure 5) the exponential is very large because the thermal voltage is very small, so the subtracted ‘1’ in the diode equation is negligible and the forward diode current is often approximated as

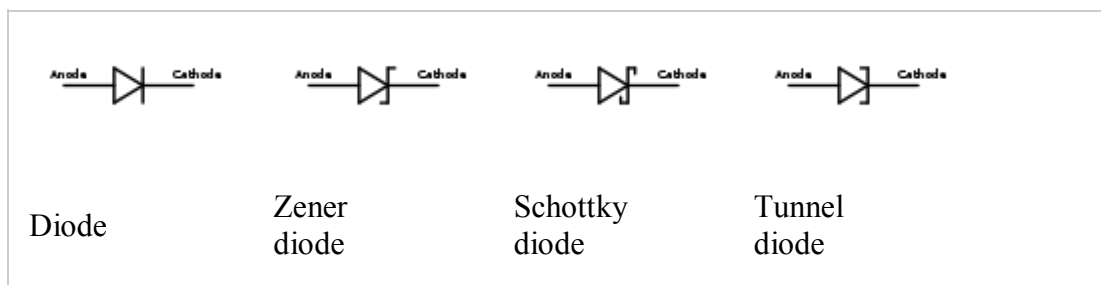
$$I = I_S e^{V_D / (nV_T)}$$

The use of the diode equation in circuit problems is illustrated in the article on diode modeling.

### Small-signal behaviour

For circuit design, a small-signal model of the diode behavior often proves useful. A specific example of diode modeling is discussed in the article on small-signal circuits.

### Types of semiconductor diode



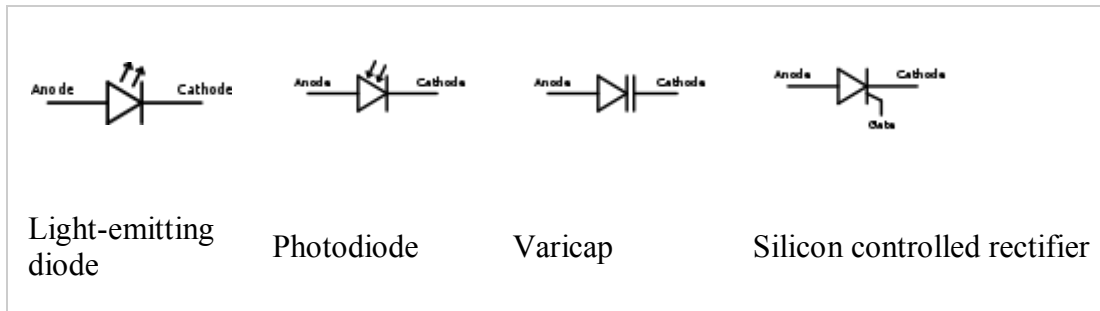


Figure 6: Some diode symbols.



Figure 7: Typical diode packages in same alignment as diode symbol. Thin bar depicts the cathode.



Figure 8: Several types of diodes. The scale is centimeters.

There are several types of junction diodes, which either emphasize a different physical aspect of a diode often by geometric scaling, doping level, choosing the right electrodes, are just an application of a diode in a special circuit, or are really different devices like the Gunn and laser diode and the MOSFET:

Normal (p-n) diodes, which operate as described above, are usually made of doped silicon or, more rarely, germanium. Before the development of modern silicon power rectifier diodes, cuprous oxide and later selenium was used; its low efficiency gave it a much higher forward voltage drop (typically 1.4–1.7 V per cell, with multiple cells stacked to increase the peak inverse voltage rating in high voltage rectifiers), and required a large heat sink (often an

extension of the diode's metal substrate), much larger than a silicon diode of the same current ratings would require. The vast majority of all diodes are the p-n diodes found in CMOS integrated circuits, which include two diodes per pin and many other internal diodes.

### Avalanche diodes

Diodes that conduct in the reverse direction when the reverse bias voltage exceeds the breakdown voltage. These are electrically very similar to Zener diodes, and are often mistakenly called Zener diodes, but break down by a different mechanism, the *avalanche effect*. This occurs when the reverse electric field across the p-n junction causes a wave of ionization, reminiscent of an avalanche, leading to a large current. Avalanche diodes are designed to break down at a well-defined reverse voltage without being destroyed. The difference between the avalanche diode (which has a reverse breakdown above about 6.2 V) and the Zener is that the channel length of the former exceeds the "mean free path" of the electrons, so there are collisions between them on the way out. The only practical difference is that the two types have temperature coefficients of opposite polarities.

### Cat's whisker or crystal diodes

These are a type of point-contact diode. The cat's whisker diode consists of a thin or sharpened metal wire pressed against a semiconducting crystal, typically galena or a piece of coal.<sup>[9]</sup> The wire forms the anode and the crystal forms the cathode. Cat's whisker diodes were also called crystal diodes and found application in crystal radio receivers. Cat's whisker diodes are generally obsolete, but may be available from a few manufacturers.<sup>[citation needed]</sup>

### Constant current diodes

These are actually a JFET<sup>[10]</sup> with the gate shorted to the source, and function like a two-terminal current-limiter analog to the Zener diode, which is limiting voltage. They allow a current through them to rise to a certain value, and then level off at a specific value. Also called *CLDs*, *constant-current diodes*, *diode-connected transistors*, or *current-regulating diodes*.

### Esaki or tunnel diodes

These have a region of operation showing negative resistance caused by quantum tunneling, thus allowing amplification of signals and very simple bistable circuits. These diodes are also the type most resistant to nuclear radiation.

### Gunn diodes

These are similar to tunnel diodes in that they are made of materials such as GaAs or InP that exhibit a region of negative differential resistance. With appropriate biasing, dipole

domains form and travel across the diode, allowing high frequency microwave oscillators to be built.

### Light-emitting diodes (LEDs)

In a diode formed from a direct band-gap semiconductor, such as gallium arsenide, carriers that cross the junction emit photons when they recombine with the majority carrier on the other side. Depending on the material, wavelengths (or colors)<sup>[11]</sup> from the infrared to the near ultraviolet may be produced.<sup>[12]</sup> The forward potential of these diodes depends on the wavelength of the emitted photons: 1.2 V corresponds to red, 2.4 V to violet. The first LEDs were red and yellow, and higher-frequency diodes have been developed over time. All LEDs produce incoherent, narrow-spectrum light; “white” LEDs are actually combinations of three LEDs of a different color, or a blue LED with a yellow scintillator coating. LEDs can also be used as low-efficiency photodiodes in signal applications. An LED may be paired with a photodiode or phototransistor in the same package, to form an opto-isolator.

### Laser diodes

When an LED-like structure is contained in a resonant cavity formed by polishing the parallel end faces, a laser can be formed. Laser diodes are commonly used in optical storage devices and for high speed optical communication.

### Peltier diodes

These diodes are used as sensors, heat engines for thermoelectric cooling. Charge carriers absorb and emit their band gap energies as heat.

### Photodiodes

All semiconductors are subject to optical charge carrier generation. This is typically an undesired effect, so most semiconductors are packaged in light blocking material. Photodiodes are intended to sense light (photodetector), so they are packaged in materials that allow light to pass, and are usually PIN (the kind of diode most sensitive to light).<sup>[13]</sup> A photodiode can be used in solar cells, in photometry, or in optical communications. Multiple photodiodes may be packaged in a single device, either as a linear array or as a two-dimensional array. These arrays should not be confused with charge-coupled devices.

### Point-contact diodes

These work the same as the junction semiconductor diodes described above, but their construction is simpler. A block of n-type semiconductor is built, and a conducting sharp-point contact made with some group-3 metal is placed in contact with the semiconductor. Some metal migrates into the semiconductor to make a small region of p-type

semiconductor near the contact. The long-popular 1N34 germanium version is still used in radio receivers as a detector and occasionally in specialized analog electronics.

### PIN diodes

A PIN diode has a central un-doped, or *intrinsic*, layer, forming a p-type/intrinsic/n-type structure.<sup>[14]</sup> They are used as radio frequency switches and attenuators. They are also used as large volume ionizing radiation detectors and as photodetectors. PIN diodes are also used in power electronics, as their central layer can withstand high voltages. Furthermore, the PIN structure can be found in many power semiconductor devices, such as IGBTs, power MOSFETs, and thyristors.

### Schottky diodes

Schottky diodes are constructed from a metal to semiconductor contact. They have a lower forward voltage drop than p-n junction diodes. Their forward voltage drop at forward currents of about 1 mA is in the range 0.15 V to 0.45 V, which makes them useful in voltage clamping applications and prevention of transistor saturation. They can also be used as low loss rectifiers although their reverse leakage current is generally higher than that of other diodes. Schottky diodes are majority carrier devices and so do not suffer from minority carrier storage problems that slow down many other diodes — so they have a faster “reverse recovery” than p-n junction diodes. They also tend to have much lower junction capacitance than p-n diodes which provides for high switching speeds and their use in high-speed circuitry and RF devices such as switched-mode power supply, mixers and detectors.

### Super barrier diodes

Super barrier diodes are rectifier diodes that incorporate the low forward voltage drop of the Schottky diode with the surge-handling capability and low reverse leakage current of a normal p-n junction diode.

### Gold-doped diodes

As a dopant, gold (or platinum) acts as recombination centers, which help a fast recombination of minority carriers. This allows the diode to operate at signal frequencies, at the expense of a higher forward voltage drop. Gold doped diodes are faster than other p-n diodes (but not as fast as Schottky diodes). They also have less reverse-current leakage than Schottky diodes (but not as good as other p-n diodes).<sup>[15][16]</sup> A typical example is the 1N914.

### Snap-off or Step recovery diodes

The term *step recovery* relates to the form of the reverse recovery characteristic of these devices. After a forward current has been passing in an SRD and the current is interrupted or reversed, the reverse conduction will cease very abruptly (as in a step waveform).

SRDs can therefore provide very fast voltage transitions by the very sudden disappearance of the charge carriers.

### Transient voltage suppression diode (TVS)

These are avalanche diodes designed specifically to protect other semiconductor devices from high-voltage transients.<sup>[17]</sup> Their p-n junctions have a much larger cross-sectional area than those of a normal diode, allowing them to conduct large currents to ground without sustaining damage.

### Varicap or varactor diodes

These are used as voltage-controlled capacitors. These are important in PLL (phase-locked loop) and FLL (frequency-locked loop) circuits, allowing tuning circuits, such as those in television receivers, to lock quickly, replacing older designs that took a long time to warm up and lock. A PLL is faster than an FLL, but prone to integer harmonic locking (if one attempts to lock to a broadband signal). They also enabled tunable oscillators in early discrete tuning of radios, where a cheap and stable, but fixed-frequency, crystal oscillator provided the reference frequency for a voltage-controlled oscillator.

### Zener diodes

Diodes that can be made to conduct backwards. This effect, called Zener breakdown, occurs at a precisely defined voltage, allowing the diode to be used as a precision voltage reference. In practical voltage reference circuits Zener and switching diodes are connected in series and opposite directions to balance the temperature coefficient to near zero. Some devices labeled as high-voltage Zener diodes are actually avalanche diodes (see above). Two (equivalent) Zeners in series and in reverse order, in the same package, constitute a transient absorber (or Transorb, a registered trademark). The Zener diode is named for Dr. Clarence Melvin Zener of Southern Illinois University, inventor of the device.

Other uses for semiconductor diodes include sensing temperature, and computing analog logarithms (see Operational amplifier applications#Logarithmic).

### **Numbering and coding schemes**

There are a number of common, standard and manufacturer-driven numbering and coding schemes for diodes; the two most common being the EIA/JEDEC standard and the European Pro Electron standard:

#### **EIA/JEDEC**

A standardized 1N-series numbering system was introduced in the US by EIA/JEDEC (Joint Electron Device Engineering Council) about 1960. Among the most popular in this series were:

1N34A/1N270 (Germanium signal), 1N914/1N4148 (Silicon signal), 1N4001-1N4007 (Silicon 1A power rectifier) and 1N54xx (Silicon 3A power rectifier)<sup>[18][19][20]</sup>

## Pro Electron

The European Pro Electron coding system for active components was introduced in 1966 and comprises two letters followed by the part code. The first letter represents the semiconductor material used for the component (A = Germanium and B = Silicon) and the second letter represents the general function of the part (for diodes: A = low-power/signal, B = Variable capacitance, X = Multiplier, Y = Rectifier and Z = Voltage reference), for example:

- AA-series germanium low-power/signal diodes (e.g.: AA119)
- BA-series silicon low-power/signal diodes (e.g.: BAT18 Silicon RF Switching Diode)
- BY-series silicon rectifier diodes (e.g.: BY127 1250V, 1A rectifier diode)
- BZ-series silicon zener diodes (e.g.: BZY88C4V7 4.7V zener diode)

Other common numbering / coding systems (generally manufacturer-driven) include:

- GD-series germanium diodes (ed: GD9) — this is a very old coding system
- OA-series germanium diodes (e.g.: OA47) — a coding sequence developed by Mullard, a UK company

As well as these common codes, many manufacturers or organisations have their own systems too — for example:

- HP diode 1901-0044 = JEDEC 1N4148
- UK military diode CV448 = Mullard type OA81 = GEC type GEX23

## Related devices

- Rectifier
- Transistor
- Thyristor or silicon controlled rectifier (SCR)
- TRIAC
- Diac
- Varistor

In optics, an equivalent device for the diode but with laser light would be the Optical isolator, also known as an Optical Diode, that allows light to only pass in one direction. It uses a Faraday rotator as the main component.

## Applications

### Radio demodulation

The first use for the diode was the demodulation of amplitude modulated (AM) radio broadcasts. The history of this discovery is treated in depth in the radio article. In summary, an AM signal consists of alternating positive and negative peaks of voltage, whose amplitude or “envelope” is proportional to the original audio signal. The diode (originally a crystal diode) rectifies the AM radio frequency signal, leaving an audio signal which is the original audio signal, minus atmospheric noise. The audio is extracted using a simple filter and fed into an audio amplifier or transducer, which generates sound waves.

### **Power conversion**

**Rectifiers** are constructed from diodes, where they are used to convert alternating current (AC) electricity into direct current (DC). Automotive alternators are a common example, where the diode, which rectifies the AC into DC, provides better performance than the commutator of earlier dynamo. Similarly, diodes are also used in **Cockcroft–Walton voltage multipliers** to convert AC into higher DC voltages.

### **Over-voltage protection**

Diodes are frequently used to conduct damaging high voltages away from sensitive electronic devices. They are usually reverse-biased (non-conducting) under normal circumstances. When the voltage rises above the normal range, the diodes become forward-biased (conducting). For example, diodes are used in (stepper motor and H-bridge) motor controller and relay circuits to de-energize coils rapidly without the damaging voltage spikes that would otherwise occur. (Any diode used in such an application is called a flyback diode). Many integrated circuits also incorporate diodes on the connection pins to prevent external voltages from damaging their sensitive transistors. Specialized diodes are used to protect from over-voltages at higher power (see Diode types above).

### **Logic gates**

Diodes can be combined with other components to construct AND and OR logic gates. This is referred to as diode logic.

### **Ionizing radiation detectors**

In addition to light, mentioned above, semiconductor diodes are sensitive to more energetic radiation. In electronics, cosmic rays and other sources of ionizing radiation cause noise pulses and single and multiple bit errors. This effect is sometimes exploited by particle detectors to detect radiation. A single particle of radiation, with thousands or millions of electron volts of energy, generates many charge carrier pairs, as its energy is deposited in the semiconductor material. If the depletion layer is large enough to catch the whole shower or to stop a heavy particle, a fairly accurate measurement of the particle’s energy can be made, simply by measuring the charge conducted and without the complexity of a magnetic spectrometer or etc. These semiconductor radiation detectors need efficient and uniform charge collection and low leakage current. They are often cooled by liquid nitrogen. For longer range (about a centimetre) particles they need a very large depletion depth and large area. For short range particles, they

need any contact or un-depleted semiconductor on at least one surface to be very thin. The back-bias voltages are near breakdown (around a thousand volts per centimetre). Germanium and silicon are common materials. Some of these detectors sense position as well as energy. They have a finite life, especially when detecting heavy particles, because of radiation damage. Silicon and germanium are quite different in their ability to convert gamma rays to electron showers.

Semiconductor detectors for high energy particles are used in large numbers. Because of energy loss fluctuations, accurate measurement of the energy deposited is of less use.

### **Temperature measurements**

A diode can be used as a temperature measuring device, since the forward voltage drop across the diode depends on temperature, as in a Silicon bandgap temperature sensor. From the Shockley ideal diode equation given above, it appears the voltage has a positive temperature coefficient (at a constant current) but depends on doping concentration and operating temperature (Sze 2007). The temperature coefficient can be negative as in typical thermistors or positive for temperature sense diodes down to about 20 kelvins. Typically, silicon diodes have approximately  $-2 \text{ mV}/^\circ\text{C}$  temperature coefficient at room temperature.

### **Current steering**

Diodes will prevent currents in unintended directions. To supply power to an electrical circuit during a power failure, the circuit can draw current from a battery. An Uninterruptible power supply may use diodes in this way to ensure that current is only drawn from the battery when necessary. Similarly, small boats typically have two circuits each with their own battery/batteries: one used for engine starting; one used for domestics. Normally both are charged from a single alternator, and a heavy duty split charge diode is used to prevent the higher charge battery (typically the engine battery) from discharging through the lower charged battery when the alternator is not running.

Diodes are also used in electronic musical keyboards. To reduce the amount of wiring needed in electronic musical keyboards, these instruments often use keyboard matrix circuits. The keyboard controller scans the rows and columns to determine which note the player has pressed. The problem with matrix circuits is that when several notes are pressed at once, the current can flow backwards through the circuit and trigger "phantom keys" that cause "ghost" notes to play. To avoid triggering unwanted notes, most keyboard matrix circuits have diodes soldered with the switch under each key of the musical keyboard. The same principle is also used for the switch matrix in solid state pinball machines.

## **p-n Junction Diode**

### **Diode:**

A pure silicon crystal or germanium crystal is known as an intrinsic semiconductor. There are

not enough free electrons and holes in an intrinsic semi-conductor to produce a usable current. The electrical action of these can be modified by doping means adding impurity atoms to a crystal to increase either the number of free holes or no of free electrons.

When a crystal has been doped, it is called a extrinsic semi-conductor. They are of two types

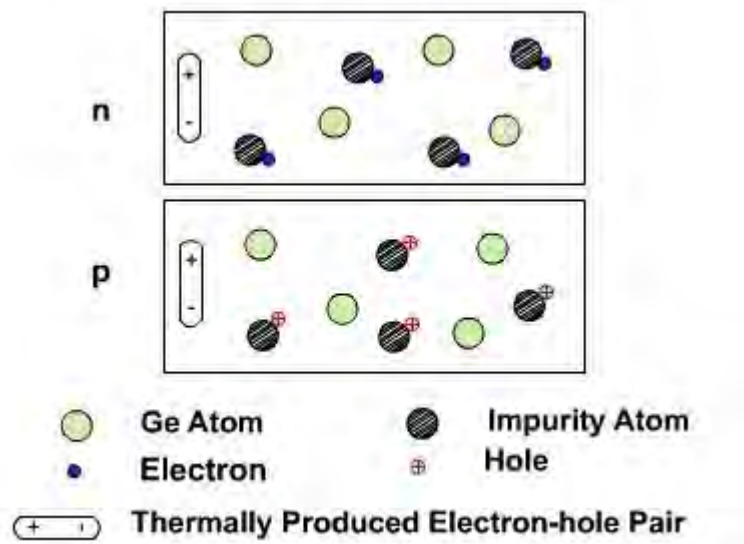
- n-type semiconductor having free electrons as majority carriers
- p-type semiconductor having free holes as majority carriers

By themselves, these doped materials are of little use. However, if a junction is made by joining p-type semiconductor to n-type semiconductor a useful device is produced known as diode. It will allow current to flow through it only in one direction. The unidirectional properties of a diode allow current flow when forward biased and disallow current flow when reversed biased. This is called rectification process and therefore it is also called rectifier.

How is it possible that by properly joining two semiconductors each of which, by itself, will freely conduct the current in any direct refuses to allow conduction in one direction.

Consider first the condition of p-type and n-type germanium just prior to joining **fig. 1**. The majority and minority carriers are in constant motion.

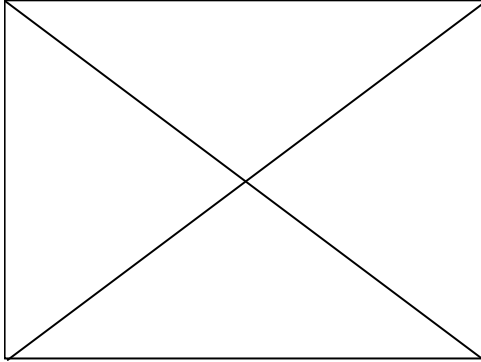
The minority carriers are thermally produced and they exist only for short time after which they recombine and neutralize each other. In the mean time, other minority carriers have been produced and this process goes on and on.



**Fig.1**

The number of these electron hole pair that exist at any one time depends upon the temperature. The number of majority carriers is however, fixed depending on the number of impurity atoms available. While the electrons and holes are in motion but the atoms are fixed in place and do not move.

As soon as, the junction is formed, the following processes are initiated **fig. 2**.

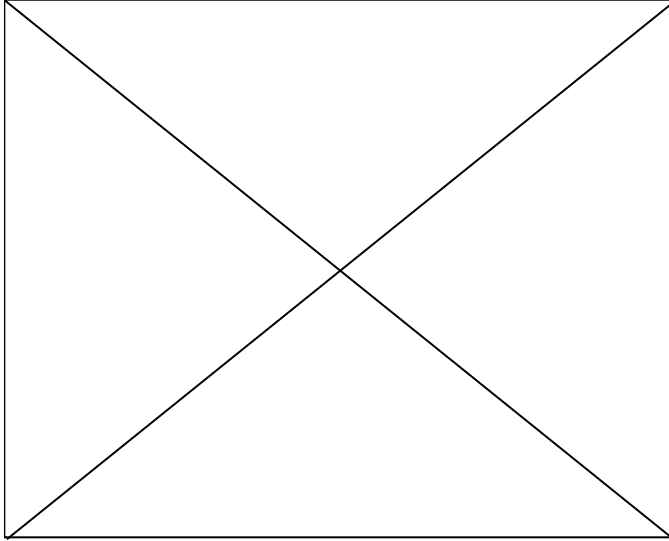


**Fig.2**

- Holes from the p-side diffuse into n-side where they recombine with free electrons.
- Free electrons from n-side diffuse into p-side where they recombine with free holes.
- The diffusion of electrons and holes is due to the fact that large no of electrons are concentrated in one area and large no of holes are concentrated in another area.
- When these electrons and holes begin to diffuse across the junction then they collide each other and negative charge in the electrons cancels the positive charge of the hole and both will lose their charges.
- The diffusion of holes and electrons is an electric current referred to as a recombination current. The recombination process decay exponentially with both time and distance from the junction. Thus most of the recombination occurs just after the junction is made and very near to junction.
- A measure of the rate of recombination is the lifetime defined as the time required for the density of carriers to decrease to 37% to the original concentration

The impurity atoms are fixed in their individual places. The atoms itself is a part of the crystal and so cannot move. When the electrons and hole meet, their individual charge is cancelled and this leaves the originating impurity atoms with a net charge, the atom that produced the electron now lack an electronic and so becomes charged positively, whereas the atoms that produced the hole now lacks a positive charge and becomes negative.

The electrically charged atoms are called ions since they are no longer neutral. These ions produce an electric field as shown in **fig. 3**. After several collisions occur, the electric field is great enough to repel rest of the majority carriers away of the junction. For example, an electron trying to diffuse from n to p side is repelled by the negative charge of the p-side. Thus diffusion process does not continue indefinitely but continues as long as the field is developed.



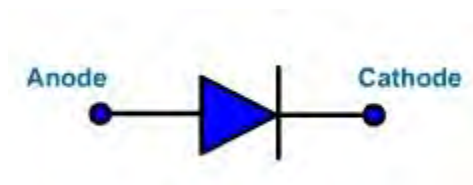
**Fig.3**

This region is produced immediately surrounding the junction that has no majority carriers. The majority carriers have been repelled away from the junction and junction is depleted from carriers. The junction is known as the barrier region or depletion region. The electric field represents a potential difference across the junction also called *space charge potential or barrier potential*. This potential is 0.7v for Si at 25° celcius and 0.3v for Ge.

The physical width of the depletion region depends on the doping level. If very heavy doping is used, the depletion region is physically thin because diffusion charge need not travel far across the junction before recombination takes place (short life time). If doping is light, then depletion is more wide (long life time).

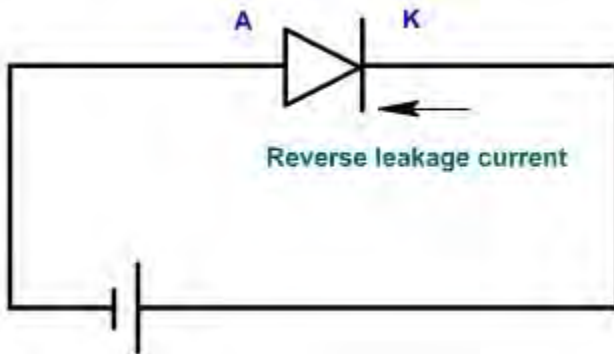
### p-n Junction Diode

The symbol of diode is shown in **fig. 4**. The terminal connected to p-layer is called anode (A) and the terminal n-layer is called cathode (K)



Reverse Bias:

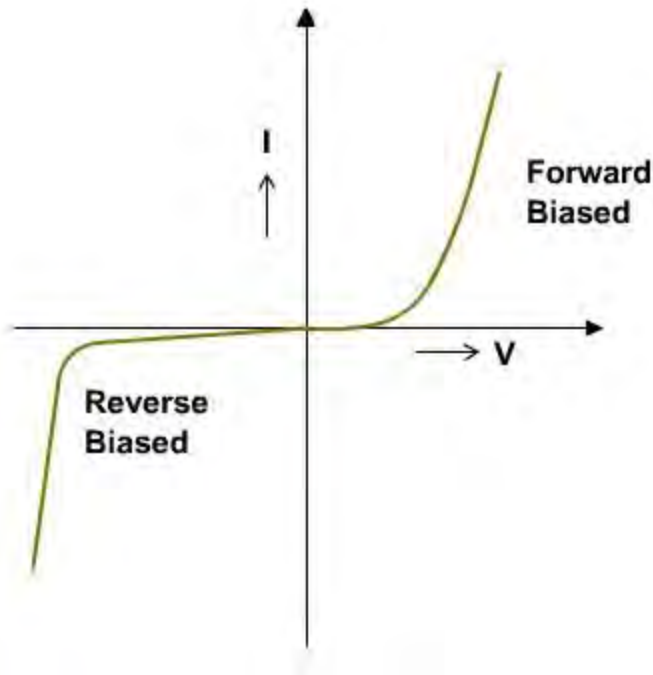
If positive terminal of dc source is connected to cathode and negative terminal is connected to anode, the diode is biased as shown in **fig. 5**.



**Fig.5**

When the diode is reverse biased then the depletion region width increases, majority carriers move away from the junction. There is no flow of current due to majority carriers but there are thermally produced electron hole pairs also. If these holes are generated in the vicinity of the junction then there is a flow of current. The negative voltage applied to the cathode attracts the holes thus generated and repels the electrons. At the same time, the positive voltage applied to the anode attracts the electrons and repels the holes. This will cause current to flow in the circuit. This current is usually very small (in the order of micro to nano amp). Since this current is due to minority carriers and the number of minority carriers is fixed at a given temperature, therefore, the current is almost constant known as reverse saturation current  $I_{CO}$ .

In an actual diode, the current is not almost constant but increases slightly with voltage. This is due to surface leakage. The surface of the diode follows Ohm's law ( $V=IR$ ). The resistance under reverse bias conditions is very high (100k to MΩ). As the reverse voltage is increased, then at a certain voltage, breakdown of the diode takes place and it conducts heavily. This is known as avalanche or Zener breakdown. The characteristic of the diode is shown in **fig. 6**.



**Fig.6**

**Forward bias:**

When the diode is forward bias, then majority carriers are pushed towards junction, when they collide and recombine. Number of majority carriers are fixed in semiconductor. Therefore as each electron is eliminated at the junction, an electron must be introduced, this comes from battery. At the same time, one hole must be created in p-layer, by extracting one electron from p-layer. Therefore, there is a flow of carriers and thus flow of current.

**Diode**

**Space charge capacitance  $C_T$  of diode:**

Reverse bias causes majority carriers to move away from the junction, thereby creating more ions. Hence the thickness of depletion region increases. This region behaves as the dielectric material used for making capacitors. The p-type and n-type conducting on each side of dielectric act as the plate. The incremental capacitance  $C_T$  is defined by

$$C_T = \left| \frac{dQ}{dV} \right|$$

Since  $i = \frac{dQ}{dt}$

Therefore,  $i = C_T \frac{dV}{dt}$  (E-1)

where,  $dQ$  is the increase in charge caused by a change  $dV$  in voltage.  $C_T$  is not constant, it depends upon applied voltage, there fore it is defined as  $dQ / dV$ .

When p-n junction is forward biased, then also a capacitance is defined called *diffusion capacitance*  $C_D$  (rate of change of injected charge with voltage) to take into account the time delay in moving the charges across the junction by the diffusion process. It is considered as a fictitious element that allow us to predict time delay.

If the amount of charge to be moved across the junction is increased, the time delay is greater, it follows that diffusion capacitance varies directly with the magnitude of forward current.

$$C_D = \frac{dQ}{dV} = \frac{I\tau}{dV} \quad (E-2)$$

### Relationship between Diode Current and Diode Voltage

An exponential relationship exists between the carrier density and applied potential of diode junction as given in equation E-3. This exponential relationship of the current  $i_D$  and the voltage  $v_D$  holds over a range of at least seven orders of magnitudes of current - that is a factor of  $10^7$ .

$$i_D = I_0 \left[ \exp\left(\frac{qV_D}{nkT}\right) - 1 \right] = I_0 \left[ e^{\left(\frac{qV_D}{nkT}\right)} - 1 \right] \quad (E-3)$$

Where,

$i_D$ = Current through the diode (dependent variable in this expression)  
 $v_D$ = Potential difference across the diode terminals (independent variable in this expression)  
 $I_0$ = Reverse saturation current (of the order of  $10^{-15}$  A for small signal diodes, but  $I_0$  is a strong function of temperature)  
 $q$  = Electron charge:  $1.60 \times 10^{-19}$  joules/volt  
 $k$  = Boltzmann's constant:  $1.38 \times 10^{-23}$  joules /° K  
 $T$  = Absolute temperature in degrees Kelvin (°K = 273 + temperature in °C)  
 $n$  = Empirical scaling constant between 0.5 and 2, sometimes referred to as the Exponential Ideality Factor

The empirical constant,  $n$ , is a number that can vary according to the voltage and current levels. It depends on electron drift, diffusion, and carrier recombination in the depletion region. Among the quantities affecting the value of  $n$  are the diode manufacture, levels of doping and purity of materials. If  $n=1$ , the value of  $k T/ q$  is 26 mV at 25°C. When  $n=2$ ,  $k T/ q$  becomes 52 mV.

For germanium diodes,  $n$  is usually considered to be close to 1. For silicon diodes,  $n$  is in the

range of 1.3 to 1.6.  $n$  is assumed 1 for all junctions all throughout unless otherwise noted.

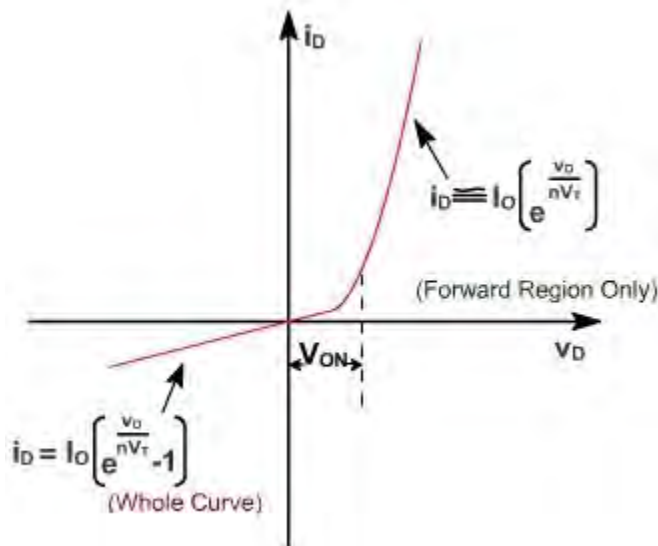
Equation (E-3) can be simplified by defining  $V_T = k T/q$ , yielding

$$i_D = I_0 \left[ \exp\left(\frac{v_D}{nV_T}\right) - 1 \right] = I_0 \left[ e^{\left(\frac{v_D}{nV_T}\right)} - 1 \right] \quad (\text{E-4})$$

At room temperature (25°C) with forward-bias voltage only the first term in the parentheses is dominant and the current is approximately given by

$$i_D = I_0 e^{\frac{v_D}{nV_T}} \quad (\text{E-5})$$

The current-voltage (I-V) characteristic of the diode, as defined by (E-3) is illustrated in **fig. 1**. The curve in the figure consists of two exponential curves. However, the exponent values are such that for voltages and currents experienced in practical circuits, the curve sections are close to being straight lines. For voltages less than  $V_{ON}$ , the curve is approximated by a straight line of slope close to zero. Since the slope is the conductance (i.e.,  $i / v$ ), the conductance is very small in this region, and the equivalent resistance is very high. For voltages above  $V_{ON}$ , the curve is approximated by a straight line with a very large slope. The conductance is therefore very large, and the diode has a very small equivalent resistance.



**Fig.1 - Diode Voltage relationship**

The slope of the curves of **fig.1** changes as the current and voltage change since the I-V characteristic follows the exponential relationship of relationship of equation (E-4). Differentiate the equation (E-4) to find the slope at any arbitrary value of  $v_D$  or  $i_D$ ,

$$\frac{di_D}{dv_D} = \frac{I_0}{nV_T} \exp\left(\frac{v_D}{nV_T}\right) = \frac{I_0}{nV_T} e^{\frac{v_D}{nV_T}} \quad (\text{E-6})$$

This slope is the equivalent conductance of the diode at the specified values of  $v_D$  or  $i_D$ .

We can approximate the slope as a linear function of the diode current. To eliminate the exponential function, we substitute equation (E-4) into the exponential of equation (E-7) to obtain

$$\exp\left(\frac{v_D}{nV_T}\right) = \frac{i_D}{I_0} + 1 = \left(\frac{di_D}{dv_D}\right) \left(\frac{nV_T}{I_0}\right) \quad (\text{E-7})$$

A realistic assumption is that  $I_0 \ll i_D$  equation (E-7) then yields,

$$\frac{di_D}{dv_D} = \frac{i_D + I_0}{nV_T} \approx \frac{i_D}{nV_T} \quad (\text{E-8})$$

The approximation applies if the diode is forward biased. The dynamic resistance is the reciprocal of this expression.

$$r_d = \frac{nV_T}{i_D + I_0} \approx \frac{nV_T}{i_D} \quad (\text{E-9})$$

Although  $r_d$  is a function of  $i_d$ , we can approximate it as a constant if the variation of  $i_D$  is small. This corresponds to approximating the exponential function as a straight line within a specific operating range.

Normally, the term  $R_f$  to denote diode forward resistance.  $R_f$  is composed of  $r_d$  and the contact resistance. The contact resistance is a relatively small resistance composed of the resistance of the actual connection to the diode and the resistance of the semiconductor prior to the junction. The reverse-bias resistance is extremely large and is often approximated as infinity.

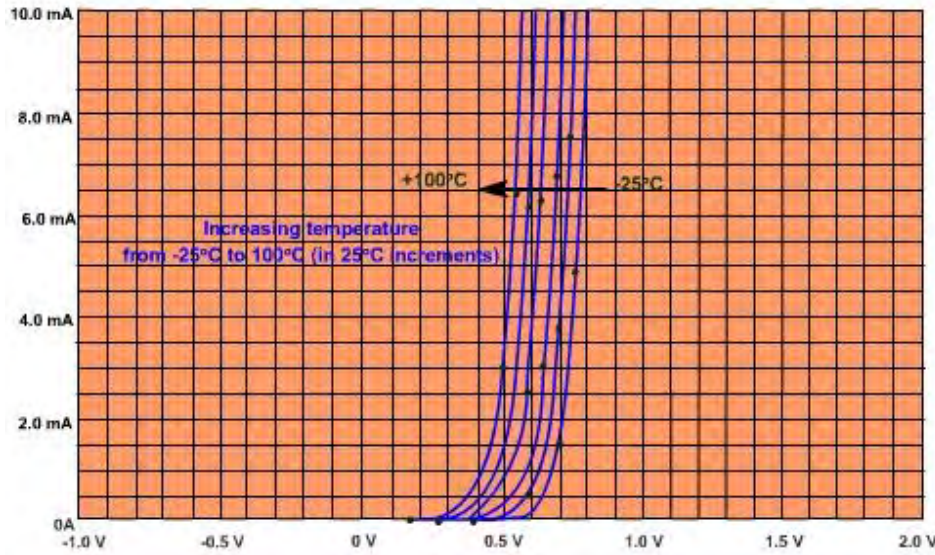
## Diode

### Temperature Effects:

Temperature plays an important role in determining the characteristic of diodes. As temperature increases, the turn-on voltage,  $v_{ON}$ , decreases. Alternatively, a decrease in temperature results in an increase in  $v_{ON}$ . This is illustrated in **fig. 2**, where  $V_{ON}$  varies linearly with temperature which is evidenced by the evenly spaced curves for increasing temperature in 25 °C increments.

The temperature relationship is described by equation

$$V_{ON}(T_{New}) - V_{ON}(T_{room}) = k_T(T_{New} - T_{room}) \quad (\text{E-10})$$



**Fig. 2 - Dependence of  $i_D$  on temperature versus  $v_D$  for real diode ( $k_T = -2.0 \text{ mV}/^\circ\text{C}$ )**

where,

$T_{room}$  = room temperature, or  $25^\circ\text{C}$ .

$T_{New}$  = new temperature of diode in  $^\circ\text{C}$ .

$V_{ON}(T_{room})$  = diode voltage at room temperature.

$V_{ON}(T_{New})$  = diode voltage at new temperature.

$k_T$  = temperature coefficient in  $\text{V}/^\circ\text{C}$ .

Although  $k_T$  varies with changing operating parameters, standard engineering practice permits approximation as a constant. Values of  $k_T$  for the various types of diodes at room temperature are given as follows:

$k_T = -2.5 \text{ mV}/^\circ\text{C}$  for germanium diodes

$k_T = -2.0 \text{ mV}/^\circ\text{C}$  for silicon diodes

The reverse saturation current,  $I_O$  also depends on temperature. At room temperature, it increases approximately 16% per  $^\circ\text{C}$  for silicon and 10% per  $^\circ\text{C}$  for germanium diodes. In other words,  $I_O$  approximately doubles for every  $5^\circ\text{C}$  increase in temperature for silicon, and for every  $7^\circ\text{C}$  for germanium. The expression for the reverse saturation current as a function of temperature can be approximated as

$$I_O(\text{at } T_2) = I_O(\text{at } T_1) \exp(k_i(T_2 - T_1)) = I_O(\text{at } T_1) e^{K_i(T_2 - T_1)} \quad (\text{E-11})$$

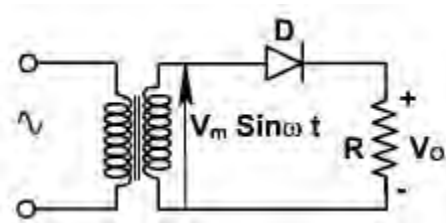
where  $K_f = 0.15/^\circ\text{C}$  ( for silicon) and  $T_1$  and  $T_2$  are two arbitrary temperatures.

## Applications of Diode

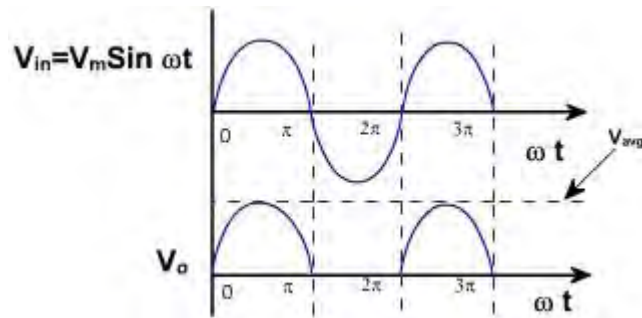
### Applications of diode:

#### Half wave Rectifier:

The single – phase half wave rectifier is shown in **fig. 8**.



**Fig. 8**



**Fig. 9**

In positive half cycle,  $D$  is forward biased and conducts. Thus the output voltage is same as the input voltage. In the negative half cycle,  $D$  is reverse biased, and therefore output voltage is zero. The output voltage waveform is shown in **fig. 9**.

The average output voltage of the rectifier is given by

$$\begin{aligned} V_{avg} &= \frac{1}{2\pi} \int_0^{\pi} V_m \sin \omega t \, d(\omega t) \\ &= \frac{V_m}{\pi} = 0.318 V_m \end{aligned}$$

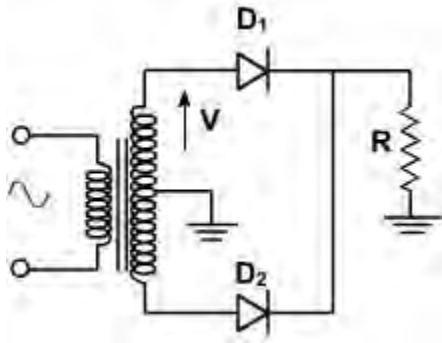
The average output current is given by

$$I_{avg} = \frac{V_m}{\pi R}$$

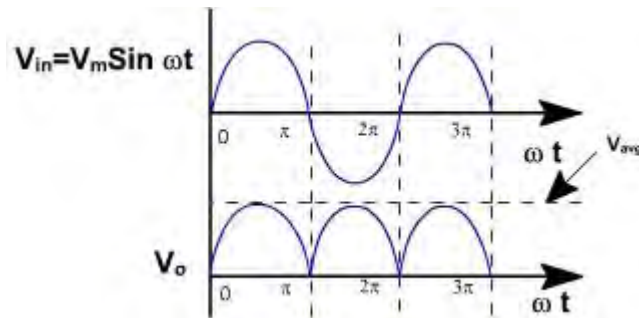
When the diode is reverse biased, entire transformer voltage appears across the diode. The maximum voltage across the diode is  $V_m$ . The diode must be capable to withstand this voltage. Therefore PIV half wave rating of diode should be equal to  $V_m$  in case of single-phase rectifiers. The average current rating must be greater than  $I_{avg}$

### Full Wave Rectifier:

A single – phase full wave rectifier using center tap transformer is shown in **fig. 10**. It supplies current in both half cycles of the input voltage.



**Fig. 10**



**Fig. 11**

In the first half cycle  $D_1$  is forward biased and conducts. But  $D_2$  is reverse biased and does not conduct. In the second half cycle  $D_2$  is forward biased, and conducts and  $D_1$  is reverse biased. It is also called 2 – pulse midpoint converter because it supplies current in both the half cycles. The output voltage waveform is shown in **fig. 11**.

The average output voltage is given by

$$V_{avg} = \frac{1}{\pi} \int_0^{\pi} V_m \sin \omega t \, d(\omega t)$$

$$= \frac{2V_m}{\pi}$$

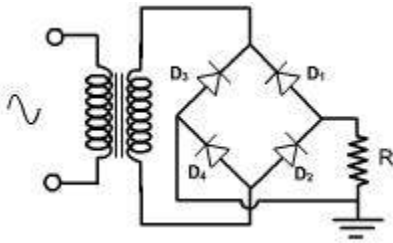
and the average load current is given by

$$I_{avg} = \frac{2V_m}{\pi R}$$

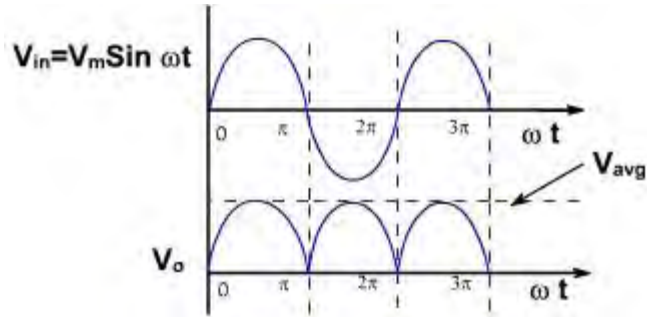
When  $D_1$  conducts, then full secondary voltage appears across  $D_2$ , therefore PIV rating of the diode should be  $2 V_m$ .

### Bridge Rectifier:

The single – phase full wave bridge rectifier is shown in **fig. 1**. It is the most widely used rectifier. It also provides currents in both the half cycle of input supply.



**Fig. 1**



**Fig. 2**

In the positive half cycle,  $D_1$  &  $D_4$  are forward biased and  $D_2$  &  $D_3$  are reverse biased. In the negative half cycle,  $D_2$  &  $D_3$  are forward biased, and  $D_1$  &  $D_4$  are reverse biased. The output voltage waveform is shown in **fig. 2** and it is same as full wave rectifier but the advantage is that PIV rating of diodes are  $V_m$  and only single secondary transformer is required.

The main disadvantage is that it requires four diodes. When low dc voltage is required then secondary voltage is low and diodes drop (1.4V) becomes significant. For low dc output, 2-pulse center tap rectifier is used because only one diode drop is there.

The ripple factor is the measure of the purity of dc output of a rectifier and is defined as

$$\text{Ripple factor} = \frac{\text{r.m.s value of the ac output voltage}}{\text{average dc output voltage}}$$

$$= \sqrt{V_0^2 + \sum_{n=1}^{\infty} V_n^2}$$

Therefore,

$$\text{Ripple factor} = \frac{\sqrt{V_{rms}^2 - V_0^2}}{V_0}$$

$$= \sqrt{\left(\frac{V_{rms}}{V_0}\right)^2 - 1}$$

### **Zener Diode:**

The diodes designed to work in breakdown region are called zener diode. If the reverse voltage exceeds the breakdown voltage, the zener diode will normally not be destroyed as long as the current does not exceed maximum value and the device closes not over load.

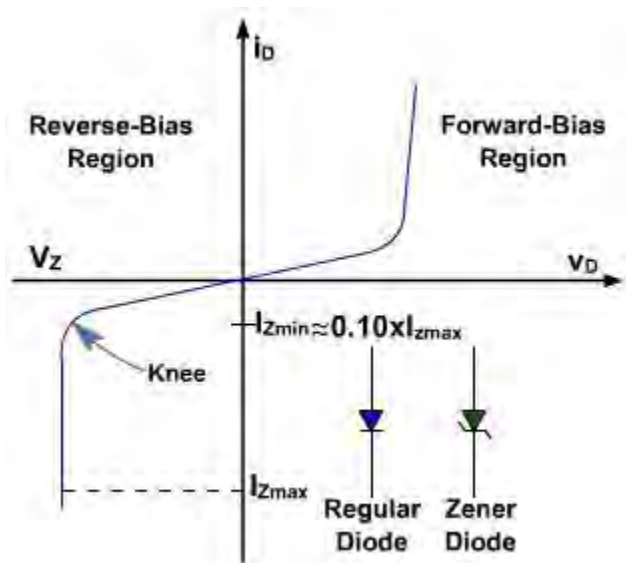
When a thermally generated carrier (part of the reverse saturation current) falls down the

junction and acquires energy of the applied potential, the carrier collides with crystal ions and imparts sufficient energy to disrupt a covalent bond. In addition to the original carrier, a new electron-hole pair is generated. This pair may pick up sufficient energy from the applied field to collide with another crystal ion and create still another electron-hole pair. This action continues and thereby disrupts the covalent bonds. The process is referred to as impact ionization, avalanche multiplication or avalanche breakdown.

There is a second mechanism that disrupts the covalent bonds. The use of a sufficiently strong electric field at the junction can cause a direct rupture of the bond. If the electric field exerts a strong force on a bound electron, the electron can be torn from the covalent bond thus causing the number of electron-hole pair combinations to multiply. This mechanism is called high field emission or Zener breakdown. The value of reverse voltage at which this occurs is controlled by the amount of doping of the diode. A heavily doped diode has a low Zener breakdown voltage, while a lightly doped diode has a high Zener breakdown voltage.

At voltages above approximately 8V, the predominant mechanism is the avalanche breakdown. Since the Zener effect (avalanche) occurs at a predictable point, the diode can be used as a voltage reference. The reverse voltage at which the avalanche occurs is called the breakdown or Zener voltage.

A typical Zener diode characteristic is shown in **fig. 1**. The circuit symbol for the Zener diode is different from that of a regular diode, and is illustrated in the figure. The maximum reverse current,  $I_{Z(max)}$ , which the Zener diode can withstand is dependent on the design and construction of the diode. A design guideline that the minimum Zener current, where the characteristic curve remains at  $V_Z$  (near the knee of the curve), is  $0.1 / I_{Z(max)}$ .



**Fig. 1 - Zener diode characteristic**

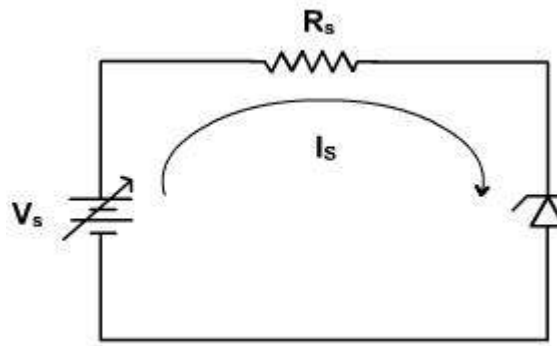
The power handling capacity of these diodes is better. The power dissipation of a zener diode equals the product of its voltage and current.

$$P_Z = V_Z I_Z$$

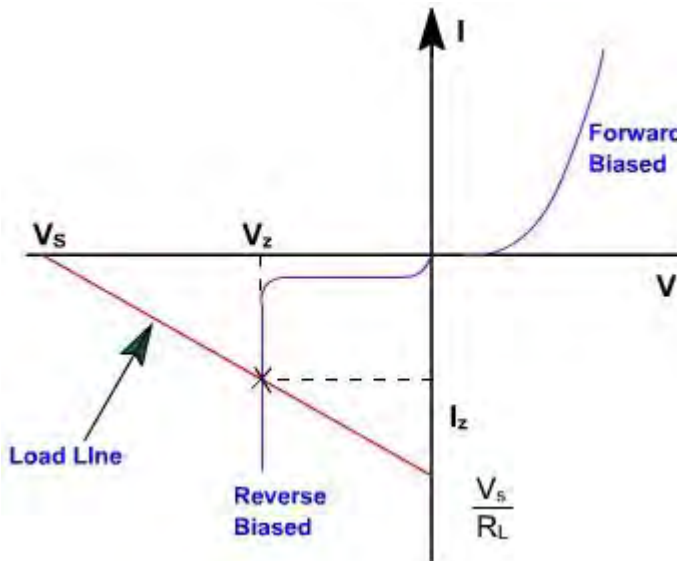
The amount of power which the zener diode can withstand ( $V_Z \cdot I_{Z(\max)}$ ) is a limiting factor in power supply design.

Zener Regulator:

When zener diode is forward biased it works as a diode and drop across it is 0.7 V. When it works in breakdown region the voltage across it is constant ( $V_Z$ ) and the current through diode is decided by the external resistance. Thus, zener diode can be used as a voltage regulator in the configuration shown in **fig. 2** for regulating the dc voltage. It maintains the output voltage constant even through the current through it changes.



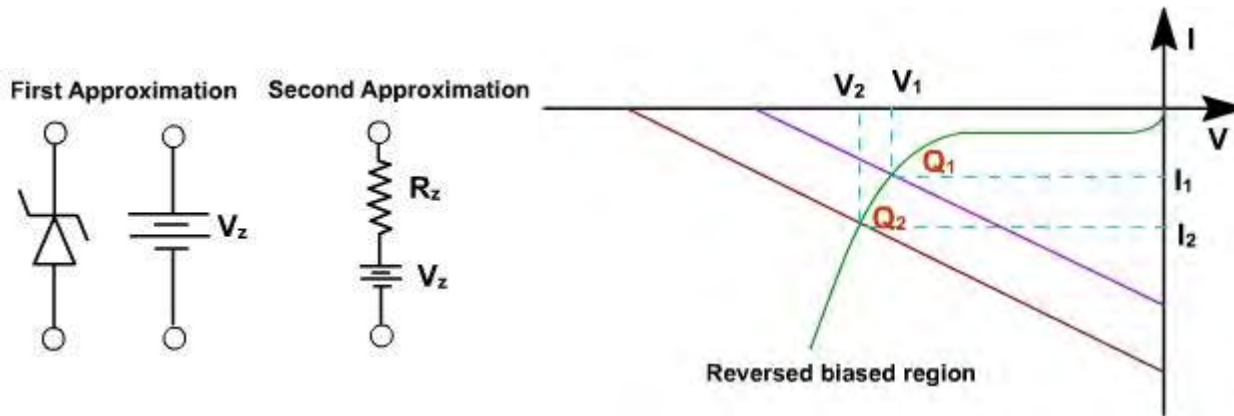
**Fig. 2**



**Fig. 3**

The load line of the circuit is given by  $V_s = I_s R_s + V_Z$ . The load line is plotted along with zener characteristic in **fig. 3**. The intersection point of the load line and the zener characteristic gives the output voltage and zener current.

To operate the zener in breakdown region  $V_s$  should always be greater than  $V_Z$ .  $R_s$  is used to limit the current. If the  $V_s$  voltage changes, operating point also changes simultaneously but voltage across zener is almost constant. The first approximation of zener diode is a voltage source of  $V_Z$  magnitude and second approximation includes the resistance also. The two approximate equivalent circuits are shown in **fig. 4**.



If second approximation of zener diode is considered, the output voltage varies slightly as shown in **fig. 5**. The zener ON state resistance produces more  $I \cdot R$  drop as the current increases. As the voltage varies from  $V_1$  to  $V_2$  the operating point shifts from  $Q_1$  to  $Q_2$ .

The voltage at  $Q_1$  is

$$V_1 = I_1 R_Z + V_Z$$

and at  $Q_2$

$$V_2 = I_2 R_Z + V_Z$$

Thus, change in voltage is

$$V_2 - V_1 = (I_2 - I_1) R_Z$$

$$\Delta V_Z = \Delta I_Z R_Z$$

## Zener Diode

Design of Zener regulator circuit:

A zener regulator circuit is shown in **fig. 6**. The varying load current is represented by a variable load resistance  $R_L$ .

The zener will work in the breakdown region only if the Thevenin voltage across zener is more than  $V_Z$ .

$$V_{TH} = V_s \frac{R_L}{R_s + R_L}$$

If zener is operating in breakdown region, the current through  $R_s$  is given by

$$I_s = \frac{V_s - V_Z}{R_s}$$

and load current  $I_L = \frac{V_Z}{R_L}$

$$I_s = I_Z + I_L$$

The circuit is designed such that the diode always operates in the breakdown region and the voltage  $V_Z$  across it remains fairly constant even though the current  $I_Z$  through it vary considerably.

If the load  $I_L$  should increase, the current  $I_Z$  should decrease by the same percentage in order to maintain load current constant  $I_s$ . This keeps the voltage drop across  $R_s$  constant and hence the output voltage.

If the input voltage should increase, the zener diode passes a larger current, that extra voltage is dropped across the resistance  $R_s$ . If input voltage falls, the current  $I_Z$  falls such that  $V_Z$  is constant.

In the practical application the source voltage,  $v_s$ , varies and the load current also varies. The design challenge is to choose a value of  $R_s$  which permits the diode to maintain a relatively constant output voltage, even when the input source voltage varies and the load current also varies.

We now analyze the circuit to determine the proper choice of  $R_s$ . For the circuit shown in figure,

$$R_s = \frac{v_s - V_Z}{i_R} = \frac{v_s - V_Z}{i_Z + i_L} \quad (E-1)$$

$$i_Z = \frac{v_s - V_Z}{R_s} - i_L \quad (E-2)$$

The variable quantities in Equation (E-2) are  $v_Z$  and  $i_L$ . In order to assure that the diode remains

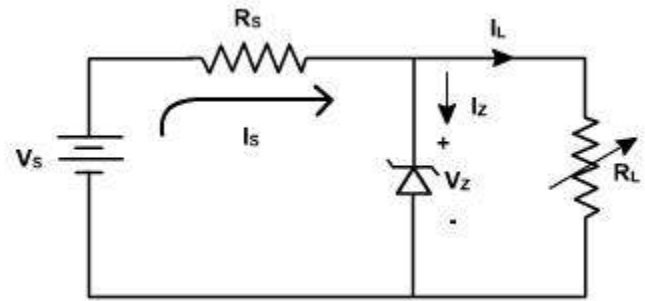


Fig. 6

in the constant voltage (breakdown) region, we examine the two extremes of input/output conditions, as follows:

- The current through the diode,  $i_Z$ , is a minimum ( $I_{Z \text{ min}}$ ) when the load current,  $i_L$  is maximum ( $I_{L \text{ max}}$ ) and the source voltage,  $v_s$  is minimum ( $V_{s \text{ min}}$ ).
- The current through the diode,  $i_Z$ , is a maximum ( $I_{Z \text{ max}}$ ) when the load current,  $i_L$ , is minimum ( $i_{L \text{ min}}$ ) and the source voltage  $v_s$  is minimum ( $V_{s \text{ max}}$ ).

When these characteristics of the two extremes are inserted into Equation (E-1),

we find 
$$R_s = \frac{V_{s \text{ min}} - V_Z}{I_{L \text{ max}} + I_{Z \text{ min}}} = \frac{V_{s \text{ max}} - V_Z}{I_{L \text{ min}} + I_{Z \text{ max}}} \quad (\text{E-3})$$

$$(V_{s \text{ min}} - V_Z)(I_{L \text{ min}} + I_{Z \text{ max}}) = (V_{s \text{ max}} - V_Z)(I_{L \text{ max}} + I_{Z \text{ min}}) \quad (\text{E-4})$$

In a practical problem, we know the range of input voltages, the range of output load currents, and the desired Zener voltage. Equation (E-4) thus represents one equation in two unknowns, the maximum and minimum Zener current. A second equation is found from the characteristic of zener. To avoid the non-constant portion of the characteristic curve, we use an accepted rule of thumb that the minimum Zener current should be 0.1 times the maximum (i.e., 10%), that is,

$$I_{Z \text{ min}} = 0.1 \times I_{Z \text{ max}} \quad (\text{E-5})$$

Solving the equations E-4 and E-5, we get,

$$I_{Z \text{ max}} = \frac{I_{L \text{ min}} (V_Z - V_{s \text{ min}}) + I_{L \text{ max}} (V_{s \text{ max}} - V_Z)}{V_{s \text{ min}} - 0.9V_Z - 0.1V_{s \text{ max}}} \quad (\text{E-6})$$

Now that we can solve for the maximum Zener current, the value of  $R_s$ , is calculated from Equation (E-3).

Zener diodes are manufactured with breakdown voltages  $V_Z$  in the range of a few volts to a few hundred volts. The manufacturer specifies the maximum power the diode can dissipate. For example, a 1W, 10 V zener can operate safely at currents up to 100mA.

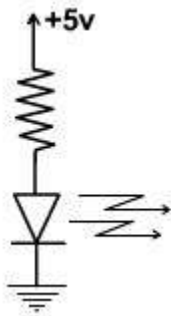
## Special Purpose Diodes

Light Emitting Diode :

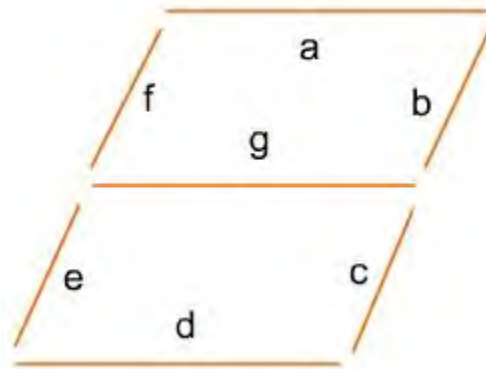
In a forward biased diode free electrons cross the junction and enter into p-layer where they

recombine with holes. Each recombination radiates energy as electron falls from higher energy level to a lower energy level. In ordinary diodes this energy is in the form of heat. In light emitting diode, this energy is in the form of light.

The symbol of LED is shown in **fig. 2**. Ordinary diodes are made of Ge or Si. This material blocks the passage of light. LEDs are made of different materials such as gallium, arsenic and phosphorus. LEDs can radiate red, green, yellow, blue, orange or infrared (invisible). The LED's forward voltage drop is more approximately 1.5V. Typical LED current is between 10 mA to 50 mA.



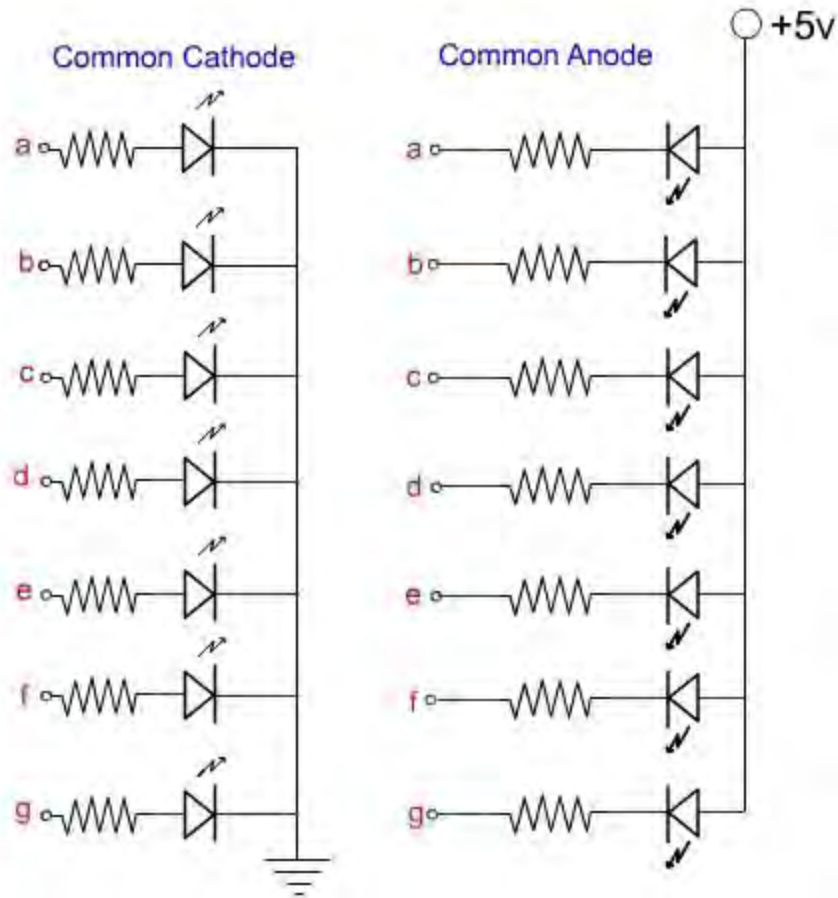
**Fig. 2**



**Fig. 3**

### Seven Segment Display :

Seven segment displays are used to display digits and few alphabets. It contains seven rectangular LEDs. Each LED is called a Segment. External resistors are used to limit the currents to safe Values. It can display any letters a, b, c, d, e, f, g, as shown in **fig. 3**.

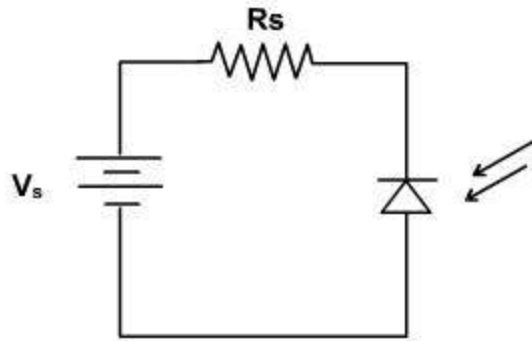


**Fig. 4**

The LEDs of seven-segment display are connected in either in common anode configuration or in common cathode configuration as shown in **fig. 4**.

**Photo diode :**

When a diode is reversed biased as shown in **fig. 5**, a reverse current flows due to minority carriers. These carriers exist because thermal energy keeps on producing free electrons and holes. The lifetime of the minority carriers is short, but while they exist they can contribute to the reverse current. When light energy bombards a p-n junction, it too can produce free electrons.



**Fig. 5**

In other words, the amount of light striking the junction can control the reverse current in a diode. A photo diode is made on the same principle. It is sensitive to the light. In this diode, through a window light falls to the junction. The stronger the light, the greater the minority carriers and larger the reverse current.

### **Voltage regulator**

A **voltage regulator** is an electrical regulator designed to automatically maintain a constant voltage level.

It may use an electromechanical mechanism, or passive or active electronic components. Depending on the design, it may be used to regulate one or more AC or DC voltages.

With the exception of passive shunt regulators, all modern electronic voltage regulators operate by comparing the actual output voltage to some internal fixed reference voltage. Any difference is amplified and used to control the regulation element in such a way as to reduce the voltage error. This forms a negative feedback control loop; increasing the open-loop gain tends to increase regulation accuracy but reduce stability (avoidance of oscillation, or ringing during step changes). There will also be a trade-off between stability and the speed of the response to changes. If the output voltage is too low (perhaps due to input voltage reducing or load current increasing), the regulation element is commanded, *up to a point*, to produce a higher output

voltage - by dropping less of the input voltage (for linear series regulators and buck switching regulators), or to draw input current for longer periods (boost-type switching regulators); if the output voltage is too high, the regulation element will normally be commanded to produce a lower voltage. However, many regulators have over-current protection, so that they will entirely stop sourcing current (or limit the current in some way) if the output current is too high, and some regulators may also shut down if the input voltage is outside a given range.

### Measures of regulator quality

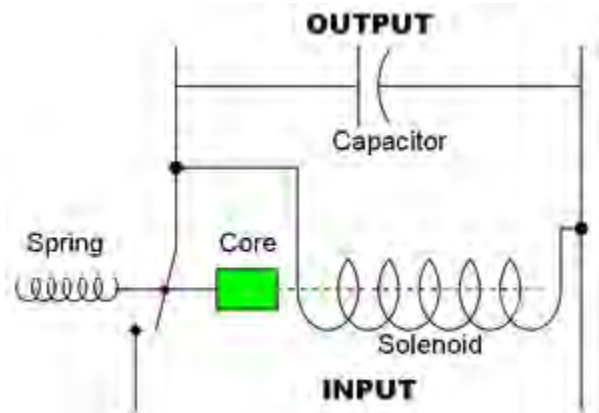
The output voltage can only be held *roughly* constant; the regulation is specified by two measurements:

- **load regulation** is the change in output voltage for a given change in load current (for example: "typically 15mV, maximum 100mV for load currents between 5mA and 1.4A, at some specified temperature and input voltage").
- **line regulation** or **input regulation** is the degree to which output voltage changes with input (supply) voltage changes - as a ratio of output to input change (for example "typically 13mV/V"), or the output voltage change over the entire specified input voltage range (for example "plus or minus 2% for input voltages between 90V and 260V, 50-60Hz").

Other important parameters are:

- **Temperature coefficient** of the output voltage is the change in output voltage with temperature (perhaps averaged over a given temperature range), while...
- **Initial accuracy** of a voltage regulator (or simply "the voltage accuracy") reflects the error in output voltage for a fixed regulator without taking into account temperature or aging effects on output accuracy.
- **Dropout voltage** - the minimum difference between input voltage and output voltage for which the regulator can still supply the specified current. A Low Drop-Out (LDO) regulator is designed to work well even with an input supply only a Volt or so above the output voltage.
- **Absolute Maximum Ratings** are defined for regulator components, specifying the continuous and peak output currents that may be used (sometimes internally limited), the maximum input voltage, maximum power dissipation at a given temperature, etc.
- **Output noise** (thermal white noise) and **output dynamic impedance** may be specified as graphs versus frequency, while output **ripple** noise (mains "hum" or switch-mode "hash" noise) may be given as peak-to-peak or RMS voltages, or in terms of their spectra.
- **Quiescent current** in a regulator circuit is the current drawn internally, not available to the load, normally measured as the input current while no load is connected (and hence a source of inefficiency; some linear regulators are, surprisingly, more efficient at very low current loads than switch-mode designs because of this).

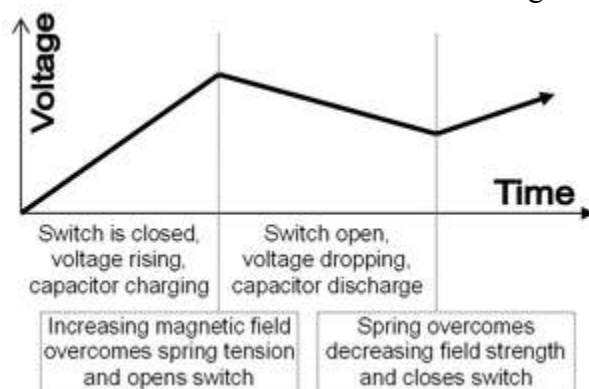
## Electromechanical regulators



□ Circuit design for a simple electromechanical voltage regulator.



□ Interior of an old electromechanical voltage regulator.



□ Graph of voltage output on a time scale.

In older electromechanical regulators, voltage regulation is easily accomplished by coiling the sensing wire to make an electromagnet. The magnetic field produced by the current attracts a moving ferrous core held back under spring tension or gravitational pull. As voltage increases, so

does the current, strengthening the magnetic field produced by the coil and pulling the core towards the field. The magnet is physically connected to a mechanical power switch, which opens as the magnet moves into the field. As voltage decreases, so does the current, releasing spring tension or the weight of the core and causing it to retract. This closes the switch and allows the power to flow once more.

If the mechanical regulator design is sensitive to small voltage fluctuations, the motion of the solenoid core can be used to move a selector switch across a range of resistances or transformer windings to gradually step the output voltage up or down, or to rotate the position of a moving-coil AC regulator.

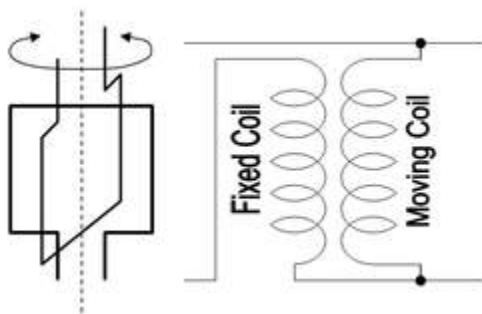
Early automobile generators and alternators had a mechanical voltage regulator using one, two, or three relays and various resistors to stabilize the generator's output at slightly more than 6 or 12 V, independent of the engine's rpm or the varying load on the vehicle's electrical system. Essentially, the relay(s) employed pulse width modulation to regulate the output of the generator, controlling the field current reaching the generator (or alternator) and in this way controlling the output voltage produced.

The regulators used for generators (but not alternators) also disconnect the generator when it was not producing electricity, thereby preventing the battery from discharging back into the generator and attempting to run it as a motor. The rectifier diodes in an alternator automatically perform this function so that a specific relay is not required; this appreciably simplified the regulator design.

More modern designs now use *solid state* technology (transistors) to perform the same function that the relays perform in electromechanical regulators.

Electromechanical regulators are used for mains voltage stabilisation—see Voltage regulator#AC voltage stabilizers below.

### Coil-rotation AC voltage regulator



Basic design principle and circuit diagram for the rotating-coil AC voltage regulator.


This is an older type of regulator used in the 1920s that uses the principle of a fixed-position field coil and a second field coil that can be rotated on an axis in parallel with the fixed coil.

When the movable coil is positioned perpendicular to the fixed coil, the magnetic forces acting on the movable coil balance each other out and voltage output is unchanged. Rotating the coil in one direction or the other away from the center position will increase or decrease voltage in the secondary movable coil.

This type of regulator can be automated via a servo control mechanism to advance the movable coil position in order to provide voltage increase or decrease. A braking mechanism or high ratio gearing is used to hold the rotating coil in place against the powerful magnetic forces acting on the moving coil.

### AC voltage stabilizers



 Magnetic mains regulator

### Electromechanical

Electromechanical regulators, usually called voltage stabilizers, have also been used to regulate the voltage on AC power distribution lines. These regulators operate by using a servomechanism to select the appropriate tap on an autotransformer with multiple taps, or by moving the wiper on a continuously variable autotransformer. If the output voltage is not in the acceptable range, the servomechanism switches connections or moves the wiper to adjust the voltage into the acceptable region. The controls provide a deadband wherein the controller will not act, preventing the controller from constantly adjusting the voltage ("hunting") as it varies by an acceptably small amount.

### Constant-voltage transformer

An alternative method is the use of a type of saturating transformer called a **ferro resonant transformer** or **constant-voltage transformer**. These transformers use a tank circuit composed of a high-voltage resonant winding and a capacitor to produce a nearly constant average output with a varying input. The ferroresonant approach is attractive due to its lack of active components, relying on the square loop saturation characteristics of the tank circuit to absorb variations in average input voltage. Older designs of ferroresonant transformers had an output with high harmonic content, leading to a distorted output waveform. Modern devices are used to

construct a perfect sinewave. The ferroresonant action is a flux limiter rather than a voltage regulator, but with a fixed supply frequency it can maintain an almost constant average output voltage even as the input voltage varies widely.

The ferroresonant transformers, which are also known as Constant Voltage Transformers (CVTs) or ferros, are also good surge suppressors, as they provide high isolation and inherent short-circuit protection.

A ferroresonant transformer can operate with an input voltage range  $\pm 40\%$  or more of the nominal voltage.

Output power factor remains in the range of 0.96 or higher from half to full load.

Because it regenerates an output voltage waveform, output distortion, which is typically less than 4%, is independent of any input voltage distortion, including notching.

Efficiency at full load is typically in the range of 89% to 93%. However, at low loads, efficiency can drop below 60% and no load losses can be as high as 20%. The current-limiting capability also becomes a handicap when a CVT is used in an application with moderate to high inrush current like motors, transformers or magnets. In this case, the CVT has to be sized to accommodate the peak current, thus forcing it to run at low loads and poor efficiency.

Minimum maintenance is required. Transformers and capacitors can be very reliable. Some units have included redundant capacitors to allow several capacitors to fail between inspections without any noticeable effect on the device's performance.

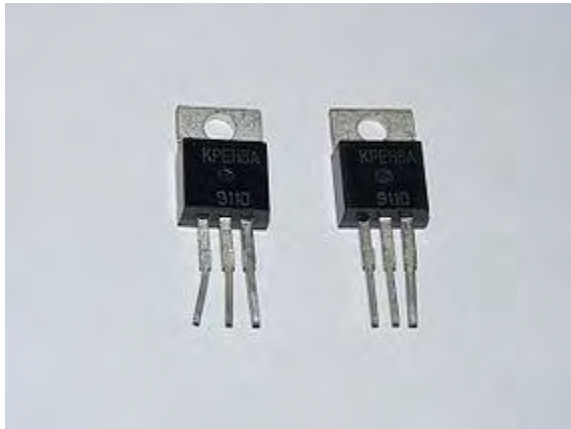
Output voltage varies about 1.2% for every 1% change in supply frequency. For example, a 2-Hz change in generator frequency, which is very large, results in an output voltage change of only 4%, which has little effect for most loads.

It accepts 100% single-phase switch-mode power supply loading without any requirement for derating, including all neutral components.

Input current distortion remains less than 8% THD even when supplying nonlinear loads with more than 100% current THD.

Drawbacks of CVTs (constant voltage transformers) are their larger size, audible humming sound, and high heat generation.

## DC voltage stabilizers



 KPEH8A stabilizers

Many simple DC power supplies regulate the voltage using a *shunt regulator* such as a zener diode, avalanche breakdown diode, or voltage regulator tube. Each of these devices begins conducting at a specified voltage and will conduct as much current as required to hold its terminal voltage to that specified voltage. The power supply is designed to only supply a maximum amount of current that is within the safe operating capability of the shunt regulating device (commonly, by using a series resistor). In shunt regulators, the voltage reference is also the regulating device.

If the stabilizer must provide more power, the shunt regulator output is only used to provide the standard voltage reference for the electronic device, known as the voltage stabilizer. The voltage stabilizer is the electronic device, able to deliver much larger currents on demand.

## Active regulators

Active regulators employ at least one active (amplifying) component such as a transistor or operational amplifier. Shunt regulators are often (but not always) passive and simple, but always inefficient because they (essentially) dump the excess current not needed by the load. When more power must be supplied, more sophisticated circuits are used. In general, these active regulators can be divided into several classes:

- Linear series regulators
- Switching regulators
- SCR regulators

## Linear regulators

Linear regulators are based on devices that operate in their linear region (in contrast, a switching regulator is based on a device forced to act as an on/off switch). In the past, one or more vacuum tubes were commonly used as the variable resistance. Modern designs use one or more transistors instead, perhaps within an Integrated Circuit. Linear designs have the advantage of

very "clean" output with little noise introduced into their DC output, but are most often much less efficient and unable to step-up or invert the input voltage like switched supplies. All linear regulators require a higher input than the output. All linear regulators are subject to the parameter of dropout voltage.

Entire linear regulators are available as integrated circuits. These chips come in either fixed or adjustable voltage types.

### **Switching regulators**

Switching regulators rapidly switch a series device on and off. The duty cycle of the switch sets how much charge is transferred to the load. This is controlled by a similar feedback mechanism as in a linear regulator. Because the series element is either fully conducting, or switched off, it dissipates almost no power; this is what gives the switching design its efficiency. Switching regulators are also able to generate output voltages which are higher than the input, or of opposite polarity — something not possible with a linear design.

Like linear regulators, nearly-complete switching regulators are also available as integrated circuits. Unlike linear regulators, these usually require one external component: an inductor that acts as the energy storage element. (Large-valued inductors tend to be physically large relative to almost all other kinds of componentry, so they are rarely fabricated within integrated circuits and IC regulators — with some exceptions.<sup>[1]</sup>)

### **Comparing linear vs. switching regulators**

The two types of regulators have their different advantages:

- Linear regulators are best when low output noise (and low RFI radiated noise) is required
- Linear regulators are best when a fast response to input and output disturbances is required.
- At low levels of power, linear regulators are cheaper and occupy less printed circuit board space.
- Switching regulators are best when power efficiency is critical (such as in portable computers), *except* linear regulators are more efficient in a small number of cases (such as a 5V microprocessor often in "sleep" mode fed from a 6V battery, *if* the complexity of the switching circuit and the junction capacitance charging current means a high quiescent current in the switching regulator).
- Switching regulators are required when the only power supply is a DC voltage, and a higher output voltage is required.
- At high levels of power (above a few watts), switching regulators are cheaper (for example, the cost of removing heat generated is less).

### **SCR regulators**

Regulators powered from AC power circuits can use silicon controlled rectifiers (SCRs) as the series device. Whenever the output voltage is below the desired value, the SCR is triggered,

allowing electricity to flow into the load until the AC mains voltage passes through zero (ending the half cycle). SCR regulators have the advantages of being both very efficient and very simple, but because they can not terminate an on-going half cycle of conduction, they are not capable of very accurate voltage regulation in response to rapidly-changing loads.

### **Combination (hybrid) regulators**

Many power supplies use more than one regulating method in series. For example, the output from a switching regulator can be further regulated by a linear regulator. The switching regulator accepts a wide range of input voltages and efficiently generates a (somewhat noisy) voltage slightly above the ultimately desired output. That is followed by a linear regulator that generates exactly the desired voltage and eliminates nearly all the noise generated by the switching regulator. Other designs may use an SCR regulator as the "pre-regulator", followed by another type of regulator. An efficient way of creating a variable-voltage, accurate output power supply is to combine a multi-tapped transformer with an adjustable linear post-regulator.

### **Voltage stabilizer**

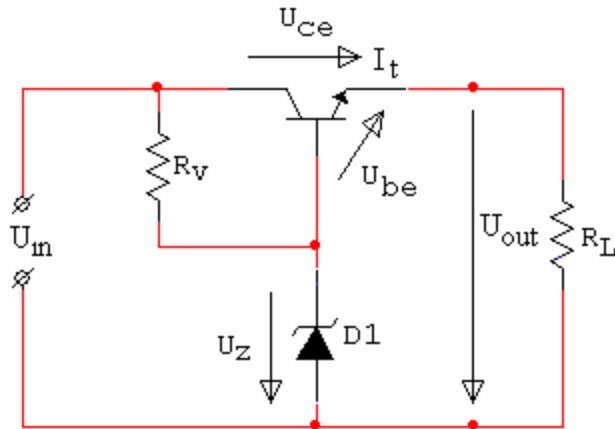
A **voltage stabilizer** is an electronic device able to deliver relatively constant output voltage while input voltage and load current changes over time.

The voltage stabilizer is the *shunt regulator* such as a Zener diode or avalanche diode. Each of these devices begins conducting at a specified voltage and will conduct as much current as required to hold its terminal voltage to that specified voltage. Hence the shunt regulator can be viewed as the limited power parallel stabilizer. The shunt regulator output is used as a voltage reference.

The Zener diode and avalanche diode have opposite threshold voltage dependence on temperature. By connecting these two devices sequentially, it is possible to construct a voltage reference with improved thermal stability. Sometimes (mostly for the voltages around 5.6 V) both effects are combined in the same diode.

### **Simple voltage stabilizer**

In the simplest case emitter follower is used, the base of the regulating transistor is directly connected to the voltage reference:

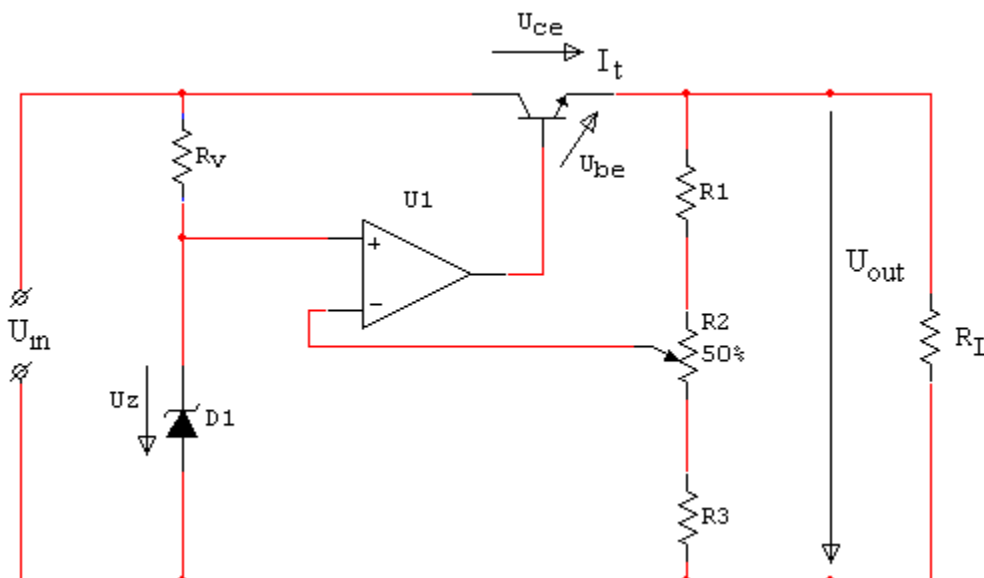


The stabilizer uses the power source, having voltage  $U_{in}$  that may vary over time. It delivers the relatively constant voltage  $U_{out}$ . The output load  $R_L$  can also vary over time. For such a device to work properly, the input voltage must be larger than the output voltage and Voltage drop must not exceed the limits of the transistor used.

The output voltage of the stabilizer is equal to  $U_Z - U_{BE}$  where  $U_{BE}$  is about 0.7V and depends on the load current. If the output voltage drops below that limit, this increases the voltage difference between the base and emitter ( $U_{be}$ ), opening the transistor and delivering more current. Delivering more current through the same output resistor  $R_L$  increases the voltage again.

### Voltage stabilizer with an operational amplifier

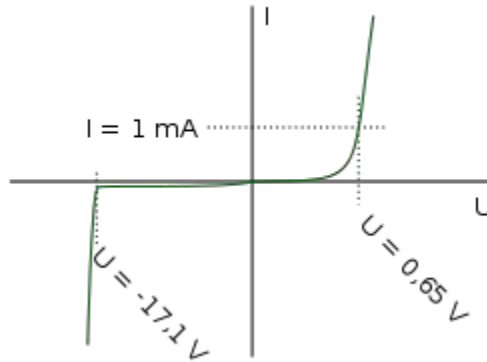
The stability of the output voltage can be significantly increased by using the operational amplifier:



In this case, the operational amplifier opens the transistor more if the voltage at its inverting

input drops significantly below the output of the voltage reference at the non-inverting input. Using the voltage divider (R1, R2 and R3) allows choice of the arbitrary output voltage between  $U_z$  and  $U_{in}$ .

### Zener diode



Current-voltage characteristic of a Zener diode with a breakdown voltage of 17 volt. Notice the change of voltage scale between the forward biased (positive) direction and the reverse biased (negative) direction.

A **Zener diode** is a type of diode that permits current not only in the forward direction like a normal diode, but also in the reverse direction if the voltage is larger than the breakdown voltage known as "Zener knee voltage" or "Zener voltage". The device was named after Clarence Zener, who discovered this electrical property.

A conventional solid-state diode will not allow significant current if it is reverse-biased below its reverse breakdown voltage. When the reverse bias breakdown voltage is exceeded, a conventional diode is subject to high current due to avalanche breakdown. Unless this current is limited by circuitry, the diode will be permanently damaged. In case of large forward bias (current in the direction of the arrow), the diode exhibits a voltage drop due to its junction built-in voltage and internal resistance. The amount of the voltage drop depends on the semiconductor material and the doping concentrations.

A Zener diode exhibits almost the same properties, except the device is specially designed so as to have a greatly reduced breakdown voltage, the so-called Zener voltage. By contrast with the conventional device, a reverse-biased Zener diode will exhibit a controlled breakdown and allow the current to keep the voltage across the Zener diode at the Zener voltage. For example, a diode with a Zener breakdown voltage of 3.2 V will exhibit a voltage drop of 3.2 V if reverse bias voltage applied across it is more than its Zener voltage. The Zener diode is therefore ideal for applications such as the generation of a reference voltage (e.g. for an amplifier stage), or as a voltage stabilizer for low-current applications.

The Zener diode's operation depends on the heavy doping of its p-n junction allowing electrons to tunnel from the valence band of the p-type material to the conduction band of the n-type material. In the atomic scale, this tunneling corresponds to the transport of valence band

electrons into the empty conduction band states; as a result of the reduced barrier between these bands and high electric fields that are induced due to the relatively high levels of dopings on both sides.<sup>[1]</sup> The breakdown voltage can be controlled quite accurately in the doping process. While tolerances within 0.05% are available, the most widely used tolerances are 5% and 10%. Breakdown voltage for commonly available zener diodes can vary widely from 1.2 volts to 200 volts.

Another mechanism that produces a similar effect is the avalanche effect as in the avalanche diode. The two types of diode are in fact constructed the same way and both effects are present in diodes of this type. In silicon diodes up to about 5.6 volts, the Zener effect is the predominant effect and shows a marked negative temperature coefficient. Above 5.6 volts, the avalanche effect becomes predominant and exhibits a positive temperature coefficient<sup>[1]</sup>. In a 5.6 V diode, the two effects occur together and their temperature coefficients neatly cancel each other out, thus the 5.6 V diode is the component of choice in temperature-critical applications. Modern manufacturing techniques have produced devices with voltages lower than 5.6 V with negligible temperature coefficients, but as higher voltage devices are encountered, the temperature coefficient rises dramatically. A 75 V diode has 10 times the coefficient of a 12 V diode.

All such diodes, regardless of breakdown voltage, are usually marketed under the umbrella term of "Zener diode".

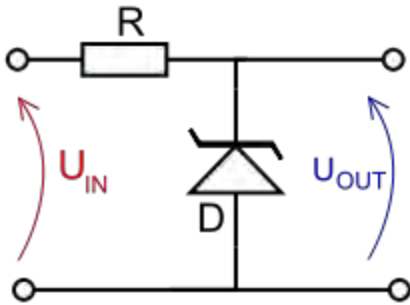
•

## Uses



Zener diode shown with typical packages. *Reverse* current  $-i_Z$  is shown.

Zener diodes are widely used as voltage references and as shunt regulators to regulate the voltage across small circuits. When connected in parallel with a variable voltage source so that it is reverse biased, a Zener diode conducts when the voltage reaches the diode's reverse breakdown voltage. From that point on, the relatively low impedance of the diode keeps the voltage across the diode at that value.



In this circuit, a typical voltage reference or regulator, an input voltage,  $U_{IN}$ , is regulated down to a stable output voltage  $U_{OUT}$ . The intrinsic voltage drop of diode  $D$  is stable over a wide current range and holds  $U_{OUT}$  relatively constant even though the input voltage may fluctuate over a fairly wide range. Because of the low impedance of the diode when operated like this, Resistor  $R$  is used to limit current through the circuit.

In the case of this simple reference, the current flowing in the diode is determined using Ohms law and the known voltage drop across the resistor  $R$ .  $I_{Diode} = (U_{IN} - U_{OUT}) / R_{\Omega}$

The value of  $R$  must satisfy two conditions:

1.  $R$  must be small enough that the current through  $D$  keeps  $D$  in reverse breakdown. The value of this current is given in the data sheet for  $D$ . For example, the common BZX79C5V6<sup>[2]</sup> device, a 5.6 V 0.5 W Zener diode, has a recommended reverse current of 5 mA. If insufficient current exists through  $D$ , then  $U_{OUT}$  will be unregulated, and less than the nominal breakdown voltage (this differs to voltage regulator tubes where the output voltage will be higher than nominal and could rise as high as  $U_{IN}$ ). When calculating  $R$ , allowance must be made for any current through the external load, not shown in this diagram, connected across  $U_{OUT}$ .
2.  $R$  must be large enough that the current through  $D$  does not destroy the device. If the current through  $D$  is  $I_D$ , its breakdown voltage  $V_B$  and its maximum power dissipation  $P_{MAX}$ , then  $I_D V_B < P_{MAX}$ .

A load may be placed across the diode in this reference circuit, and as long as the zener stays in reverse breakdown, the diode will provide a stable voltage source to the load.

A Zener diode used in this way is known as a *shunt voltage regulator* (*shunt*, in this context, meaning connected in parallel, and *voltage regulator* being a class of circuit that produces a stable voltage across any load). In a sense, a portion of the current through the resistor is shunted through the Zener diode, and the rest is through the load. Thus the voltage that the load sees is controlled by causing some fraction of the current from the power source to bypass it—hence the name, by analogy with locomotive switching points.

Shunt regulators are simple, but the requirements that the ballast resistor be small enough to avoid excessive voltage drop during worst-case operation (low input voltage concurrent with high load current) tends to leave a lot of current flowing in the diode much of the time, making

for a fairly wasteful regulator with high quiescent power dissipation, only suitable for smaller loads.

Zener diodes in this configuration are often used as stable references for more advanced voltage regulator circuits.

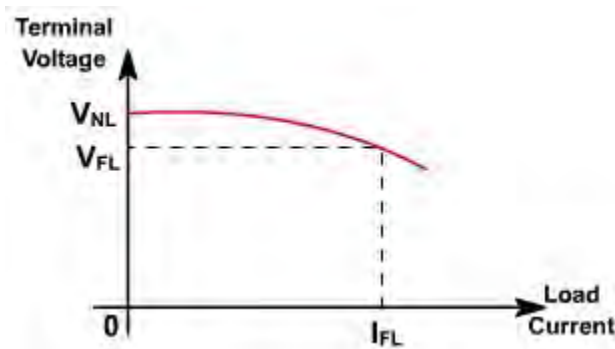
These devices are also encountered, typically in series with a base-emitter junction, in transistor stages where selective choice of a device centered around the avalanche/Zener point can be used to introduce compensating temperature co-efficient balancing of the transistor PN junction. An example of this kind of use would be a DC error amplifier used in a stabilized power supply circuit feedback loop system.

Zener diodes are also used in surge protectors to limit transient voltage spikes.

Another notable application of the zener diode is the use of noise caused by its avalanche breakdown in a random number generator that never repeats.

### Voltage Regulators:

An ideal power supply maintains a constant voltage at its output terminals under all operating conditions. The output voltage of a practical power supply changes with load generally dropping as load current increases as shown in **fig. 1**.



**Fig. 1**

The terminal voltage when full load current is drawn is called full load voltage ( $V_{FL}$ ). The no load voltage is the terminal voltage when zero current is drawn from the supply, that is, the open circuit terminal voltage.

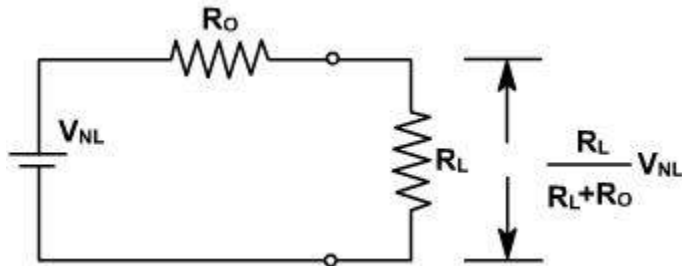
Power supply performance is measured in terms of percent voltage regulation, which indicates its ability to maintain a constant voltage. It is defined as

$$VR = \frac{V_{NL} - V_{FL}}{V_{FL}} \times 100 \%$$

The Thevenin's equivalent of a power supply is shown in **fig. 2**. The Thevenin voltage is the no-

load voltage  $V_{NL}$  and the Thevenin resistance is called the output resistance  $R_o$ . Let the full load current be  $I_{FL}$ . Therefore, the full load resistance  $R_{FL}$  is given by

$$R_{FL} = \frac{V_{FL}}{I_{FL}}$$



**Fig. 2**

From the equivalent circuit, we have

$$V_{FL} = \left( \frac{R_{FL}}{R_{FL} + R_o} \right) V_{NL}$$

and the voltage regulation is given by

$$VR = \frac{V_{NL} - \left( \frac{R_{FL}}{R_{FL} + R_o} \right) V_{NL}}{\left( \frac{R_{FL}}{R_{FL} + R_o} \right) V_{NL}} \times 100 \%$$

$$VR = \frac{R_o}{R_{FL}} \times 100 \%$$

$$\text{or } VR = R_o \left( \frac{I_{FL}}{V_{FL}} \right) \times 100 \%$$

It is clear that the ideal power supply has zero output resistance.

### Example-1

A power supply having output resistance  $1.5\Omega$  supplies a full load current of  $500\text{mA}$  to a  $50\Omega$  load. Determine:

1. percent voltage regulation of the supply
2. no load output voltage.

**Solution:**

(a). Full load output voltage  $V_{FL} = (500\text{mA}) (50\Omega) = 25\text{V}$ .

$$\begin{aligned}VR &= R_o \left( \frac{I_{FL}}{V_{FL}} \right) \times 100 \% \\ &= 1.5 \left( \frac{0.5}{25} \right) \times 100 \% \\ &= 3.0 \%\end{aligned}$$

Therefore,

$$\begin{aligned}V_{NL} &= \left( \frac{R_{FL} + R_o}{R_{FL}} \right) V_{FL} \\ &= \frac{25 (50 + 1.5)}{50}\end{aligned}$$

(b). The no load voltage = 25.75 V

### Voltage Regulators:

An unregulated power supply consists of a transformer (step down), a rectifier and a filter. These power supplies are not good for some applications where constant voltage is required irrespective of external disturbances. The main disturbances are:

1. As the load current varies, the output voltage also varies because of its poor regulation.
2. The dc output voltage varies directly with ac input supply. The input voltage may vary over a wide range thus dc voltage also changes.
3. The dc output voltage varies with the temperature if semiconductor devices are used.

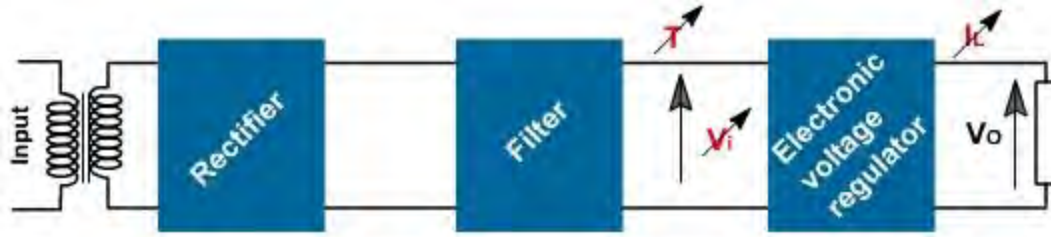
An electronic voltage regulator is essentially a controller used along with unregulated power supply to stabilize the output dc voltage against three major disturbances

- a. Load current ( $I_L$ )
- b. Supply voltage ( $V_i$ )
- c. Temperature ( $T$ )

**Fig. 3**, shows the basic block diagram of voltage regulator. where

$V_i$  = unregulated dc voltage.

$V_o$  = regulated dc voltage.



**Fig. 3**

Since the output dc voltage  $V_{L0}$  depends on the input unregulated dc voltage  $V_i$ , load current  $I_L$  and the temperature  $t$ , then the change  $\Delta V_o$  in output voltage of a power supply can be expressed as follows

$$V_o = V_o(V_i, I_L, T)$$

Take partial derivative of  $V_o$ , we get,

$$\Delta V_o = \frac{\partial V_o}{\partial V_i} \Delta V_i + \frac{\partial V_o}{\partial I_L} \Delta I_L + \frac{\partial V_o}{\partial T} \Delta T$$

or

$$\Delta V_o = S_V \Delta V_i + R_L \Delta I_L + S_T \Delta T$$

$$S_V = \left. \frac{\Delta V_o}{\Delta V_i} \right|_{\substack{\Delta I_L=0 \\ \Delta T=0}}$$

$$R_o = \left. \frac{\Delta V_o}{\Delta I_L} \right|_{\substack{\Delta V_i=0 \\ \Delta T=0}}$$

$$S_T = \left. \frac{\Delta V_o}{\Delta T} \right|_{\substack{\Delta V_i=0 \\ \Delta I_L=0}}$$

$S_V$  gives variation in output voltage only due to unregulated dc voltage.  $R_o$  gives the output voltage variation only due to load current.  $S_T$  gives the variation in output voltage only due to temperature.

The smaller the value of the three coefficients, the better the regulations of power supply. The input voltage variation is either due to input supply fluctuations or presence of ripples due to inadequate filtering.

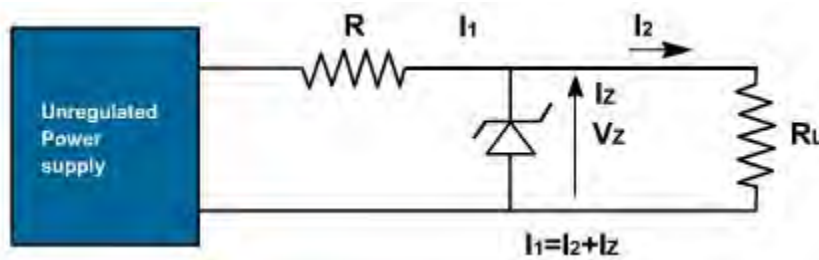
### **Voltage Regulator:**

A voltage regulator is a device designed to maintain the output voltage of power supply nearly constant. It can be regarded as a closed loop system because it monitors the output voltage and generates the control signal to increase or decrease the supply voltage as necessary to compensate for any change in the output voltage. Thus the purpose of voltage regulator is to eliminate any output voltage variation that might occur because of changes in load, changes in

supply voltage or changes in temperature.

### Zener Voltage Regulator:

The regulated power supply may use zener diode as the voltage controlling device as shown in **fig. 4**. The output voltage is determined by the reverse breakdown voltage of the zener diode. This is nearly constant for a wide range of currents. The load voltage can be maintained constant by controlling the current through zener.

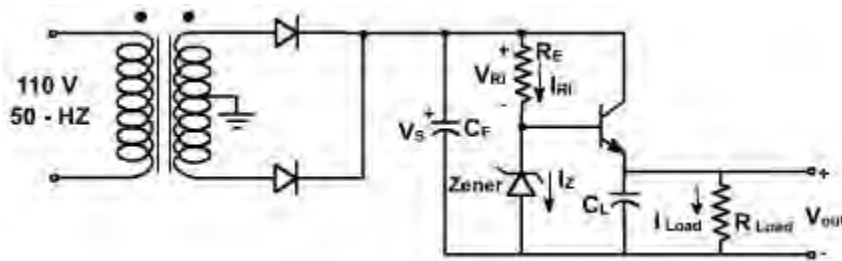


**Fig. 4**

The zener diode regulator has limitations of range. The load current range for which regulation is maintained, is the difference between maximum allowable zener current and minimum current required for the zener to operate in breakdown region. For example, if zener diode requires a minimum current of 10 mA and is limited to a maximum of 1A (to prevent excessive dissipation), the range is  $1 - 0.01 = 0.99A$ . If the load current variation exceeds 0.99A, regulation may be lost.

### Emitter Follower Regulator:

To obtain better voltage regulation in shunt regulator, the zener diode can be connected to the base circuit of a power transistor as shown in **fig. 5**. This amplifies the zener current range. It is also known as emitter follower regulation.



**Fig. 5**

This configuration reduces the current flow in the diode. The power transistor used in this configuration is known as pass transistor. The purpose of  $C_L$  is to ensure that the variations in one of the regulated power supply loads will not be fed to other loads. That is, the capacitor effectively shorts out high-frequency variations.

Because of the current amplifying property of the transistor, the current in the zener diode is small. Hence there is little voltage drop across the diode resistance, and the zener approximates an ideal constant voltage source.

**Operation of the circuit:**

The current through resistor R is the sum of zener current  $I_Z$  and the transistor base current  $I_B (= I_L / \beta)$ .

$$I_L = I_Z + I_B$$

The output voltage across  $R_L$  resistance is given by

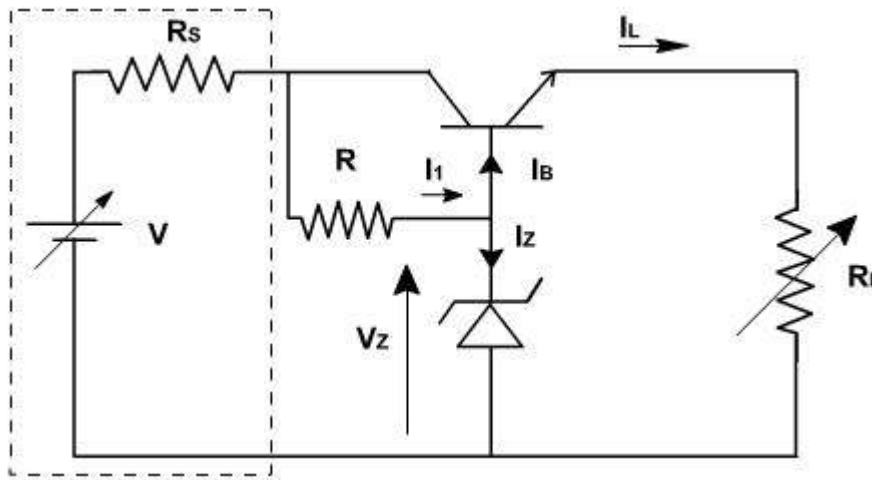
$$V_O = V_Z - V_{BE}$$

Where  $V_{BE} \approx 0.7 \text{ V}$

Therefore,  $V_O = \text{constant}$ .

The emitter current is same as load current. The current  $I_R$  is assumed to be constant for a given supply voltage. Therefore, if  $I_L$  increases, it needs more base currents, to increase base current  $I_Z$  decreases. The difference in this regulator with zener regulator is that in later case the zener current decreases (increases) by same amount by which the load current increases (decreases). Thus the current range is less, while in the shunt regulators, if  $I_L$  increases by  $\Delta I_L$  then  $I_B$  should increase by  $\Delta I_L / \beta$  or  $I_Z$  should decrease by  $\Delta I_L / \beta$ . Therefore the current range control is more for the same rating zener.

The simplified circuit of the shunt regulator is shown in **fig. 6**.



**Fig. 6**

In a power supply the power regulation is basically, because of its high internal impedance. In the circuit discussed, the unregulated supply has resistance  $R_S$  of the order of 100 ohm. The use of emitter follower is to reduce the output resistance and it becomes approximately.

$$R_O = (R_Z + h_{ie}) / (1 + h_{fe})$$

Where  $R_Z$  represents the dynamic zener resistance. The voltage stabilization ratio  $S_V$  is approximately

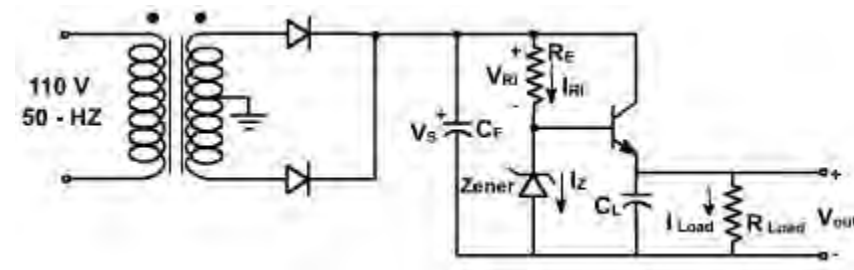
$$S_V = \partial V_o / \partial V_I = R_Z / (R_Z + R)$$

$S_V$  can be improved by increasing  $R$ . This increases  $V_{CE}$  and power dissipated in the transistor. Other disadvantages of the circuit are.

1. No provision for varying the output voltage since it is almost equal to the zener voltage.
2. Change in  $V_{BE}$  and  $V_Z$  due to temperature variations appear at the output since the transistor is connected in series with load, it is called series regulator and transistor is called series pass transistor.

### Design of Series Voltage Regulator:

**Fig. 1** shows the basic circuit of a series voltage regulator. The operation of this regulator has been discussed in previous lecture. It consists of series (pass) transistor to control the output voltage.



**Fig. 1**

The circuit can be designed taking two extreme operating conditions,

1.  $V_{S \max}, I_{Z \max}, I_{load \min} / \beta$
2.  $V_{S \min}, I_{Z \min}, I_{load \max} / \beta$

We calculate  $R_s$  for both conditions and since  $R_{si}$  is constant, we equate these two expressions as in Equation E-1.

$$R_i = \frac{V_{s \max} - V_Z}{I_{Z \max} + I_{\text{load min}} / \beta} = \frac{V_{s \min} - V_Z}{I_{Z \min} + I_{\text{load max}} / \beta} \quad (\text{E-1})$$

A design guideline that set  $I_{Z \min} = 0.1 I_{Z \max}$ . Then we equate the expressions for Equation (E-1) to obtain,

$$(V_{s \max} - V_Z) \left( 0.1 I_{Z \max} + \frac{I_{\text{load max}}}{\beta} \right) = (V_{s \min} - V_Z) \left( I_{Z \max} + \frac{I_{\text{load min}}}{\beta} \right) \quad (\text{E-2})$$

Solving for  $I_{Z \max}$ , we obtain,

$$I_{Z \max} = \frac{I_{\text{load min}} (V_Z - V_{s \min}) + I_{\text{load max}} (V_{s \max} - V_Z)}{\beta (V_{s \min} - 0.1 V_{s \max} - 0.9 V_Z)} \quad (\text{E-3})$$

We estimate the load resistance by taking the ratio of the minimum source voltage to the maximum load current. Since  $R_{\text{load}}$  is large and in parallel, it can be ignored. This is the worst case since it represents the smallest load and therefore the maximum load current.

$$R_{\text{load (equivalent)}} = R_{\text{load (worst case)}} = \frac{V_{s \min}}{I_{\text{load max}}} \quad (\text{E-4})$$

The output filter capacitor size can be estimated according to the permissible output voltage variation and ripple voltage frequency and is given by

$$C_F = \frac{V_{s \max} - V_Z}{\Delta V_{f_p} R_{\text{load (equivalent)}}} \quad (\text{E-5})$$

Since the voltage gain of an EF amplifier is unity, the output voltage of the regulated power supply is,

$$V_{\text{load}} = V_Z - V_{BE} \quad (\text{E-6})$$

The percent regulation of the power supply is given by

$$\begin{aligned} \% \text{ reg} &= \frac{V_{Z \max} - V_{Z \min}}{V_Z} \times 100 \\ &= \frac{R_Z (I_{Z \max} - I_{Z \min})}{V_Z} \times 100 \end{aligned} \quad (\text{E-7})$$

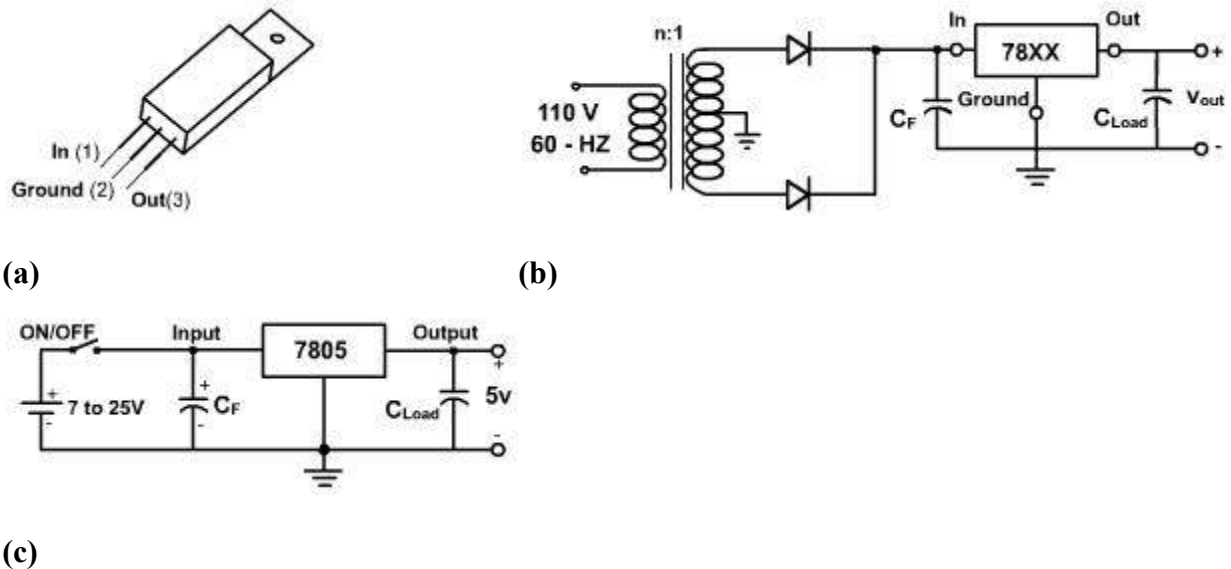
## Voltage Regulator

The maximum power dissipated in this type of series regulator is the power dissipated in the

internal pass transistor, which is approximately  $(V_{S \max} - V_{out}) I_{L \max}$ . Hence, as the load current increases, the power dissipated in the internal pass transistor increases. If  $I_{Load}$  exceeds 0.75 A, the IC package should be secured to a heat sink. When this is done,  $I_{Load}$  can increase to about 1.5 A.

We now focus our attention on the 78XX series of regulators. The last two digits of the IC part number denote the output voltage of the device. Thus, for example, a 7808 IC package produces an 8V regulated output. These packages, although internally complex, are inexpensive and easy to use.

There are a number of different voltages that can be obtained from the 78XX series IC; they are 5, 6, 8, 8.5, 10, 12, 15, 18, and 24 V. In order to design a regulator around one of these ICs, we need only select a transformer, diodes, and filter. The physical configuration is shown in **fig. 3(a)**. The ground lead and the metal tab are connected together. This permits direct attachment to a heat sink for cooling purposes. A typical circuit application is shown in **fig. 3(c)**.



**Fig. 3**

The specification sheet for this IC indicates that there must be a common ground between the input and output, and the minimum voltage at the IC input must be above the regulated output. In order to assure this last condition, it is necessary to filter the output from the rectifier. The  $C_F$  in **fig. 3(b)** performs this filtering when combined with the input resistance to the IC. We use an  $n:1$  step down transformer, with the secondary winding center-tapped, to drive a full-wave rectifier.

The minimum and maximum input voltages for the 78XX family of regulators are shown in **Table-1**.

Type	Min	Max
7805	7	25
7806	8	25
7808	10.5	25
7885	10.5	25
7810	12.5	28
7812	14.5	30
7815	17.5	30
7818	21	33
7824	27	38

**Table - 1**

We use **Table -1** to select the turns ratio,  $n$ , for a 78XX regulator. As a design guide, we will take the average of  $V_{\max}$  and  $V_{\min}$  of the particular IC regulator to calculate  $n$ . For example, using a 7805 regulator, we obtain

$$\frac{V_{\max} + V_{\min}}{2} = \frac{7 + 25}{2} = 16$$

The center tap provides division by 2 so the peak voltage out of the rectifier is  $115 \sqrt{2} / 2n = 16$ . Therefore,  $n = 5$ . This is a conservative method of selecting the transformer ratio.

The filter capacitor,  $C_F$ , is chosen to maintain the voltage input range to the regulator as specified in Table 8.1.

The output capacitor,  $C_{Load}$ , aids in isolating the effect of the transients that may appear on the regulated supply line.  $C_{Load}$  should be a high quality tantalum capacitor with a capacitance of 1.0  $\mu\text{F}$ . It should be connected close to the 78XX regulator using short leads in order to improve the

stability performance.

This family of regulators can also be used for battery powered systems. **Fig. 3(c)** shows a battery powered application. The value of  $C_F$  is chosen in the same manner as for the standard filter.

The 79XX series regulator is identical to the 78XX series except that it provides negative regulated voltages instead of positive.

## UNIVERSITY QUESTIONS

1. What is meant by depletion region?
2. Define the transition capacitance of a diode.
3. Differentiate a PN junction diode and zener diode.
4. Give the diode current equation.
5. Define cut-in voltage.
6. What assumptions are made while analyzing the motion of an electron in an electric field?
7. How do you increase the conductivity of intrinsic semiconductor?
8. What is Hall effect?
9. How do the transition region width and contact potential across a pn junction vary with the applied bias voltage?
10. Mention the two mechanisms of breakdown in a pn junction.
11. State law of mass action.
12. Why is the mobility of electrons greater than the holes?
13. The reverse saturation current of a silicon PN junction diode is  $10 \mu\text{A}$ . Calculate the diode current for the forward-bias voltage of  $0.6 \text{ V}$  at  $25 \text{ C}$ .
14. Give the equation for diode current under reverse bias.
15. Compare LED and LCD.
16. Define the potential-energy barrier.
17. What is meant by Diffusion capacitance?

18. Name any two material used to manufacture LEDs.
19. What is precision rectifier?
20. Find the ripple factor for a FWR with capacitor filter with the output waveform as shown in the figure. Assume  $R_L=100\Omega$  with capacitor  $C=1000\mu F$ .
21. Derive the ripple factor for a HWR and FWR .
22. Why a series resistor is necessary when a diode is forward biased?
23. List any four applications of light emitting diode.

### **BIG QUESTIONS**

1. With the volt-ampere characteristics, explain the working principle of the diode and also explain the static ,dynamic resistance of the diodes? (16)
2. Write a detailed note on:
  - i) diode switching times (6)
  - ii) applications of diodes. (4)
  - iii) capacitance of diodes.(6)
3. Zener diode can be used as a voltage regulator-Justify it.(8)
4. Explain the working of a PN junction diode under various biasing conditions using the relevant circuit sketch. (16)
5. Explain how a PN junction is formed? (8).
6. Write a note on diode capacitance.(8)
7. Write a detailed note on: LED (8)
8. Describe the conduction of current in an intrinsic semiconductor.(8)
9. Find the conductivity and resistivity of an intrinsic semiconductor at temperature of  $300^\circ K$ . It is given that  
 $N_i=2.5 \times 10^{13}/cm^3, \mu_n=3800 cm^2/Sv;$   
 $\mu_p=1800 cm^2/Sv, q=1.6 \times 10^{-19} C$  (8)
10. Derive the continuity equation for a semiconductor. (10)

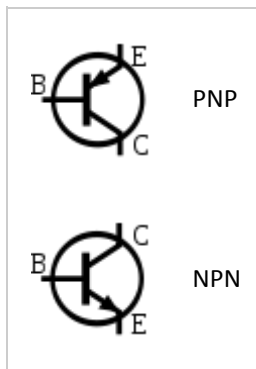
11. In an N-type semiconductor, the Fermi-level lies 0.3 eV below the conduction band at 27°C. If the temperature is increased to 55°C, find the new position of the Fermi level. (6)
12. Give the theory of current components of pn junction diode. (10)
13. Determine the ac resistance for a semiconductor diode having a forward bias of 200 mV and reverse saturation current of  $1\mu\text{A}$  at room temperature. (6)
14. Derive an expression for the diffusion capacitance of a pn junction diode. (8)
15. What is zener effect? Explain the function of a zener diode and draw its characteristics. (10)
16. Write a note on temperature dependence of breakdown voltages. (6)
17. Derive an expression for total current in a semiconductor due to drift and diffusion phenomena.
18. What is meant by carrier life time? Discuss. (6)
19. Describe the action of the PN junction diode under forward and reverse biased conditions and hence draw the volt amp characteristics of the diode. (10)
20. Distinguish between avalanche breakdown and zener breakdown. (8)
21. Explain the switching characteristics of PN junction diode. (8)
22. Derive an expression for the current under forward bias and reverse bias. (10)
23. The diode current is 0.6 mA, when the applied voltage is 400 mV. 20 mA when the applied voltage is 500 mV. Determine  $\eta$ . Assume  $kt/q = 25\text{mV}$ .
24. Explain the effect of temperature on PN junction diodes. (5)
25. Explain the use of zener diode as voltage regulator. (6)
26. With neat diagram explain the operation of LCD.
27. Discuss the V-I characteristics of P-N junction diode and zener diode.
28. Explain the operation of full wave rectifier. Also derive the expression for its average output voltage.
29. Derive the ripple factor for FWR with capacitor filter. (10)
30. Explain in detail about the operation of the following type of filters and derive the ripple factor for all i) C-filter ii) L-filter iii) LC-filter iv) CLC-filter (16)
31. Explain in detail the operation of the electronic voltage regulators.

## UNIT 2

# Bipolar Junction Transistor And Its Applications

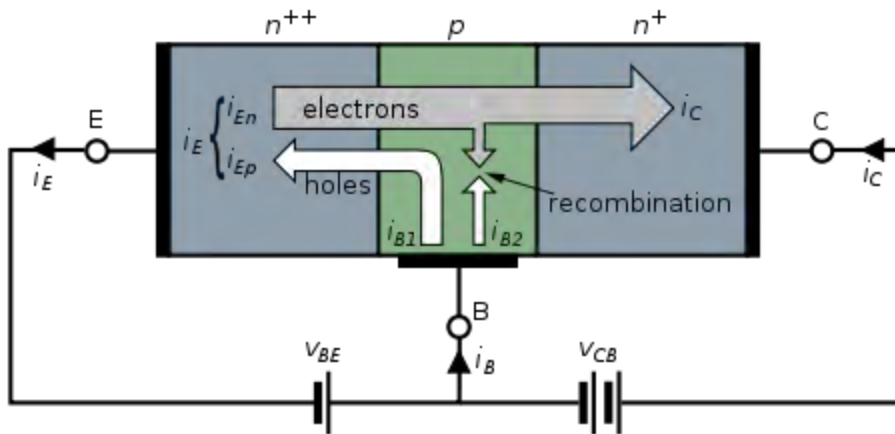
### Introduction

A **bipolar (junction) transistor (BJT)** is a three-terminal electronic device constructed of [doped semiconductor](#) material and may be used in [amplifying](#) or switching applications. *Bipolar* transistors are so named because their operation involves both [electrons](#) and [holes](#). Charge flow in a BJT is due to bidirectional [diffusion](#) of [charge carriers](#) across a junction between two regions of different charge concentrations. This mode of operation is contrasted with *unipolar transistors*, such as [field-effect transistors](#), in which only one carrier type is involved in charge flow due to [drift](#). By design, most of the BJT collector current is due to the flow of charges injected from a high-concentration emitter into the base where they are [minority carriers](#) that diffuse toward the collector, and so BJTs are classified as *minority-carrier* devices.



Schematic symbols for PNP- and NPN-type BJTs.

-



### NPN BJT with forward-biased E–B junction and reverse-biased B–C junction

An NPN transistor can be considered as two [diodes](#) with a shared [anode](#). In typical operation, the base–emitter [junction](#) is [forward biased](#) and the base–collector junction is [reverse biased](#). In an NPN transistor, for example, when a positive voltage is applied to the base–emitter junction, the equilibrium between thermally generated [carriers](#) and the repelling electric field of the [depletion region](#) becomes unbalanced, allowing thermally excited electrons to inject into the base region. These electrons wander (or "[diffuse](#)") through the base from the region of high concentration near the emitter towards the region of low concentration near the collector. The electrons in the base are called [minority carriers](#) because the base is [doped](#) p-type which would make holes the [majority carrier](#) in the base.

To minimize the percentage of carriers that [recombine](#) before reaching the collector–base junction, the transistor's base region must be thin enough that carriers can diffuse across it in much less time than the semiconductor's minority carrier lifetime. In particular, the thickness of the base must be much less than the [diffusion length](#) of the electrons. The collector–base junction is reverse-biased, and so little electron injection occurs from the collector to the base, but electrons that diffuse through the base towards the collector are swept into the collector by the electric field in the depletion region of the collector–base junction. The thin *shared* base and asymmetric collector–emitter doping is what differentiates a bipolar transistor from two *separate* and oppositely biased diodes connected in series.

### Voltage, current, and charge control

The collector–emitter current can be viewed as being controlled by the base–emitter current (current control), or by the base–emitter voltage (voltage control). These views are related by the current–voltage relation of the base–emitter junction, which is just the usual exponential current–voltage curve of a [p-n junction](#) (diode).<sup>[1]</sup>

The physical explanation for collector current is the amount of minority-carrier charge in the base region.<sup>[1][2][3]</sup> Detailed models of transistor action, such as the [Gummel–Poon model](#),

account for the distribution of this charge explicitly to explain transistor behavior more exactly.<sup>[4]</sup> The charge-control view easily handles [phototransistors](#), where minority carriers in the base region are created by the absorption of [photons](#), and handles the dynamics of turn-off, or recovery time, which depends on charge in the base region recombining. However, because base charge is not a signal that is visible at the terminals, the current- and voltage-control views are generally used in circuit design and analysis.

In [analog circuit](#) design, the current-control view is sometimes used because it is approximately linear. That is, the collector current is approximately  $\beta_F$  times the base current. Some basic circuits can be designed by assuming that the emitter–base voltage is approximately constant, and that collector current is beta times the base current. However, to accurately and reliably design production BJT circuits, the voltage-control (for example, [Ebers–Moll](#)) model is required<sup>[1]</sup>. The voltage-control model requires an exponential function to be taken into account, but when it is linearized such that the transistor can be modelled as a transconductance, as in the [Ebers–Moll model](#), design for circuits such as differential amplifiers again becomes a mostly linear problem, so the voltage-control view is often preferred. For [translinear circuits](#), in which the exponential I–V curve is key to the operation, the transistors are usually modelled as voltage controlled with [transconductance](#) proportional to collector current. In general, transistor level circuit design is performed using [SPICE](#) or a comparable analogue circuit simulator, so model complexity is usually not of much concern to the designer.

## Turn-on, turn-off, and storage delay

The Bipolar transistor exhibits a few delay characteristics when turning on and off. Most transistors, and especially power transistors, exhibit long base storage time that limits maximum frequency of operation in switching applications. One method for reducing this storage time is by using a [Baker clamp](#).

## Transistor 'alpha' and 'beta'

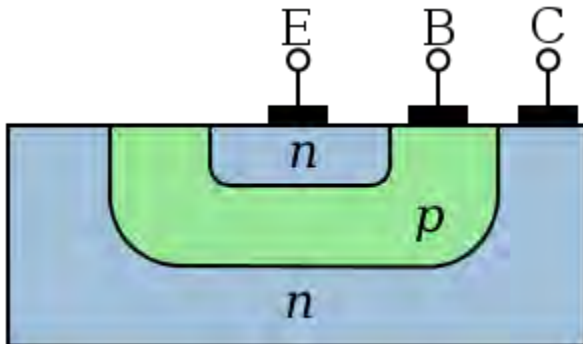
The proportion of electrons able to cross the base and reach the collector is a measure of the BJT efficiency. The heavy doping of the emitter region and light doping of the base region cause many more electrons to be injected from the emitter into the base than holes to be injected from the base into the emitter. The [common-emitter current gain](#) is represented by  $\beta_F$  or  $h_{fe}$ ; it is approximately the ratio of the DC collector current to the DC base current in forward-active region. It is typically greater than 100 for small-signal transistors but can be smaller in transistors designed for high-power applications. Another important parameter is the [common-base current gain](#),  $\alpha_F$ . The common-base current gain is approximately the gain of current from emitter to collector in the forward-active region. This ratio usually has a value close to unity; between 0.98 and 0.998. Alpha and beta are more precisely related by the following identities (NPN transistor):

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

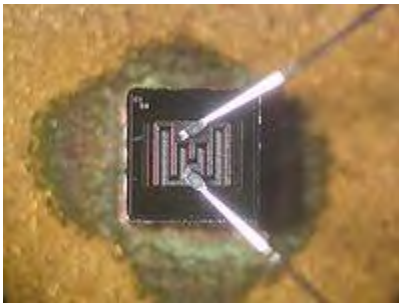
$$\beta_F = \frac{I_C}{I_B}$$

$$\beta_F = \frac{\alpha_F}{1 - \alpha_F} \iff \alpha_F = \frac{\beta_F}{\beta_F + 1}$$

## Structure



Simplified cross section of a planar *NPN* bipolar junction transistor



Die of a KSY34 high-frequency *NPN* transistor, base and emitter connected via bonded wires

A BJT consists of three differently [doped](#) semiconductor regions, the *emitter* region, the *base* region and the *collector* region. These regions are, respectively, *p* type, *n* type and *p* type in a [PNP](#), and *n* type, *p* type and *n* type in a [NPN transistor](#). Each semiconductor region is connected to a terminal, appropriately labeled: *emitter* (E), *base* (B) and *collector* (C).

The *base* is physically located between the *emitter* and the *collector* and is made from lightly doped, high resistivity material. The collector surrounds the emitter region, making it almost impossible for the electrons injected into the base region to escape being collected, thus making the resulting value of  $\alpha$  very close to unity, and so, giving the transistor a large  $\beta$ . A cross section view of a BJT indicates that the collector–base junction has a much larger area than the emitter–base junction.

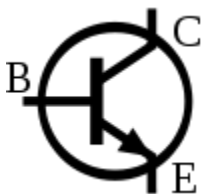
The bipolar junction transistor, unlike other transistors, is usually not a symmetrical device. This means that interchanging the collector and the emitter makes the transistor leave the forward active mode and start to operate in reverse mode. Because the transistor's internal structure is usually optimized for forward-mode operation, interchanging the collector and the emitter makes the values of  $\alpha$  and  $\beta$  in reverse operation much smaller than those in forward operation; often the  $\alpha$  of the reverse mode is lower than 0.5. The lack of symmetry is primarily due to the doping ratios of the emitter and the collector. The emitter is heavily doped, while the collector is lightly doped, allowing a large reverse bias voltage to be applied before the collector–base junction breaks down. The collector–base junction is reverse biased in normal operation. The reason the emitter is heavily doped is to increase the emitter injection efficiency: the ratio of carriers injected by the emitter to those injected by the base. For high current gain, most of the carriers injected into the emitter–base junction must come from the emitter.

The low-performance "lateral" bipolar transistors sometimes used in [CMOS](#) processes are sometimes designed symmetrically, that is, with no difference between forward and backward operation.

Small changes in the voltage applied across the base–emitter terminals causes the current that flows between the *emitter* and the *collector* to change significantly. This effect can be used to amplify the input voltage or current. BJTs can be thought of as voltage-controlled [current sources](#), but are more simply characterized as current-controlled current sources, or current amplifiers, due to the low impedance at the base.

Early transistors were made from [germanium](#) but most modern BJTs are made from [silicon](#). A significant minority are also now made from [gallium arsenide](#), especially for very high speed applications (see HBT, below).

## NPN



The symbol of an NPN bipolar junction transistor

NPN is one of the two types of bipolar transistors, consisting of a layer of P-[doped](#) semiconductor (the "base") between two N-doped layers. A small current entering the base is amplified in the collector output. That is, an NPN transistor is "on" when its base is pulled **high** relative to the emitter.

Most of the NPN current is carried by electrons, moving from emitter to collector as [minority carriers](#) in the P-type base region. Most bipolar transistors used today are NPN, because [electron mobility](#) is higher than [hole mobility](#) in semiconductors, allowing greater currents and faster operation.

## PNP



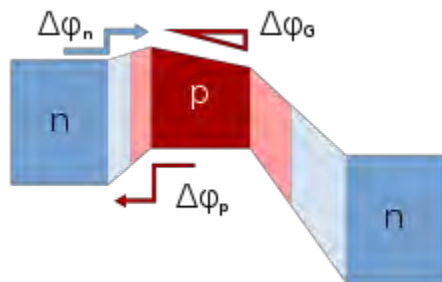
The symbol of a PNP Bipolar Junction Transistor.

The other type of BJT is the PNP, consisting of a layer of N-[doped](#) semiconductor between two layers of P-doped material. A small current leaving the base is amplified in the collector output. That is, a PNP transistor is "on" when its base is pulled **low** relative to the emitter.

The arrows in the NPN and PNP transistor symbols are on the emitter legs and point in the direction of the [conventional current](#) flow when the device is in forward active mode.

A [mnemonic](#) device for the NPN / PNP distinction, based on the arrows in their symbols and the letters in their names, is *not pointing in* for NPN and *pointing in* for PNP.<sup>[5]</sup>

## Heterojunction bipolar transistor



Bands in graded heterojunction NPN bipolar transistor. Barriers indicated for electrons to move from emitter to base, and for holes to be injected backward from base to emitter; Also, grading of bandgap in base assists electron transport in base region; Light colors indicate [depleted regions](#)

The [heterojunction bipolar transistor](#) (HBT) is an improvement of the BJT that can handle signals of very high frequencies up to several hundred [GHz](#). It is common in modern ultrafast circuits, mostly [RF](#) systems.<sup>[6][7]</sup> Heterojunction transistors have different semiconductors for the elements of the transistor. Usually the emitter is composed of a larger bandgap material than the base. The figure shows that this difference in bandgap allows the barrier for holes to inject

backward into the base, denoted in figure as  $\Delta\phi_p$ , to be made large, while the barrier for electrons to inject into the base  $\Delta\phi_n$  is made low. This barrier arrangement helps reduce minority carrier injection from the base when the emitter-base junction is under forward bias, and thus reduces base current and increases emitter injection efficiency.

The improved injection of carriers into the base allows the base to have a higher doping level, resulting in lower resistance to access the base electrode. In the more traditional BJT, also referred to as homojunction BJT, the efficiency of carrier injection from the emitter to the base is primarily determined by the doping ratio between the emitter and base, which means the base must be lightly doped to obtain high injection efficiency, making its resistance relatively high. In addition, higher doping in the base can improve figures of merit like the [Early voltage](#) by lessening base narrowing.

The grading of composition in the base, for example, by progressively increasing the amount of germanium in a [SiGe](#) transistor, causes a gradient in bandgap in the neutral base, denoted in the figure by  $\Delta\phi_G$ , providing a "built-in" field that assists electron transport across the base. That [drift component](#) of transport aids the normal diffusive transport, increasing the frequency response of the transistor by shortening the transit time across the base.

Two commonly used HBTs are silicon–germanium and aluminum gallium arsenide, though a wide variety of semiconductors may be used for the HBT structure. HBT structures are usually grown by [epitaxy](#) techniques like [MOCVD](#) and [MBE](#).

## Regions of operation

Applied voltages	B-E Junction Bias (NPN)	B-C Junction Bias (NPN)	Mode (NPN)
$E < B < C$	Forward	Reverse	Forward active
$E < B > C$	Forward	Forward	Saturation
$E > B < C$	Reverse	Reverse	Cut-off
$E > B > C$	Reverse	Forward	Reverse-active

Bipolar transistors have five distinct regions of operation, defined by BJT junction biases.

The modes of operation can be described in terms of the applied voltages (this description applies to NPN transistors; polarities are reversed for PNP transistors):

Forward active: base higher than emitter, collector higher than base (in this mode the collector current is proportional to base current by  $\beta_F$ ).

- Saturation: base higher than emitter, but collector is not higher than base.
- Cut-Off: base lower than emitter, but collector is higher than base. It means the transistor is not letting conventional current to go through collector to emitter.
- Reverse-action: base lower than emitter, collector lower than base: reverse conventional current goes through transistor.

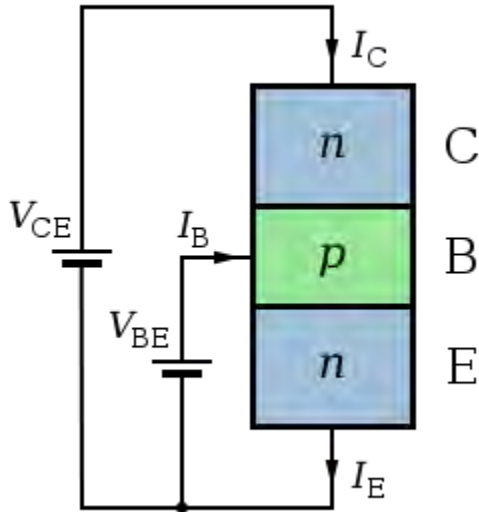
Applied voltages	B-E Junction Bias (PNP)	B-C Junction Bias (PNP)	Mode (PNP)
$E < B < C$	Reverse	Forward	Reverse-active
$E < B > C$	Reverse	Reverse	Cut-off
$E > B < C$	Forward	Forward	Saturation
$E > B > C$	Forward	Reverse	Forward active

In terms of junction biasing: ('reverse biased base–collector junction' means  $V_{bc} < 0$  for NPN, opposite for PNP)

- **Forward-active** (or simply, **active**): The base–emitter junction is forward biased and the base–collector junction is reverse biased. Most bipolar transistors are designed to afford the greatest common-emitter current gain,  $\beta_F$ , in forward-active mode. If this is the case, the collector–emitter current is approximately [proportional](#) to the base current, but many times larger, for small base current variations.
- **Reverse-active** (or **inverse-active** or **inverted**): By reversing the biasing conditions of the forward-active region, a bipolar transistor goes into reverse-active mode. In this mode, the emitter and collector regions switch roles. Because most BJTs are designed to maximize current gain in forward-active mode, the  $\beta_F$  in inverted mode is several (2–3 for the ordinary germanium transistor) times smaller. This transistor mode is seldom used, usually being considered only for failsafe conditions and some types of bipolar logic. The reverse bias breakdown voltage to the base may be an order of magnitude lower in this region.
- **Saturation**: With both junctions forward-biased, a BJT is in saturation mode and facilitates high current conduction from the emitter to the collector. This mode corresponds to a logical "on", or a closed switch.
- **Cut-off**: In cut-off, biasing conditions opposite of saturation (both junctions reverse biased) are present. There is very little current, which corresponds to a logical "off", or an open switch.
- [Avalanche breakdown region](#)

Although these regions are well defined for sufficiently large applied voltage, they overlap somewhat for small (less than a few hundred millivolts) biases. For example, in the typical grounded-emitter configuration of an NPN BJT used as a pulldown switch in digital logic, the "off" state never involves a reverse-biased junction because the base voltage never goes below ground; nevertheless the forward bias is close enough to zero that essentially no current flows, so this end of the forward active region can be regarded as the cutoff region.

## Active-mode NPN transistors in circuits



Structure and use of NPN transistor. Arrow according to schematic.

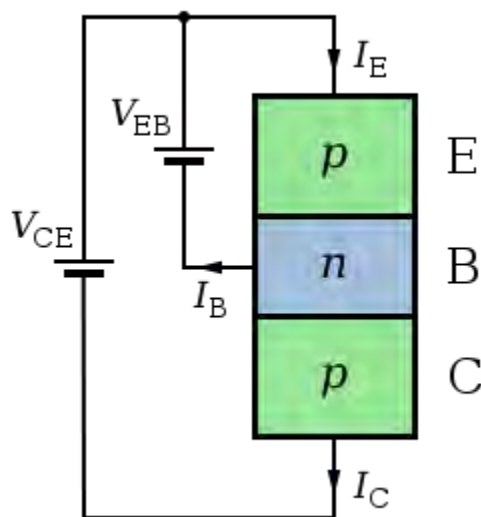
The diagram opposite is a schematic representation of an NPN transistor connected to two voltage sources. To make the transistor conduct appreciable current (on the order of 1 mA) from C to E,  $V_{BE}$  must be above a minimum value sometimes referred to as the **cut-in voltage**. The cut-in voltage is usually about 600 mV for silicon BJTs at [room temperature](#) but can be different depending on the type of transistor and its [biasing](#). This applied voltage causes the lower P-N junction to 'turn-on' allowing a flow of electrons from the emitter into the base. In active mode, the electric field existing between base and collector (caused by  $V_{CE}$ ) will cause the majority of these electrons to cross the upper P-N junction into the collector to form the collector current  $I_C$ . The remainder of the electrons recombine with holes, the majority carriers in the base, making a current through the base connection to form the base current,  $I_B$ . As shown in the diagram, the emitter current,  $I_E$ , is the total transistor current, which is the sum of the other terminal currents (i.e.,  $I_E = I_B + I_C$ ).

In the diagram, the arrows representing current point in the direction of [conventional current](#) – the flow of electrons is in the opposite direction of the arrows because electrons carry negative [electric charge](#). In active mode, the ratio of the collector current to the base current is called the *DC current gain*. This gain is usually 100 or more, but robust circuit designs do not depend on

the exact value (for example see [op-amp](#)). The value of this gain for DC signals is referred to as  $h_{FE}$ , and the value of this gain for AC signals is referred to as  $h_{fe}$ . However, when there is no particular frequency range of interest, the symbol  $\beta$  is used<sup>[[citation needed](#)]</sup>.

It should also be noted that the emitter current is related to  $V_{BE}$  exponentially. At [room temperature](#), an increase in  $V_{BE}$  by approximately 60 mV increases the emitter current by a factor of 10. Because the base current is approximately proportional to the collector and emitter currents, they vary in the same way.

### Active-mode PNP transistors in circuits



Structure and use of PNP transistor.

The diagram opposite is a schematic representation of a PNP transistor connected to two voltage sources. To make the transistor conduct appreciable current (on the order of 1 mA) from E to C,  $V_{EB}$  must be above a minimum value sometimes referred to as the **cut-in voltage**. The cut-in voltage is usually about 600 mV for silicon BJTs at [room temperature](#) but can be different depending on the type of transistor and its [biasing](#). This applied voltage causes the upper P-N junction to 'turn-on' allowing a flow of [holes](#) from the emitter into the base. In active mode, the electric field existing between the emitter and the collector (caused by  $V_{CE}$ ) causes the majority of these holes to cross the lower P-N junction into the collector to form the collector current  $I_C$ . The remainder of the holes recombine with electrons, the majority carriers in the base, making a current through the base connection to form the base current,  $I_B$ . As shown in the diagram, the emitter current,  $I_E$ , is the total transistor current, which is the sum of the other terminal currents (i.e.,  $I_E = I_B + I_C$ ).

In the diagram, the arrows representing current point in the direction of [conventional current](#) –

the flow of holes is in the same direction of the arrows because holes carry positive [electric charge](#). In active mode, the ratio of the collector current to the base current is called the *DC current gain*. This gain is usually 100 or more, but robust circuit designs do not depend on the exact value. The value of this gain for DC signals is referred to as  $h_{FE}$ , and the value of this gain for AC signals is referred to as  $h_{fe}$ . However, when there is no particular frequency range of interest, the symbol  $\beta$  is used<sup>[[citation needed](#)]</sup>.

It should also be noted that the emitter current is related to  $V_{EB}$  exponentially. At [room temperature](#), an increase in  $V_{EB}$  by approximately 60 mV increases the emitter current by a factor of 10. Because the base current is approximately proportional to the collector and emitter currents, they vary in the same way.

## History

The bipolar point-contact transistor was [invented in December 1947](#) at the [Bell Telephone Laboratories](#) by [John Bardeen](#) and [Walter Brattain](#) under the direction of [William Shockley](#). The junction version known as the bipolar junction transistor, [invented by Shockley in 1948](#), enjoyed three decades as the device of choice in the design of discrete and [integrated circuits](#). Nowadays, the use of the BJT has declined in favor of [CMOS](#) technology in the design of digital integrated circuits.

## Germanium transistors

The [germanium](#) transistor was more common in the 1950s and 1960s, and while it exhibits a lower "cut off" voltage, typically around 0.2 V, making it more suitable for some applications, it also has a greater tendency to exhibit [thermal runaway](#).

## Early manufacturing techniques

Various methods of manufacturing bipolar junction transistors were developed<sup>[8]</sup>.

- [Point-contact transistor](#) – first type to demonstrate transistor action, limited commercial use due to high cost and noise.
- [Grown junction transistor](#) – first type of bipolar junction transistor made<sup>[9]</sup>. Invented by [William Shockley](#) at [Bell Labs](#). Invented on June 23, 1948<sup>[10]</sup>. Patent filed on June 26, 1948.
- [Alloy junction transistor](#) – emitter and collector alloy beads fused to base. Developed at [General Electric](#) and [RCA](#)<sup>[11]</sup> in 1951.
  - [Micro alloy transistor](#) – high speed type of alloy junction transistor. Developed at [Philco](#)<sup>[12]</sup>.
  - [Micro alloy diffused transistor](#) – high speed type of alloy junction transistor. Developed at [Philco](#).
  - [Post alloy diffused transistor](#) – high speed type of alloy junction transistor. Developed at [Philips](#).
- [Tetrode transistor](#) – high speed variant of grown junction transistor<sup>[13]</sup> or alloy junction

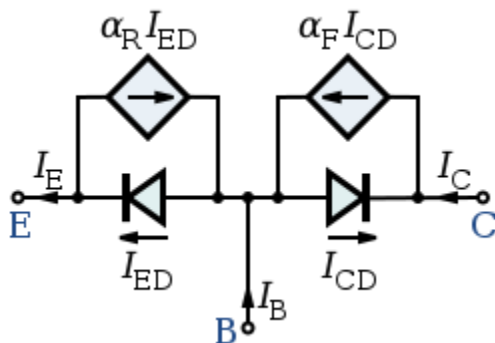
- transistor<sup>[14]</sup> with two connections to base.
- [Surface barrier transistor](#) – high speed metal barrier junction transistor. Developed at [Philco](#)<sup>[15]</sup> in 1953<sup>[16]</sup>.
  - [Drift-field transistor](#) – high speed bipolar junction transistor. Invented by [Herbert Kroemer](#)<sup>[17][18]</sup> at the Central Bureau of Telecommunications Technology of the German Postal Service, in 1953.
  - [Diffusion transistor](#) – modern type bipolar junction transistor. Prototypes<sup>[19]</sup> developed at Bell Labs in 1954.
    - [Diffused base transistor](#) – first implementation of diffusion transistor.
    - [Mesa transistor](#) – Developed at [Texas Instruments](#) in 1957.
    - [Planar transistor](#) – the bipolar junction transistor that made mass produced monolithic [integrated circuits](#) possible. Developed by Dr. [Jean Hoerni](#)<sup>[20]</sup> at [Fairchild](#) in 1959.
  - Epitaxial transistor – a bipolar junction transistor made using vapor phase deposition. See [epitaxy](#). Allows very precise control of doping levels and gradients.

## Theory and modeling

In the discussion below, focus is on the NPN bipolar transistor. In the NPN transistor in what is called **active mode** the base-emitter voltage  $V_{BE}$  and collector-base voltage  $V_{CB}$  are positive, forward biasing the emitter-base junction and reverse-biasing the collector-base junction. In active mode of operation, electrons are injected from the forward biased n-type emitter region into the p-type base where they diffuse to the reverse biased n-type collector and are swept away by the electric field in the reverse biased collector-base junction. For a figure describing forward and reverse bias, see the end of the article [semiconductor diodes](#).

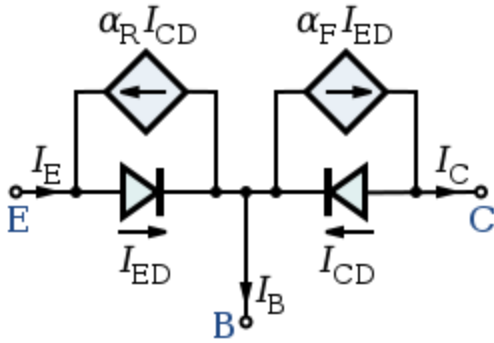
## Large-signal models

### Ebers–Moll model



Ebers–Moll Model for an NPN transistor.<sup>[21]</sup>

- $I_B, I_C, I_E$ : base, collector and emitter currents
- $I_{CD}, I_{ED}$ : collector and emitter diode currents
- $\alpha_F, \alpha_R$ : forward and reverse common-base current gains



Ebers–Moll Model for a PNP transistor.

The DC emitter and collector currents in active mode are well modeled by an approximation to the Ebers–Moll model:

$$I_E = I_{ES} \left( e^{\frac{V_{BE}}{V_T}} - 1 \right)$$

$$I_C = \alpha_T I_{ES} \left( e^{\frac{V_{BE}}{V_T}} - 1 \right)$$

The base internal current is mainly by diffusion (see [Fick's law](#)) and

$$J_n(\text{base}) = \frac{qD_n n_{bo}}{W} e^{\frac{V_{BE}}{V_T}}$$

where

- $V_T$  is the [thermal voltage](#)  $kT / q$  (approximately 26 mV at 300 K  $\approx$  room temperature).
- $I_E$  is the emitter current
- $I_C$  is the collector current
- $\alpha_T$  is the common base forward short circuit current gain (0.98 to 0.998)
- $I_{ES}$  is the reverse saturation current of the base–emitter diode (on the order of  $10^{-15}$  to  $10^{-12}$  amperes)
- $V_{BE}$  is the base–emitter voltage
- $D_n$  is the diffusion constant for electrons in the p-type base
- $W$  is the base width

The  $\alpha$  and forward  $\beta$  parameters are as described previously. A reverse  $\beta$  is sometimes included in the model.

The unapproximated Ebers–Moll equations used to describe the three currents in any operating region are given below. These equations are based on the transport model for a bipolar junction transistor.<sup>[22]</sup>

$$i_C = I_S \left( e^{\frac{V_{BE}}{V_T}} - e^{\frac{V_{BC}}{V_T}} \right) - \frac{I_S}{\beta_R} \left( e^{\frac{V_{BC}}{V_T}} - 1 \right)$$

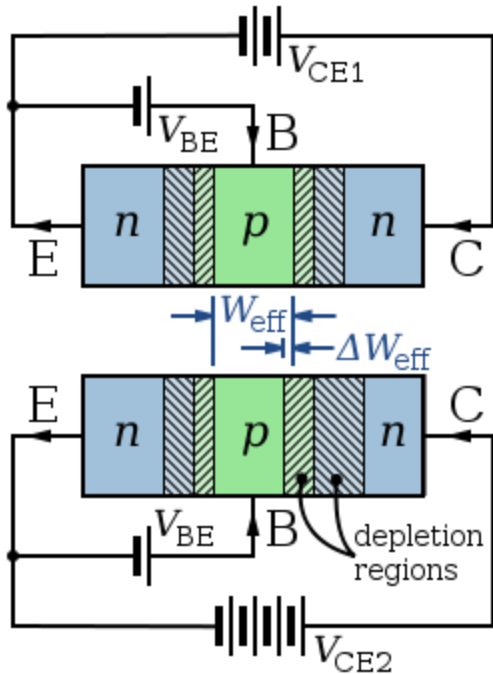
$$i_B = \frac{I_S}{\beta_F} \left( e^{\frac{V_{BE}}{V_T}} - 1 \right) + \frac{I_S}{\beta_R} \left( e^{\frac{V_{BC}}{V_T}} - 1 \right)$$

$$i_E = I_S \left( e^{\frac{V_{BE}}{V_T}} - e^{\frac{V_{BC}}{V_T}} \right) + \frac{I_S}{\beta_F} \left( e^{\frac{V_{BE}}{V_T}} - 1 \right)$$

where

- $i_C$  is the collector current
- $i_B$  is the base current
- $i_E$  is the emitter current
- $\beta_F$  is the forward common emitter current gain (20 to 500)
- $\beta_R$  is the reverse common emitter current gain (0 to 20)
- $I_S$  is the reverse saturation current (on the order of  $10^{-15}$  to  $10^{-12}$  amperes)
- $V_T$  is the [thermal voltage](#) (approximately 26 mV at 300 K  $\approx$  room temperature).
- $V_{BE}$  is the base–emitter voltage
- $V_{BC}$  is the base–collector voltage

#### Base-width modulation



Top: NPN base width for low collector-base reverse bias; Bottom: narrower NPN base width for large collector-base reverse bias. Hashed regions are [depleted regions](#).

Main article: [Early Effect](#)

As the applied collector–base voltage ( $V_{CB} = V_{CE} - V_{BE}$ ) varies, the collector–base depletion region varies in size. An increase in the collector–base voltage, for example, causes a greater reverse bias across the collector–base junction, increasing the collector–base depletion region width, and decreasing the width of the base. This variation in base width often is called the "[Early effect](#)" after its discoverer [James M. Early](#).

Narrowing of the base width has two consequences:

- There is a lesser chance for recombination within the "smaller" base region.
- The charge gradient is increased across the base, and consequently, the current of minority carriers injected across the emitter junction increases.

Both factors increase the collector or "output" current of the transistor in response to an increase in the collector–base voltage.

In the [forward-active region](#), the Early effect modifies the collector current ( $i_C$ ) and the forward common emitter current gain ( $\beta_F$ ) as given by: <sup>[[citation needed](#)]</sup>

$$i_C = I_S e^{\frac{v_{BE}}{V_T}} \left( 1 + \frac{V_{CB}}{V_A} \right)$$

$$\beta_F = \beta_{F0} \left( 1 + \frac{V_{CB}}{V_A} \right)$$

$$r_o = \frac{V_A}{I_C}$$

where:

- $V_{CB}$  is the collector–base voltage
- $V_A$  is the Early voltage (15 V to 150 V)
- $\beta_{F0}$  is forward common-emitter current gain when  $V_{CB} = 0$  V
- $r_o$  is the output impedance
- $I_C$  is the collector current

### Current–voltage characteristics

The following assumptions are involved when deriving ideal current-voltage characteristics of the BJT

- Low level injection
- Uniform doping in each region with abrupt junctions
- One-dimensional current
- Negligible recombination-generation in space charge regions
- Negligible electric fields outside of space charge regions.

It is important to characterize the minority diffusion currents induced by injection of carriers.

With regard to pn-junction diode, a key relation is the diffusion equation.

$$\frac{d^2 \Delta p_B(x)}{dx^2} = \frac{\Delta p_B(x)}{L_B^2}$$

A solution of this equation is below, and two boundary conditions are used to solve and find  $C_1$  and  $C_2$ .

$$\Delta p_B(x) = C_1 e^{x/L_B} + C_2 e^{-x/L_B}$$

The following equations apply to the emitter and collector region, respectively, and the origins  $0$ ,  $0'$ , and  $0''$  apply to the base, collector, and emitter.

$$\Delta n_B(x'') = A_1 e^{x''/L_B} + A_2 e^{-x''/L_B}$$

$$\Delta n_c(x') = B_1 e^{x'/L_B} + B_2 e^{-x'/L_B}$$

A boundary condition of the emitter is below:

$$\Delta n_E(0'') = n_{E0}(\exp(qV_{EB}/kT) - 1)$$

The values of the constants  $A_1$  and  $B_1$  are zero due to the following conditions of the emitter and collector regions as  $x'' \rightarrow 0$  and  $x' \rightarrow 0$ .

$$\Delta n_E(x'') \rightarrow 0$$

$$\Delta n_c(x') \rightarrow 0$$

Because  $A_1 = B_1 = 0$ , the values of  $\Delta n_E(0'')$  and  $\Delta n_c(0')$  are  $A_2$  and  $B_2$ , respectively.

$$\Delta n_E(x'') = n_{E0}(\exp(qV_{EB}/kT) - 1) \exp(-x''/L_E)$$

$$\Delta n_c(x') = n_{C0}(\exp(qV_{CB}/kT) - 1) \exp(-x'/L_C)$$

Expressions of  $I_{En}$  and  $I_{Cn}$  can be evaluated.

$$I_{En} = -qAD_E \frac{d\Delta n_E(x'')}{dx} \Big|_{x''=0''}$$

$$I_{Cn} = -qA \frac{D_C}{L_C} n_{C0}(\exp(qV_{CB}/kT) - 1)$$

Because insignificant recombination occurs, the second derivative of  $\Delta p_B(x)$  is zero. There is therefore a linear relationship between excess hole density and  $x$ .

$$\Delta p_B(x) = D_1 x + D_2$$

The following are boundary conditions of  $\Delta p_B$ .

$$\Delta p_B(0) = D_2$$

$$\Delta p_B(W) = D_1 W + \Delta p_B(0)$$

Substitute into the above linear relation.

$$\Delta p_B(x) = - \left( \frac{\Delta p_B(0) - \Delta p_B(W)}{W} \right) x + \Delta p_B(0)$$

With this result, derive value of  $I_{Ep}$ .

$$I_{Ep}(0) = -qAD_B \frac{d\Delta p_B}{dx} \Big|_{x=0}$$

$$I_{Ep}(0) = \frac{qAD_B}{W} (\Delta p_B(0) - \Delta p_B(W))$$

Use the expressions of  $I_{Ep}$ ,  $I_{En}$ ,  $\Delta p_B(0)$ , and  $\Delta p_B(W)$  to develop an expression of the emitter current.

$$\Delta p_B(W) = p_{B0} \exp(qV_{CB}/kT)$$

$$\Delta p_B(0) = p_{B0} \exp(qV_{EB}/kT)$$

$$I_E = qA \left( \left( \frac{D_E n_{E0}}{L_E} + \frac{D_B p_{B0}}{W} \right) \left( \exp \left( \frac{qV_{EB}}{kT} \right) - 1 \right) - \left( \frac{D_B}{W} p_{B0} \right) \right) (e$$

Similarly, an expression of the collector current is derived.

$$I_{Cp}(W) = I_{Ep}(0)$$

$$I_C = I_{Cp}(W) + I_{Cn}(0')$$

$$I_C = qA \left( \left( \frac{D_B}{W} p_{B0} \right) \left( \exp(qV_{EB}/kT) - 1 \right) - \left( \frac{D_C n_{C0}}{L_C} + \frac{D_B p_{B0}}{W} \right) \right) (ex$$

An expression of the base current is found with the previous results.

$$I_B = I_E - I_C$$

$$I_B = qA \left( \frac{D_E}{L_E} n_{E0} \left( \exp(qV_{EB}/kT) - 1 \right) + \frac{D_C}{L_C} n_{C0} \left( \exp(qV_{CB}/kT) - 1 \right) \right)$$

### Punchthrough

When the base–collector voltage reaches a certain (device specific) value, the base–collector depletion region boundary meets the base–emitter depletion region boundary. When in this state the transistor effectively has no base. The device thus loses all gain when in this state.

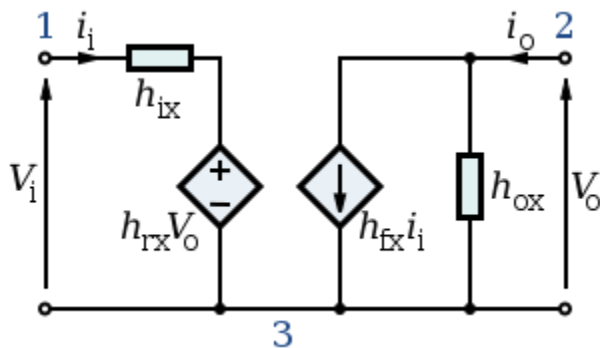
## Gummel–Poon charge-control model

The [Gummel–Poon model](#) is a detailed charge-controlled model of BJT dynamics, which has been adopted and elaborated by others to explain transistor dynamics in greater detail than the terminal-based models typically do. This model also includes the dependence of transistor  $\beta$ -values upon the direct current levels in the transistor, which are assumed current-independent in the Ebers–Moll model.

## Small-signal models

### hybrid-pi model

### h- parameter model



Generalized h-parameter model of an NPN BJT.

Replace  $x$  with  $e$ ,  $b$  or  $c$  for CE, CB and CC topologies respectively.

Another model commonly used to analyze BJT circuits is the "[h-parameter](#)" model, closely related to the [hybrid-pi model](#) and the [y-parameter two-port](#), but using input current and output voltage as independent variables, rather than input and output voltages. This two-port network is particularly suited to BJTs as it lends itself easily to the analysis of circuit behaviour, and may be used to develop further accurate models. As shown, the term " $x$ " in the model represents a different BJT lead depending on the topology used. For common-emitter mode the various symbols take on the specific values as:

- $x = 'e'$  because it is a common-emitter topology
- Terminal 1 = Base
- Terminal 2 = Collector
- Terminal 3 = Emitter
- $i_i =$  Base current ( $i_b$ )
- $i_o =$  Collector current ( $i_c$ )
- $V_{in} =$  Base-to-emitter voltage ( $V_{BE}$ )
- $V_o =$  Collector-to-emitter voltage ( $V_{CE}$ )

and the h-parameters are given by:

- $h_{ix} = h_{ie}$  – The input impedance of the transistor (corresponding to the emitter resistance  $r_e$ ).
- $h_{rx} = h_{re}$  – Represents the dependence of the transistor's  $I_B$ – $V_{BE}$  curve on the value of  $V_{CE}$ . It is usually very small and is often neglected (assumed to be zero).
- $h_{fx} = h_{fe}$  – The current-gain of the transistor. This parameter is often specified as  $h_{FE}$  or the DC current-gain ( $\beta_{DC}$ ) in datasheets.
- $h_{ox} = h_{oe}$  – The output impedance of transistor. This term is usually specified as an admittance and has to be inverted to convert it to an impedance.

As shown, the h-parameters have lower-case subscripts and hence signify AC conditions or analyses. For DC conditions they are specified in upper-case. For the CE topology, an approximate h-parameter model is commonly used which further simplifies the circuit analysis. For this the  $h_{oe}$  and  $h_{re}$  parameters are neglected (that is, they are set to infinity and zero, respectively). It should also be noted that the h-parameter model as shown is suited to low-frequency, small-signal analysis. For high-frequency analyses the inter-electrode capacitances that are important at high frequencies must be added.

## Applications

The BJT remains a device that excels in some applications, such as discrete circuit design, due to the very wide selection of BJT types available, and because of its high [transconductance](#) and output resistance compared to [MOSFETs](#). The BJT is also the choice for demanding analog circuits, especially for [very-high-frequency](#) applications, such as [radio-frequency](#) circuits for wireless systems. Bipolar transistors can be combined with MOSFETs in an integrated circuit by using a [BiCMOS](#) process of wafer fabrication to create circuits that take advantage of the application strengths of both types of transistor.

## Temperature sensors

*Main article: [Silicon bandgap temperature sensor](#)*

Because of the known temperature and current dependence of the forward-biased base–emitter junction voltage, the BJT can be used to measure temperature by subtracting two voltages at two different bias currents in a known ratio.

## Logarithmic converters

Because base–emitter voltage varies as the log of the base–emitter and collector–emitter currents, a BJT can also be used to compute [logarithms](#) and anti-logarithms. A diode can also perform these nonlinear functions, but the transistor provides more circuit flexibility.

## Vulnerabilities

Exposure of the transistor to [ionizing radiation](#) causes [radiation damage](#). Radiation causes a buildup of 'defects' in the base region that act as [recombination centers](#). The resulting reduction in minority carrier lifetime causes gradual loss of gain of the transistor.

Power BJTs are subject to a failure mode called [secondary breakdown](#), in which excessive current and normal imperfections in the silicon die cause portions of the silicon inside the device to become disproportionately hotter than the others. The doped silicon has a negative [temperature coefficient](#), meaning that it conducts more current at higher temperatures. Thus, the hottest part of the die conducts the most current, causing its conductivity to increase, which then causes it to become progressively hotter again, until the device fails internally. The thermal runaway process associated with secondary breakdown, once triggered, occurs almost instantly and may catastrophically damage the transistor package.

If the emitter-base junction is reverse biased into avalanche (zener) mode and current flows for a short period of time, the current gain of the BJT will be permanently degraded.

### **Why transistor(BJT) is called a current controlled device and why it is called a nonlinear device?**

A bipolar junction transistor (BJT) is a type of transistor. It is a three-terminal device constructed of doped semiconductor material and may be used in amplifying or switching applications. Bipolar transistors are so named because their operation involves both electrons and holes.

Although a small part of the transistor current is due to the flow of majority carriers, most of the transistor current is due to the flow of minority carriers and so BJTs are classified as 'minority-carrier' devices.

An NPN transistor can be considered as two diodes with a shared anode region. In typical operation, the emitter–base junction is forward biased and the base–collector junction is reverse biased. In an NPN transistor, for example, when a positive voltage is applied to the base–emitter junction, the equilibrium between thermally generated carriers and the repelling electric field of the depletion region becomes unbalanced, allowing thermally excited electrons to inject into the base region. These electrons wander (or "diffuse") through the base from the region of high concentration near the emitter towards the region of low concentration near the collector. The electrons in the base are called minority carriers because the base is doped p-type which would make holes the majority carrier in the base.

The base region of the transistor must be made thin, so that carriers can diffuse across it in much less time than the semiconductor's minority carrier lifetime, to minimize the percentage of carriers that recombine before reaching the collector–base junction. The thickness of the base should be less than the diffusion length of the electrons. The collector–base junction is reverse-biased, so little electron injection occurs from the collector to the base, but electrons that diffuse through the base towards the collector are swept into the collector by the electric field in the depletion region of the collector–base junction.

### **Voltage, current, and charge control**

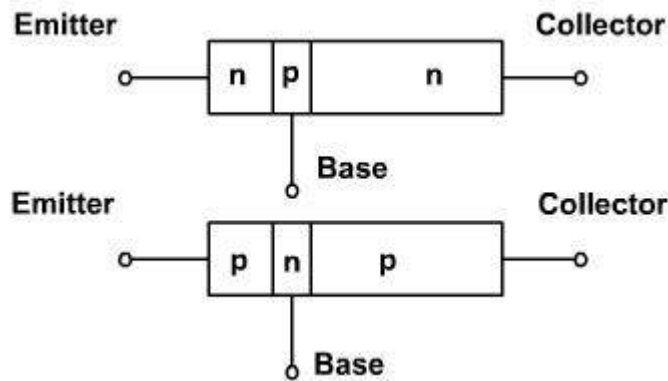
The collector–emitter current can be viewed as being controlled by the base–emitter current (current control), or by the base–emitter voltage (voltage control). These views are related by the current–voltage relation of the base–emitter junction, which is just the usual exponential current–voltage curve of a p-n junction (diode).

The physical explanation for collector current is the amount of minority-carrier charge in the base region. Detailed models of transistor action, such as the Gummel–Poon model, account for the distribution of this charge explicitly to explain transistor behavior more exactly. The charge-control view easily handles photo-transistors, where minority carriers in the base region are created by the absorption of photons, and handles the dynamics of turn-off, or recovery time, which depends on charge in the base region recombining. However, since base charge is not a signal that is visible at the terminals, the current- and voltage-control views are usually used in circuit design and analysis.

In analog circuit design, the current-control view is sometimes used since it is approximately linear. That is, the collector current is approximately  $\beta F$  times the base current. Some basic circuits can be designed by assuming that the emitter–base voltage is approximately constant, and that collector current is beta times the base current. However, to accurately and reliably design production bjt circuits, the voltage-control (for example, Ebers–Moll) model is required. The voltage-control model requires an exponential function to be taken into account, but when it is linearized such that the transistor can be modelled as a transconductance, as in the Ebers–Moll model, design for circuits such as differential amplifiers again becomes a mostly linear problem, so the voltage-control view is often preferred. For translinear circuits, in which the exponential I–V curve is key to the operation, the transistors are usually modelled as voltage controlled with transconductance proportional to collector current. In general, transistor level circuit design is performed using SPICE or a comparable analogue circuit simulator, so model complexity is usually not of much concern to the designer.

### Bipolar transistor:

A transistor is basically a Si on Ge crystal containing three separate regions. It can be either NPN or PNP type [fig. 1](#). The middle region is called the base and the outer two regions are called emitter and the collector. The outer layers although they are of same type but their functions cannot be changed. They have different physical and electrical properties.



[fig. 1](#)

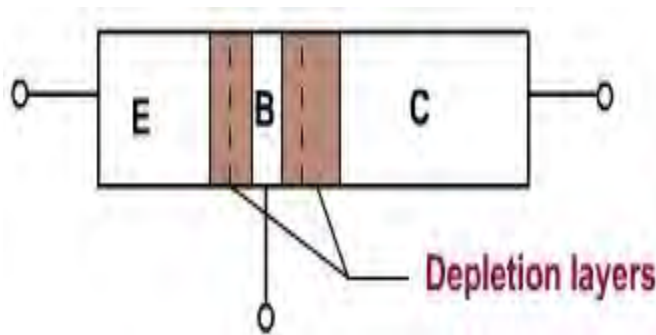
In most transistors, emitter is heavily doped. Its job is to emit or inject electrons into the base. These bases are lightly doped and very thin, it passes most of the emitter-injected electrons on to the collector. The doping level of

collector is intermediate between the heavy doping of emitter and the light doping of the base.

The collector is so named because it collects electrons from base. The collector is the largest of the three regions; it must dissipate more heat than the emitter or base. The transistor has two junctions. One between emitter and the base and other between the base and the collector. Because of this the transistor is similar to two diodes, one emitter diode and other collector base diode.

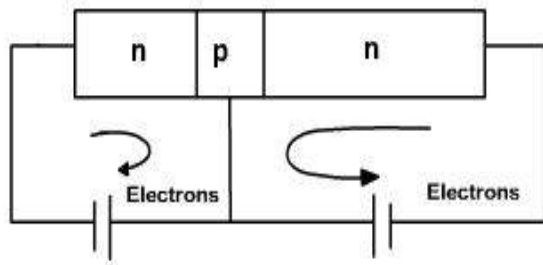
The depletion layers do not have the same width, because different regions have different doping levels. The more heavily doped a region is, the greater the concentration of ions near the junction. This means the depletion layer penetrates more deeply into the base and slightly into emitter. Similarly, it penetrates more into collector. The thickness of collector depletion layer is large while the base depletion layer is small as shown in [fig. 2](#).

When transistor is made, the diffusion of free electrons across the junction produces two depletion layers. For each of these depletion layers, the barrier potential is 0.7 V for Si transistor and 0.3 V for Ge transistor.

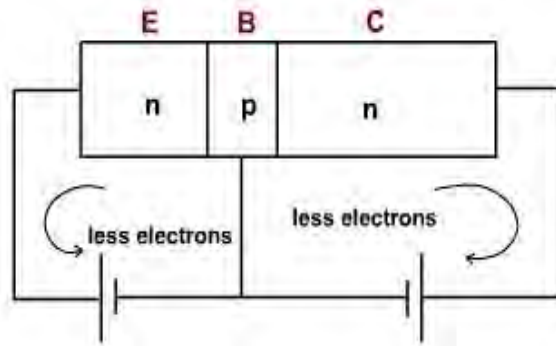


[fig. 2](#)

If both the junctions are forward biased using two d.c sources, as shown in [fig. 3a](#). free electrons (majority carriers) enter the emitter and collector of the transistor, joins at the base and come out of the base. Because both the diodes are forward biased, the emitter and collector currents are large.



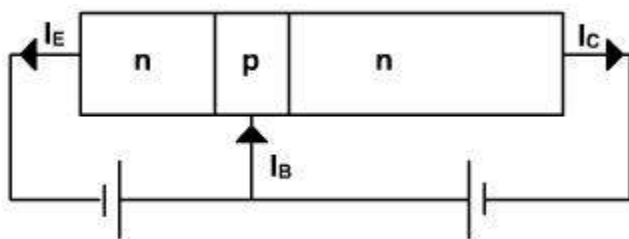
**Fig. 3a**



**Fig. 3b**

If both the junction are reverse biased as shown in [fig. 3b](#), then small currents flows through both junctions only due to thermally produced minority carriers and surface leakage. Thermally produced carriers are temperature dependent it approximately doubles for every 10 degree celsius rise in ambient temperature. The surface leakage current increases with voltage.

When the emitter diode is forward biased and collector diode is reverse biased as shown in [fig. 4](#) then one expect large emitter current and small collector current but collector current is almost as large as emitter current.



**Fig. 4**

When emitter diodes forward biased and the applied voltage is more than 0.7 V (barrier potential) then larger number of majority carriers (electrons in n-type) diffuse across the junction.

Once the electrons are injected by the emitter enter into the base, they become minority carriers. These electrons do not have separate identities from those, which are thermally generated, in the base region itself. The base is made very thin and is very lightly doped. Because of this only few electrons traveling from the emitter to base region recombine with holes. This gives rise to recombination current. The rest of the electrons exist for more time. Since the collector diode is reverse biased, (n is connected to positive supply) therefore most of the electrons are pushed into collector layer. These collector elections can then flow into the external collector lead.

Thus, there is a steady stream of electrons leaving the negative source terminal and entering the emitter region. The  $V_{EB}$  forward bias forces these emitter electrons to enter the base region. The

thin and lightly doped base gives almost all those electrons enough lifetime to diffuse into the depletion layer. The depletion layer field pushes a steady stream of electron into the collector region. These electrons leave the collector and flow into the positive terminal of the voltage source. In most transistor, more than 95% of the emitter injected electrons flow to the collector, less than 5% fall into base holes and flow out the external base lead. But the collector current is less than emitter current.

### Relation between different currents in a transistor:

The total current flowing into the transistor must be equal to the total current flowing out of it. Hence, the emitter current  $I_E$  is equal to the sum of the collector ( $I_C$ ) and base current ( $I_B$ ). That is,

$$I_E = I_C + I_B$$

The currents directions are positive directions. The total collector current  $I_C$  is made up of two components.

1. The fraction of emitter (electron) current which reaches the collector ( $\alpha_{dc} I_E$ )
2. The normal reverse leakage current  $I_{CO}$

$$\therefore I_C = \alpha_{dc} I_E + I_{CO}$$

or 
$$\alpha_{dc} = \frac{I_C - I_{CO}}{I_E}$$

$\alpha_{dc}$  is known as large signal current gain or dc alpha. It is always positive. Since collector current is almost equal to the  $I_E$  therefore  $\alpha_{dc} I_E$  varies from 0.9 to 0.98. Usually, the reverse leakage current is very small compared to the total collector current.

Neglecting  $I_{CO}$ , 
$$\alpha_{dc} = \frac{I_C}{I_E}$$

NOTE: The forward bias on the emitter diode controls the number of free electrons injected into the base. The larger ( $V_{BE}$ ) forward voltage, the greater the number of injected electrons. The reverse bias on the collector diode has little influence on the number of electrons that enter the collector. Increasing  $V_{CB}$  does not change the number of free electrons arriving at the collector junction layer.

The symbol of npn and pnp transistors are shown in [fig. 5](#).

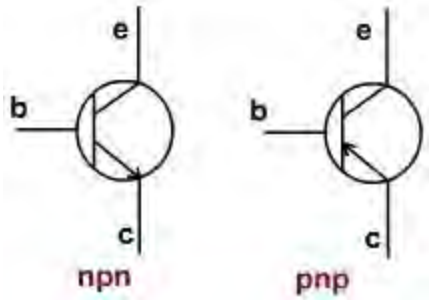


Fig. 5

### Breakdown Voltages:

Since the two halves of a transistor are diodes, too much reverse voltage on either diode can cause breakdown. The breakdown voltage depends on the width of the depletion layer and the doping levels. Because of the heavy doping level, the emitter diode has a low breakdown voltage approximately 5 to 30 V. The collector diode is less heavily doped so its breakdown voltage is higher around 20 to 300 V.

### Common Base Configuration

If the base is common to the input and output circuits, it is known as common base configuration as shown in [fig. 1](#).

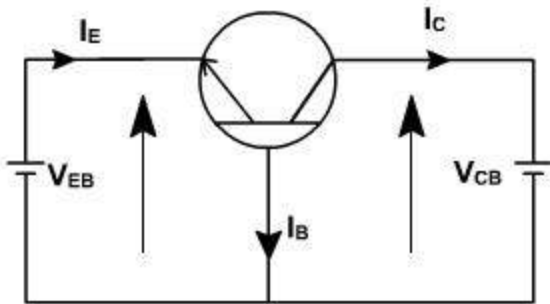


Fig. 1

For a pnp transistor the largest current components are due to holes. Holes flow from emitter to collector and few holes flow down towards ground out of the base terminal. The current directions are shown in [fig. 1](#).

$$(I_E = I_C + I_B).$$

For a forward biased junction,  $V_{EB}$  is positive and for a reverse biased junction  $V_{CB}$  is negative. The complete transistor can be described by the following two relations, which give the input

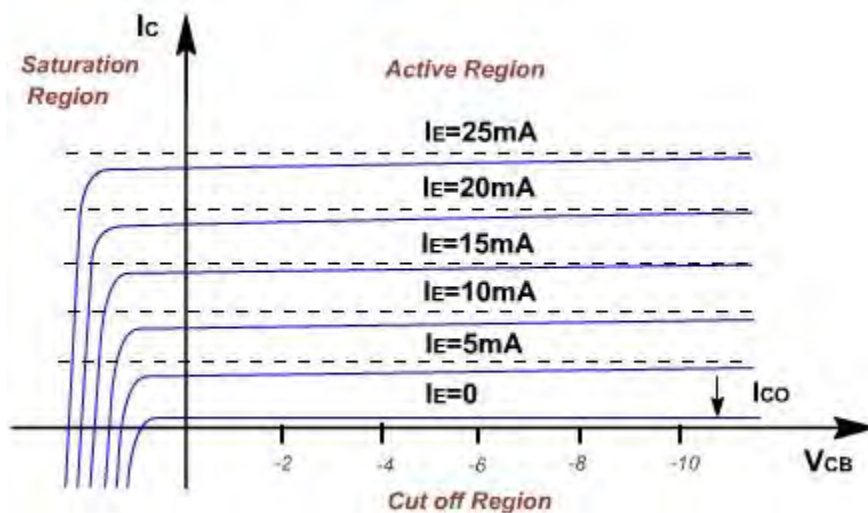
voltage  $V_{EB}$  and output current  $I_C$  in terms of the output voltage ( $V_{CB}$ ) and input current  $I_E$ .

$$V_{EB} = f_1(V_{CB}, I_E)$$

$$I_C = f_2(V_{CB}, I_E)$$

The output characteristic:

The collector current  $I_C$  is completely determined by the input current  $I_E$  and the  $V_{CB}$  voltage. The relationship is given in [fig. 2](#). It is a plot of  $I_C$  versus  $V_{CB}$ , with emitter current  $I_E$  as parameter. The curves are known as the output or collector or static characteristics. The transistor consists of two diodes placed in series back to back (with two cathodes connected together). The complete characteristic can be divided in three regions.



**Figure 7.2**

**(1). Active region:**

In this region the collector diode is reverse biased and the emitter diode is forward biased. Consider first that the emitter current is zero. Then the collector current is small and equals the reverse saturation current  $I_{CO}$  of the collector junction considered as a diode.

If the forward current  $I_B$  is increased, then a fraction of  $I_E$  ie.  $\alpha_{dc}I_E$  will reach the collector. In the active region, the collector current is essentially independent of collector voltage and depends only upon the emitter current. Because  $\alpha_{dc}$  is, less than one but almost equal to unity, the magnitude of the collector current is slightly less that of emitter current. The collector current is almost constant and work as a current source.

The collector current slightly increases with voltage. This is due to early effect. At higher voltage collector gathers in a few more electrons. This reduces the base current. The difference is so small, that it is usually neglected. If the collector voltage is increased, then space charge width increases; this decreased the effective base width. Then there is less chance for recombination

within the base region.

### (2). Saturation region:

The region to the left of the ordinate  $V_{CB} = 0$ , and above the  $I_E = 0$ , characteristic in which both emitter and collector junction are forward biased, is called saturation region.

When collector diode is forward biased, there is large change in collector current with small changes in collector voltage. A forward bias means, that p is made positive with respect to n, there is a flow of holes from p to n. This changes the collector current direction. If diode is sufficiently forward biased the current changes rapidly. It does not depend upon emitter current.

### (3). Cut off region:

The region below  $I_E = 0$  and to the right of  $V_{CB}$  for which emitter and collector junctions are both reversed biased is referred to cutoff region. The characteristics  $I_E = 0$ , is similar to other characteristics but not coincident with horizontal axis. The collector current is same as  $I_{CO}$ .  $I_{CBO}$  is frequently used for  $I_{CO}$ . It means collector to base current with emitter open. This is also temperature dependent.

## Common Base Amplifier

The common base amplifier circuit is shown in [Fig. 1](#). The  $V_{EE}$  source forward biases the emitter diode and  $V_{CC}$  source reverse biased collector diode. The ac source  $v_{in}$  is connected to emitter through a coupling capacitor so that it blocks dc. This ac voltage produces small fluctuation in currents and voltages. The load resistance  $R_L$  is also connected to collector through coupling capacitor so the fluctuation in collector base voltage will be observed across  $R_L$ .

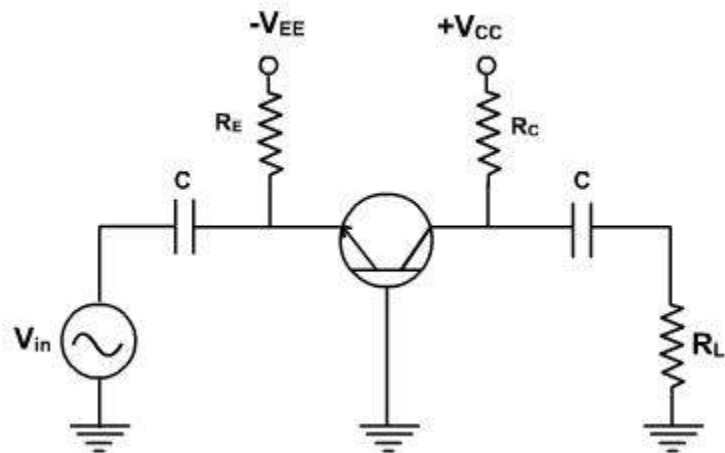


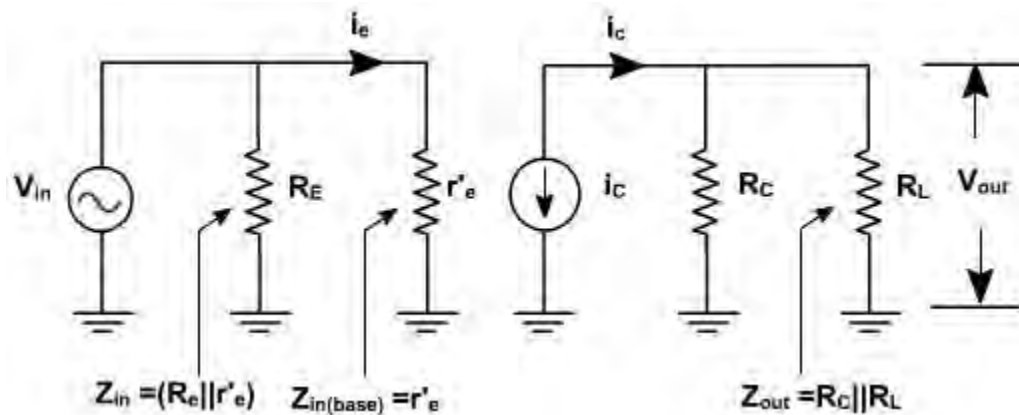
Fig. 1

The dc equivalent circuit is obtained by reducing all ac sources to zero and opening all capacitors. The dc collector current is same as  $I_E$  and

$V_{CB}$  is given by

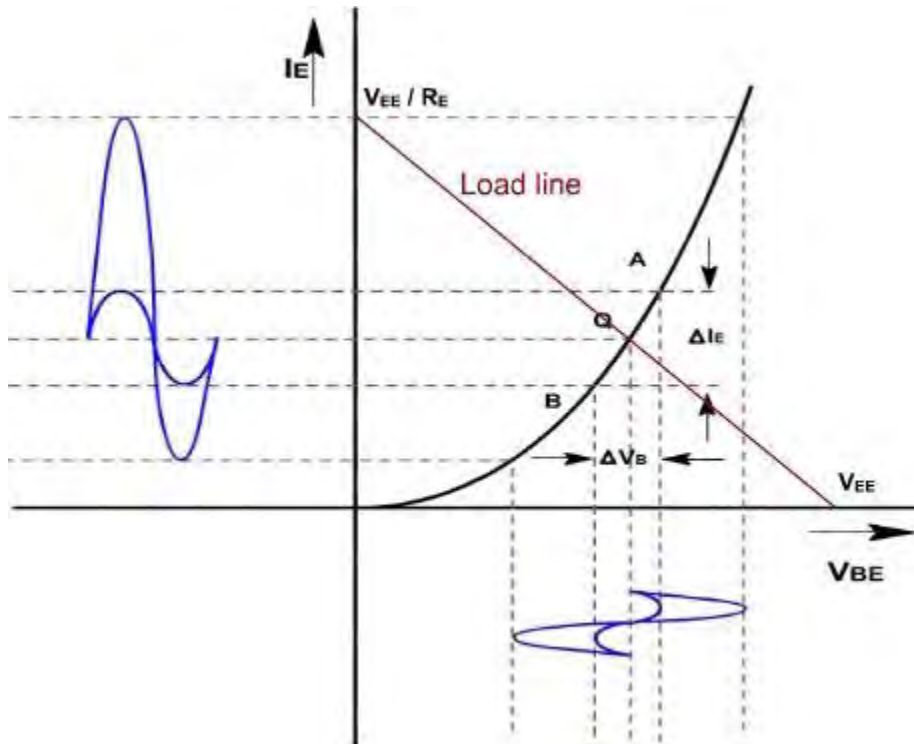
$$V_{CB} = V_{CC} - I_C R_C.$$

These current and voltage fix the Q point. The ac equivalent circuit is obtained by reducing all dc sources to zero and shorting all coupling capacitors.  $r'_e$  represents the ac resistance of the diode as shown in [Fig. 2](#).



**Fig. 2**

[Fig. 3](#), shows the diode curve relating  $I_E$  and  $V_{BE}$ . In the absence of ac signal, the transistor operates at Q point (point of intersection of load line and input characteristic). When the ac signal is applied, the emitter current and voltage also change. If the signal is small, the operating point swings sinusoidally about Q point (A to B).



**Fig .3**

If the ac signal is small, the points A and B are close to Q, and arc A B can be approximated by a straight line and diode appears to be a resistance given by

$$r'_e = \frac{\Delta V_{BE}}{\Delta I_E} \Big|_{\text{small change}}$$

$$= \frac{V_{be}}{i_e} = \frac{\text{ac voltage across base and emitter}}{\text{ac current through emitter}}$$

If the input signal is small, input voltage and current will be sinusoidal but if the input voltage is large then current will no longer be sinusoidal because of the non linearity of diode curve. The emitter current is elongated on the positive half cycle and compressed on negative half cycle. Therefore the output will also be distorted.

$r'_e$  is the ratio of  $\Delta V_{BE}$  and  $\Delta I_E$  and its value depends upon the location of Q. Higher up the Q point small will be the value of  $r'_e$  because the same change in  $V_{BE}$  produces large change in  $I_E$ . The slope of the curve at Q determines the value of  $r'_e$ . From calculation it can be proved that.

$$r'_e = 25\text{mV} / I_E$$

## Common Base Amplifier

### Proof:

In general, the current through a diode is given by

$$I = I_{CO} (e^{\frac{qV}{KT}} - 1)$$

Where  $q$  is the charge on electron,  $V$  is the drop across diode,  $T$  is the temperature and  $K$  is a constant.

On differentiating w.r.t  $V$ , we get,

$$\frac{dI}{dV} = I_{CO} * e^{\frac{qV}{KT}} * \frac{q}{KT}$$

The value of  $(q / KT)$  at  $25^\circ\text{C}$  is approximately 40.

$$\frac{dI}{dV} = 40 * I_{CO} * e^{\frac{qV}{KT}}$$

Therefore,  $= 40 * (I + I_{CO})$

$$\text{or, } \frac{dV}{dI} = \frac{1}{40 * (I + I_{CO})} \approx \frac{1}{40 * I}$$

$$\text{Therefore, ac resistance of the emitter diode} = \frac{dV}{dI} = \frac{25\text{mV}}{I} \text{ Ohms}$$

To a close approximation the small changes in collector current equal the small changes in emitter current. In the ac equivalent circuit, the current  $\underline{i}_C'$  is shown upward because if  $\underline{i}_e'$  increases, then  $\underline{i}_C'$  also increases in the same direction.

### **Voltage gain:**

Since the ac input voltage source is connected across  $r'_e$ . Therefore, the ac emitter current is given by

$$i_e = V_{in} / r'_e$$

$$\text{or, } V_{in} = i_e r'_e$$

The output voltage is given by  $V_{out} = i_c (R_C \parallel R_L)$

Therefore, voltage gain  $A_V = \frac{v_{out}}{v_{in}} = \frac{(R_C \parallel R_L)}{r'_e}$   
 $= \frac{R_C}{r}$

Under open circuit condition  $v_{out} = i_c R_C$

Therefore, voltage gain in open circuit condition  $= A_V = \frac{R_C}{r'_e}$

## Common Base Amplifier

### Example-1

Find the voltage gain and output of the amplifier shown in [fig. 4](#), if input voltage is 1.5mV.

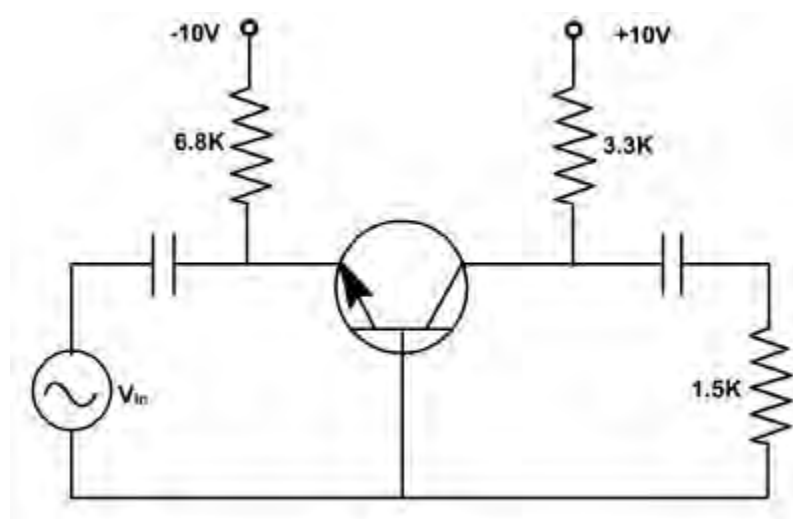


Fig. 4

**Solution:**

The emitter dc current  $I_E$  is given by  $I_E = \frac{10 - 0.7}{6.8k} = 1.37mA$

Therefore, emitter ac resistance  $= \frac{r_c}{r'_e} = \frac{3.3k \parallel 1.5k}{18.2\Omega}$

or,  $A_V = 56.6$

and,  $V_{out} = 1.5 \times 56.6 = 84.9 \text{ mV}$

### Example-2

Repeat example-1 if ac source has resistance  $R_s = 100 \Omega$ .

#### Solution:

The ac equivalent circuit with ac source resistance is shown in [fig. 5](#).

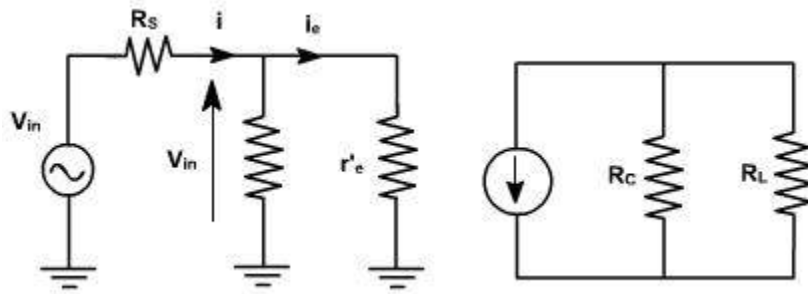


Fig. 5

The emitter ac current is given by 
$$i_e = \frac{v_{in}}{R_s + (R_E \parallel r'_e)} \times \frac{R_E}{R_E + r'_e}$$

or, 
$$i_e = \frac{v_{in}}{(R_s + r'_e) R_E + R_s r'_e} \times R_E ; \frac{v_{in}}{R_s + r'_e}$$

Therefore, voltage gain of the amplifier = 
$$A_V = \frac{v_{out}}{v_{in}} = \frac{i_e r_c}{i_e (R_s + r'_e)} = \frac{r_c}{R_s + r'_e}$$

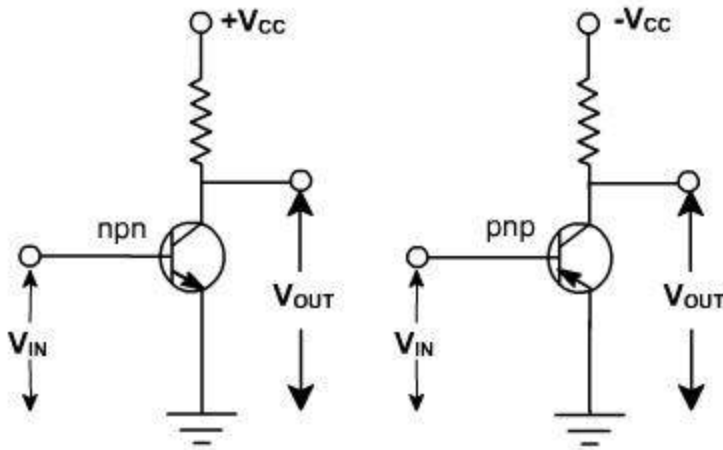
$$A_V = \frac{3.3k \parallel 1.5k}{100\Omega + 18.2\Omega} = 8.71$$

and,  $V_{out} = 1.5 \times 8.71 = 13.1 \text{ mV}$

## Common Emitter Configuration

### Common Emitter Curves:

The common emitter configuration of BJT is shown in [fig. 1](#).



**Fig. 1**

In C.E. configuration the emitter is made common to the input and output. It is also referred to as grounded emitter configuration. It is most commonly used configuration. In this, base current and output voltages are taken as independent parameters and input voltage and output current as dependent parameters

$$V_{BE} = f_1 ( I_B, V_{CE} )$$

$$I_C = f_2( I_B, V_{CE} )$$

Input Characteristic:

The curve between  $I_B$  and  $V_{BE}$  for different values of  $V_{CE}$  are shown in [fig. 2](#). Since the base emitter junction of a transistor is a diode, therefore the characteristic is similar to diode one. With higher values of  $V_{CE}$  collector gathers slightly more electrons and therefore base current reduces. Normally this effect is neglected. (Early effect). When collector is shorted with emitter then the input characteristic is the characteristic of a forward biased diode when  $V_{BE}$  is zero and  $I_B$  is also zero.

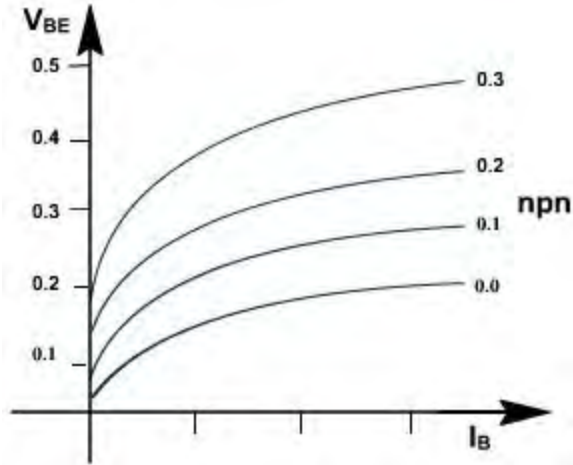


Fig. 2

## Common Emitter Configuration

### Output Characteristic:

The output characteristic is the curve between  $V_{CE}$  and  $I_C$  for various values of  $I_B$ . For fixed value of  $I_B$  and is shown in [fig. 3](#). For fixed value of  $I_B$ ,  $I_C$  is not varying much dependent on  $V_{CE}$  but slopes are greater than CE characteristic. The output characteristics can again be divided into three parts.

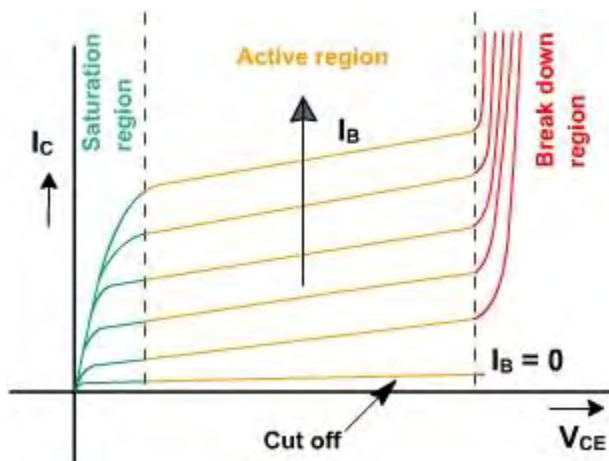


Fig. 3

### (1) Active Region:

In this region collector junction is reverse biased and emitter junction is forward biased. It is the

area to the right of  $V_{CE} = 0.5$  V and above  $I_B = 0$ . In this region transistor current responds most sensitively to  $I_B$ . If transistor is to be used as an amplifier, it must operate in this region.

$$I_E = I_C + I_B$$

Since,  $I_C = I_{C0} + \alpha_{dc} I_E$

$$I_C = I_{C0} + \alpha_{dc} (I_C + I_B)$$

or  $(1 - \alpha_{dc}) I_C = \alpha_{dc} I_B + I_{C0}$

or  $I_C = \left( \frac{\alpha_{dc}}{1 - \alpha_{dc}} \right) I_B + \left( \frac{1}{1 - \alpha_{dc}} \right) I_{C0}$

Let,  $\beta_{dc} = \frac{\alpha_{dc}}{1 - \alpha_{dc}}$

$$\therefore I_C = (1 + \beta_{dc}) I_{C0} + \beta_{dc} I_B$$

$\beta_{dc}$  is defined as current gain of the transistor is given by

$$\beta_{dc} = \frac{I_C - I_{C0}}{I_B + I_{C0}}$$

If  $\alpha_{dc}$  is truly constant then  $I_C$  would be independent of  $V_{CE}$ . But because of early effect,  $\alpha_{dc}$  increases by 0.1% (0.001) e.g. from 0.995 to 0.996 as  $V_{CE}$  increases from a few volts to 10V. Then  $\beta_{dc}$  increases from  $0.995 / (1 - 0.995) = 200$  to  $0.996 / (1 - 0.996) = 250$  or about 25%. This shows that small change in  $\alpha$  reflects large change in  $\beta$ . Therefore the curves are subjected to large variations for the same type of transistors.

## (2) Cut Off:

Cut off in a transistor is given by  $I_B = 0$ ,  $I_C = I_{C0}$ . A transistor is not at cut off if the base current is simply reduced to zero (open circuited) under this condition,

$$I_C = I_E = I_{C0} / (1 - \alpha_{dc}) = I_{CEO}$$

The actual collector current with base open is designated as  $I_{CEO}$ . Since even in the neighborhood of cut off,  $\alpha_{dc}$  may be as large as 0.9 for Ge, then  $I_C = 10 I_{C0}$  (approximately), at zero base current. Accordingly in order to cut off transistor it is not enough to reduce  $I_B$  to zero, but it is necessary to reverse bias the emitter junction slightly. It is found that reverse voltage of 0.1 V is sufficient for cut off a transistor. In Si, the  $\alpha_{dc}$  is very nearly equal to zero, therefore,  $I_C = I_{C0}$ . Hence even with  $I_B = 0$ ,  $I_C = I_E = I_{C0}$  so that transistor is very close to cut off.

In summary, cut off means  $I_E = 0$ ,  $I_C = I_{C0}$ ,  $I_B = -I_C = -I_{C0}$ , and  $V_{BE}$  is a reverse voltage whose magnitude is of the order of 0.1 V for Ge and 0 V for Si.

## Reverse Collector Saturation Current $I_{CBO}$ :

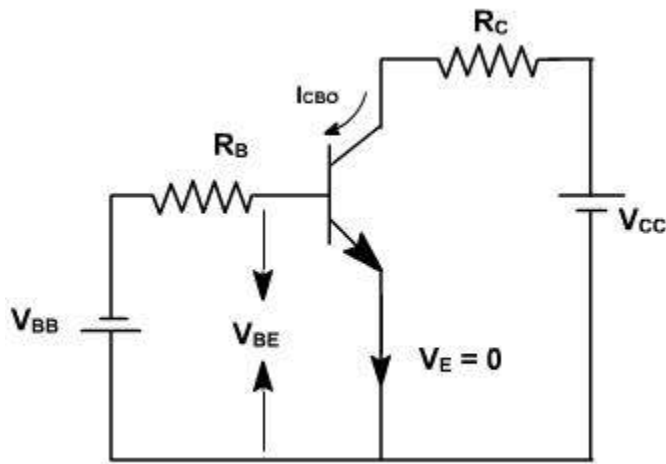
When in a physical transistor emitter current is reduced to zero, then the collector current is known as  $I_{CBO}$  (approximately equal to  $I_{C0}$ ). Reverse collector saturation current  $I_{CBO}$  also varies with temperature, avalanche multiplication and variability from sample to sample. Consider the

circuit shown in [fig. 4](#).  $V_{BB}$  is the reverse voltage applied to reduce the emitter current to zero.

$$I_E = 0, \quad I_B = -I_{CBO}$$

If we require,  $V_{BE} = -0.1 \text{ V}$

$$\text{Then } -V_{BB} + I_{CBO} R_B < -0.1 \text{ V}$$



**Fig. 4**

If  $R_B = 100 \text{ K}$ ,  $I_{CBO} = 100 \text{ m A}$ , Then  $V_{BB}$  must be 10.1 Volts. Hence transistor must be capable to withstand this reverse voltage before breakdown voltage exceeds.

### **(3).Saturation Region:**

In this region both the diodes are forward biased by at least cut in voltage. Since the voltage  $V_{BE}$  and  $V_{BC}$  across a forward is approximately 0.7 V therefore,  $V_{CE} = V_{CB} + V_{BE} = -V_{BC} + V_{BE}$  is also few tenths of volts. Hence saturation region is very close to zero voltage axis, where all the current rapidly reduces to zero. In this region the transistor collector current is approximately given by  $V_{CC} / R_C$  and independent of base current. Normal transistor action is lost and it acts like a small ohmic resistance.

## **Common Emitter Configuration**

### **Large Signal Current Gain $\beta_{dc}$ :-**

*The ratio  $I_c / I_B$  is defined as transfer ratio or large signal current gain  $\beta_{dc}$*

$$\beta_{dc} = \frac{I_C}{I_B}$$

Where  $I_C$  is the collector current and  $I_B$  is the base current. The  $\beta_{dc}$  is an indication of how well the transistor works. The typical value of  $\beta_{dc}$  varies from 50 to 300.

In terms of  $h$  parameters,  $\beta_{dc}$  is known as  $dc$  current gain and is designated  $h_{FE}$  ( $\beta_{dc} = h_{FE}$ ). Knowing the maximum collector current and  $\beta_{dc}$  the minimum base current can be found which will be needed to saturate the transistor.

$$I_{C(max)} = \frac{V_{CC} - V_{CE(sat)}}{R_C} = I_{C(sat)}$$

$$I_{B(min)} = \frac{I_{C(sat)}}{\beta_{dc}}$$

This expression of  $\beta_{dc}$  is defined neglecting reverse leakage current ( $I_{CO}$ ).

Taking reverse leakage current ( $I_{CO}$ ) into account, the expression for the  $\beta_{dc}$  can be obtained as follows:

$\beta_{dc}$  in terms of  $\alpha_{dc}$  is given by

$$\begin{aligned} \beta_{dc} &= \frac{\alpha_{dc}}{1 - \alpha_{dc}} \\ &= \frac{\frac{I_C - I_{CO}}{I_E}}{1 - \frac{I_C - I_{CO}}{I_E}} = \frac{I_C - I_{CO}}{I_E - I_C + I_{CO}} \\ &= \frac{I_C - I_{CO}}{I_B + I_{CO}} \end{aligned}$$

Since,  $I_{CO} = I_{CBO}$

$$\therefore \beta_{dc} = \frac{I_C - I_{CBO}}{I_B + I_{CBO}}$$

Cut off of a transistor means  $I_E = 0$ , then  $I_C = I_{CBO}$  and  $I_B = -I_{CBO}$ . Therefore, the above expression  $\beta_{dc}$  gives the collector current increment to the base current change from cut off to  $I_B$  and hence it represents the large signal current gain of all common emitter transistor.

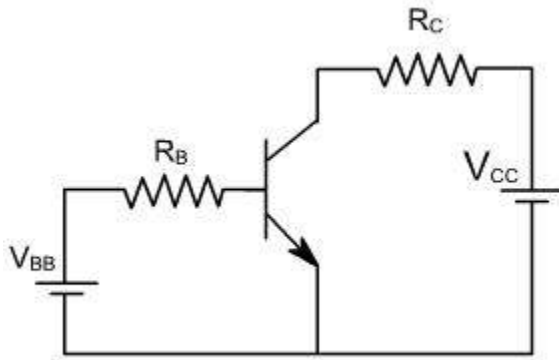
## Biasing Techniques for CE Amplifiers

Biasing Circuit Techniques or Locating the Q - Point:

### Fixed Bias or Base Bias:

In order for a transistor to amplify, it has to be properly biased. This means forward biasing the base emitter junction and reverse biasing collector base junction. For linear amplification, the transistor should operate in active region ( If  $I_E$  increases,  $I_C$  increases,  $V_{CE}$  decreases proportionally).

The source  $V_{BB}$ , through a current limit resistor  $R_B$  forward biases the emitter diode and  $V_{CC}$  through resistor  $R_C$  (load resistance) reverse biases the collector junction as shown in [fig. 1](#).



**Fig. 1**

The dc base current through  $R_B$  is given by

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

$$\text{or } V_{BE} = V_{BB} - I_B R_B$$

Normally  $V_{BE}$  is taken 0.7V or 0.3V. If exact voltage is required, then the input characteristic ( $I_B$  vs  $V_{BE}$ ) of the transistor should be used to solve the above equation. The load line for the input circuit is drawn on input characteristic. The two points of the load line can be obtained as given below

$$\text{For } I_B = 0, \quad V_{BE} = V_{BB}.$$

$$\text{and For } V_{BE} = 0, \quad I_B = V_{BB} / R_B.$$

The intersection of this line with input characteristic gives the operating point Q as shown in

[fig. 2](#). If an ac signal is connected to the base of the transistor, then variation in  $V_{BE}$  is about

Q - point. This gives variation in  $I_B$  and hence  $I_C$ .

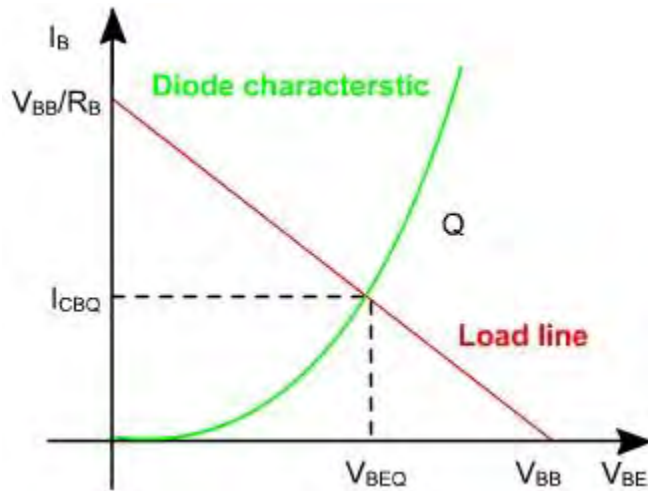


Fig. 2

## Biasing Techniques for CE Amplifiers

In the output circuit, the load equation can be written as

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

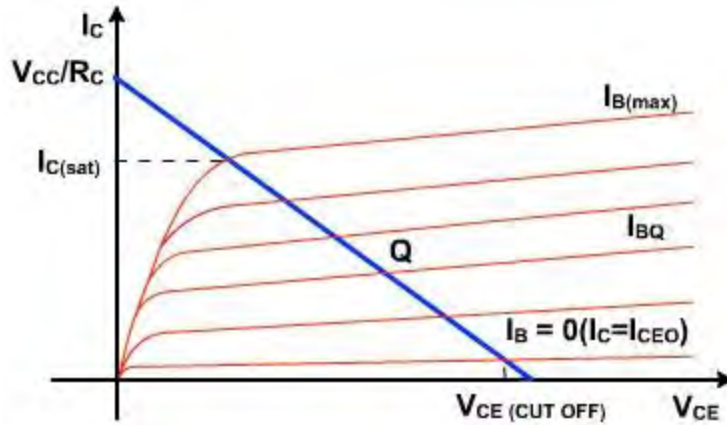
This equation involves two unknown  $V_{CE}$  and  $I_C$  and therefore can not be solved. To solve this equation output characteristic ( $I_C$  vs  $V_{CE}$ ) is used.

The load equation is the equation of a straight line and given by two points:

$$I_C = 0, \quad V_{CE} = V_{CC}$$

$$\& \quad V_{CE} = 0, \quad I_C = V_{CC} / R_C$$

The intersection of this line which is also called dc load line and the characteristic gives the operating point Q as shown in [fig. 3](#).



**Fig. 3**

The point at which the load line intersects with  $I_B = 0$  characteristic is known as cut off point. At this point base current is zero and collector current is almost negligibly small. At cut off the emitter diode comes out of forward bias and normal transistor action is lost. To a close approximation.

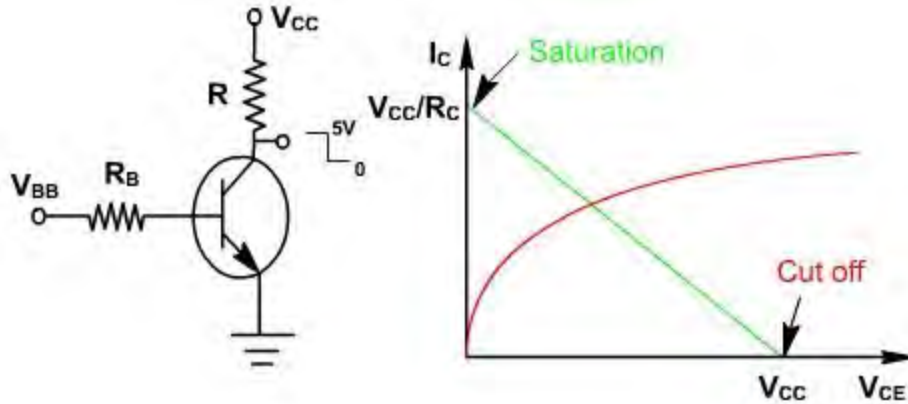
$$V_{CE} (\text{cut off}) \approx V_{CC} (\text{approximately}).$$

The intersection of the load line and  $I_B = I_{B(\text{max})}$  characteristic is known as saturation point. At this point  $I_B = I_{B(\text{max})}$ ,  $I_C = I_{C(\text{sat})}$ . At this point collector diodes comes out of reverse bias and again transistor action is lost. To a close approximation,

$$I_{C(\text{sat})} \approx V_{CC} / R_C (\text{approximately}).$$

The  $I_{B(\text{sat})}$  is the minimum current required to operate the transistor in saturation region. If the  $I_B$  is less than  $I_{B(\text{sat})}$ , the transistor will operate in active region. If  $I_B > I_{B(\text{sat})}$  it always operates in saturation region.

If the transistor operates at saturation or cut off points and no where else then it is operating as a switch is shown in [fig. 4](#).



**Fig. 4**

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

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

If  $I_B > I_{B(sat)}$ , then it operates at saturation, If  $I_B = 0$ , then it operates at cut off.

If a transistor is operating as an amplifier then Q point must be selected carefully. Although we can select the operating point any where in the active region by choosing different values of  $R_B$  &  $R_C$  but the various transistor ratings such as maximum collector dissipation  $P_{C(max)}$  maximum collector voltage  $V_{C(max)}$  and  $I_{C(max)}$  &  $V_{BE(max)}$  limit the operating range.

Once the Q point is established an ac input is connected. Due to this the ac source the base current varies. As a result of this collector current and collector voltage also varies and the amplified output is obtained.

If the Q-point is not selected properly then the output waveform will not be exactly the input waveform. i.e. It may be clipped from one side or both sides or it may be distorted one.

## Biasing Techniques for CE Amplifiers

### Example-1

Find the transistor current in the circuit shown in [fig. 5](#), if  $I_{CO} = 20nA$ ,  $\beta = 100$ .

**Solution:**

For the base circuit,  $5 = 200 \times I_B + 0.7$

Therefore,  $I_B = \frac{5 - 0.7}{200k} = 0.0215mA$

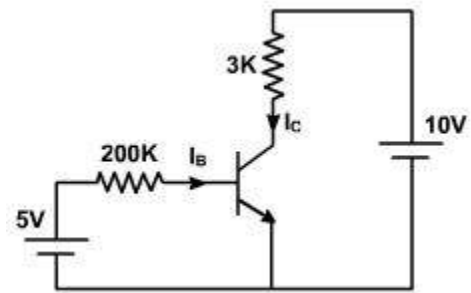
Since  $I_{CO} \ll I_B$ , therefore,  $I_C = \beta I_B = 2.15 \text{ mA}$

From the collector circuit,  $V_{CE} = 10 - 3 \times 2.15 = 3.55 \text{ V}$

Since,  $V_{CE} = V_{CB} + V_{BE}$

Thus,  $V_{CB} = 3.55 - 0.7 = 2.55 \text{ V}$

Therefore, collector junction is reverse biased and transistor is operating in its active region.



**Fig. 5**

**Example - 2**

If a resistor of 2K is connected in series with emitter in the circuit as shown in [fig. 6](#), find the currents. Given  $I_{CO} = 20 \text{ nA}$ ,  $\beta = 100$ .

**Solution:**

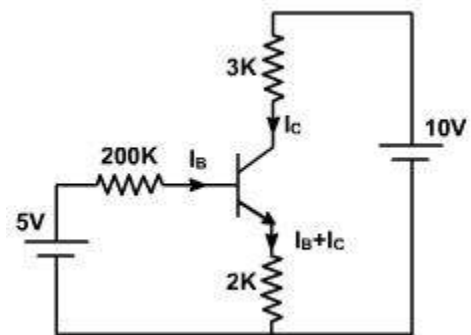
$I_E = I_B + I_C = I_B + 100 I_B = 101 I_B$

For the base circuit,  $5 = 200 \times I_B + 0.7 + 2k \times 101 I_B$

Therefore,  $I_B = \frac{5 - 0.7}{402k} = 0.0107mA$

Since  $I_{CO} \ll I_B$ , therefore,  $I_C = \beta I_B = 1.07 \text{ mA}$

From the collector circuit,  $V_{CB} = 10 - 3 \times 1.07 - 0.7 - 2 \times 101 \times 0.0107 = 3.93 \text{ V}$



**Fig. 6**

Therefore collector junction is reverse biased and transistor is operating in its active region.

### Example - 3

Repeat the **example-1** if  $R_B$  is replaced by 50k.

#### Solution:

The circuit is shown in [fig. 7](#).

Since the base resistance is reduced, the base current must have increased and there is a possibility that the transistor has entered into saturation region.

Assuming transistor is operating in its saturation region,

$$V_{BE(sat)} = 0.8 \text{ V and } V_{CE(sat)} = 0.2 \text{ V}$$

$$\text{Therefore, } I_B = \frac{5 - 0.8}{50k} = 0.0840 \text{ mA}$$

$$\text{and } I_C = \frac{10 - 0.2}{3k} = 3.267 \text{ mA}$$

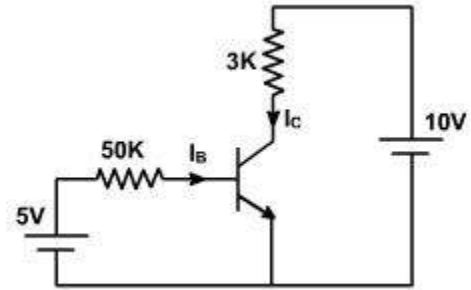


Fig. 7

The minimum base current required for operating the transistor in saturation region is

$$I_{B(\min)} = \frac{I_C}{\beta} = 0.03267 \text{ mA}$$

Since  $I_B > I_{B(\min)}$ , therefore, transistor is operating in its saturation region.

### Example - 4

Repeat the **example-2** if  $R_B$  is replaced by 50k.

#### Solution:

The circuit is shown in [fig. 8](#).

Since the base resistance is reduced, the base current must have increased and there is a possibility that the transistor has entered into saturation region.

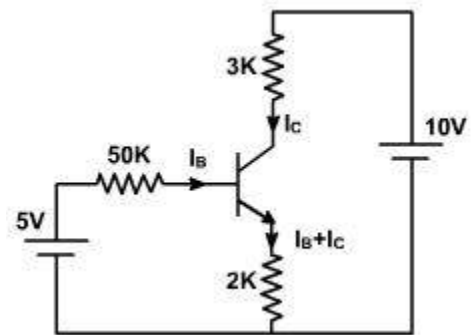


Fig. 8

Assuming transistor is operating in its saturation region,

$$5 = 50I_B + 0.8 + 2 \times (I_B + I_C)$$

$$10 = 3I_C + 0.2 + 2 \times (I_B + I_C)$$

Solving these equations, we get,

$$I_C = 1.96\text{mA} \text{ and } I_B = 0.0035\text{mA}$$

The minimum base current required for operating the transistor in saturation region is

$$I_{B(\text{min})} = \frac{I_C}{\beta} = 0.0196\text{mA}$$

Since  $I_B < I_{B(\text{min})}$ , therefore, transistor is operating in its **active** region and not in saturation. The base and the collector currents can be recalculated assuming the transistor to be in active region.

For the base circuit,  $5 = 50 \times I_B + 0.7 + 2\text{k} \times 101 I_B$

Therefore, 
$$I_B = \frac{5 - 0.7}{252\text{k}} = 0.0171\text{mA}$$

$$I_C = 1.71\text{mA}$$

From the collector circuit,  $V_{CB} = 10 - 3 \times 1.71 - 0.7 - 2 \times 101 \times 0.0171 = 0.716 \text{ V}$

## Stability of Operating Point

Let us consider three operating points of transistor operating in common emitter amplifier.

1. Near cut off
2. Near saturation
3. In the middle of active region

If the operating point is selected near the cutoff region, the output is clipped in negative half cycle as shown in [fig. 1](#).

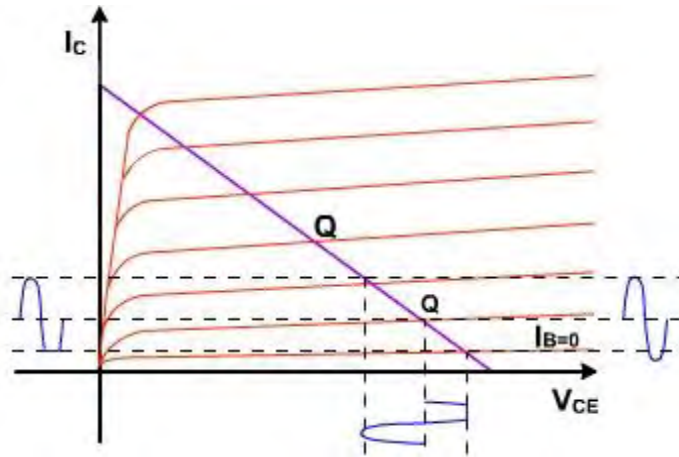


Fig. 1

If the operating point is selected near saturation region, then the output is clipped in positive cycle as shown in [fig. 2](#).

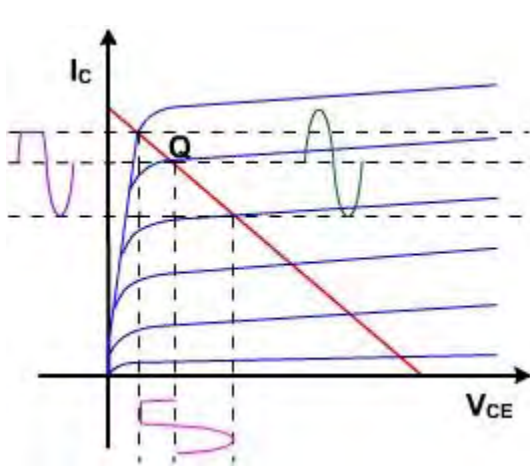


Fig. 2

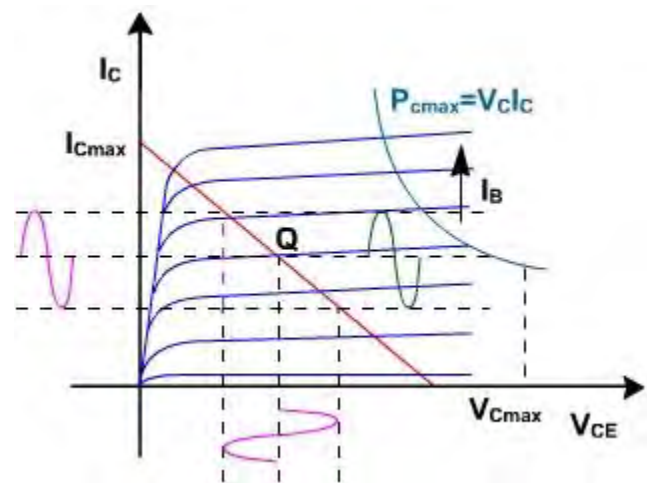


Fig. 3

If the operating point is selected in the middle of active region, then there is no clipping and the output follows input faithfully as shown in [fig. 3](#). If input is large then clipping at both sides will take place. The first circuit for biasing the transistor in CE configuration is fixed bias.

In biasing circuit shown in [fig. 4\(a\)](#), two different power supplies are required. To avoid the use of two supplies the base resistance  $R_B$  is connected to  $V_{CC}$  as shown in [fig. 4\(b\)](#).

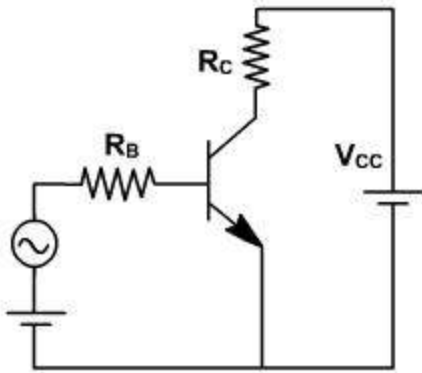


Fig. 4(a)

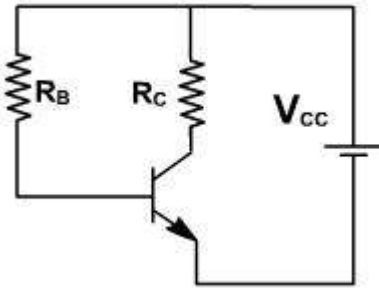


Fig. 4(b)

Now  $V_{CC}$  is still forward biasing emitter diode. In this circuit Q point is very unstable. The base resistance  $R_B$  is selected by noting the required base current  $I_B$  for operating point Q.

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

Voltage across base emitter junction is approximately 0.7 V. Since  $V_{CC}$  is usually very high

i.e.  $I_B = V_{CC} / R_B$

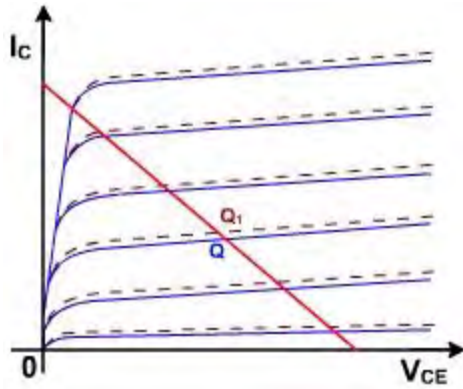
Since  $I_B$  is constant therefore it is called fixed bias circuit.

## Biassing

### Stability of quiescent operating point:

Let us assume that the transistor is replaced by an other transistor of same type. The  $\beta_{dc}$  of the two transistors of same type may not be same. Therefore, if  $\beta_{dc}$  increases then for same  $I_B$ , output characteristic shifts upward. If  $\beta_{dc}$  decreases, the output characteristic shifts downward. Since  $I_B$  is maintained constant, therefore the operating point shifts from Q to  $Q_1$  as shown in [fig. 5](#). The new operating point may be completely unsatisfactory.

Therefore, to maintain operating point stable,  $I_B$  should be allowed to change so as to maintain  $V_{CE}$  &  $I_C$  constant as  $\beta_{dc}$  changes.



**Fig. 5**

A second cause for bias instability is a variation in temperature. The reverse saturation current changes with temperature. Specifically,  $I_{CO}$  doubles for every  $10^\circ\text{C}$  rise in temperature. The collector current  $I_C$  causes the collector junction temperature to rise, which in turn increases  $I_{CO}$ . As a result of this growth  $I_{CO}$ ,  $I_C$  will increase  $(\beta_{dc} I_B + (1 + \beta_{dc}) I_{CO})$  and so on. It may be possible that this process goes on and the ratings of the transistors are exceeded. This increase in  $I_C$  changes the characteristic and hence the operating point.

### **Stability Factor:**

The operating point can be made stable by keeping  $I_C$  and  $V_{CE}$  constant. There are two techniques to make Q point stable.

1. stabilization techniques
2. compensation techniques

In first, resistor biasing circuits are used which allow  $I_B$  to vary so as to keep  $I_C$  relatively constant with variations in  $\beta_{dc}$ ,  $I_{CO}$  and  $V_{BE}$ .

In second, temperature sensitive devices such as diodes, transistors are used which provide compensating voltages and currents to maintain the operating point constant.

To compare different biasing circuits, stability factor  $S$  is defined as the rate of change of collector current with respect to the  $I_{CO}$ , keeping  $\beta_{dc}$  and  $V_{CE}$  constant

$$S = \partial I_C / \partial I_{CO}$$

If  $S$  is large, then circuit is thermally unstable.  $S$  cannot be less than unity. The other stability factors are,  $\partial I_C / \partial \beta_{dc}$  and  $\partial I_C / \partial V_{BE}$ . The bias circuit, which provide stability with  $I_{CO}$ , also show stability even if  $\beta$  and  $V_{BE}$  changes.

$$I_C = \beta_{dc} I_B + (1 + \beta_{dc}) I_{CO}$$

Differentiating with respect to  $I_C$ ,

$$1 = \beta_{dc} \frac{\partial I_B}{\partial I_C} + \frac{(1 + \beta_{dc})}{S}$$

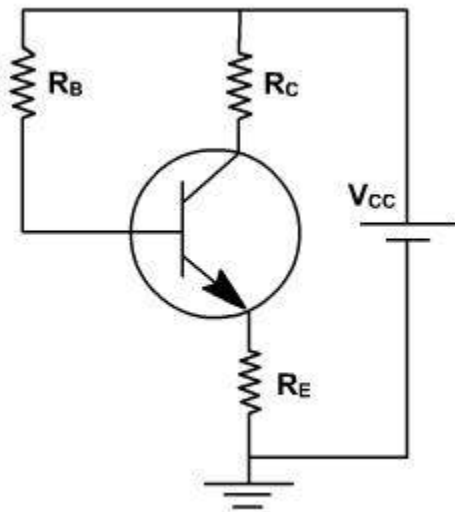
$$\therefore S = \frac{1 + \beta_{dc}}{1 - \beta_{dc} \frac{\partial I_B}{\partial I_C}}$$

In fixed bias circuit,  $I_B$  &  $I_C$  are independent. Therefore  $\frac{\partial I_B}{\partial I_C} = 0$  and  $S = 1 + \beta_{dc}$ . If  $\beta_{dc} = 100$ ,  $S = 101$ , which means  $I_C$  increases 101 times as fast as  $I_{C0}$ . Such a large change definitely operate the transistor in saturation.

## Biassing Techniques

### Emitter Feedback Bias:

**Fig. 1**, shows the emitter feedback bias circuit. In this circuit, the voltage across resistor  $R_E$  is used to offset the changes in  $\beta_{dc}$ . If  $\beta_{dc}$  increases, the collector current increases. This increases the emitter voltage which decrease the voltage across base resistor and reduces base current. The reduced base current result in less collector current, which partially offsets the original increase in  $\beta_{dc}$ . The feedback term is used because output current ( $I_C$ ) produces a change in input current ( $I_B$ ).  $R_E$  is common in input and output circuits.



**Fig. 1**

In this case

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

Since  $I_E = I_C + I_B$

$$\therefore \frac{\partial I_B}{\partial I_C} = -\frac{R_E}{R_B + R_E}$$

Therefore,

$$S = \frac{1 + \beta_{dc}}{1 + \beta_{dc} \left( \frac{R_E}{R_B + R_E} \right)} \ll (1 + \beta_{dc})$$

In this case, S is less compared to fixed bias circuit. Thus the stability of the Q point is better.

Further,

$$I_E \approx I_C = \frac{V_{CC} - V_{BE}}{R_E + \frac{R_B}{\beta_{dc}}}$$

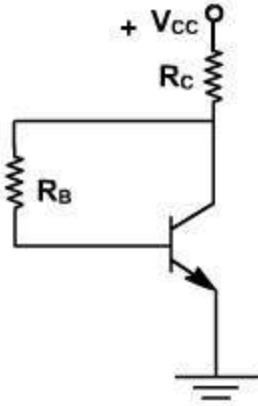
If  $I_C$  is to be made insensitive to  $\beta_{dc}$  than

$$R_E \gg \frac{R_B}{\beta_{dc}}$$

$R_E$  cannot be made large enough to swamp out the effects of  $\beta_{dc}$  without saturating the transistor.

### **Collector Feedback Bias:**

In this case, the base resistor is returned back to collector as shown in [fig. 2](#). If temperature increases,  $\beta_{dc}$  increases. This produces more collectors current. As  $I_C$  increases, collector emitter voltage decreases. It means less voltage across  $R_B$  and causes a decrease in base current this decreasing  $I_C$ , and compensating the effect of  $\beta_{dc}$ .



**Fig. 2**

In this circuit, the voltage equation is given by

$$V_{CC} = (I_B + I_C) R_C + I_B R_B + V_{BE}$$

Circuit is stiff sensitive to changes in  $\beta_{dc}$ . The advantage is only two resistors are used.

Then,

$$\frac{\partial I_B}{\partial I_C} = - \frac{R_C}{R_B + R_C}$$

Therefore,

$$S = \frac{1 + \beta_{dc}}{1 + \frac{R_C \beta_{dc}}{R_B + R_C}} < \frac{1 + \beta_{dc}}{1 + \beta \frac{R_E}{R_B}} < 1 + \beta_{dc}$$

It is better as compared to fixed bias circuit.

Further,

$$I_C = \frac{V_{CC} V_{BE}}{R_C + \frac{R_B}{\beta_{dc}}}$$

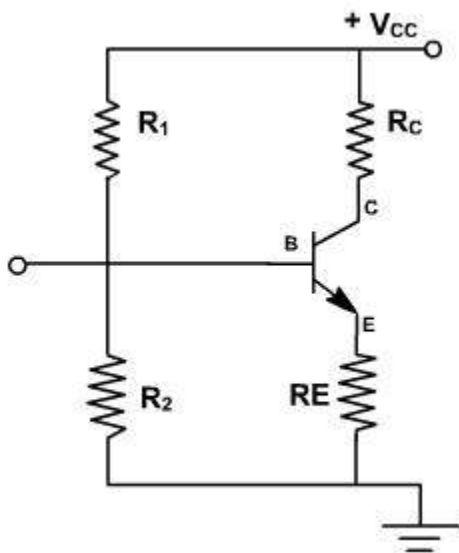
Circuit is still sensitive to changes in  $\beta_{dc}$ . The advantage is only two resistors are used.

## Biasing Techniques

### Voltage Divider Bias:

If the load resistance  $R_C$  is very small, e.g. in a transformer coupled circuit, then there is no improvement in stabilization in the collector to base bias circuit over fixed bias circuit. A circuit which can be used even if there is no dc resistance in series with the collector, is the voltage divider bias or self bias. [fig. 3](#).

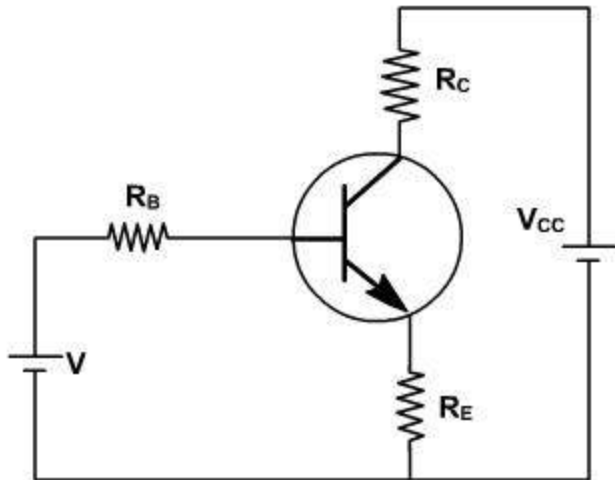
The current in the resistance  $R_E$  in the emitter lead causes a voltage drop which is in the direction to reverse bias the emitter junction. Since this junction must be forward biased, the base voltage is obtained from the supply through  $R_1, R_2$  network. If  $R_b = R_1 \parallel R_2$  equivalent resistance is very – very small, then  $V_{BE}$  voltage is independent of  $I_{C0}$  and  $\frac{\partial I_C}{\partial I_{C0}} \approx 0$ . For best stability  $R_1$  &  $R_2$  must be kept small.



**Fig. 3**

If  $I_C$  tends to increase, because of  $I_{C0}$ , then the current in  $R_C$  increases, hence base current is decreased because of more reverse biasing and it reduces  $I_C$ .

To analysis this circuit, the base circuit is replaced by its thevenin's equivalent as shown in [fig. 4](#).



**Fig. 4**

Thevenin's voltage is

$$V = \frac{R_2}{R_1 + R_2} V_{CC},$$

$$R_b = \frac{R_1 R_2}{R_1 + R_2}$$

$R_b$  is the effective resistance seen back from the base terminal.

$$V = I_B R_b + V_{BE} + (I_B + I_C) R_E$$

If  $V_{BE}$  is considered to be independent of  $I_C$ , then

$$\frac{\partial I_B}{\partial I_C} = - \frac{R_E}{R_b + R_E}$$

$$S = \frac{1 + \beta_{dc}}{1 + \frac{\beta_{dc} R_E}{R_b + R_E}}$$

$$\text{If } \frac{R_b}{R_c} \rightarrow 0 \text{ then } S \rightarrow 1$$

$$\text{If } \frac{R_b}{R_c} \rightarrow \infty \text{ then } S \rightarrow (1 + \beta_{dc}).$$

The smaller the value of  $R_b$ , the better is the stabilization but  $S$  cannot be reduced to unity.

Hence  $I_C$  always increases more than  $I_{C0}$ . If  $R_b$  is reduced, then current drawn from the supply increases. Also if  $R_E$  is increased then to operate at same Q-point, the magnitude of  $V_{CC}$  must be

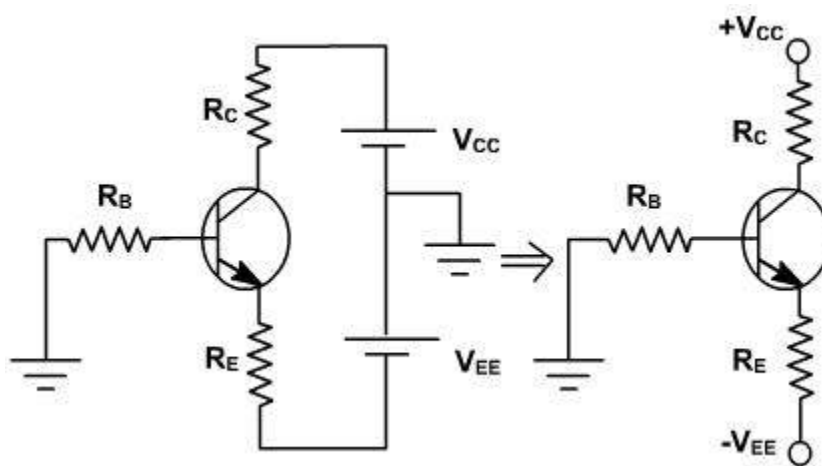
increased. In both the cases the power loss increased and reduced  $h$ .

In order to avoid the loss of ac signal because of the feedback caused by  $R_E$ , this resistance is often bypassed by a large capacitance ( $> 10 \text{ m F}$ ) so that its reactance at the frequency under consideration is very small.

## Biassing Techniques

### Emitter Bias:

[Fig. 5](#), shown the emitter bias circuit. The circuit gets this name because the negative supply  $V_{EE}$  is used to forward bias the emitter junction through resistor  $R_E$ .  $V_{CC}$  still reverse biases collector junction. This also gives the same stability as voltage divider circuit but it is used only if split supply is available.



**Fig. 5**

In this circuit, the voltage equation is given by

$$I_B R_B + V_{BE} + I_E R_E = V_{EE}$$

therefore,  $I_B = \frac{I_C}{\beta_{dc}} = \frac{I_E}{\beta_{dc}}$

and  $I_E = \frac{V_{EE} - V_{BE}}{R_E + \frac{R_B}{\beta_{dc}}}$

If  $I_C$  or  $I_E$  is to be independent of  $\beta$  then  $R_E \gg \frac{R_B}{\beta_{dc}}$

and then  $I_E = \frac{V_{EE} - V_{BE}}{R_E}$

This shown that emitter is virtually at ground potential.

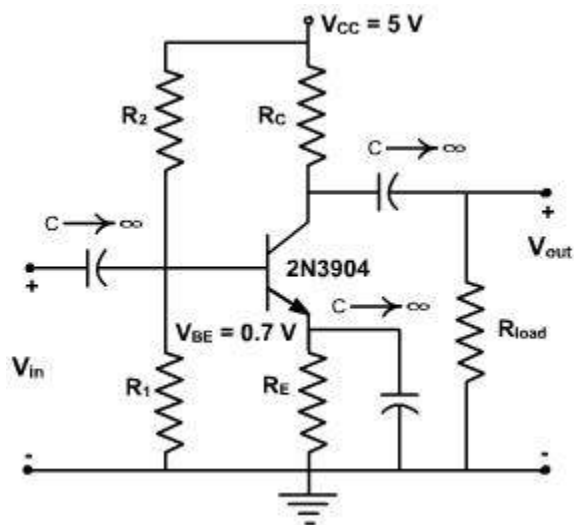
$$V_{CE} = V_{CC} - I_C R_C$$

Normally  $R_B$  is selected less than  $0.01 \beta_{dc} R_E$

## Biasing Techniques

### Example-1

Determine the Q-point for the CE amplifier given in [fig. 1](#), if  $R_1 = 1.5K \Omega$  and  $R_s = 7K \Omega$ . A 2N3904 transistor is used with  $\beta = 180$ ,  $R_E = 100\Omega$  and  $R_C = R_{load} = 1K \Omega$ . Also determine the  $P_{out(ac)}$  and the dc power delivered to the circuit by the source.



**Fig. 1**

**Solution:**

We first obtain the Thevenin equivalent.

$$V_{BB} = \frac{R_1}{R_1 + R_2} V_{CC} = \frac{1500}{1500 + 7000} \cdot 5 = 0.882 \text{ V}$$

and 
$$R_B = \frac{R_1 R_2}{R_1 + R_2} = 1.24 \text{ k}\Omega$$

$$I_{CQ} = \frac{V_{BB} - V_{BE}}{R_B / \beta + R_E} = \frac{0.882 - 0.7}{1240/180 + 100} = 1.70 \text{ mA}$$

Note that this is not a desirable Q-point location since  $V_{BB}$  is very close to  $V_{BE}$ . Variation in  $V_{BE}$  therefore significantly change  $I_C$ . We find  $R_{ac} = R_C \parallel R_{load} = 500 \text{ }\Omega$  and  $R_{dc} = R_C + R_E = 1.1 \text{ k}\Omega$ . The value of  $V_{CEQ}$  representing the quiescent value associated with  $I_{CQ}$  is found as follows,

$$V_{CEQ} = V_{CC} - I_{CQ} R_{dc} = 5 - (1.70 \times 10^{-3}) (1.1 \times 10^3) = 3.13 \text{ V}$$

Then

$$V_{CC} = V_{CEQ} + I_{CQ} R_{ac} = 3.13 + (1.7 \times 10^{-3}) (500) = 3.98 \text{ V}$$

Since the Q-point is on the lower half of the ac load line, the maximum possible symmetrical output voltage swing is

$$2(I_{CQ} - 0) (R_C \parallel R_{load}) = 2 (1.70 \times 10^{-3}) (500) = 1.70 \text{ V}_{\text{peak-peak}}$$

The ac power output can be calculated as

$$P_{out(ac)} = \frac{1}{2} i_{load}^2 R_{load} = \frac{1}{2} \left( 1.70 \times 10^{-3} \times \frac{1000}{2000} \right)^2 \times 1000 = 0.361 \text{ mW}$$

The power drawn from the dc source is given by

$$P_{V_{CC}(dc)} = I_{CQ} V_{CC} + \frac{V_{CC}^2}{R_1 + R_2} = 11.4 \text{ mW}$$

The power loss in the transistor is given by

$$P_{transistor} = V_{CEQ} I_{CQ} = 3.13 \text{ V} \times 1.70 \text{ mA} = 5.32 \text{ mW}$$

The Q-point in this example is not in the middle of the load line so that output swing is not as great as possible. However, if the input signal is small and maximum output is not required, a small  $I_C$  can be used to reduce the power dissipated in the circuit.

## Biasing Techniques

### Moving Ground Around:

Ground is a reference point that can be moved around. e.g. consider a collector feedback bias circuit. The various stages of moving ground are shown in [fig. 2](#).

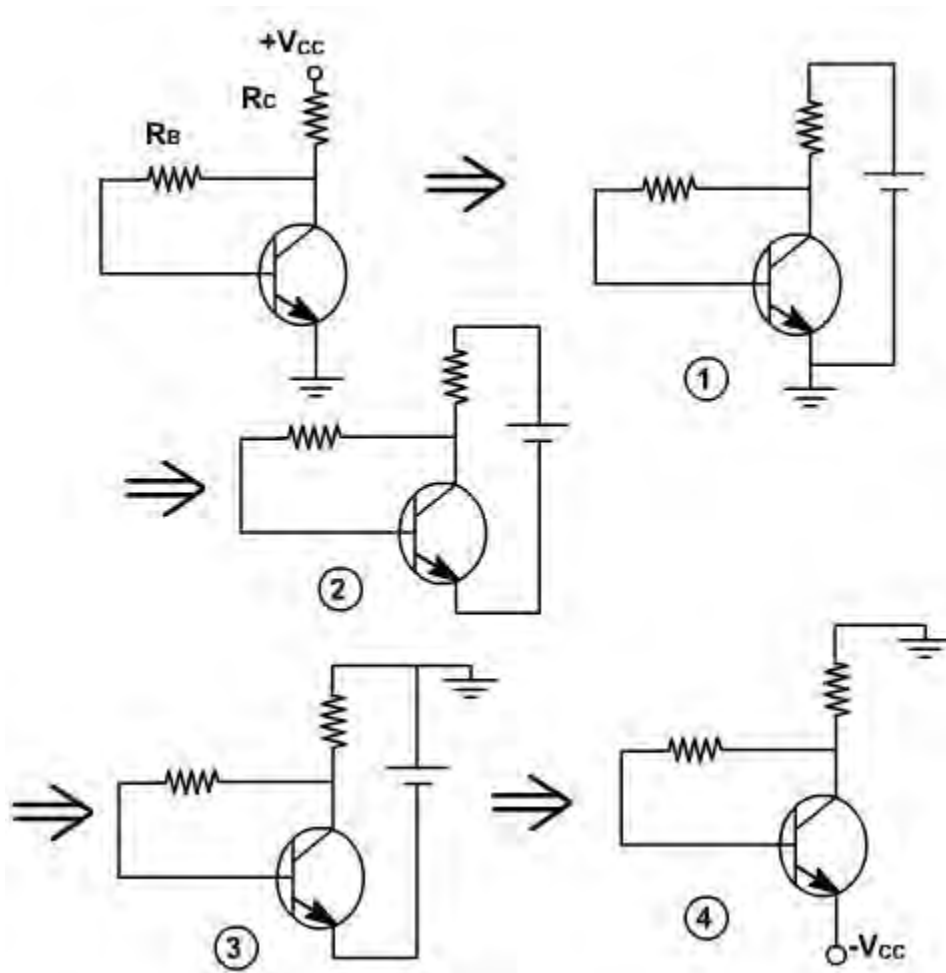
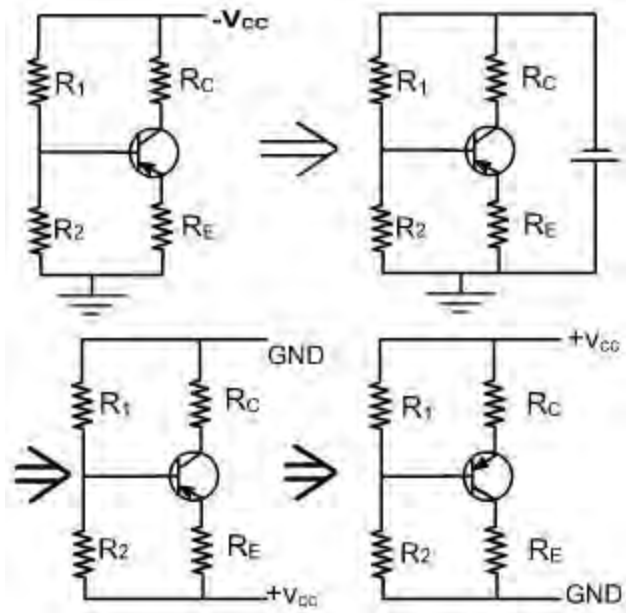


Fig. 2

### Biasing a pnp Transistor:

The biasing of pnp transistor is done similar to npn transistor except that supply is of opposite polarity. The various biasing circuits of pnp transistor are shown in [fig. 3](#).



**Fig. 3**

Example 2:

For the circuit shown in [fig. 4](#), calculate  $I_C$  and  $V_{CE}$

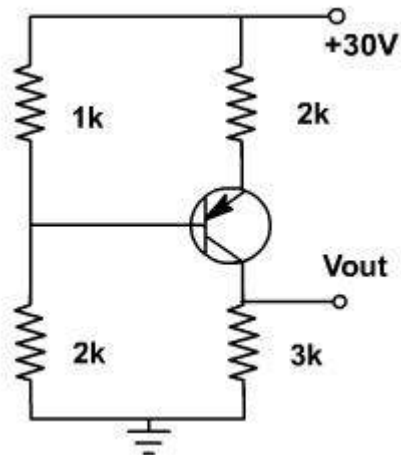
**Solution:**

$$\text{Voltage across } 1\text{K ohm resistor} = \frac{1}{3} \times 30 = 10\text{V}$$

Therefore,

$$I_C \approx I_E = \frac{10 - 0.7}{2\text{K}} = \frac{9.3}{2\text{K}} = 4.65\text{mA}$$

$$\text{Therefore, } V_C = 465 \times 3\text{K} = 13.95\text{V}$$



**Fig. 4**

## Small Signal CE Amplifiers

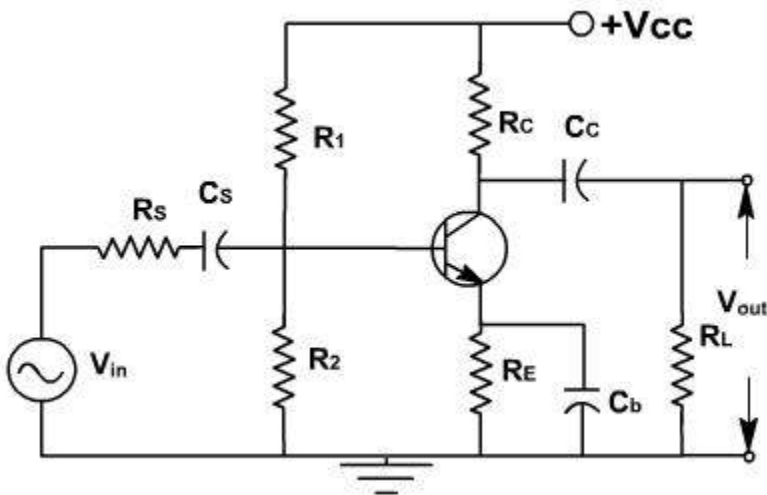
### Small Signal CE Amplifiers:

CE amplifiers are very popular to amplify the small signal ac. After a transistor has been biased with a Q point near the middle of a dc load line, ac source can be coupled to the base. This produces fluctuations in the base current and hence in the collector current of the same shape and frequency. The output will be enlarged sine wave of same frequency.

The amplifier is called linear if it does not change the wave shape of the signal. As long as the input signal is small, the transistor will use only a small part of the load line and the operation will be linear.

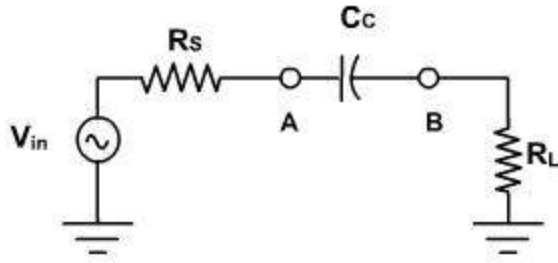
On the other hand, if the input signal is too large. The fluctuations along the load line will drive the transistor into either saturation or cut off. This clips the peaks of the input and the amplifier is no longer linear.

The CE amplifier configuration is shown in [fig. 1](#).



**Fig. 1**

The coupling capacitor ( $C_C$ ) passes an ac signal from one point to another. At the same time it does not allow the dc to pass through it. Hence it is also called blocking capacitor.



**Fig. 2**

For example in [fig. 2](#), the ac voltage at point A is transmitted to point B. For this series reactance  $X_C$  should be very small compared to series resistance  $R_S$ . The circuit to the left of A may be a source and a series resistor or may be the Thevenin equivalent of a complex circuit. Similarly  $R_L$  may be the load resistance or equivalent resistance of a complex network. The current in the loop is given by

$$i = \frac{V_{in}}{\sqrt{(R_S + R_L)^2 + X_C^2}}$$

$$= \frac{V_{in}}{\sqrt{R^2 + X^2}}$$

As frequency increases,  $X_C \left( = \frac{1}{2\pi f C} \right)$  decreases, and current increases until it reaches to its maximum value  $v_{in} / R$ . Therefore the capacitor couples the signal properly from A to B when  $X_C \ll R$ . The size of the coupling capacitor depends upon the lowest frequency to be coupled. Normally, for lowest frequency  $X_C \leq 0.1R$  is taken as design rule.

The coupling capacitor acts like a switch, which is open to dc and shorted for ac.

The bypass capacitor  $C_b$  is similar to a coupling capacitor, except that it couples an ungrounded point to a grounded point. The  $C_b$  capacitor looks like a short to an ac signal and therefore emitter is said ac grounded. A bypass capacitor does not disturb the dc voltage at emitter because it looks open to dc current. As a design rule  $X_{C_b} \leq 0.1R_E$  at lowest frequency.

## Small Signal CE Amplifiers

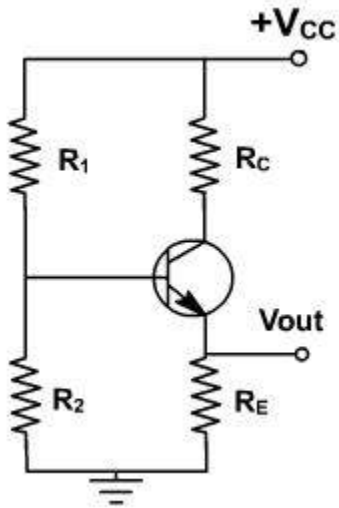
Analysis of CE amplifier:

In a transistor amplifier, the dc source sets up quiescent current and voltages. The ac source then produces fluctuations in these current and voltages. The simplest way to analyze this circuit is to split the analysis in two parts: dc analysis and ac analysis. One can use superposition theorem for

analysis .

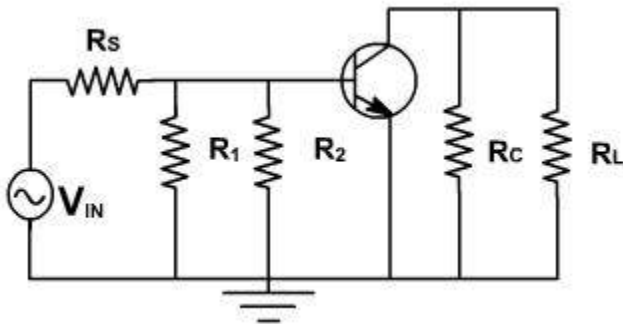
### AC & DC Equivalent Circuits:

For dc equivalent circuit, reduce all ac voltage sources to zero and open all ac current sources and open all capacitors. With this reduced circuit shown in [fig. 3](#) dc current and voltages can be calculated.



**Fig. 3**

For ac equivalent circuits reduce dc voltage sources to zero and open current sources and short all capacitors. This circuit is used to calculate ac currents and voltage as shown in [fig. 4](#).



**Fig. 4**

The total current in any branch is the sum of dc and ac currents through that branch. The total voltage across any branch is the sum of the dc voltage and ac voltage across that branch.

### Phase Inversion:

Because of the fluctuation in base current; collector current and collector voltage also swings

above and below the quiescent voltage. The ac output voltage is inverted with respect to the ac input voltage, meaning it is  $180^\circ$  out of phase with input.

During the positive half cycle base current increase, causing the collector current to increase. This produces a large voltage drop across the collector resistor; therefore, the voltage output decreases and negative half cycle of output voltage is obtained. Conversely, on the negative half cycle of input voltage less collector current flows and the voltage drop across the collector resistor decreases, and hence collector voltage increases we get the positive half cycle of output voltage as shown in [fig. 5](#).

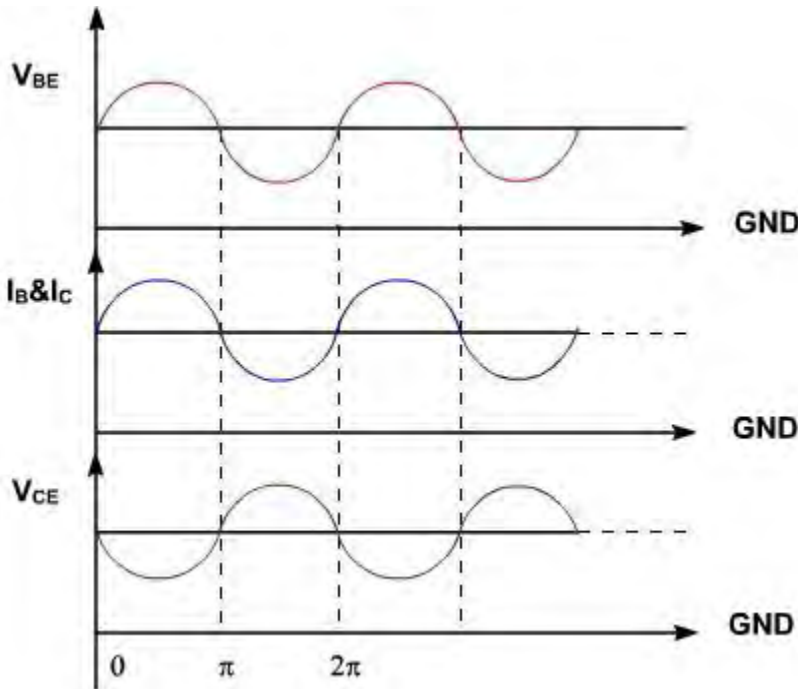
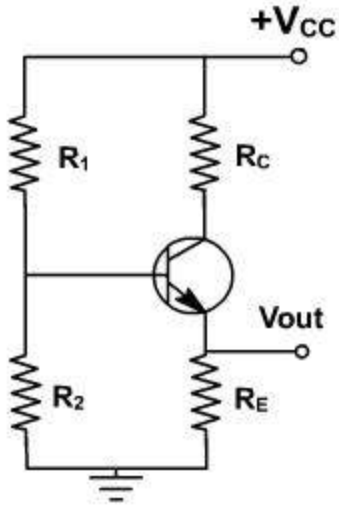


Fig. 5

## Analysis of CE amplifier

AC Load line:

Consider the dc equivalent circuit [fig. 1](#).



**Fig. 1**

Assuming  $I_C = I_C(\text{approx})$ , the output circuit voltage equation can be written as

$$V_{CE} = V_{CC} - I_C(R_C + R_E)$$

and  $I_C = -\frac{V_{CE}}{R_C + R_E} + \frac{V_{CC}}{R_C + R_E}$

$$V_{CE} = 0, I_C = \frac{V_{CC}}{R_C + R_E}$$

and  $I_C = 0, V_{CE} = V_{CC}$

The slope of the d.c load line is  $-\frac{1}{R_C + R_E}$ .

When considering the ac equivalent circuit, the output impedance becomes  $R_C \parallel R_L$  which is less than  $(R_C + R_E)$ .

In the absence of ac signal, this load line passes through Q point. Therefore ac load line is a line of slope  $(-1 / (R_C \parallel R_L))$  passing through Q point. Therefore, the output voltage fluctuations will now be corresponding to ac load line as shown in [fig. 2](#). Under this condition, Q-point is not in the middle of load line, therefore Q-point is selected slightly upward, means slightly shifted to saturation side.

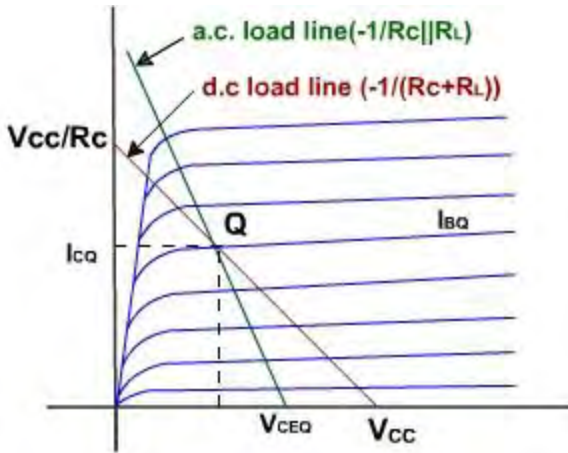


Fig. 2

## Analysis of CE amplifier

### Voltage gain:

To find the voltage gain, consider an unloaded CE amplifier. The ac equivalent circuit is shown in [fig. 3](#). The transistor can be replaced by its collector equivalent model i.e. a current source and emitter diode which offers ac resistance  $r'_e$ .

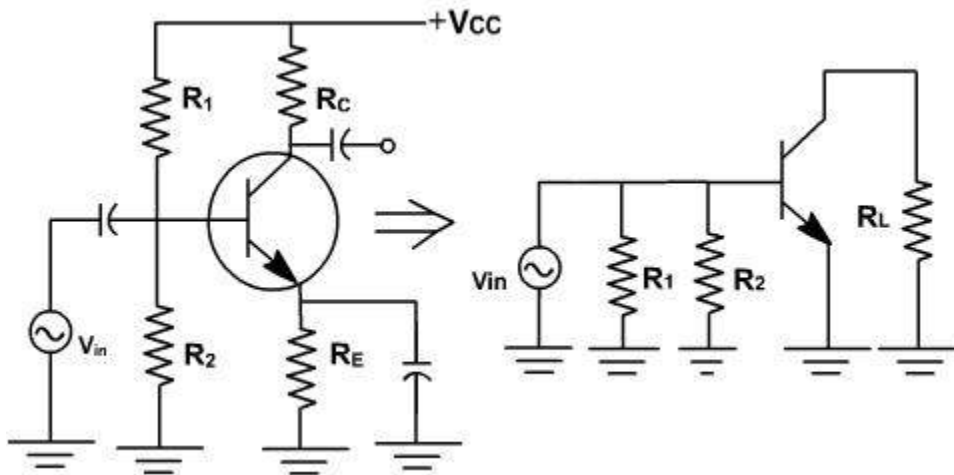


Fig. 3

The input voltage appears directly across the emitter diode.

Therefore emitter current  $i_e = V_{in} / r'_e$ .

Since, collector current approximately equals emitter current and  $i_C = i_e$  and  $v_{out} = - i_e R_C$  (The minus sign is used here to indicate phase inversion)

$$\text{Further } v_{out} = - (V_{in} R_C) / r'_e$$

$$\text{Therefore voltage gain } A = v_{out} / v_{in} = -R_C / r'_e$$

The ac source driving an amplifier has to supply alternating current to the amplifier. The input impedance of an amplifier determines how much current the amplifier takes from the ac source.

In a normal frequency range of an amplifier, where all capacitors look like ac shorts and other reactance are negligible, the ac input impedance is defined as

$$Z_{in} = v_{in} / i_{in}$$

Where  $v_{in}$ ,  $i_{in}$  are peak to peak values or rms values

The impedance looking directly into the base is symbolized  $Z_{in (base)}$  and is given by

$$Z_{in(base)} = v_{in} / i_b ,$$

$$\text{Since, } v_{in} = i_e r'_e$$

$$\gg b_i b r'_e$$

$$Z_{in (base)} = b r'_e .$$

From the ac equivalent circuit, the input impedance  $z_{in}$  is the parallel combination of  $R_1$  ,  $R_2$  and  $b r'_e$ .

$$Z_{in} = R_1 \parallel R_2 \parallel b r'_e$$

The Thevenin voltage appearing at the output is

$$v_{out} = A v_{in}$$

The Thevenin impedance is the parallel combination of  $R_C$  and the internal impedance of the current source. The collector current source is an ideal source, therefore it has an infinite internal impedance.

$$Z_{out} = R_C .$$

The simplified ac equivalent circuit is shown in [fig. 4](#).

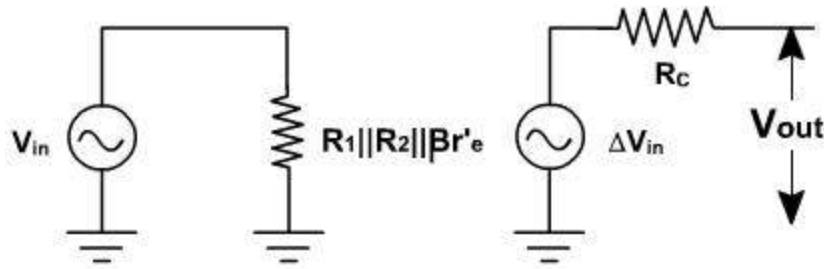


Fig. 4

## h-Parameters

Small signal low frequency transistor Models:

All the transistor amplifiers are two port networks having two voltages and two currents. The positive directions of voltages and currents are shown in [fig. 1](#).

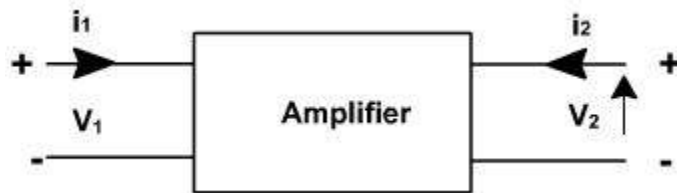


Fig. 1

Out of four quantities two are independent and two are dependent. If the input current  $i_1$  and output voltage  $v_2$  are taken independent then other two quantities  $i_2$  and  $v_1$  can be expressed in terms of  $i_1$  and  $V_2$ .

$$v_1 = f_1(i_1, v_2)$$

$$i_2 = f_2(i_1, v_2)$$

The equations can be written as

$$v_1 = h_{11} i_1 + h_{12} v_2$$

$$i_2 = h_{21} i_1 + h_{22} v_2$$

where  $h_{11}$ ,  $h_{12}$ ,  $h_{21}$  and  $h_{22}$  are called h-parameters.

$$h_{11} = \left. \frac{v_1}{i_1} \right|_{v_2 = 0}$$

=  $h_i$  = input impedance with output short circuit to ac.

$$h_{12} = \frac{v_1}{v_2} \Big|_{i_2 = 0}$$

=  $h_r$  = fraction of output voltage at input with input open circuited or reverse voltage gain with input open circuited to ac (dimensions).

$$h_{21} = \frac{i_2}{i_1} \Big|_{v_2 = 0}$$

=  $h_f$  = negative of current gain with output short circuited to ac.

The current entering the load is negative of  $I_2$ . This is also known as forward short circuit current gain.

$$h_{22} = \frac{i_2}{i_1} \Big|_{i_2 = 0}$$

=  $h_o$  = output admittance with input open circuited to ac.

If these parameters are specified for a particular configuration, then suffixes e,b or c are also included, e.g.  $h_{fe}$ ,  $h_{ib}$  are h parameters of common emitter and common collector amplifiers

Using two equations the generalized model of the amplifier can be drawn as shown in [fig. 2](#).

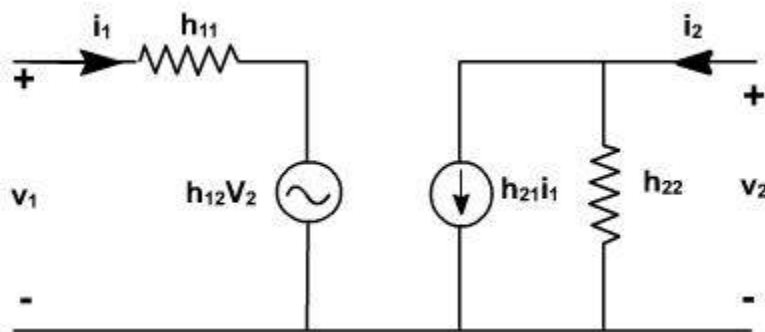
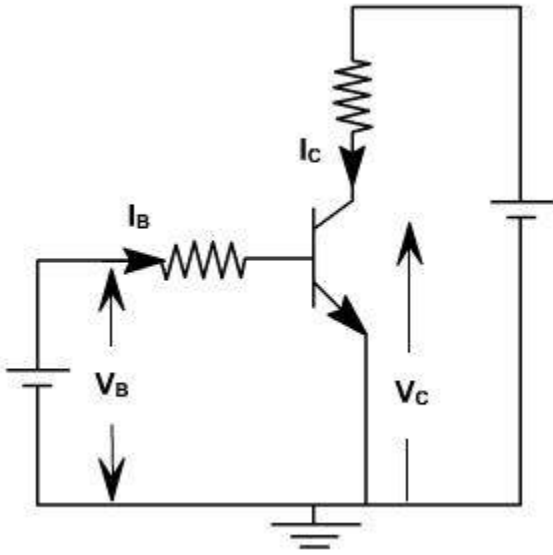


Fig. 2

## h-Parameters

The hybrid model for a transistor amplifier can be derived as follow:

Let us consider CE configuration as show in [fig. 3](#). The variables,  $i_B$ ,  $i_C$ ,  $v_C$ , and  $v_B$  represent total instantaneous currents and voltages  $i_B$  and  $v_C$  can be taken as independent variables and  $v_B$ ,  $I_C$  as dependent variables.



**Fig. 3**

$$v_B = f_1 (i_B, v_C)$$

$$I_C = f_2 (i_B, v_C).$$

Using Taylor 's series expression, and neglecting higher order terms we obtain.

$$\Delta v_B = \left. \frac{\partial f_1}{\partial i_B} \right|_{v_C} \Delta i_B + \left. \frac{\partial f_1}{\partial v_C} \right|_{i_B} \Delta v_C$$

$$\Delta i_C = \left. \frac{\partial f_2}{\partial i_B} \right|_{v_C} \Delta i_B + \left. \frac{\partial f_2}{\partial v_C} \right|_{i_B} \Delta v_C$$

The partial derivatives are taken keeping the collector voltage or base current constant. The  $\Delta v_B$ ,  $\Delta v_C$ ,  $\Delta i_B$ ,  $\Delta i_C$  represent the small signal (incremental) base and collector current and voltage and can be represented as  $v_b$ ,  $i_b$ ,  $v_c$ ,  $i_c$ .

$$\therefore v_b = h_{ie} i_B + h_{re} v_C$$

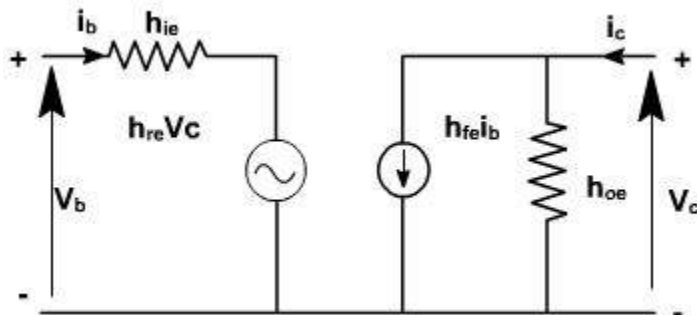
$$i_C = h_{fe} i_B + h_{oe} v_b$$

where

$$h_{ie} = \left. \frac{\partial i_B}{\partial v_B} \right|_{v_C} = \left. \frac{\partial v_B}{\partial i_B} \right|_{v_C}; \quad h_{re} = \left. \frac{\partial i_C}{\partial v_C} \right|_{i_B} = \left. \frac{\partial v_B}{\partial v_C} \right|_{i_B}$$

$$h_{fe} = \left. \frac{\partial i_C}{\partial i_B} \right|_{v_C} = \left. \frac{\partial i_C}{\partial i_B} \right|_{v_C}; \quad h_{oe} = \left. \frac{\partial i_C}{\partial v_C} \right|_{i_B} = \left. \frac{\partial v_B}{\partial v_C} \right|_{i_B}$$

The model for CE configuration is shown in [fig. 4](#).



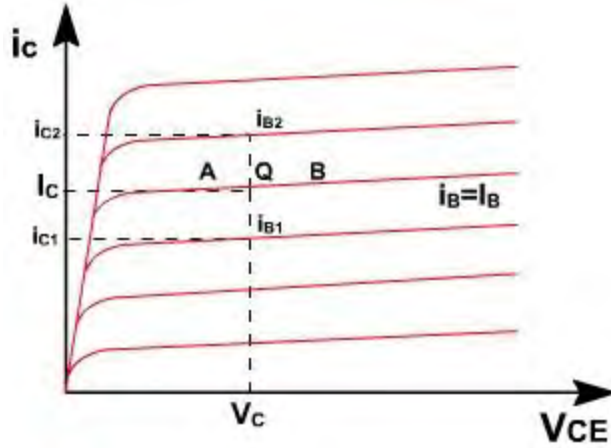
**Fig. 4**

## h-Parameters

### Δετερμινατιον οφ η – παραμετερσ:

To determine the four h-parameters of transistor amplifier, input and output characteristic are used. Input characteristic depicts the relationship between input voltage and input current with output voltage as parameter. The output characteristic depicts the relationship between output voltage and output current with input current as parameter. [Fig. 5](#), shows the output characteristics of CE amplifier.

$$h_{fe} = \left. \frac{\partial i_C}{\partial i_B} \right|_{v_C} = \frac{i_{C2} - i_{C1}}{i_{B2} - i_{B1}}$$



**Fig. 5**

The current increments are taken around the quiescent point Q which corresponds to  $i_B = I_B$  and to the collector voltage  $V_{CE} = V_C$

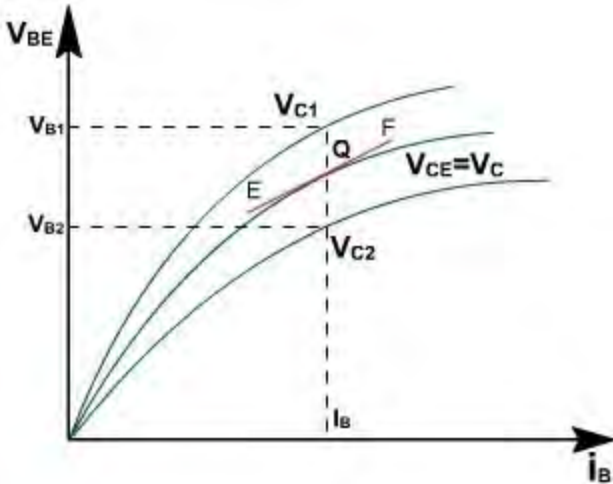
$$h_{oe} = \left. \frac{\partial i_C}{\partial V_C} \right|_{i_B}$$

The value of  $h_{oe}$  at the quiescent operating point is given by the slope of the output characteristic at the operating point (i.e. slope of tangent AB).

$$h_{ie} = \frac{\partial V_B}{\partial i_B} \approx \left. \frac{\Delta V_B}{\Delta i_B} \right|_{V_C}$$

$h_{re}$  is the slope of the appropriate input on [fig. 6](#), at the operating point (slope of tangent EF at Q).

$$h_{re} = \frac{\partial V_B}{\partial V_C} = \left. \frac{\Delta V_B}{\Delta V_C} \right|_{i_B} = \frac{V_{B2} - V_{B1}}{V_{C2} - V_{C1}}$$



**Fig. 6**

A vertical line on the input characteristic represents constant base current. The parameter  $h_{re}$  can be obtained from the ratio  $(V_{B2} - V_{B1})$  and  $(V_{C2} - V_{C1})$  for at  $Q$ .

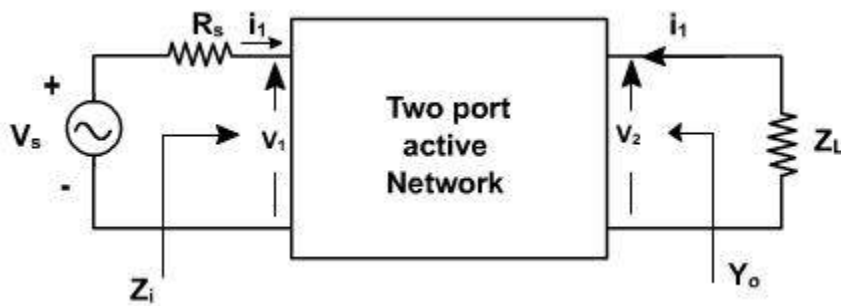
Typical CE h-parameters of transistor 2N1573 are given below:

- $h_{ie} = 1000 \text{ ohm.}$
- $h_{re} = 2.5 * 10^{-4}$
- $h_{fe} = 50$
- $h_{oe} = 25 \text{ m A / V}$

## h-parameters

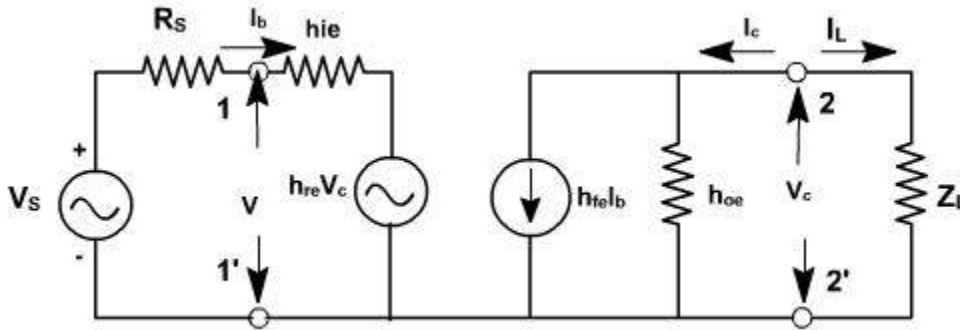
Analysis of a transistor amplifier using h-parameters:

To form a transistor amplifier it is only necessary to connect an external load and signal source as indicated in [fig. 1](#) and to bias the transistor properly.



**Fig. 1**

Consider the two-port network of CE amplifier.  $R_S$  is the source resistance and  $Z_L$  is the load impedance. h-parameters are assumed to be constant over the operating range. The ac equivalent circuit is shown in [fig. 2](#). (Phasor notations are used assuming sinusoidal voltage input). The quantities of interest are the current gain, input impedance, voltage gain, and output impedance.



**Fig. 2**

**Current gain:**

For the transistor amplifier stage,  $A_i$  is defined as the ratio of output to input currents.

$$A_i = \frac{i_L}{i_b} = \frac{-i_c}{i_b} \quad (i_L + i_c = 0 \therefore i_L = -i_c)$$

$$i_c = h_{fe} i_b + h_{oe} V_c$$

$$V_c = i_L Z_L = -i_c Z_L$$

$$\therefore i_c = h_{fe} i_b + h_{oe} (-i_c Z_L)$$

$$\text{or } \frac{i_c}{i_b} = \frac{h_{fe}}{1 + h_{oe} Z_L}$$

$$\therefore A_i = - \frac{h_{fe}}{1 + h_{oe} Z_L}$$

**Input Impedence:**

The impedance looking into the amplifier input terminals ( 1,1' ) is the input impedance  $Z_i$

$$Z_i = \frac{V_b}{I_b}$$

$$V_b = h_{ie} I_b + h_{re} V_c$$

$$\frac{V_b}{I_b} = h_{ie} + h_{re} \frac{V_c}{I_b}$$

$$= h_{ie} - \frac{h_{re} I_c Z_L}{I_b}$$

$$\therefore Z_i = h_{ie} + h_{re} A_i Z_L$$

$$= h_{ie} - \frac{h_{re} h_{fe} Z_L}{1 + h_{oe} Z_L}$$

$$\therefore Z_i = h_{ie} - \frac{h_{re} h_{fe}}{Y_L + h_{oe}} \quad (\text{since } Y_L = \frac{1}{Z_L})$$

## h-parameters

### Voltage gain:

The ratio of output voltage to input voltage gives the gain of the transistors.

$$A_v = \frac{V_c}{V_b} = - \frac{I_c Z_L}{V_b}$$

$$\therefore A_v = \frac{I_b A_i Z_L}{V_b} = \frac{A_i Z_L}{Z_i}$$

### Output Admittance:

It is defined as

$$Y_0 = \left. \frac{i_c}{V_c} \right|_{V_s=0} = 0$$

$$i_c = h_{fe} i_b + h_{oe} V_c$$

$$\frac{i_c}{V_c} = h_{fe} \frac{i_b}{V_c} + h_{oe}$$

when  $V_s = 0$ ,  $R_s \cdot i_b + h_{ie} \cdot i_b + h_{re} V_c = 0$ .

$$\frac{i_b}{V_c} = -\frac{h_{re}}{R_s + h_{ie}}$$

$$\therefore Y_0 = h_{oe} - \frac{h_{re} h_{fe}}{R_s + h_{ie}}$$

Voltage amplification taking into account source impedance ( $R_s$ ) is given by

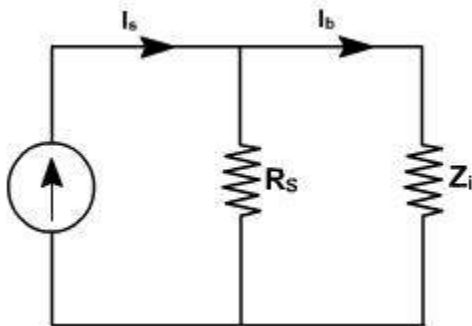
$$A_{vS} = \frac{V_c}{V_s} = \frac{V_c}{V_b} \cdot \frac{V_b}{V_s} \quad \left( V_b = \frac{V_s}{R_s + Z_i} \cdot Z_i \right)$$

$$= A_v \cdot \frac{Z_i}{Z_i + R_s}$$

$$= \frac{A_i Z_L}{Z_i + R_s}$$

$A_v$  is the voltage gain for an ideal voltage source ( $R_s = 0$ ).

Consider input source to be a current source  $I_s$  in parallel with a resistance  $R_s$  as shown in [fig. 3](#).



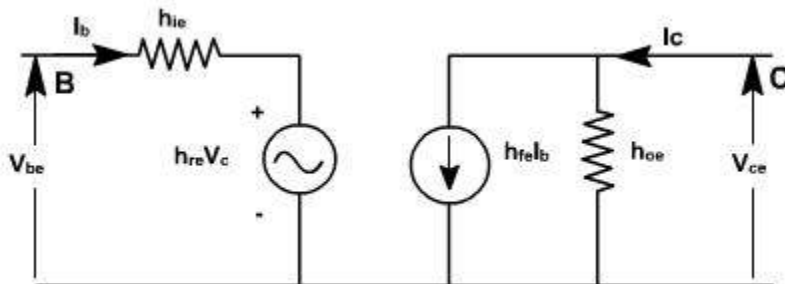
**Fig. 3**

In this case, overall current gain  $A_{IS}$  is defined as

$$\begin{aligned}
 A_{I_s} &= \frac{I_L}{I_s} \\
 &= -\frac{I_c}{I_s} \\
 &= -\frac{I_c + I_b}{I_s} \quad \left( I_b = \frac{I_s * R_s}{R_s + Z_i} \right) \\
 &= A_T * \frac{R_s}{R_s + Z_i} \\
 \text{If } R_s \rightarrow \infty, \quad A_{I_s} &\rightarrow A_T
 \end{aligned}$$

## h-parameters

To analyze multistage amplifier the h-parameters of the transistor used are obtained from manufacture data sheet. The manufacture data sheet usually provides h-parameter in CE configuration. These parameters may be converted into CC and CB values. For example [fig. 4](#)  $h_{rc}$  in terms of CE parameter can be obtained as follows.



**Fig. 4**

For CE transistor configuration

$$V_{be} = h_{ie} I_b + h_{re} V_{ce}$$

$$I_c = h_{fe} I_b + h_{oe} V_{ce}$$

The circuit can be redrawn like CC transistor configuration as shown in [fig. 5](#).

$$V_{bc} = h_{ie} I_b + h_{rc} V_{ec}$$

$$I_c = h_{fe} I_b + h_{oe} V_{ec}$$

$$h_{rc} = \left. \frac{V_{be}}{V_{ec}} \right|_{I_b=0}$$

$$= \left. \frac{V_{be} + V_{ec}}{V_{ec}} \right|_{I_b=0}$$

$$= \left. \left( \frac{V_{be}}{V_{ec}} + 1 \right) \right|_{I_b=0}$$

Since  $I_b = 0$ ,  $V_{be} = h_{re} V_{ec} = -h_{re} V_{ec}$

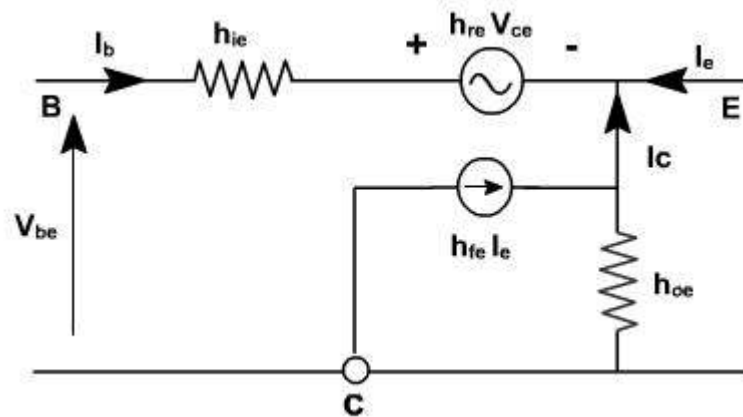
$$\therefore h_{rc} = 1 + \left( \frac{h_{re} V_{ec}}{V_{ec}} \right)$$

$$= 1 - h_{re}$$

Similarly

$$h_{fc} = \left. \frac{I_e}{I_b} \right|_{V_{ec}=0} = \left. \frac{-(I_b + I_c)}{I_b} \right|_{V_{ec}=0}$$

$$= -(1 + h_{fe})$$



## h-parameters

### Example - 1

For the circuits shown in [fig. 1](#). (CE-CC configuration) various h-parameters are given

$h_{ie} = 2K$ ,  $h_{fe} = 50$ ,  $h_{re} = 6 * 10^{-4}$ ,  $h_{oc} = 25 \text{ m A / V}$ .

$h_{ic} = 2K$ ,  $h_{fc} = -51$ ,  $h_{rc} = 1$ ,  $h_{oc} = 25 \text{ m A / V}$ .

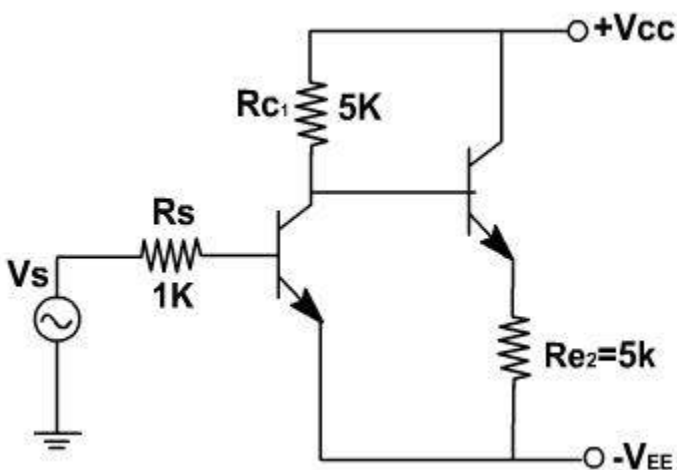
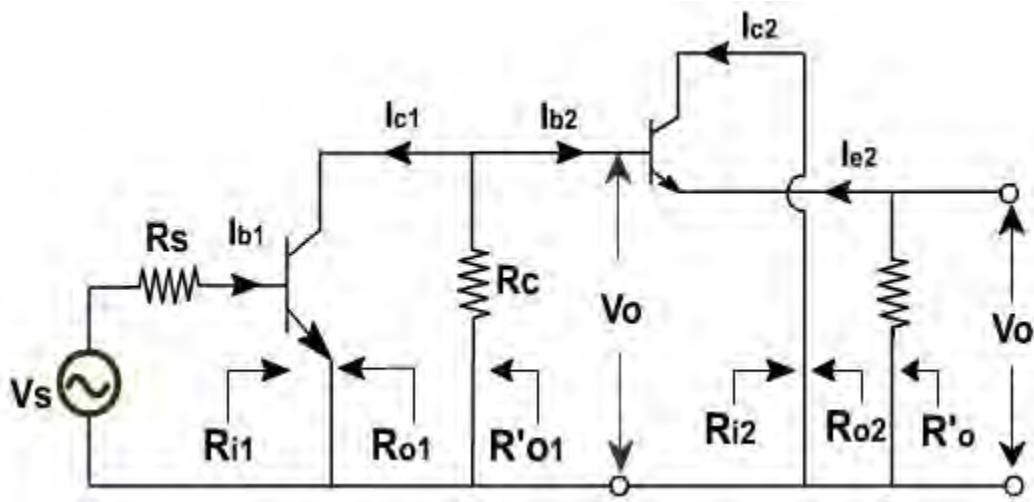


Fig. 1

The small signal model of the transistor amplifier is shown in [fig. 2](#).



**Fig. 2**

In the circuit, the collector resistance of first stage is shunted by the input impedance of last stage. Therefore the analysis is started with last stage. It is convenient; to first compute current gain, input impedance and voltage gain. Then output impedance is calculated starting from first stage and moving towards end.

$$A_{i2} = \frac{-h_{fe}}{1+h_{oc}Z_L} = \frac{51}{1+25 \cdot 10^{-6} \cdot 5 \cdot 10^3}$$

$$= 45.3$$

$$R_{i2} = h_{ic} + h_{rc} A_{i2} Z_L$$

$$= 2 \cdot 10^3 + 1 \cdot 45.3 \cdot 5 \cdot 10^3$$

$$= 228.5K \text{ (high input impedance)}$$

$$A_{v2} = \frac{V_o}{V_2} = \frac{A_v Z_L}{Z_{i2}}$$

$$= \frac{45.3 \cdot 5}{228.5} = 0.99 \approx 1$$

$$R_{L1} = R_{C1} \parallel R_{i2}$$

$$= \frac{5 \cdot 228.5}{5 + 228.5} = 4.9K$$

$$A_{i1} = -\frac{h_{fe}}{1+h_{oe} R_L} = \frac{-50}{1+25 \cdot 10^{-6} \cdot 4.9 \cdot 10^3}$$

$$= 44.5$$

$$R_{i1} = h_{ie} + h_{re} A_{i1} R_{L1}$$

$$= 2 \cdot 6 \cdot 10^{-4} \cdot 44.5 \cdot 4.9$$

$$= 1.87K$$

Voltage gain of first stage is

$$A_{v1} = \frac{A_{i1} R_{L1}}{R_{i1}} = \frac{-44.5 \cdot 4.9}{1.87}$$

$$= -116.6$$

$$Y_{o1} = h_{oe} - \frac{h_{fe} h_{re}}{h_{ie} + R_s}$$

$$= 25 \cdot 10^{-6} - \frac{50 \cdot 6 \cdot 10^{-4}}{2 \cdot 10^3 + 1 \cdot 10^3}$$

$$= 15 \cdot 10^{-6} \text{ mho}$$

$$R_{o1} = \frac{1}{Y_{o1}} = 66.7K$$

$$R'_{o1} = R_{o1} \parallel R_{C1}$$

$$= 66.7 \parallel 5$$

$$= 4.65K$$

The effective source resistance  $R'_{S2}$  for the second stage is  $R_{o1} \parallel R_{C1}$ . Thus  $R_{S2} = R'_{o1} = 4.65K$

## Frequency response

$$\begin{aligned}
 Y_{02} &= h_{oe} - \frac{h_{fe} h_{rc}}{h_{ic} + R_{s2}} \\
 &= 25 \times 10^{-6} - \frac{(-51)(1)}{2 \times 10^3 + 4.65 \times 10^3} \\
 &= 7.70 \times 10^{-3} \text{ A/V}
 \end{aligned}$$

$$R_{02} = \frac{1}{Y_{02}} = 130 \Omega$$

$$\begin{aligned}
 R'_{02} &= R_{02} \parallel R_{c2} \\
 &= 0.13 \parallel 5K \\
 &= 127 \Omega
 \end{aligned}$$

Overall current gain of the amplifier is  $A_i$  and is given by

$$\begin{aligned}
 A_i &= -\frac{i_{e2}}{i_{b1}} \\
 &= -\frac{i_{e2}}{i_{b2}} \cdot \frac{i_{b2}}{i_{c1}} \cdot \frac{i_{c1}}{i_{b1}} \\
 &= -A_{i2} \cdot \frac{i_{b2}}{i_{c1}} \cdot A_{i1}
 \end{aligned}$$

The equivalent circuit of the amplifier is shown in [fig. 3](#). From the circuit it is clear that the current  $i_{c1}$  is divided into two parts.

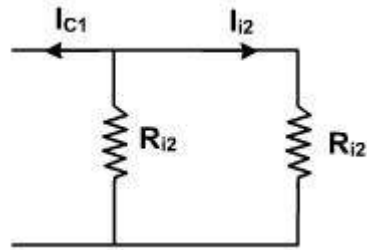


Fig.3

Therefore,

$$\frac{i_{b2}}{i_{c1}} = \frac{-R_{c1}}{R_{c1} + R_{i2}}$$

and

$$\begin{aligned}
 \therefore A_i &= A_{i2} \cdot A_{i1} \cdot \frac{R_{c1}}{R_{c1} + R_{i2}} \\
 &= 45.3 \cdot (-44.5) \cdot \frac{5}{228.5 + 5} = -43.2
 \end{aligned}$$

$$\begin{aligned}
 A_v &= \frac{V_o}{V_i} = \frac{V_o}{V_2} \cdot \frac{V_2}{V_i} \\
 &= A_{v2} \cdot A_{v1} \\
 &= 0.99 \cdot (-11.6) \\
 &= -11.5
 \end{aligned}$$

Overall voltage gain of the amplifier is given by

$$\begin{aligned}
 A_{Vs} &= \frac{V_0}{V_s} = A_V \cdot \frac{R_{i1}}{R_{i1} + R_s} \\
 &= -115 \cdot \frac{1.87}{1.87 + 1} \\
 &= -75.3
 \end{aligned}$$

## h-parameters

### Simplified common emitter hybrid model:

In most practical cases it is appropriate to obtain approximate values of  $A_V$ ,  $A_i$  etc rather than calculating exact values. How the circuit can be modified without greatly reducing the accuracy. [Fig. 4](#) shows the CE amplifier equivalent circuit in terms of h-parameters. Since  $1/h_{oe}$  in parallel with  $R_L$  is approximately equal to  $R_L$  if  $1/h_{oe} \gg R_L$  then  $h_{oe}$  may be neglected. Under these conditions.

$$I_c = h_{fe} I_b .$$

$$h_{re} V_c = h_{re} I_c R_L = h_{re} h_{fe} I_b R_L .$$

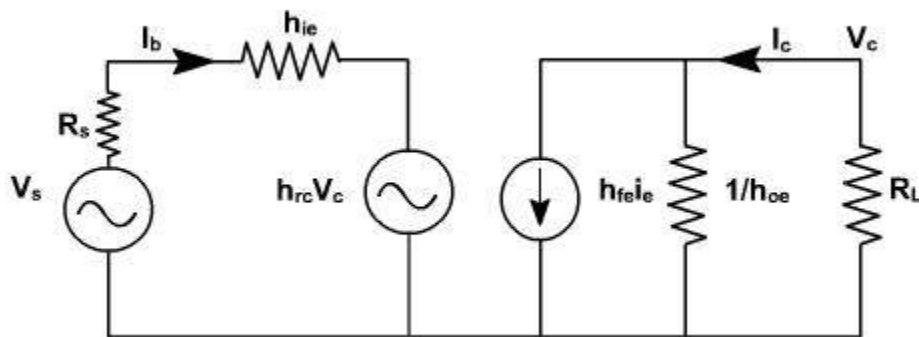


Fig. 4

Since  $h_{fe} \cdot h_{re} \approx 0.01$ , this voltage may be neglected in comparison with  $h_{ie} I_b$  drop across  $h_{ie}$  provided  $R_L$  is not very large. If load resistance  $R_L$  is small than  $h_{oe}$  and  $h_{re}$  can be neglected.

$$A_i = -\frac{h_{fe}}{1+h_{oe} R_L} \approx -h_{fe}$$

$$R_i = h_{ie}$$

$$A_v = \frac{A_i R_L}{R_i} = -\frac{h_{fe} R_L}{h_{ie}}$$

Output impedance seems to be infinite. When  $V_s = 0$ , and an external voltage is applied at the output we find  $I_b = 0$ ,  $I_c = 0$ . True value depends upon  $R_s$  and lies between 40 K and 80K.

On the same lines, the calculations for CC and CB can be done.

### CE amplifier with an emitter resistor:

The voltage gain of a CE stage depends upon  $h_{fe}$ . This transistor parameter depends upon temperature, aging and the operating point. Moreover,  $h_{fe}$  may vary widely from device to device, even for same type of transistor. To stabilize voltage gain  $A_v$  of each stage, it should be independent of  $h_{fe}$ . A simple and effective way is to connect an emitter resistor  $R_e$  as shown in [fig. 5](#). The resistor provides negative feedback and provide stabilization.

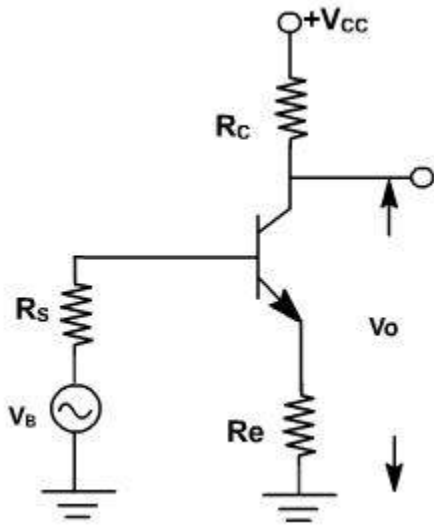


Fig. 5

An approximate analysis of the circuit can be made using the simplified model.

$$\begin{aligned} \text{Current gain } A_i &= \frac{I_L}{I_b} = -\frac{I_C}{I_b} = -\frac{h_{fe} I_b}{I_b} \\ &= -h_{fe} \end{aligned}$$

It is unaffected by the addition of  $R_C$ .

Input resistance is given by

$$\begin{aligned} R_i &= \frac{V_i}{I_b} \\ &= \frac{h_{ie} I_b + (1+h_{fe}) I_b R_e}{I_b} \\ &= h_{ie} + (1+h_{fe}) R_e \end{aligned}$$

The input resistance increases by  $(1+h_{fe}) R_e$

$$A_v = \frac{A_i R_L}{R_i} = \frac{-h_{fe} R_L}{h_{ie} + (1+h_{fe}) R_e}$$

Clearly, the addition of  $R_e$  reduces the voltage gain.

If  $(1+h_{fe}) R_e \gg h_{ie}$  and  $h_{fe} \gg 1$

then

$$A_v = \frac{-h_{fe} R_L}{(1+h_{fe}) R_e} \approx -\frac{R_L}{R_e}$$

Subject to above approximation  $A_v$  is completely stable. The output resistance is infinite for the approximate model.

### Opto Coupler:

It combines a LED and a photo diode in a single package as shown in [fig. 6](#). LED radiates the light depending on the current through LED. This light falls on photo diode and this sets up a reverse current. The advantage of an opto coupler is the electrical isolation between the input and output circuits. The only contact between the input and output is a beam of light. Because of this, it is possible to have an insulation resistance between the two circuits of the order of thousands of mega ohms. They can be used to isolate two circuits of different voltage levels.

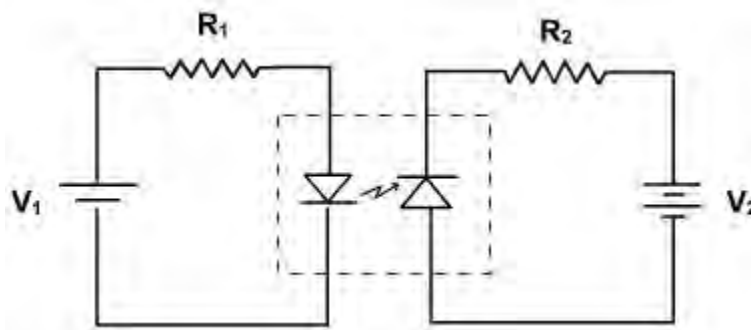


Fig. 6

## TRANSISTOR BIAS STABILITY

### **BJT: Bipolar Junction Transistors:**

Bipolar transistor amplifiers must be properly biased to operate correctly. In circuits made with individual devices (discrete circuits), biasing networks consisting of resistors are commonly employed. Much more elaborate biasing arrangements are used in integrated circuits, for example, bandgap voltage references and current mirrors.

### **Bias circuit requirements:**

For analog circuit operation, the Q-point is placed so the transistor stays in active mode (does not shift to operation in the saturation region or cut-off region) when input is applied. For digital operation, the Q-point is placed so the transistor does the contrary - switches from "on" to "off" state. Often, Q-point is established near the center of active region of transistor characteristic to allow similar signal swings in positive and negative directions. Q-point should be stable. In particular, it should be insensitive to variations in transistor parameters (for example, should not shift if transistor is replaced by another of the same type), variations in temperature, variations in power supply voltage and so forth. The circuit must be practical: easily implemented and cost-effective.

### **Load line:**

The concept of load line is very important in understanding the working of a transistor. It is defined as the locus of operating points on the output characteristics of the transistor. It is the line on which the operating point moves when ac signal is applied to the transistor.

The dc load line gives the value of  $I_C$  and  $V_{CE}$  corresponding to zero signal conditions. The ac load line gives the value of  $I_C$  and  $V_{CE}$  when ac signal is applied. AC load line is steeper than the dc load line but the two intersect at the Q-point determined by biasing dc voltages and currents. AC load line takes into account the ac load resistance while dc load line considers only the dc load resistance.

### **Q Point:**

The operating point of a device, also known as bias point, quiescent point, or Q-point, is the point on the output characteristics that shows the DC collector-emitter voltage ( $V_{ce}$ ) and the collector current ( $I_c$ ) with no input signal applied. The term is normally used in connection with devices such as transistors.

For bipolar junction transistors the bias point is chosen to keep the transistor operating in the active mode, using a variety of circuit techniques, establishing the Q-point DC voltage and current. A small signal is then applied on top of the Q-point bias voltage, thereby either modulating or switching the current, depending on the purpose of the circuit.

The quiescent point of operation is typically near the middle of DC load line. The process of obtaining certain DC collector current at a certain DC collector voltage by setting up operating point is called biasing.

After establishing the operating point, when input signal is applied, the output signal should not move the transistor either to saturation or to cut-off. However, this unwanted shift still might occur, due to the following reasons:

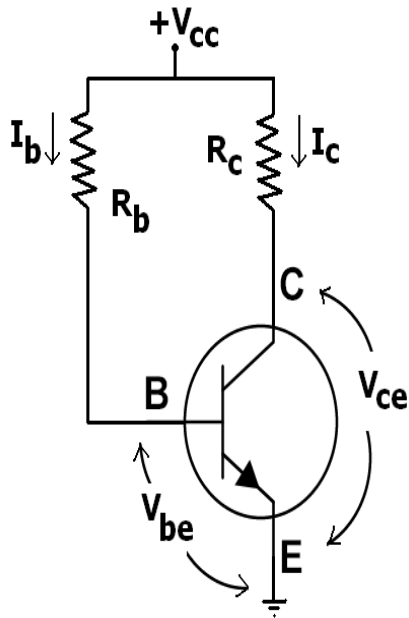
1. Parameters of transistors depend on junction temperature. As junction temperature increases, leakage current due to minority charge carriers ( $I_{CBO}$ ) increases. As  $I_{CBO}$  increases,  $I_{CEO}$  also increases, causing an increase in collector current  $I_C$ . This produces heat at the collector junction. This process repeats, and, finally, Q-point may shift into the saturation region. Sometimes, the excess heat produced at the junction may even burn the transistor. This is known as thermal runaway.
2. When a transistor is replaced by another of the same type, the Q-point may shift, due to changes in parameters of the transistor, such as current gain ( $\beta$ ) which varies slightly for each unique transistor.

To avoid a shift of Q-point, bias-stabilization is necessary. Various biasing circuits can be used for this purpose.

## **VARIOUS METHODS FOR TRANSISTOR BIASING**

- 1. Fixed bias**
- 2. Collector-to-base bias**
- 3. Voltage divider bias**
- 4. Emitter bias**

### 1) Fixed bias (base bias):



This form of biasing is also called *base bias*. In the example image on the right, the single power source (for example, a battery) is used for both collector and base of transistor, although separate batteries can also be used.

In the given circuit,

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

Therefore,

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

For a given transistor,  $V_{be}$  does not vary significantly during use. As  $V_{CC}$  is of fixed value, on selection of  $R_B$ , the base current  $I_B$  is fixed. Therefore this type is called *fixed bias* type of circuit.

Also for given circuit,

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

Therefore,

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

The common-emitter current gain of a transistor is an important parameter in circuit design, and is specified on the data sheet for a particular transistor. It is denoted as  $\beta$  on this page.

Because  $I_C = \beta I_B$

we can obtain  $I_C$  as well. In this manner, operating point given as  $(V_{CE}, I_C)$  can be set for given transistor.

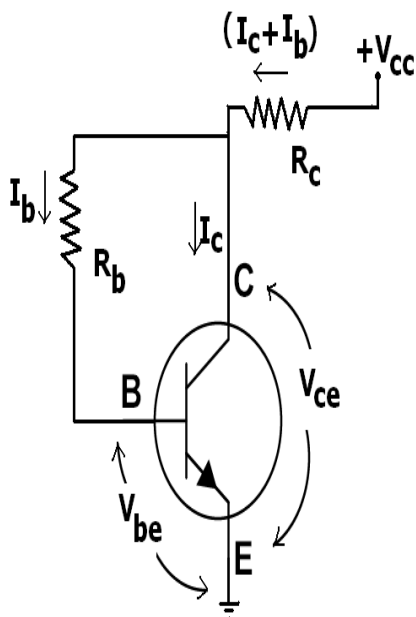
### Merits:

- It is simple to shift the operating point anywhere in the active region by merely changing the base resistor ( $R_B$ ).
- A very small number of components are required.

### Demerits:

- The collector current does not remain constant with variation in temperature or power supply voltage. Therefore the operating point is unstable.
- Changes in  $V_{be}$  will change  $I_B$  and thus cause  $R_E$  to change. This in turn will alter the gain of the stage.
- When the transistor is replaced with another one, considerable change in the value of  $\beta$  can be expected. Due to this change the operating point will shift.

### 2)Collector-to-base bias:



This configuration employs negative feedback to prevent thermal runaway and stabilize the operating point. In this form of biasing, the base resistor  $R_B$  is connected to the collector instead of connecting it to the DC source  $V_{CC}$ . So any thermal runaway will induce a voltage drop across the  $R_C$  resistor that will throttle the transistor's base current.

If  $V_{be}$  is held constant and temperature increases, then the collector current  $I_C$  increases. However, a larger  $I_C$  causes the voltage drop across resistor  $R_C$  to increase, which in turn reduces the voltage across the base resistor  $R_b$ . A lower base-resistor voltage drop reduces the base current  $I_b$ , which results in less collector current  $I_C$ . Because an increase in collector current with temperature is opposed, the operating point is kept stable.

**Merits:**

- Circuit stabilizes the operating point against variations in temperature and  $\beta$  (ie. replacement of transistor)

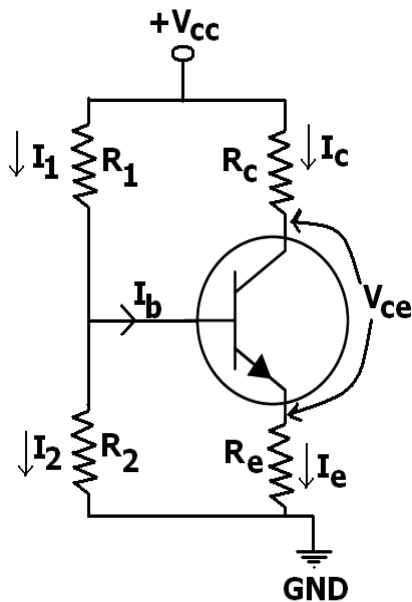
**Demerits:**

- In this circuit, to keep  $I_C$  independent of  $\beta$ , the following condition must be met:

which is the case when

- As  $\beta$ -value is fixed (and generally unknown) for a given transistor, this relation can be satisfied either by keeping  $R_C$  fairly large or making  $R_b$  very low.
  - If  $R_C$  is large, a high  $V_{cc}$  is necessary, which increases cost as well as precautions necessary while handling.
  - If  $R_b$  is low, the reverse bias of the collector–base region is small, which limits the range of collector voltage swing that leaves the transistor in active mode.
- The resistor  $R_b$  causes an AC feedback, reducing the voltage gain of the amplifier. This undesirable effect is a trade-off for greater Q-point stability.

### 3) Voltage divider bias



The voltage divider is formed using external resistors  $R_1$  and  $R_2$ . The voltage across  $R_2$  forward biases the emitter junction. By proper selection of resistors  $R_1$  and  $R_2$ , the operating point of the transistor can be made independent of  $\beta$ . In this circuit, the voltage divider holds the base voltage fixed independent of base current provided the divider current is large compared to the base current. However, even with a fixed base voltage, collector current varies with temperature (for example) so an emitter resistor is added to stabilize the Q-point, similar to the above circuits with emitter resistor.

#### Merits:

- Unlike above circuits, only one dc supply is necessary.
- Operating point is almost independent of  $\beta$  variation.
- Operating point stabilized against shift in temperature.

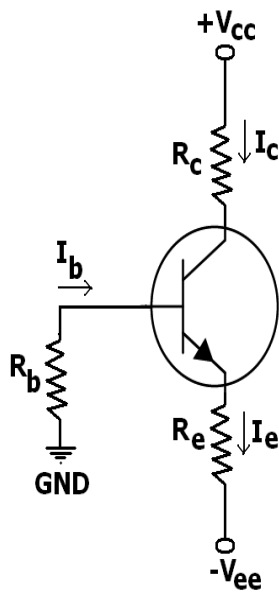
#### Demerits:

- As  $\beta$ -value is fixed for a given transistor, this relation can be satisfied either by keeping  $R_E$  fairly large, or making  $R_1 \parallel R_2$  very low.
  - If  $R_E$  is of large value, high  $V_{CC}$  is necessary. This increases cost as well as precautions necessary while handling.
  - If  $R_1 \parallel R_2$  is low, either  $R_1$  is low, or  $R_2$  is low, or both are low. A low  $R_1$  raises  $V_B$  closer to  $V_C$ , reducing the available swing in collector voltage, and limiting how large  $R_C$  can be made without driving the transistor out of active mode. A low  $R_2$  lowers  $V_{be}$ , reducing the allowed collector current. Lowering both resistor

values draws more current from the power supply and lowers the input resistance of the amplifier as seen from the base.

- AC as well as DC feedback is caused by  $R_E$ , which reduces the AC voltage gain of the amplifier. A method to avoid AC feedback while retaining DC feedback is discussed below.

#### 4)Emitter bias:



When a split supply (dual power supply) is available, this biasing circuit is the most effective, and provides zero bias voltage at the emitter or collector for load. The negative supply  $V_{EE}$  is used to forward-bias the emitter junction through  $R_E$ . The positive supply  $V_{CC}$  is used to reverse-bias the collector junction. Only two resistors are necessary for the common collector stage and four resistors for the common emitter or common base stage.

We know that,

$$V_B - V_E = V_{be}$$

If  $R_B$  is small enough, base voltage will be approximately zero. Therefore emitter current is,

$$I_E = (V_{EE} - V_{be})/R_E$$

The operating point is independent of  $\beta$  if  $R_E \gg R_B/\beta$

**Merit:**

Good stability of operating point similar to voltage divider bias.

**Demerit:**

This type can only be used when a split (dual) power supply is available.

**Stability Factor:**

The degree of success achieved in stabilizing  $I_C$  in the face of variations in  $I_{CO}$  is expressed in terms of stability factor  $S$  and it is defined as the rate of change of  $I_C$  w.r.t.  $I_{CO}$ , keeping  $\beta$  and  $V_{BE}$  constant.

i.e.  $S = dI_C/dI_{CO}$  at constant  $\beta$  and  $V_{BE}$  (or  $I_B$ )

From above expression it is obvious that smaller is the value of  $S$ , higher is the stability. It is desirable to have as low stability factor as possible so as to achieve greater thermal stability. The ideal value of  $S$  is unity (because  $I_C$  includes  $I_{CO}$ ) but it is never possible to achieve it in practice.

Stability Factor in case of CB circuit,

$$S = dI_C/dI_{CO} = d(\beta I_E + I_{CO})/dI_{CO} = 0 + 1 \quad \text{or } S = 1$$

Stability factor in case of CE circuit

$$S = dI_C/dI_{CO} = d[\beta I_B + (1 + \beta) I_{CO}]/dI_{CO} = 1 + \beta \quad \text{taking } I_B \text{ constant.}$$

**Stabilization:**

It is desirable that once selected, the operating (or Q) point should remain stable i.e. the operating point should not shift its position owing to change in temperature etc. Unfortunately it is not possible in practice unless special efforts are made to achieve it. The maintenance of the operating point stable is called the stabilization.

The stabilization of operating point is essential because of

- a) Temperature dependence of  $I_C$
- (b) Individual variations and
- (c) Thermal runaway.

With the increase in temperature, the collector leakage current  $I_{CO}$ , the current gain  $\beta$  tend to increase and  $V_{BE}$  required to produce a given collector current  $I_C$  tends to decrease. Thus increase in temperature tends to cause increase in  $I_C$ .

The value of  $\beta$  and  $V_{BE}$  are not exactly the same for any two transistors even of the same type. So when a transistor is replaced by another one (even of the same type) the operating point (zero signals  $I_C$  and  $V_{CE}$ ) is shifted.

The collector current  $I_C$ , being equal to  $\beta I_B + (1 + \beta) I_{CO}$ , increases with the increase in temperature. This leads to increased power dissipation with further increase in temperature. Being a cumulative process, it can lead to thermal runaway resulting in burn out of the transistor.

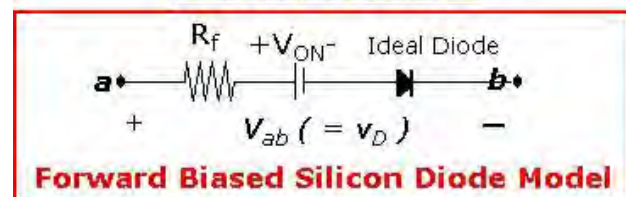
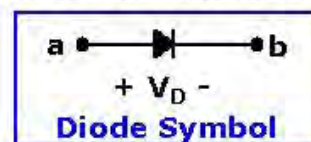
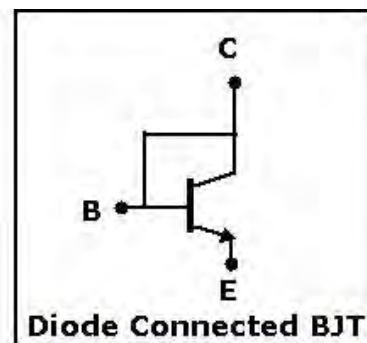
However, if by some modification,  $I_C$  is made to fall with increase in temperature automatically, then decrease in the term  $\beta$  can be made to neutralize the increase in the term  $(1 + \beta) I_{CO}$ , thereby keeping  $I_C$  almost constant. This will achieve thermal stability resulting in bias stability.

The biasing network associated with the transistor should fulfill the requirements of (i) ensuring proper zero signal collector current, (ii) ensuring  $V_{CE}$  not falling below 0.5 V for Ge transistors and 1 V for Si transistors at any instant and (iii) ensuring stabilization of operating point (zero signal  $I_C$  and  $V_{CE}$ )

### Bias Compensation techniques:

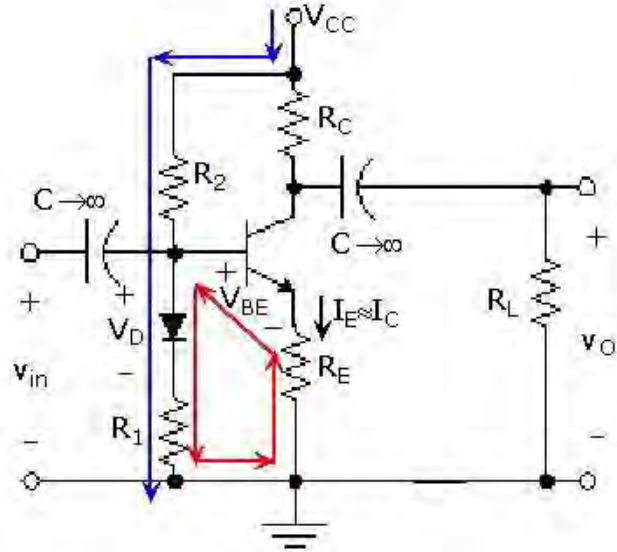
#### Diode Compensation:

In the previous section, we looked at how four parameters ( $V_{CC}$ ,  $V_{BE}$ ,  $I_{CBO}$ , and  $\beta$ ) can affect the total collector current and, therefore, the Q-point of the amplifier. Diode compensation is a technique that is used to reduce the Q-point variations by selecting a diode that has temperature characteristics similar to the transistor. To make sure that the diode and transistor have the same temperature characteristics, we can use a BJT with the same specifications as the amplifier transistor that is diode-connected. This simply involves shorting the collector to the base as shown in the figure to the right for an npn BJT (this can also be done with pnp BJTs). As a reminder, the diode symbol and the model for the forward biased silicon diode are also shown to the right – remember that the current flows in the direction of the arrow and that, unless



an ideal diode is specified, the diode forward resistance and turn-on voltage must be included in all calculations.

The diode is connected in the base circuit of the amplifier as shown in Figure 7.9a, reproduced to the right. The addition of the diode in this manner allows temperature compensation since the  $V_{ON}$  of the diode varies in the same fashion as the  $V_{BE}$  of the transistor. If the transistors used in the amplifier and to construct the diode are matched, the diode characteristics and base-emitter junction characteristics will be the same. Under these circumstances, variations in  $V_{BE}$  due to changes in bias parameters will effectively be significantly reduced or cancelled.



### Sensistor & Thermistor Compensation:

Sensistor is a resistor whose resistance changes with temperature.

The resistance increases exponentially with temperature[1], that is the temperature coefficient is positive (eg. 0.7% per degree Celsius).[2]

Sensistors are used in electronic circuits for compensation of temperature influence or as sensors of temperature for other circuits.[3]

Sensistors are made by using semiconducting silicon and in their operation are similar to PTC-type thermistors.

## 2MARKS

1. Why do we choose q point at the center of the loadline?

The operating point of a transistor is kept fixed usually at the center of the active region in order that the input signal is well amplified. If the point is fixed in the saturation region or the cut off region the positive and negative half cycle gets clipped off respectively.

2. Name the two techniques used in the stability of the q point .explain.

Stabilization technique: This refers to the use of resistive biasing circuit which allows  $I_B$  to vary so as to keep  $I_C$  relatively constant with variations in  $I_{CO}$ ,  $\beta$ , and  $V_{BE}$ .  
Compensation techniques: This refers to the use of temperature sensitive devices such as thermistors diodes. They provide compensating voltages & currents to maintain operating point constant.

3. List out the different types of biasing.

- 1 ) Voltage divider bias
- 2 ) Base bias
- 3 ) Emitter feed back bias
- 4 ) Collector feedback bias

5. What do you meant by thermal runaway?

Due to the self heating at the collector junction, the collector current rises. This causes damage to the device. This phenomenon is called thermal runaway.

6. Why is the transistor called a current controlled device?

The output characteristics of the transistor depend on the input current. So the transistor is called a current controlled device.

7. Define current amplification factor?

It is defined as the ratio of change in output current to the change in input current at constant other side voltage.

8. What are the requirements for biasing circuits?

The Q point must be taken at the Centre of the active region of the output characteristics.

9. When does a transistor act as a switch?

The transistor acts as a switch when it is operated at either cutoff region or saturation region

10. What is biasing?

To use the transistor in any application it is necessary to provide sufficient voltage and current to operate the transistor. This is called biasing.

11. What is operating point?

For the proper operation of the transistor a fixed level of current and voltages are required. This values of currents and voltages defined at a point at which the transistor operate is called operating point.

12. What is stability factor?

Stability factor is defined as the rate of change of collector current with respect to the rate of change of reverse saturation current.

13. What is d.c load line?

The d.c load line is defined as a line on the output characteristics of the transistor which gives the value of  $I_c$  &  $V_{ce}$  corresponding to zero signal condition.

14. What are the advantages of fixed bias circuit?

This is simple circuit which uses a few components. The operating point can be fixed any where on the Centre of the active region

15. Explain about the various regions in a transistor?

The three regions are active region saturation region cutoff region.

16. Explain about the characteristics of a transistor?

Input characteristics: it is drawn between input voltage & input current while keeping output voltage as constant.

Output characteristics: It is drawn between the output voltage & output current while keeping input current as constant.

## UNIT III

### Field Effect Transistor And Its Applications

#### Field Effect Transistor:

The field effect transistor is a semiconductor device, which depends for its operation on the control of current by an electric field. There are two of field effect transistors:

1. JFET (Junction Field Effect Transistor)
2. MOSFET (Metal Oxide Semiconductor Field Effect Transistor)

The FET has several advantages over conventional transistor.

1. In a conventional transistor, the operation depends upon the flow of majority and minority carriers. That is why it is called bipolar transistor. In FET the operation depends upon the flow of majority carriers only. It is called unipolar device.
2. The input to conventional transistor amplifier involves a forward biased PN junction with its inherently low dynamic impedance. The input to FET involves a reverse biased PN junction hence the high input impedance of the order of M-ohm.
3. It is less noisy than a bipolar transistor.
4. It exhibits no offset voltage at zero drain current.
5. It has thermal stability.
6. It is relatively immune to radiation.

The main disadvantage is its relatively small gain bandwidth product in comparison with conventional transistor.

### Operation of FET:

Consider a sample bar of N-type semiconductor. This is called N-channel and it is electrically equivalent to a resistance as shown in [fig. 1](#).

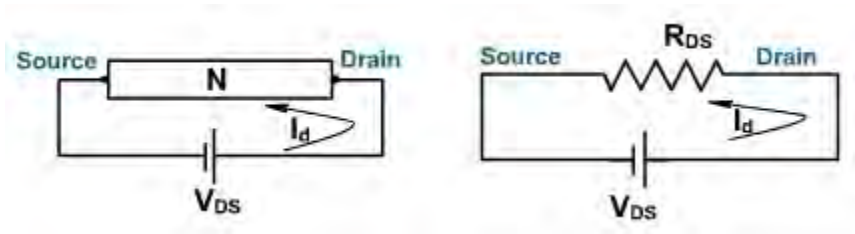


Fig. 1

Ohmic contacts are then added on each side of the channel to bring the external connection. Thus if a voltage is applied across the bar, the current flows through the channel.

The terminal from where the majority carriers (electrons) enter the channel is called source designated by S. The terminal through which majority carriers leaves the channel is called drain and designated by D. For an N-channel device, electrons are the majority carriers. Hence the circuit behaves like a dc voltage  $V_{DS}$  applied across a resistance  $R_{DS}$ . The resulting current is the drain current  $I_D$ . If  $V_{DS}$  increases,  $I_D$  increases proportionally.

Now on both sides of the n-type bar heavily doped regions of p-type impurity have been formed by any method for creating pn junction. These impurity regions are called gates (gate1 and gate2) as shown in [fig. 2](#).

Both the gates are internally connected and they are grounded yielding zero gate source voltage ( $V_{GS} = 0$ ). The word gate is used because the potential applied between gate and source controls the channel width and hence the current.

As with all PN junctions, a depletion region is formed on the two sides of the reverse biased PN junction. The current carriers have diffused across the junction, leaving only uncovered positive ions on the n side and negative ions on the p side. The depletion region width increases with the magnitude of reverse bias. The conductivity of this channel is normally zero because of the unavailability of current carriers.

The potential at any point along the channel depends on the distance of that point from the drain, points close to the drain are at a higher positive potential, relative to ground, then points close to the source. Both depletion regions are therefore subject to greater reverse voltage near the drain. Therefore the depletion region width increases as we move towards drain. The flow of electrons from source to drain is now restricted to the narrow channel between the non-conducting depletion regions. The width of this channel determines the resistance between drain and source.

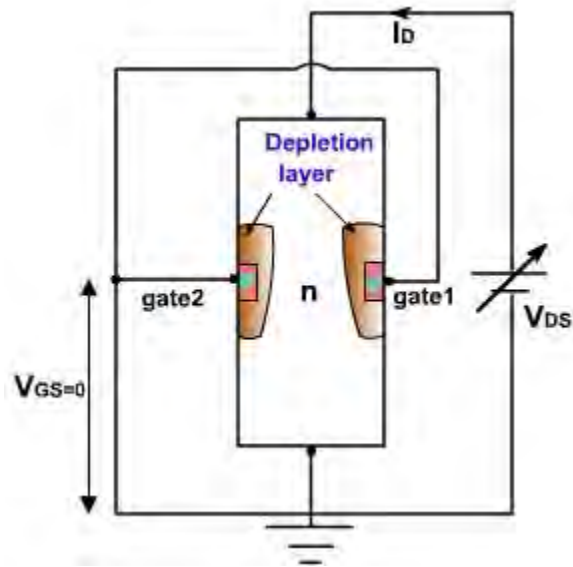


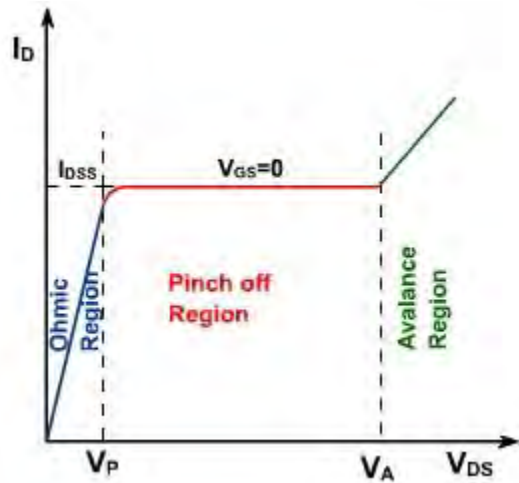
Fig. 2

### Field Effect Transistor

Consider now the behavior of drain current  $I_D$  vs drain source voltage  $V_{DS}$ . The gate source voltage is zero therefore  $V_{GS} = 0$ . Suppose that  $V_{DS}$  is gradually linearly increased from 0V.  $I_D$  also increases.

Since the channel behaves as a semiconductor resistance, therefore it follows ohm's law. The region is called ohmic region, with increasing current, the ohmic voltage drop between the source and the channel region reverse biased the junction, the conducting portion of the channel begins to constrict and  $I_D$  begins to level off until a specific value of  $V_{DS}$  is reached, called the **pinch of voltage  $V_P$** .

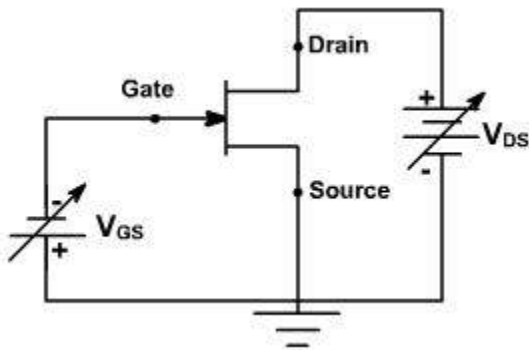
At this point further increase in  $V_{DS}$  do not produce corresponding increase in  $I_D$ . Instead, as  $V_{DS}$  increases, both depletion regions extend further into the channel, resulting in a no more cross section, and hence a higher channel resistance. Thus even though, there is more voltage, the resistance is also greater and the current remains relatively constant. This is called pinch off or saturation region. The current in this region is maximum current that FET can produce and designated by  $I_{DSS}$ . (Drain to source current with gate shorted).



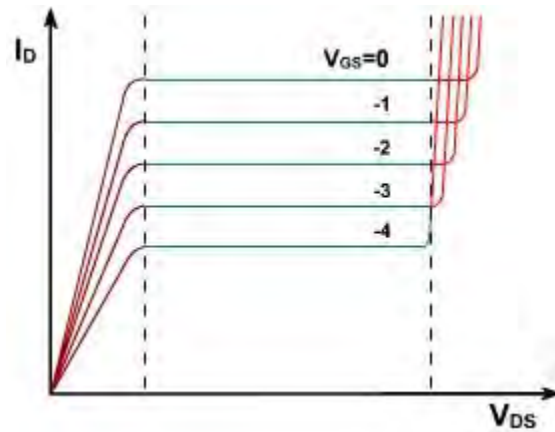
**Fig. 3**

As with all pn junctions, when the reverse voltage exceeds a certain level, avalanche breakdown of pn junction occurs and  $I_D$  rises very rapidly as shown in [fig. 3](#).

Consider now an N-channel JFET with a reverse gate source voltage as shown in [fig. 4](#).



**Fig. 4**



**Fig. 5**

The additional reverse bias, pinch off will occur for smaller values of  $|V_{DS}|$ , and the maximum drain current will be smaller. A family of curves for different values of  $V_{GS}$ (negative) is shown in [fig. 5](#).

Suppose that  $V_{GS} = 0$  and that due of  $V_{DS}$  at a specific point along the channel is +5V with

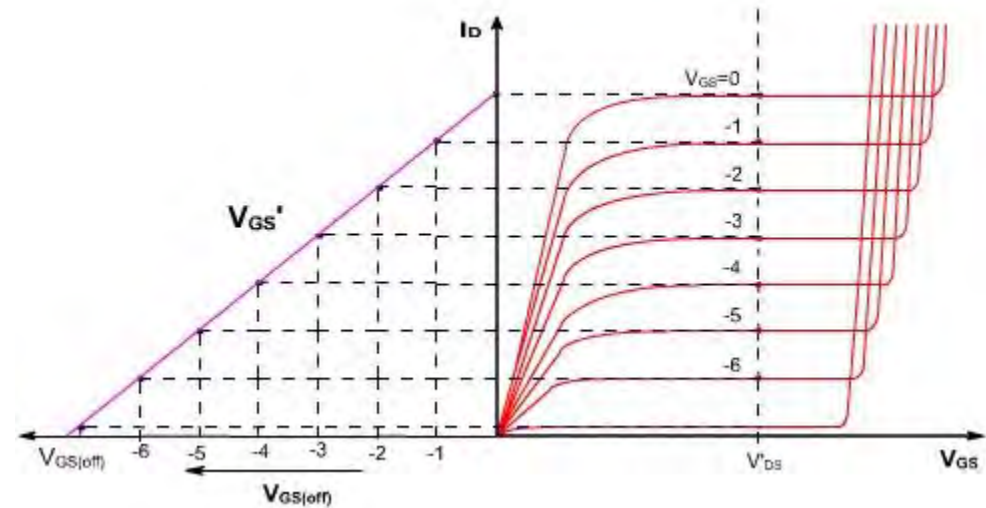
respect to ground. Therefore reverse voltage across either p-n junction is now 5V. If  $V_{GS}$  is decreased from 0 to  $-1V$  the net reverse bias near the point is  $5 - (-1) = 6V$ . Thus for any fixed value of  $V_{DS}$ , the channel width decreases as  $V_{GS}$  is made more negative.

Thus  $I_D$  value changes correspondingly. When the gate voltage is negative enough, the depletion layers touch each other and the conducting channel pinches off (disappears). In this case the drain current is cut off. The gate voltage that produces cut off is symbolized  $V_{GS(off)}$ . It is same as pinch off voltage.

Since the gate source junction is a reverse biased silicon diode, only a very small reverse current flows through it. Ideally gate current is zero. As a result, all the free electrons from the source go to the drain i.e.  $I_D = I_S$ . Because the gate draws almost negligible reverse current the input resistance is very high 10's or 100's of M ohm. Therefore where high input impedance is required, JFET is preferred over BJT. The disadvantage is less control over output current i.e. FET takes larger changes in input voltage to produce changes in output current. For this reason, JFET has less voltage gain than a bipolar amplifier.

**Biassing the Field Effect Transistor**  
**Transductance Curves:**

The transductance curve of a JFET is a graph of output current ( $I_D$ ) vs input voltage ( $V_{GS}$ ) as shown in [fig. 1](#).



**Fig. 1**

By reading the value of  $I_D$  and  $V_{GS}$  for a particular value of  $V_{DS}$ , the transductance curve can be plotted. The transductance curve is a part of parabola. It has an equation of

$$I_D = I_{DSS} \left( 1 - \frac{V_{GS}}{V_{GS(off)}} \right)^2$$

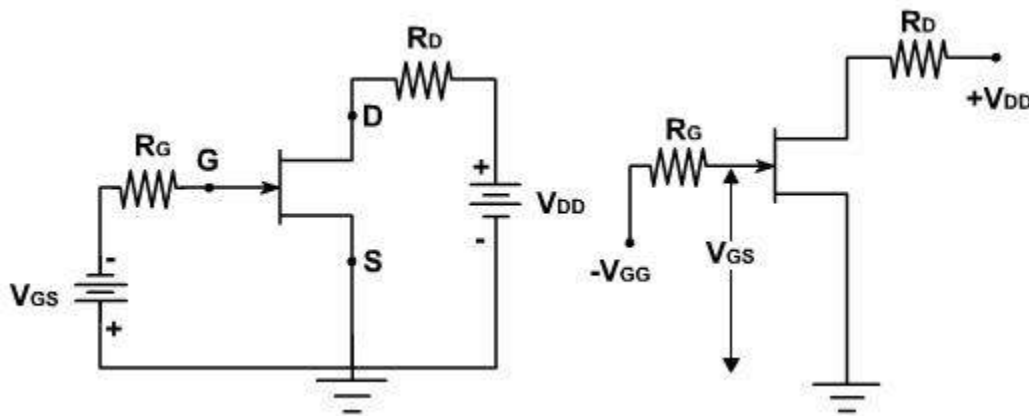
Data sheet provides only  $I_{DSS}$  and  $V_{GS(off)}$  value. Using these values the transconductance curve can be plotted.

### Biasing the FET:

The FET can be biased as an amplifier. Consider the common source drain characteristic of a JFET. For linear amplification, Q point must be selected somewhere in the saturation region. Q point is selected on the basis of ac performance i.e. gain, frequency response, noise, power, current and voltage ratings.

#### Gate Bias:

[Fig. 2](#), shows a simple gate bias circuit.



**Fig. 2**

Separate  $V_{GS}$  supply is used to set up Q point. This is the worst way to select Q point. The reason is that there is considerable variation between the maximum and minimum values of FET parameters e.g.

	$I_{DSS}$	$V_{GS(off)}$
Minimum	4mA	-2V
Maximum	13mA	-8V

This implies that the minimum and maximum transconductance curves are displaced as shown in [fig. 3](#).

Gate bias applies a fixed voltage to the gate. This fixed voltage results in a Q point that is highly sensitive to the particular JFET used. For instance, if  $V_{GS} = -1V$  the Q point may vary from  $Q_1$  to  $Q_2$  depending upon the JFET parameter is use.

At  $Q_1$ ,  $I_D = 0.016 (1 - (1/8))^2 = 12.3 \text{ mA}$

At  $Q_2$ ,  $I_D = 0.004 (1 - (1/2))^2 = 1 \text{ mA}$ .

The variation in drain current is very large.

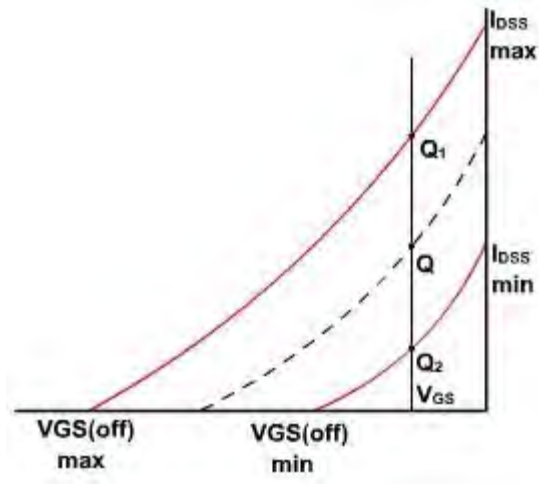


Fig. 3

### Biasing the Field Effect Transistor

#### Self Bias:

Fig. 4, shows a self bias circuit another way to bias a FET. Only a drain supply is used and no gate supply. The idea is to use the voltage across  $R_S$  to produce the gate source reverse voltage.

This is a form of a local feedback similar to that used with bipolar transistors. If drain current increases, the voltage drop across  $R_S$  increases because the  $I_D R_S$  increases. This increases the gate source reverse voltage which makes the channel narrow and reduces the drain current. The overall effect is to partially offset the original increase in drain current. Similarly, if  $I_D$  decreases, drop across  $R_S$  decreases, hence reverse bias decreases and  $I_D$  increases.

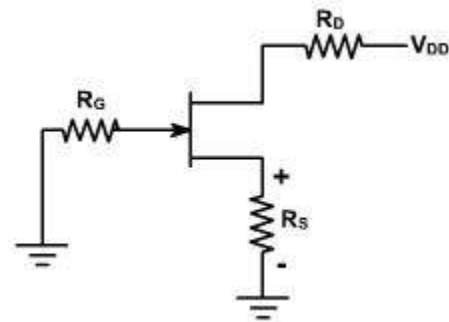


Fig. 4

Since the gate source junction is reverse biased, negligible gate current flows through  $R_G$  and so the gate voltage with respect to ground is zero.

$V_G = 0;$

The source to ground voltage equals the product of the drain current and the source resistance.

$$V_S = I_D R_S.$$

The gate source voltage is the difference between the gate voltage and the source voltage.

$$V_{GS} = V_G - V_S = 0 - I_D R_S$$

$$V_{GS} = -I_D R_S.$$

This means that the gate source voltage equals the negative of the voltage across the source resistor. The greater the drain current, the more negative the gate source voltage becomes.

Rearranging the equation:

$$I_D = -V_{GS} / R_S$$

The graph of this equation is called self bias line as shown in [Fig. 5](#).

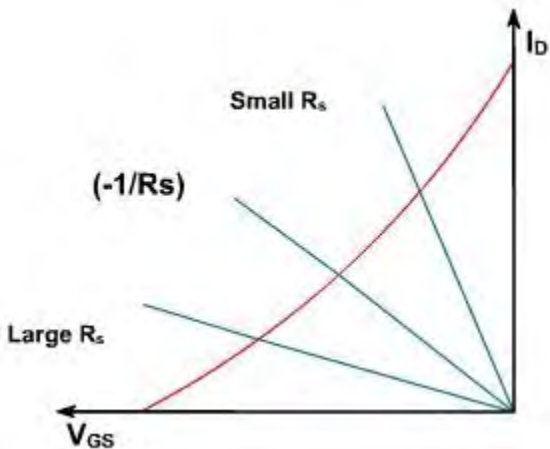
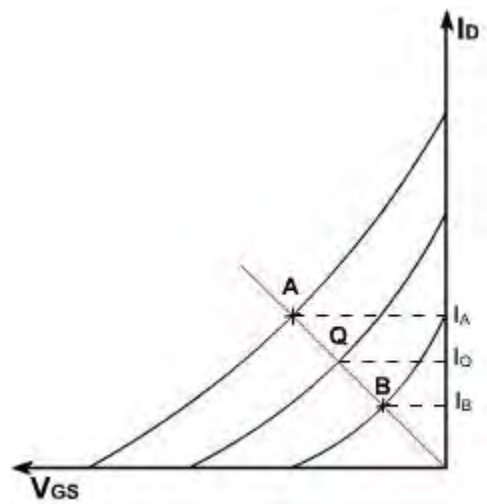


Fig. 5

The operating point on transconductance curve is the intersection of self bias line and transconductance curve. The slope of the line is  $(-1 / R_S)$ . If the source resistance is very large ( $-1 / R_S$  is small) then Q-point is far down the transconductance curve and the drain current is small. When  $R_S$  is small, the Q point is far up the transconductance curve and the drain current is large. In between there is an optimum value of  $R_S$  that sets up a Q point near the middle of the transconductance curve.

The transconductance curve varies widely for FET (because of variation in  $I_{DSS}$  and  $V_{GS(off)}$ ) as shown in [fig. 6](#). The actual curve may be in between there extremes. A and B are the optimum points for the two extreme curves. To find the optimum resistance  $R_S$ , so that Q-point is correct for all the curves, A and B points are joined such that it passes through origin.

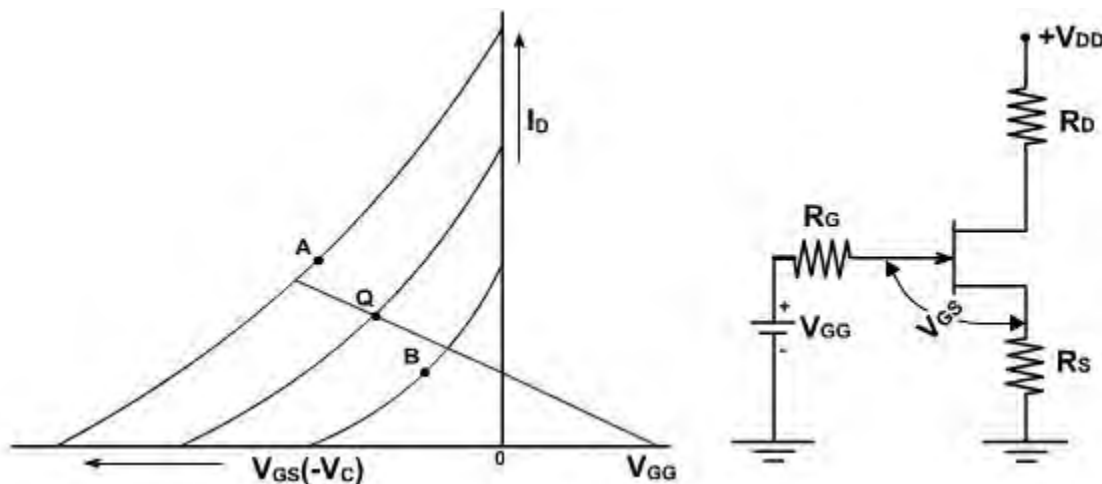


**Fig. 6**

The slope of this line gives the resistance value  $R_S$  ( $V_{GS} = -I_D R_S$ ). The current  $I_Q$  is such that  $I_A > I_Q > I_B$ . Here A, Q and B all points are in straight line.

Consider the case where a line drawn to pass between points A and B does not pass through the origin. The equation  $V_{GS} = -I_D R_S$  is not valid. The equation of this line is  $V_{GS} = V_{GG} - I_D R_S$ .

Such a bias relationship may be obtained by adding a fixed bias to the gate in addition to the source self bias as shown in [fig. 7](#).



**Fig. 7**

In this circuit.

$$V_{GG} = R_S I_G + V_{GS} + I_D R_S$$

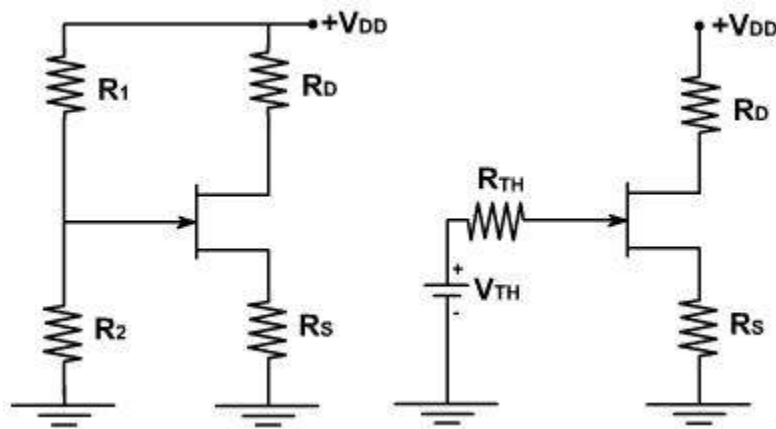
Since  $R_S I_G = 0$ ;

$$V_{GG} = V_{GS} + I_D R_S$$

or  $V_{GS} = V_{GG} - I_D R_S$

### Biasing the Field Effect Transistor Voltage Divider Bias :

The biasing circuit based on single power supply is shown in [fig. 1](#). This is similar to the voltage divider bias used with a bipolar transistor.



**Fig. 1**

The Thevenin voltage  $V_{TH}$  applied to the gate is

$$V_{TH} = \frac{R_2}{R_1 + R_2} V_{DD}$$

The Thevenin resistance is given as

$$R_{TH} = \frac{R_2 R_1}{R_1 + R_2}$$

The gate current is assumed to be negligible.  $V_{TH}$  is the dc voltage from gate to ground.

$$V_{TH} = V_{GS} + V_S \text{ (neglecting } I_G)$$

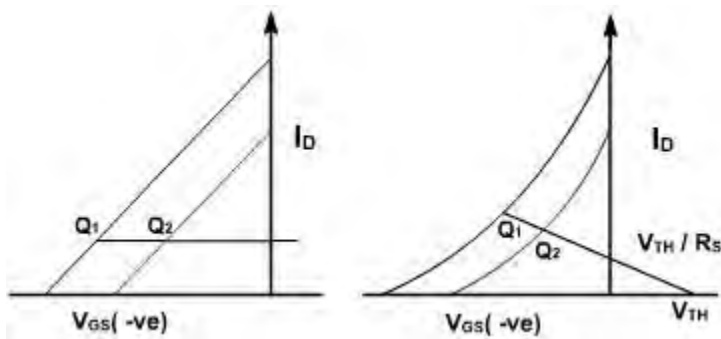
$$\therefore V_S = V_{TH} = V_{GS}$$

The drain current  $I_D$  is given by

$$I_D = \frac{V_{TH} - V_{GS}}{R_S}$$

and the dc voltage from the drain to ground is  $V_D = V_{DD} - I_D R_D$ .

If  $V_{TH}$  is large enough to swamp out  $V_{GS}$  the drain current is approximately constant for any JFET as shown in [fig. 2](#).



**Fig. 2**

There is a problem in JFET. In a BJT,  $V_{BE}$  is approximately 0.7V, with only minor variations from one transistor to another. In a FET,  $V_{GS}$  can vary several volts from one JFET to another. It is therefore, difficult to make  $V_{TH}$  large enough to swamp out  $V_{GS}$ . For this reason, voltage divider bias is less effective with, FET than BJT. Therefore,  $V_{GS}$  is not negligible. The current increases slightly from Q2 to Q1. However, voltage divider bias maintains  $I_D$  nearly constant.

Consider a voltage divider bias circuit shown in [fig. 3](#).

$$V_{GS(\min)} = -1, \quad V_{GS(\max)} = -5V$$

$$V_{TH} = 15V$$

$$I_{D(\min)} = \frac{15 - (-1)}{7.5K} = 2.13 \text{ mA}$$

$$I_{D(\max)} = \frac{15 - (-5)}{7.5K} = 2.67 \text{ mA}$$

Difference in  $I_{D(\min)}$  and  $I_{D(\max)}$  is less

$$V_{D(\max)} = 30 - 2.13 * 4.7 = 20 \text{ V}$$

$$V_{D(\min)} = 30 - 2.67 * 4.7 = 17.5 \text{ V}$$

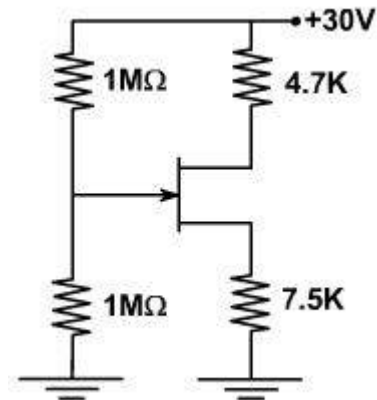


Fig. 3

### Biassing the Field Effect Transistor

#### Current Source Bias:

This is another way to produce solid Q point. The aim is to produce a drain current that is independent of  $V_{GS}$ . Voltage divider bias and self bias attempt to do this by swamping out of variations in  $V_{GS}$ .

#### Using two power supplies:

The current source bias can be used to make  $I_D$  constant [fig. 4](#).

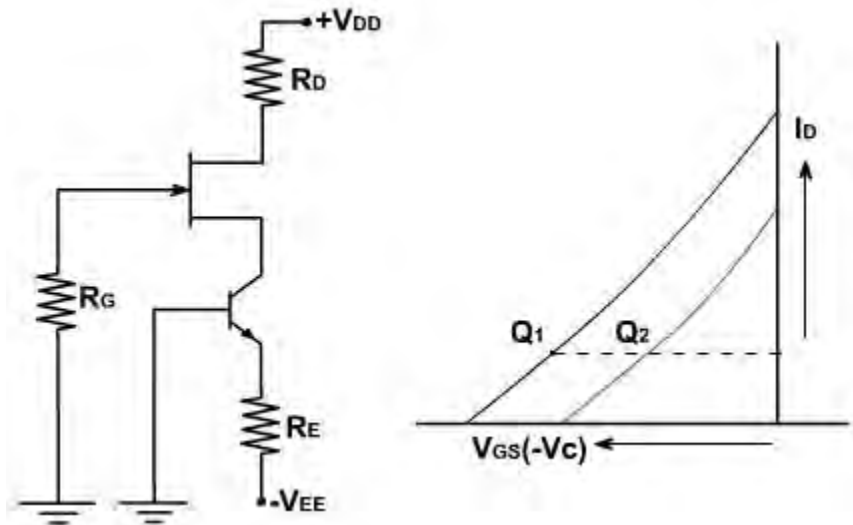


Fig. 4

The bipolar transistor is emitter biased; its collector current is given by

$$I_C = (V_{EE} - V_{BE}) / R_E.$$

Because the bipolar transistor acts like a current source, it forces the drain current to equal the bipolar collector current.

$$I_D = I_C$$

Since  $I_C$  is constant, both Q points have the same value of drain current. The current source effectively wipes out the influence of  $V_{GS}$ . Although  $V_{GS}$  is different for each Q point, it no longer influences the value of drain current.

### Using One power supply:

When only a positive supply is available, the circuit shown in [fig. 5](#), can be used to set up a constant drain current.

In this case, the bipolar transistor is voltage divider biased. Assuming a stiff voltage divider, the emitter and collector currents are constant for all bipolar transistors. This forces the FET drain current equal the bipolar collector current.

$$V_{TH} = \frac{R_2 V_{DD}}{R_1 + R_2}$$

$$I_E = \frac{V_{TH} - V_{BE}}{R_E}$$

Since  $V_{TH}$  is constant,  $I_E$  is also constant  
 $I_C = I_S = I_D = \text{constant}$

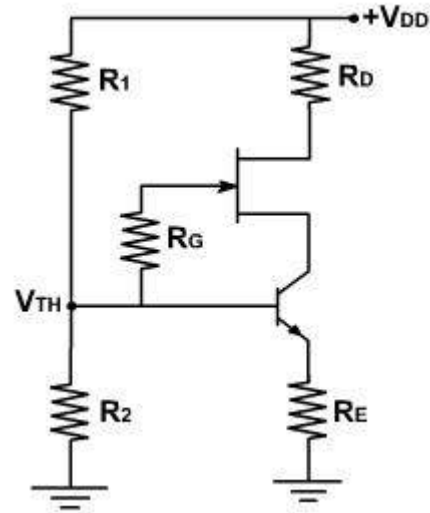


Fig. 5

### Biassing the Field Effect Transistor Transductance:

The transductance of a FET is defined as

$$g_m = \left. \frac{\Delta I_D}{\Delta V_{gs}} \right|_{V_{ds}=0} \quad \mu A/Volts$$

Because the changes in  $I_D$  and  $V_{GS}$  are equivalent to ac current and voltage. This equation can be written as

$$g_m = \left. \frac{i_d}{V_{gs}} \right|_{V_{ds}=0}$$

The unit of  $g_m$  is mho or siemens.

Typical value of  $g_m$  is 2000 m A / V.

The value of  $g_m$  can be obtained from the transductance curve as shown in [fig. 6](#).

If A and B points are considered, than a change in  $V_{GS}$  produces a change in  $I_D$ . The ratio of  $I_D$

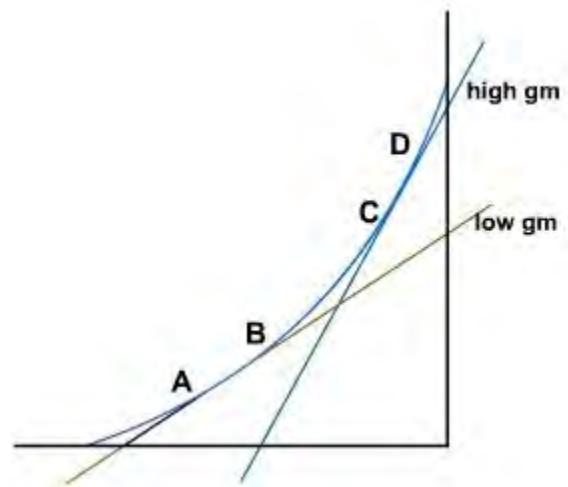


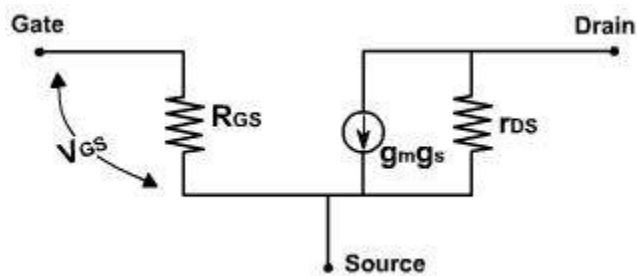
Fig. 6

and  $V_{GS}$  is the value of  $g_m$  between A and B points. If C and D points are considered, then same change in  $V_{GS}$  produces more change in  $I_D$ . Therefore,  $g_m$  value is higher. In a nutshell,  $g_m$  tells us how much control gate voltage has over drain current. Higher the value of  $g_m$ , the more effective is gate voltage in controlling gate current. The second parameter  $r_d$  is the drain resistance.

$$r_d = \left. \frac{V_{ds}}{i_d} \right|_{V_{gs}=0} \quad (r_d \text{ is negligible})$$

### FET a amplifier

Similar to Bipolar junction transistor. JFET can also be used as an amplifier. The ac equivalent circuit of a JFET is shown in [fig. 1](#).



**Fig. 1**

The resistance between the gate and the source  $R_{GS}$  is very high. The drain of a JFET acts like a current source with a value of  $g_m V_{gs}$ . This model is applicable at low frequencies.

From the ac equivalent model

$$i_d = g_m V_{gs} + \frac{V_{ds}}{r_d}$$

$$\text{When } i_d = 0, \quad \frac{V_{ds}}{V_{gs}} = -g_m r_d$$

The amplification factor  $\mu$  for FET is defined as

$$\mu = \left. \frac{V_{ds}}{V_{gs}} \right|_{i_d=0} \quad \therefore \mu = g_m r_d$$

When  $V_{GS} = 0$ ,  $g_m$  has its maximum value. The maximum value is designated as  $g_{m0}$ .

Again consider the equation,

$$I_D = I_{DSS} \left[ 1 - \frac{V_{GS}}{V_{GS(off)}} \right]^2$$
$$g_m = \frac{\partial I_D}{\partial V_{GS}} = 2I_{DSS} \left[ 1 - \frac{V_{GS}}{V_{GS(off)}} \right] \left[ \frac{-1}{V_{GS(off)}} \right]$$
$$g_m = \frac{-2I_{DSS}}{V_{GS(off)}} \left[ 1 - \frac{V_{GS}}{V_{GS(off)}} \right]$$

$$\text{When } V_{GS} = 0, g_m = g_{m0} = \frac{-2I_{DSS}}{V_{GS(off)}}$$

$$\therefore g_m = g_{m0} \left[ 1 - \frac{V_{GS}}{V_{GS(off)}} \right]$$

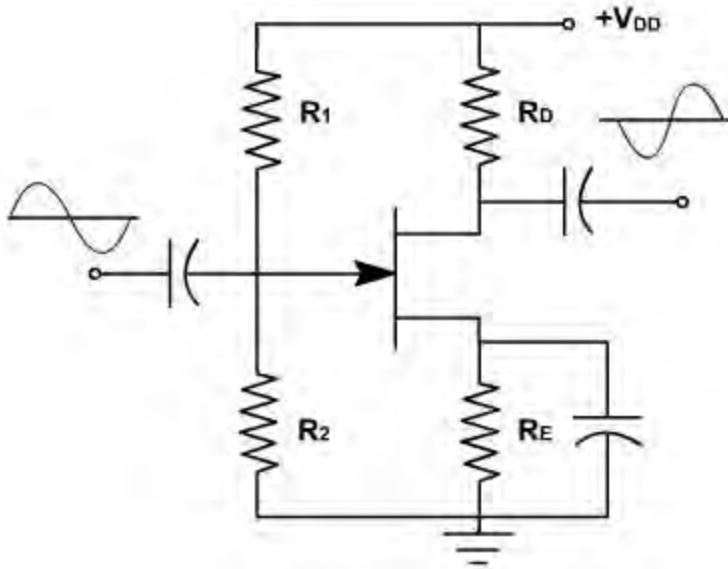
As  $V_{GS}$  increases,  $g_m$  decreases linearly.

$$V_{GS(off)} = \frac{-2I_{DSS}}{g_{m0}}$$

Measuring  $I_{DSS}$  and  $g_m$ ,  $V_{GS(off)}$  can be determined

## FET as amplifier

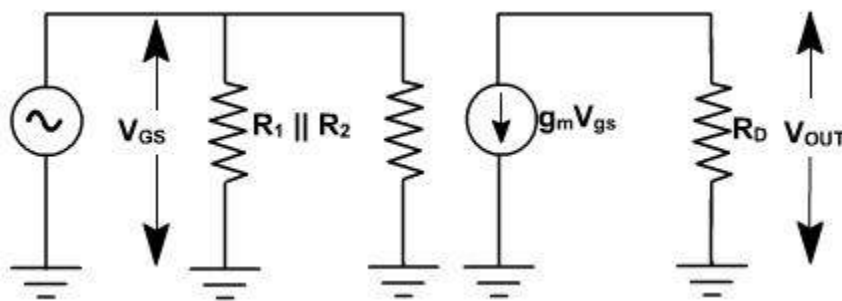
[Fig. 2](#), shows a common source amplifier.



**Fig. 2**

When a small ac signal is coupled into the gate it produces variations in gate source voltage. This produces a sinusoidal drain current. Since an ac current flows through the drain resistor. An amplified ac voltage is obtained at the output. An increase in gate source voltage produces more drain current, which means that the drain voltage is decreasing. Since the positive half cycle of input voltage produces the negative half cycle of output voltage, we get phase inversion in a CS amplifier.

The ac equivalent circuit is shown in [fig. 3](#).



**Fig. 3**

The ac output voltage is

$$V_{out} = - g_m V_{gs} R_D$$

Negative sign means phase inversion. Because the ac source is directly connected between the

gate source terminals therefore ac input voltage equals

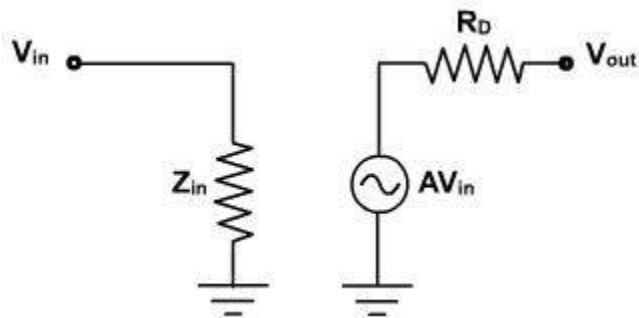
$$V_{in} = V_{gs}$$

The voltage gain is given by

$$A_V = \frac{v_{out}}{v_{in}} = -g_m R_D$$

$A_V$  = unloaded voltage gain

The further simplified model of the amplifier is shown in [fig. 4](#).



**Fig. 4**

$Z_{in}$  is the input impedance. At low frequencies, this is parallel combination of  $R_1 || R_2 || R_{GS}$ . Since  $R_{GS}$  is very large, it is parallel combination of  $R_1$  &  $R_2$ .  $AV_{in}$  is output voltage and  $R_D$  is the output impedance.

### **FET as amplifier**

Because of nonlinear transductance curve, a JFET distorts large signals, as shown in [fig. 5](#).

Given a sinusoidal input voltage, we get a non-sinusoidal output current in which positive half cycle is elongated and negative cycle is compressed. This type of distortion is called Square law distortion because the transductance curve is parabolic.

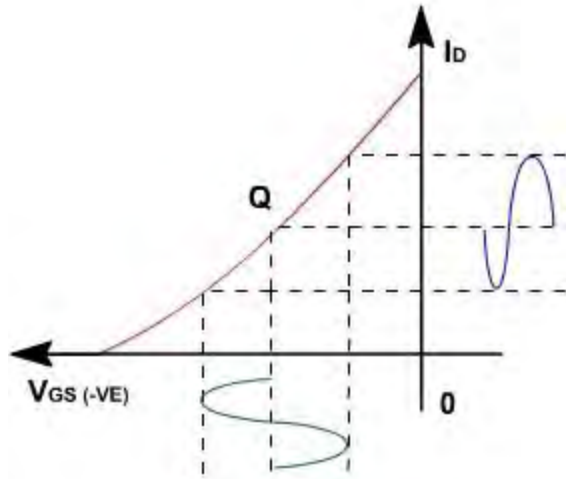


Fig. 5

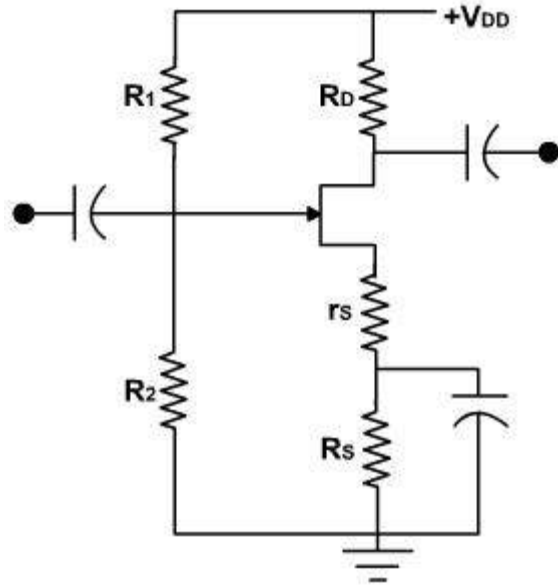


Fig. 6

This distortion is undesirable for an amplifier. One way to minimize this is to keep the signal small. In that case a part of the curve is used and operation is approximately linear. Some times swamping resistor is used to minimize distortion and gain constant. Now the source is no longer ac ground as shown in [fig. 6](#).

The drain current through  $r_s$  produces an ac voltage between the source and ground. If  $r_s$  is large enough the local feedback can swamp out the non-linearity of the curve. Then the voltage gain approaches an ideal value of  $R_D / r_s$ .

Since  $R_{GS}$  approaches infinity therefore, all the drain current flows through  $r_s$  producing a voltage drop of  $g_m V_{gs} r_s$ . The ac equivalent circuit is shown in [fig. 7](#).

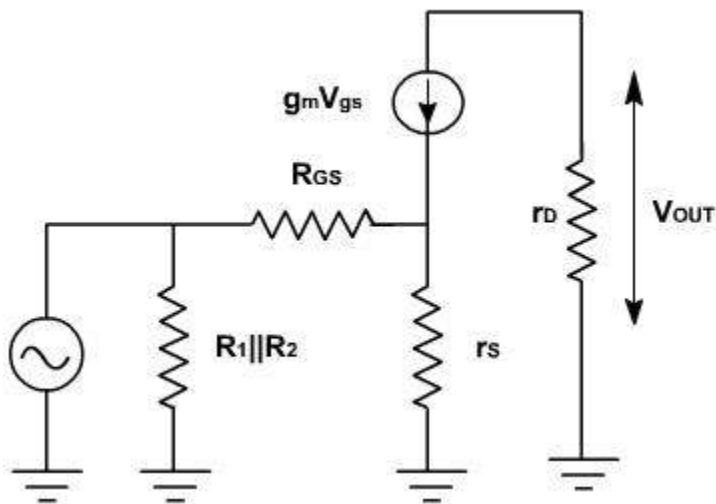


Fig. 7

$$v_{gs} + g_m v_{gs} \cdot r_s - v_{in} = 0$$

$$v_{in} = (1 + g_m r_s) v_{gs}$$

$$v_{out} = -g_m R_D v_{gs}$$

$$A = \frac{-g_m R_D}{1 + g_m r_s} = \frac{-R_D}{r_s + 1/g_m}$$

The voltage gain reduces but voltage gain is less effective by change in  $g_m$ .  $r_s$  must be greater than  $1 / g_m$  only then

$$v_{gs} = -\frac{R_D}{r_s} v_{out}$$

## JFET Applications

### Example-1:

Determine  $g_m$  for an n-channel JFET with characteristic curve shown in [fig. 1](#).

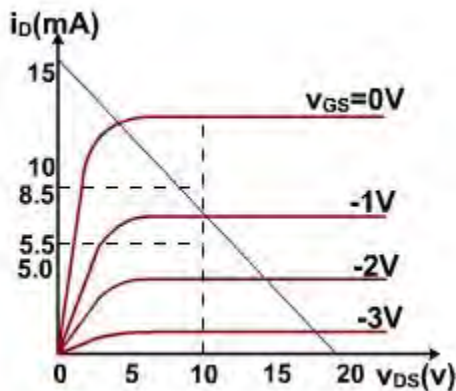


Fig. 1

### Solution:

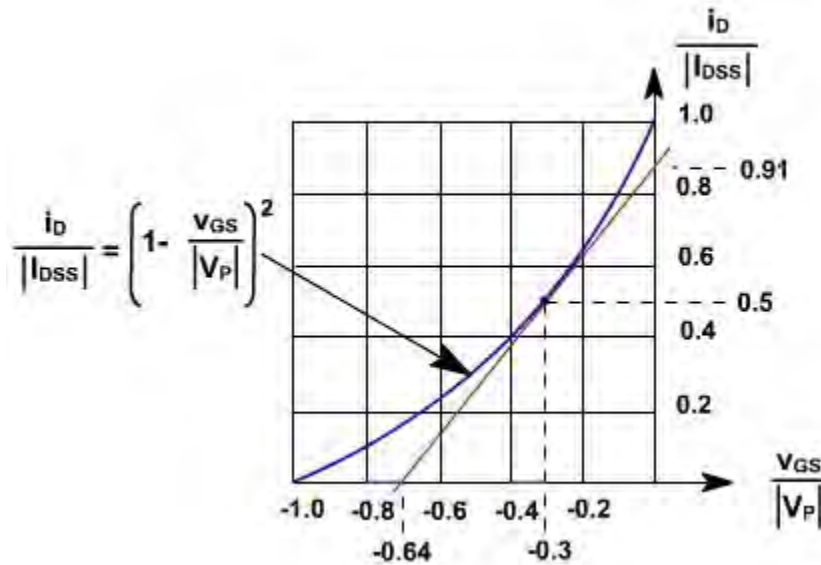
We select an operating region which is approximately in the middle of the curves; that is, between  $v_{GS} = -0.8$  V and  $v_{GS} = -1.2$  V;  $i_D = 8.5$  mA and  $i_D = 5.5$  mA. Therefore, the transconductance of the JFET is given by

$$g_m = \left. \frac{\Delta i_D}{\Delta v_{GS}} \right|_{v_{GS} = \text{constant}} = 7.5 \text{ m}\Omega^{-1}$$

### Design of JFET amplifier:

To design a JFET amplifier, the Q point for the dc bias current can be determined graphically. The dc bias current at the Q point should lie between 30% and 70% of  $I_{DSS}$ . This locates the Q point in the linear region of the characteristic curves.

The relationship between  $i_D$  and  $v_{GS}$  can be plotted on a dimensionless graph (i.e., a normalized curve) as shown in [fig. 2](#).



**Fig. 2**

The vertical axis of this graph is  $i_D / I_{DSS}$  and the horizontal axis is  $v_{GS} / V_P$ . The slope of the curve is  $g_m$ .

A reasonable procedure for locating the quiescent point near the center of the linear operating region is to select  $I_{DQ} \approx I_{DSS} / 2$  and  $V_{GSQ} \approx 0.3V_P$ . Note that this is near the midpoint of the curve. Next we select  $v_{DS} \approx V_{DD} / 2$ . This gives a wide range of values for  $v_{ds}$  that keep the transistor in the pinch-off mode.

The transductance at the Q-point can be found from the slope of the curve of [fig.2](#) and is given by

$$g_m = \frac{1.41 I_{DSS}}{V_P}$$

### Example-2

Determine  $g_m$  for a JFET where  $I_{DSS} = 7 \text{ mA}$ ,  $V_P = -3.5 \text{ V}$  and  $V_{DD} = 15\text{V}$ . Choose a reasonable

location for the Q-point.

**Solution:**

Let us select the Q-point as given below:

$$I_{DQ} = \frac{I_{DSS}}{2} = 3.5 \text{ mA}$$

$$V_{D SQ} = \frac{V_{DD}}{2} = 7.5 \text{ V}$$

$$V_{GSQ} = 0.3V_p = -1.05 \text{ V}$$

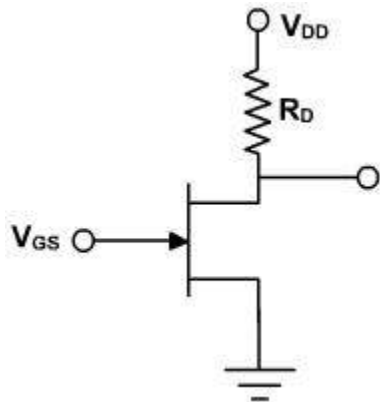
The transconductance,  $g_m$ , is found from the slope of the curve at the point  $i_D / I_{DSS} = 0.5$  and  $v_{GS} / V_p = 0.3$ . Hence,

$$g_m = \frac{1.41 I_{DSS}}{V_p} = 2840 \mu\Omega^{-1}$$

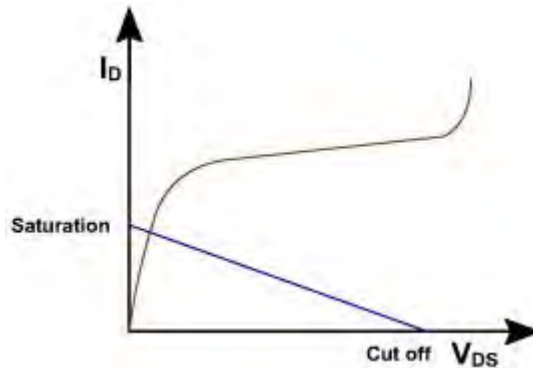
**JFET Applications**

**JFET as Analog Switch:**

JFET can be used as an analog switch as shown in [fig. 3](#). It is the major application of a JFET. The idea is to use two points on the load line: cut off and saturation. When JFET is cut off, it is like an open switch. When it is saturated, it is like a closed switch.



**Fig. 3**



**Fig. 4**

When  $V_{GS} = 0$ , the JFET is saturated and operates at the upper end of the load line. When  $V_{GS}$  is equal to or more negative than  $V_{GS(off)}$ , it is cut off and operates at lower end of the load line (open and closed switch). This is shown in [fig. 4](#).

Only these two points are used for operation when used as a switch. The JFET is normally saturated well below the knee of the drain curve. For this reason the drain current is much smaller than  $I_{DSS}$ .

**FET as a Shunt Switch:**

FET can be used as a shunt switch as shown in [fig. 5](#). When  $V_{cont}=0$ , the JFT is saturated and the switch is closed. When  $V_{cont}$  is more negative FET is like an open switch. The equivalent circuit is also shown in [fig. 5](#).

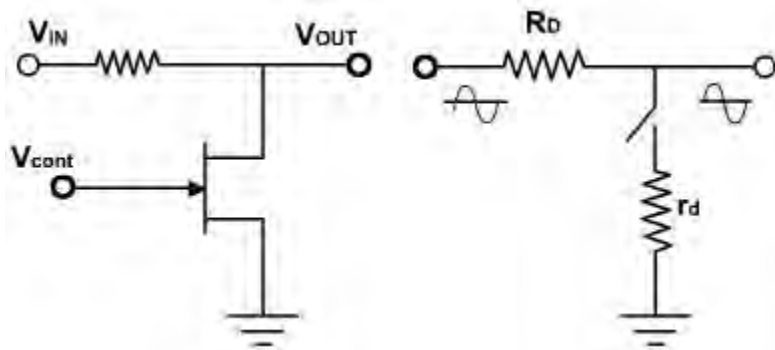


Fig. 5

**FET as a series switch:**

JFET can also be used as series switch as shown in [fig. 6](#). When control is zero, the FET is a closed switch. When  $V_{con}=$  negative, the FET is an open switch. It is better than shunt switch.

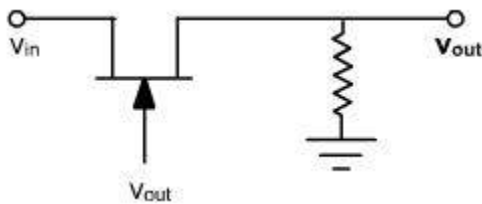
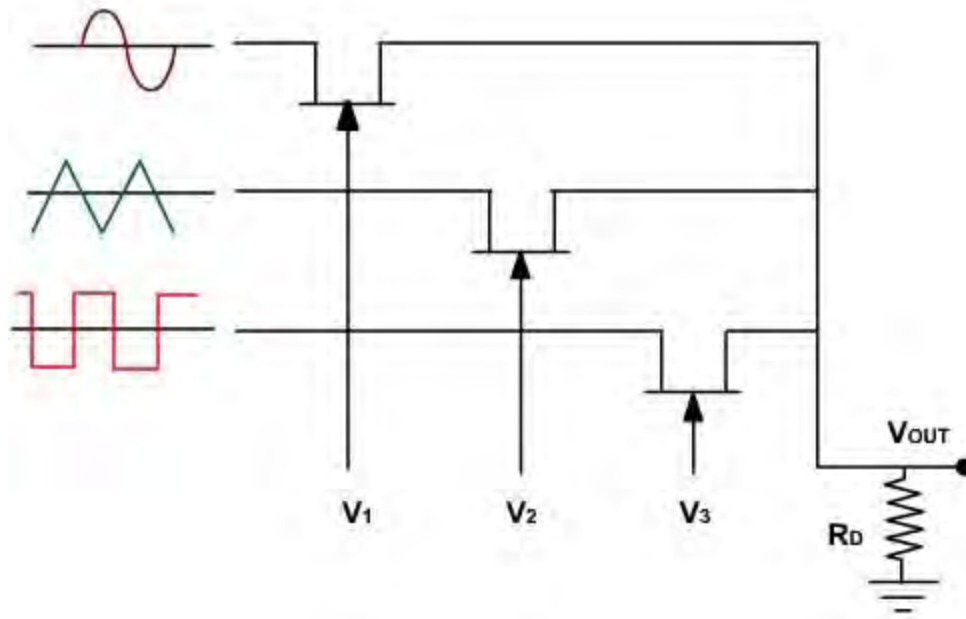


Fig. 6

**Multiplexing:**

One of the important application of FET is in analog multiplexer. Analog multiplexer is a circuit that selects one of the output lines as shown in [fig. 7](#). When control voltages are more negative all switches are open and output is zero. When any control voltage becomes zero the input is transmitted to the output.



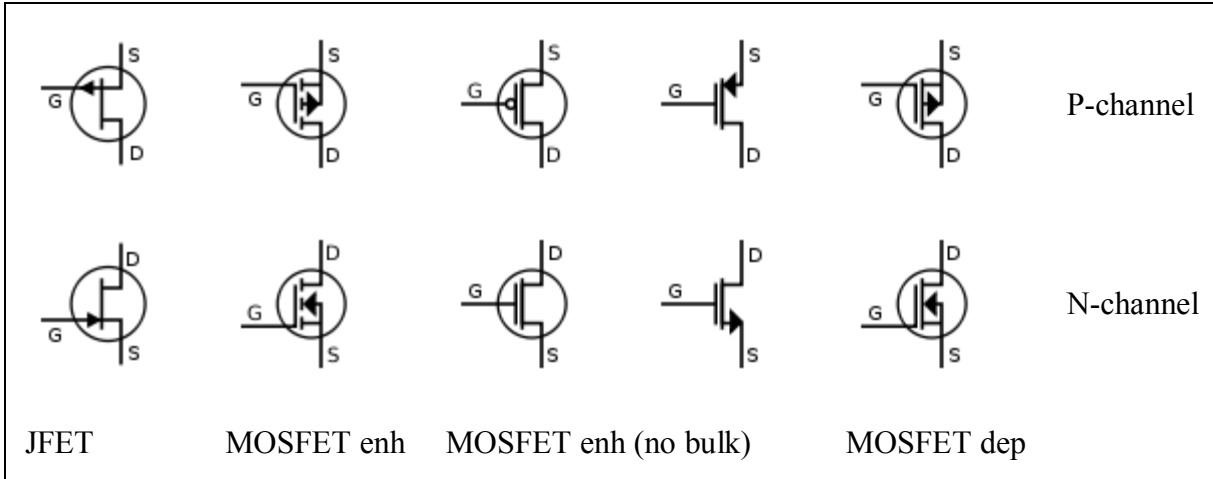
**Fig. 7**

### Circuit symbols

A variety of symbols are used for the MOSFET. The basic design is generally a line for the channel with the source and drain leaving it at right angles and then bending back at right angles into the same direction as the channel. Sometimes three line segments are used for [enhancement mode](#) and a solid line for depletion mode. Another line is drawn parallel to the channel for the gate.

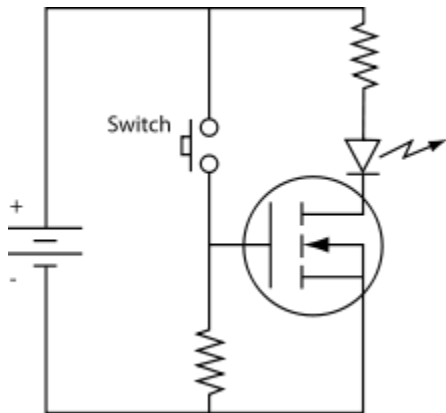
The bulk connection, if shown, is shown connected to the back of the channel with an arrow indicating PMOS or NMOS. Arrows always point from P to N, so an NMOS (N-channel in P-well or P-substrate) has the arrow pointing in (from the bulk to the channel). If the bulk is connected to the source (as is generally the case with discrete devices) it is sometimes angled to meet up with the source leaving the transistor. If the bulk is not shown (as is often the case in IC design as they are generally common bulk) an inversion symbol is sometimes used to indicate PMOS, alternatively an arrow on the source may be used in the same way as for bipolar transistors (out for nMOS, in for pMOS).

Comparison of enhancement-mode and depletion-mode MOSFET symbols, along with [JFET](#) symbols (drawn with source and drain ordered such that higher voltages appear higher on the page than lower voltages):

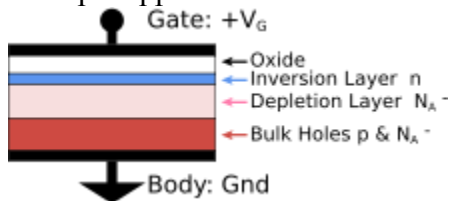


For the symbols in which the bulk, or body, terminal is shown, it is here shown internally connected to the source. This is a typical configuration, but by no means the only important configuration. In general, the MOSFET is a four-terminal device, and in integrated circuits many of the MOSFETs share a body connection, not necessarily connected to the source terminals of all the transistors.

### MOSFET operation



Example application of an N-Channel MOSFET. When the switch is pushed the LED lights up. [\[2\]](#)



Metal-oxide-semiconductor structure on P-type silicon

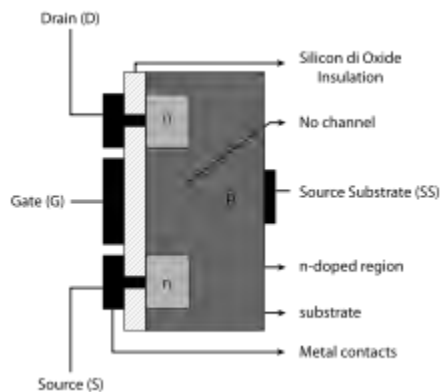
### Metal-oxide-semiconductor structure


A traditional metal–oxide–semiconductor (MOS) structure is obtained by growing a layer of [silicon dioxide](#) ( $\text{SiO}_2$ ) on top of a silicon substrate and depositing a layer of metal or [polycrystalline silicon](#) (the latter is commonly used). As the silicon dioxide is a [dielectric](#) material, its structure is equivalent to a planar [capacitor](#), with one of the electrodes replaced by a semiconductor.

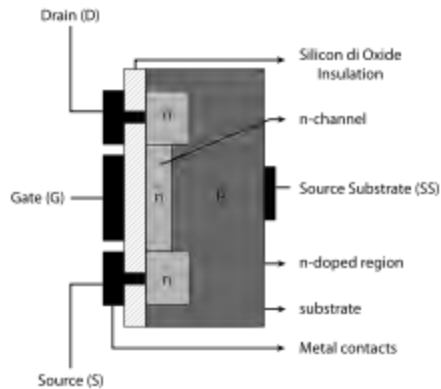
When a voltage is applied across a MOS structure, it modifies the distribution of charges in the semiconductor. If we consider a P-type semiconductor (with  $N_A$  the density of [acceptors](#),  $p$  the density of holes;  $p = N_A$  in neutral bulk), a positive voltage,  $V_{GB}$ , from gate to body (see figure) creates a [depletion layer](#) by forcing the positively charged holes away from the gate-insulator/semiconductor interface, leaving exposed a carrier-free region of immobile, negatively charged acceptor ions (see [doping \(semiconductor\)](#)). If  $V_{GB}$  is high enough, a high concentration of negative charge carriers forms in an **inversion layer** located in a thin layer next to the interface between the semiconductor and the insulator. Unlike the MOSFET, where the inversion layer electrons are supplied rapidly from the source/drain electrodes, in the MOS capacitor they are produced much more slowly by thermal generation through [carrier generation and recombination](#) centers in the depletion region. Conventionally, the gate voltage at which the volume density of electrons in the inversion layer is the same as the volume density of holes in the body is called the [threshold voltage](#).


This structure with P-type body is the basis of the N-type MOSFET, which requires the addition of an N-type source and drain regions.

### MOSFET structure and channel formation



 Cross section of an NMOS without channel formed: OFF state




 Cross section of an NMOS with channel formed: ON state

A metal–oxide–semiconductor field-effect transistor (MOSFET) is based on the modulation of charge concentration by a MOS capacitance between a **body** electrode and a **gate** electrode located above the body and insulated from all other device regions by a gate dielectric layer which in the case of a MOSFET is an oxide, such as silicon dioxide. If dielectrics other than an oxide such as silicon dioxide (often referred to as oxide) are employed the device may be referred to as a metal–insulator–semiconductor FET ([MISFET](#)). Compared to the MOS capacitor, the MOSFET includes two additional terminals (**source** and **drain**), each connected to individual highly doped regions that are separated by the body region. These regions can be either p or n type, but they must both be of the same type, and of opposite type to the body region. The source and drain (unlike the body) are highly doped as signified by a '+' sign after the type of doping.

If the MOSFET is an n-channel or nMOS FET, then the source and drain are 'n+' regions and the body is a 'p' region. As described above, with sufficient gate voltage, holes from the body are driven away from the gate, forming an inversion layer or *n-channel* at the interface between the p region and the oxide. This conducting channel extends between the source and the drain, and current is conducted through it when a voltage is applied between source and drain.

For gate voltages below the threshold value, the channel is lightly populated, and only a very small [subthreshold leakage](#) current can flow between the source and the drain.

If the MOSFET is a p-channel or pMOS FET, then the source and drain are 'p+' regions and the body is a 'n' region. When a negative gate-source voltage (positive source-gate) is applied, it creates a *p-channel* at the surface of the n region, analogous to the n-channel case, but with opposite polarities of charges and voltages. When a voltage less negative than the threshold value (a negative voltage for p-channel) is applied between gate and source, the channel disappears and only a very small subthreshold current can flow between the source and the drain.

The source is so named because it is the source of the charge carriers (electrons for n-channel, holes for p-channel) that flow through the channel; similarly, the drain is where the charge carriers leave the channel.

The device may comprise a Silicon On Insulator (SOI) device in which a Buried OXide (BOX) is formed below a thin semiconductor layer. If the channel region between the gate dielectric and a

Buried Oxide (BOX) region is very thin, the very thin channel region is referred to as an Ultra Thin Channel (UTC) region with the source and drain regions formed on either side thereof in and/or above the thin semiconductor layer. Alternatively, the device may comprise a SEMiconductor On Insulator (SEMOI) device in which semiconductors other than silicon are employed. Many alternative semiconductor materials may be employed.

When the source and drain regions are formed above the channel in whole or in part, they are referred to as Raised Source/Drain (RSD) regions.

## Modes of operation

The operation of a MOSFET can be separated into three different modes, depending on the voltages at the terminals. In the following discussion, a simplified algebraic model is used that is accurate only for old technology. Modern MOSFET characteristics require computer models that have rather more complex behavior.

For an **enhancement-mode, n-channel MOSFET**, the three operational modes are:

Cutoff, subthreshold, or weak-inversion mode

**When  $V_{GS} < V_{th}$ :**

where  $V_{th}$  is the [threshold voltage](#) of the device.

According to the basic threshold model, the transistor is turned off, and there is no conduction between drain and source. In reality, the [Boltzmann distribution](#) of electron energies allows some of the more energetic electrons at the source to enter the channel and flow to the drain, resulting in a subthreshold current that is an exponential function of gate–source voltage. While the current between drain and source should ideally be zero when the transistor is being used as a turned-off switch, there is a weak-inversion current, sometimes called [subthreshold leakage](#).

In weak inversion the current varies exponentially with gate-to-source bias  $V_{GS}$  as given approximately by:<sup>[3][4]</sup>

$$I_D \approx I_{D0} e^{\frac{V_{GS} - V_{th}}{nV_T}},$$

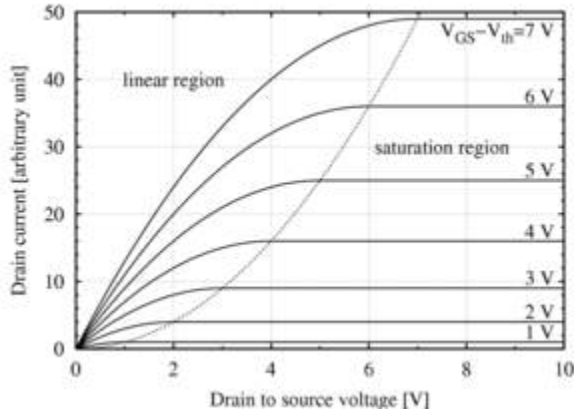
where  $I_{D0}$  = current at  $V_{GS} = V_{th}$  and the slope factor  $n$  is given by

$$n = 1 + C_D / C_{OX},$$

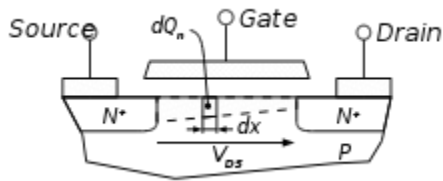
with  $C_D$  = capacitance of the [depletion layer](#) and  $C_{OX}$  = capacitance of the oxide layer. In a long-channel device, there is no drain voltage dependence of the current once  $V_{DS} \gg V_T$ , but as channel length is reduced [drain-induced barrier lowering](#) introduces drain voltage dependence that depends in a complex way upon the device geometry (for example, the channel doping, the junction doping and so on). Frequently, threshold voltage  $V_{th}$  for this mode is defined as the gate voltage at which a selected value of current  $I_{D0}$  occurs, for example,  $I_{D0} = 1 \mu A$ , which may not be the same  $V_{th}$ -value used in the equations for the following modes.

Some micropower analog circuits are designed to take advantage of subthreshold conduction.<sup>[5][6][7]</sup> By working in the weak-inversion region, the MOSFETs in these circuits deliver the highest possible transconductance-to-current ratio, namely:  $g_m / I_D = 1 / (nV_T)$ , almost that of a bipolar transistor.<sup>[8]</sup>

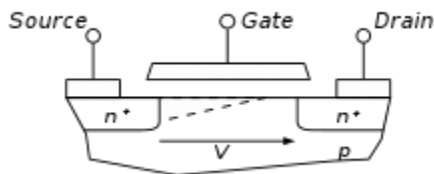
The subthreshold *I-V curve* depends exponentially upon threshold voltage, introducing a strong dependence on any manufacturing variation that affects threshold voltage; for example: variations in oxide thickness, junction depth, or body doping that change the degree of [drain-induced barrier lowering](#). The resulting sensitivity to fabrication variations complicates optimization for leakage and performance. [\[9\]\[10\]](#)



MOSFET drain current vs. drain-to-source voltage for several values of  $V_{GS} - V_{th}$ ; the boundary between **linear (Ohmic)** and **saturation (active)** modes is indicated by the upward curving parabola.



Cross section of a MOSFET operating in the linear (Ohmic) region; strong inversion region present even near drain



Cross section of a MOSFET operating in the saturation (active) region; channel exhibits **pinch-off** near drain

Triode mode or linear region (also known as the ohmic mode [\[11\]\[12\]](#))

**When  $V_{GS} > V_{th}$  and  $V_{DS} < (V_{GS} - V_{th})$**

The transistor is turned on, and a channel has been created which allows current to flow between the drain and the source. The MOSFET operates like a resistor, controlled by the gate voltage relative to both the source and drain voltages. The current from drain to source is modeled as:

$$I_D = \mu_n C_{ox} \frac{W}{L} \left( (V_{GS} - V_{th}) V_{DS} - \frac{V_{DS}^2}{2} \right)$$

where  $\mu_n$  is the charge-carrier effective mobility,  $W$  is the gate width,  $L$  is the gate length and  $C_{ox}$  is the gate oxide capacitance per unit area. The transition from the exponential subthreshold region to the triode region is not as sharp as the equations suggest.

Saturation or active mode [\[13\]\[14\]](#)

**When  $V_{GS} > V_{th}$  and  $V_{DS} > (V_{GS} - V_{th})$**

The switch is turned on, and a channel has been created, which allows current to flow between the drain and source. Since the drain voltage is higher than the gate voltage, the electrons spread out, and conduction is not through a narrow channel but through a broader, two- or three-dimensional current distribution extending away from the interface and deeper in the substrate. The onset of this region is also known as **pinch-off** to indicate the lack of channel region near the drain. The drain current is now weakly dependent upon drain voltage and controlled primarily by the gate–source voltage, and modeled very approximately as:

$$I_D = \frac{\mu_n C_{ox} W}{2 L} (V_{GS} - V_{th})^2 (1 + \lambda V_{DS}).$$

The additional factor involving  $\lambda$ , the [channel-length modulation](#) parameter, models current dependence on drain voltage due to the [Early effect](#), or channel length modulation. According to this equation, a key design parameter, the MOSFET transconductance is:

$$g_m = \frac{2I_D}{V_{GS} - V_{th}} = \frac{2I_D}{V_{ov}},$$

where the combination  $V_{ov} = V_{GS} - V_{th}$  is called the **overdrive voltage**.[\[15\]](#) Another key design parameter is the MOSFET output resistance  $r_o$  given by:

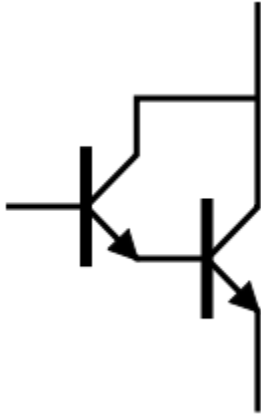
$$r_{OUT} = \frac{1}{\lambda I_D}.$$

$r_{out}$  is the inverse of  $g_{ds}$  where  $g_{DS} = \frac{\partial I_{DS}}{\partial V_{DS}}$ .  $V_{DS}$  is the expression in saturation region.

If  $\lambda$  is taken as zero, an infinite output resistance of the device results that leads to unrealistic circuit predictions, particularly in analog circuits.

As the channel length becomes very short, these equations become quite inaccurate. New physical effects arise. For example, carrier transport in the active mode may become limited by [velocity saturation](#). When velocity saturation dominates, the saturation drain current is more nearly linear than quadratic in  $V_{GS}$ . At even shorter lengths, carriers transport with near zero scattering, known as quasi-[ballistic transport](#). In addition, the output current is affected by [drain-induced barrier lowering](#) of the threshold voltage.

## Darlington transistor



Circuit diagram of a Darlington pair using NPN transistors

In [electronics](#), the **Darlington transistor** (often called a **Darlington pair**) is a compound structure consisting of two [bipolar transistors](#) (either integrated or separated devices) connected in such a way that the current amplified by the first transistor is amplified further by the second one.<sup>[1]</sup> This configuration gives a much higher [current gain](#) (written  $\beta$ ,  $h_{fe}$ , or  $h_{FE}$ ) than each transistor taken separately and, in the case of integrated devices, can take less space than two individual transistors because they can use a *shared* collector. Integrated Darlington pairs come packaged singly in transistor-like packages or as an array of devices (usually eight) in an [integrated circuit](#).

The Darlington configuration was invented by [Bell Laboratories](#) engineer [Sidney Darlington](#) in 1953. He [patented](#) the idea of having two or three transistors on a single chip, sharing a collector.<sup>[2]</sup>

A similar configuration but with transistors of opposite type (NPN and PNP) is the [Sziklai pair](#), sometimes called the "complementary Darlington."

### Behaviour

A Darlington pair behaves like a single transistor with a high current gain (approximately the product of the gains of the two transistors). In fact, integrated devices have three leads (B, C and E), broadly equivalent to those of a standard transistor.

A general relation between the compound current gain and the individual gains is given by:

$$\beta_{\text{Darlington}} = \beta_1 \cdot \beta_2 + \beta_1 + \beta_2$$

If  $\beta_1$  and  $\beta_2$  are high enough (hundreds), this relation can be approximated with:

$$\beta_{\text{Darlington}} \approx \beta_1 \cdot \beta_2$$

A typical modern device has a current gain of 1000 or more, so that only a small base current is needed to make the pair switch on. However, this high current gain comes with several drawbacks.

One drawback is an approximate doubling of base-emitter voltage. Since there are two junctions between the base and emitter of the Darlington transistor, the equivalent base-emitter voltage is the sum of both base-emitter voltages:

$$V_{BE} = V_{BE1} + V_{BE2} \approx 2V_{BE1}$$

For silicon-based technology, where each  $V_{BEi}$  is about 0.65 V when the device is operating in the active or saturated region, the necessary base-emitter voltage of the pair is 1.3 V.

Another drawback of the Darlington pair is its increased saturation voltage. The output transistor is not allowed to saturate (i.e. its base-collector junction must remain reverse-biased) because its collector-emitter voltage is now equal to the sum of its own base-emitter voltage and the collector-emitter voltage of the first transistor, both positive quantities in normal operation. (In symbols,  $V_{CE2} = V_{BE2} + V_{CE1}$ , so  $V_{C2} > V_{B2}$  always.) Thus the saturation voltage of a Darlington transistor is one  $V_{BE}$  (about 0.65 V in silicon) higher than a single transistor saturation voltage, which is typically 0.1 - 0.2 V in silicon. For equal collector currents, this drawback translates to an increase in the dissipated power for the Darlington transistor over a single transistor.

Another problem is a reduction in switching speed, because the first transistor cannot actively inhibit the base current of the second one, making the device slow to switch off. To alleviate this, the second transistor often has a resistor of a few hundred ohms connected between its base and emitter terminals.<sup>[1]</sup> This resistor provides a low impedance discharge path for the charge accumulated on the base-emitter junction, allowing a faster transistor turn-off.

The Darlington pair has more phase shift at high frequencies than a single transistor and hence can more easily become unstable with [negative feedback](#) (i.e., systems that use this configuration can have poor [phase margin](#) due to the extra transistor delay).

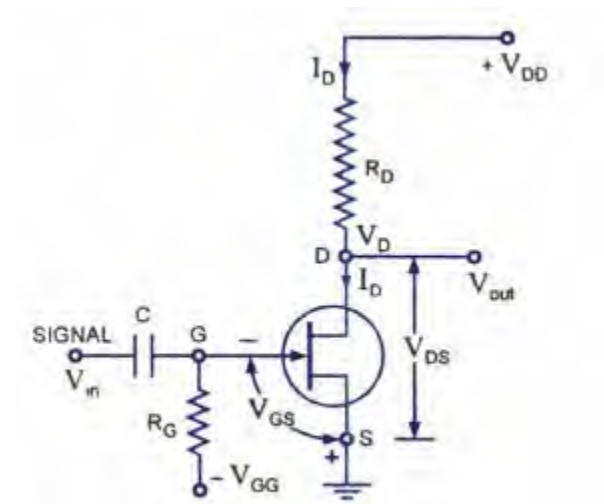
Darlington pairs are available as integrated packages or can be made from two discrete transistors;  $Q_1$  (the left-hand transistor in the diagram) can be a low power type, but normally  $Q_2$  (on the right) will need to be high power. The maximum collector current  $I_C(\text{max})$  of the pair is that of  $Q_2$ . A typical integrated power device is the 2N6282, which includes a switch-off resistor and has a current gain of 2400 at  $I_C=10A$ .

A Darlington pair can be sensitive enough to respond to the current passed by skin contact even at safe voltages. Thus it can form the input stage of a touch-sensitive switch.

#### **FET BIASING:**

Unlike BJTs, thermal runaway does not occur with FETs, as already discussed in our blog. However, the wide differences in maximum and minimum [transfer characteristics](#) make  $I_D$  levels unpredictable with simple fixed-gate bias voltage. To obtain reasonable limits on quiescent drain currents  $I_D$  and drain-source voltage  $V_{DS}$ , source resistor and potential divider bias techniques must be used. With few exceptions, MOSFET bias circuits are similar to those used for [JFETs](#). Various FET biasing circuits are discussed below:

### Fixed Bias.



*Fixed Biasing Circuit For JFET*

DC bias of a FET device needs setting of gate-source voltage  $V_{GS}$  to give desired drain current  $I_D$ . For a JFET drain current is limited by the saturation current  $I_{DS}$ . Since the FET has such a high input impedance that no gate current flows and the dc voltage of the gate set by a voltage divider or a fixed battery voltage is not affected or loaded by the FET.

Fixed dc bias is obtained using a battery  $V_{QG}$ . This battery ensures that the gate is always negative with respect to source and no current flows through resistor  $R_G$  and gate terminal that is  $I_G = 0$ . The battery provides a voltage  $V_{GS}$  to bias the N-channel JFET, but no resulting current is drawn from the battery  $V_{GG}$ . Resistor  $R_G$  is included to allow any ac signal applied through capacitor  $C$  to develop across  $R_G$ . While any ac signal will develop across  $R_G$ , the dc voltage drop across  $R_G$  is equal to  $I_G R_G$  i.e. 0 volt.

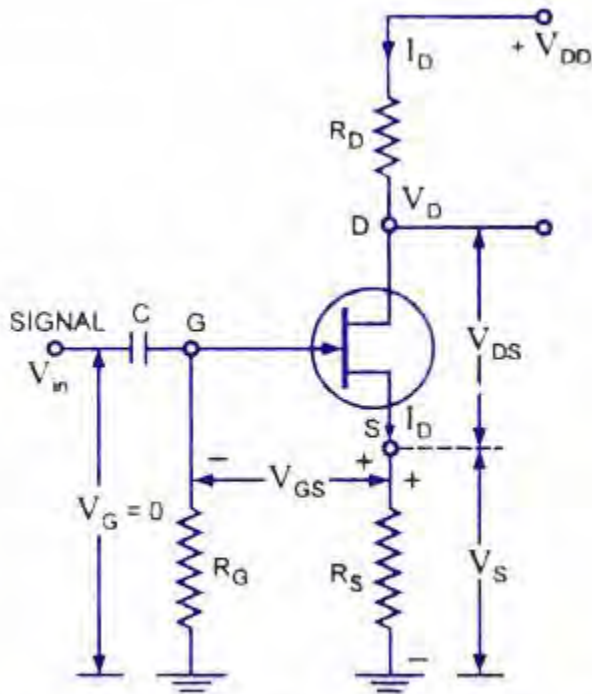
The gate-source voltage  $V_{GS}$  is then

$$V_{GS} = -v_G - v_s = -v_{GG} - 0 = -V_{GG}$$

The drain -source current  $I_D$  is then fixed by the gate-source voltage as determined by equation.

This current then causes a voltage drop across the drain resistor  $R_D$  and is given as  $V_{RD} = I_D R_D$  and output voltage,  $V_{out} = V_{DD} - I_D R_D$

### Self-Bias:



*Self-Bias Circuit For N-Channel JFET*

This is the most common method for biasing a JFET. Self-bias circuit for N-channel JFET is shown in figure.

Since no gate current flows through the reverse-biased gate-source, the gate current  $I_G = 0$  and, therefore,  $V_G = I_G R_G = 0$

The gate-source voltage is then

$$V_{GS} = V_G - V_S = 0 - I_D R_S = -I_D R_S$$

So voltage drop across resistance  $R_S$  provides the biasing voltage  $V_{GS}$  and no external source is required for biasing and this is the reason that it is called self-biasing.



The gate is reverse biased so that  $I_G = 0$  and gate voltage

$$V_G = V_2 = (V_{DD}/R_{G1} + R_{G2}) * R_{G2}$$

And

$$V_{GS} = v_G - v_s = V_G - I_D R_s$$

The operating point can be determined as

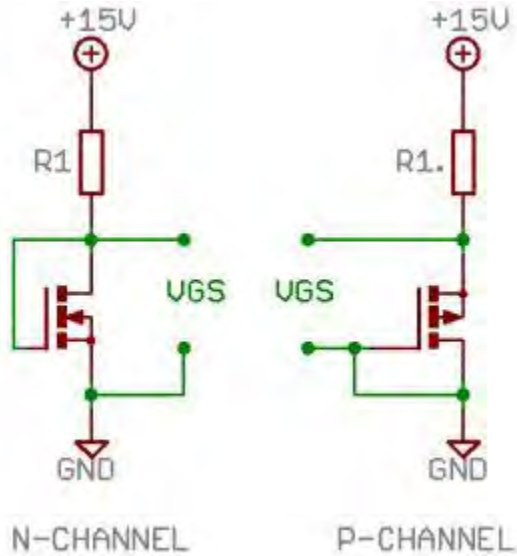
$$I_D = (V_2 - V_{GS})/ R_s$$

And

$$V_{DS} = V_{DD} - I_D (R_D + R_s)$$

### MOSFET-Metal Oxide Semiconductor Field Effect Transistor:

Transistor,



MOSFET-Schematic symbol

Metal-oxide-semiconductor field-effect transistor (MOSFET) is an important semiconductor device and is widely employed in many circuit applications. Since it is constructed with the gate terminal insulated from the channel, it is sometimes called insulated gate FET (IGFET). Like, a [JFET](#), a MOSFET is also a three terminal (source, gate and drain) device and drain current in it is

also controlled by gate bias. The operation of MOSFET is similar to that of JFET. It can be employed in any of the circuits covered for the JFET and, therefore, all the equations apply equally well to the MOSFET and JFET in amplifier connections. However, MOSFET has lower capacitance and input impedance much more than that of a JFET owing to small leakage current. In case of a MOSFET the positive voltage may be applied to the gate and still the gate current remains zero.

MOSFETs are of two types namely

- (i) Enhancement type MOSFET or E-MOSFET and
- (ii) Depletion enhancement MOSFET or DE-MOSFET.

In the depletion-mode construction a channel is physically constructed and a current between drain and source is due to voltage applied across the drain-source terminals. The enhancement MOSFET structure has no channel formed during its construction. Voltage is applied to the gate, in this case, to develop a channel of charge carriers so that a current results when a voltage is applied across the drain-source terminals.

**Advantages of FET over BJT are:**

- a) No minority carriers
- b) High input impedance
- c) It is a voltage controlled device
- d) Better thermal stability

**Effect of Source and Drain Series Resistance**

- The analysis so far neglects the effects of the source/drain series resistance, and the entire voltage is assumed to drop along the channel.
- However, for modern day MOSFETs, this effect cannot be ignored, due to smaller diffusion cross-sections and smaller drain currents.
- The extrinsic (measured) voltages  $V_{gs}$  and  $V_{ds}$  can be related to the intrinsic (device) voltages  $V_g$  and  $V_D$  by the following equations:

$$V_g = V_{gs} - I_D R_s \tag{5.27}$$

and

$$V_D = V_{ds} - I_D (R_s + R_d) \tag{5.28}$$

where  $R_s$  and  $R_d$  are the source and drain resistances respectively.

- The extrinsic transconductance  $g_m (= \partial I_D / \partial V_{gs})$  is related to the intrinsic transconductance  $g_{m0} (= \partial I_D / \partial V_a)$  by

$$g_m = \frac{g_{m0}}{1 + g_{m0}R_s + g_{d0}(R_s + R_d)} \quad (5.29)$$

where  $g_{d0} (= \partial I_D / \partial V_D)$  is the intrinsic drain conductance.

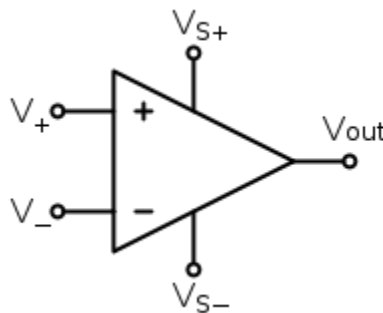
- Similarly, the extrinsic drain conductance  $g_d (= \partial I_D / \partial V_{Ds})$  is related to the intrinsic drain conductance  $g_{d0}$  by

## UNIT IV

### AMPLIFIERS AND OSCILLATORS

#### Differential amplifier

A **differential amplifier** is a type of [electronic amplifier](#) that multiplies the difference between two inputs by some constant factor (the differential [gain](#)).



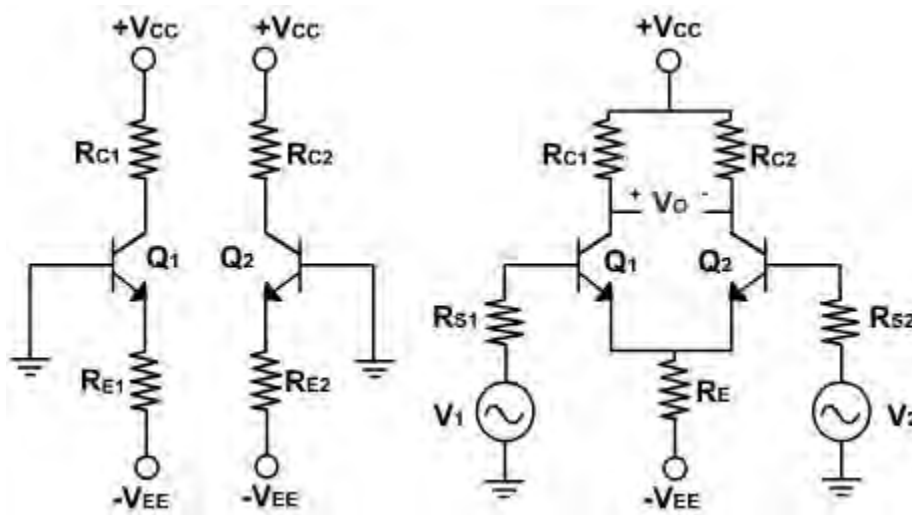
Differential amplifier symbol

The inverting and non-inverting inputs are distinguished by "-" and "+" symbols (respectively) placed in the amplifier triangle.  $V_{s+}$  and  $V_{s-}$  are the power supply voltages; they are often omitted from the diagram for simplicity, but of course must be present in the actual circuit.

## Differential Amplifiers:

Differential amplifier is a basic building block of an op-amp. The function of a differential amplifier is to amplify the difference between two input signals.

How the differential amplifier is developed? Let us consider two emitter-biased circuits as shown in [fig. 1](#).



**Fig. 1**

The two transistors  $Q_1$  and  $Q_2$  have identical characteristics. The resistances of the circuits are equal, i.e.  $R_{E1} = R_{E2}$ ,  $R_{C1} = R_{C2}$  and the magnitude of  $+V_{CC}$  is equal to the magnitude of  $-V_{EE}$ . These voltages are measured with respect to ground.

To make a differential amplifier, the two circuits are connected as shown in [fig. 1](#). The two  $+V_{CC}$  and  $-V_{EE}$  supply terminals are made common because they are same. The two emitters are also connected and the parallel combination of  $R_{E1}$  and  $R_{E2}$  is replaced by a resistance  $R_E$ . The two input signals  $v_1$  &  $v_2$  are applied at the base of  $Q_1$  and at the base of  $Q_2$ . The output voltage is taken between two collectors. The collector resistances are equal and therefore denoted by  $R_C = R_{C1} = R_{C2}$ .

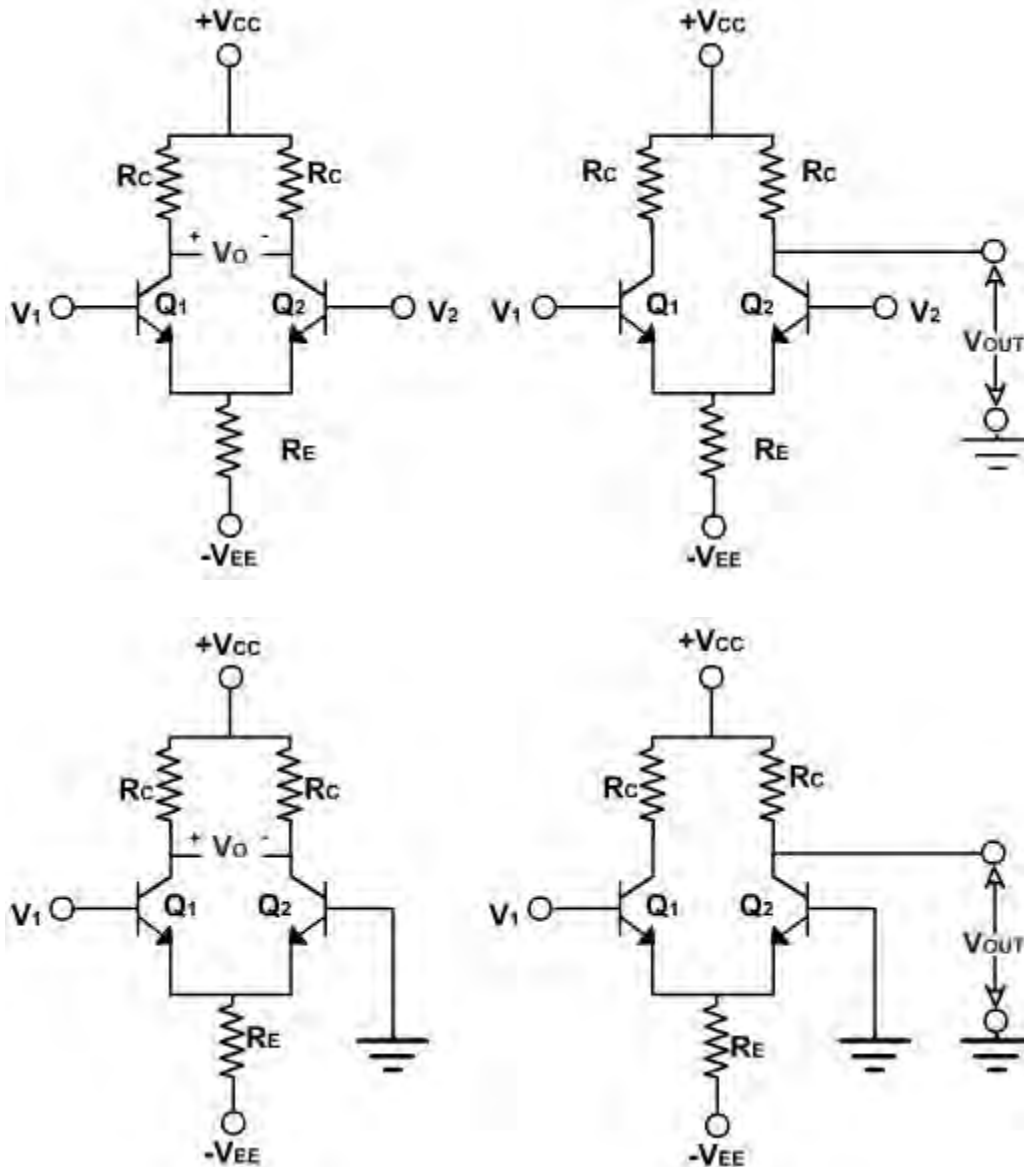
Ideally, the output voltage is zero when the two inputs are equal. When  $v_1$  is greater than  $v_2$  the output voltage with the polarity shown appears. When  $v_1$  is less than  $v_2$ , the output voltage has the opposite polarity.

The differential amplifiers are of different configurations.

The four differential amplifier configurations are following:

1. Dual input, balanced output differential amplifier.
2. Dual input, unbalanced output differential amplifier.

3. Single input balanced output differential amplifier.
4. Single input unbalanced output differential amplifier.



**Fig. 2**

These configurations are shown in [fig. 2](#), and are defined by number of input signals used and the way an output voltage is measured. If use two input signals, the configuration is said to be dual input, otherwise it is a single input configuration. On the other hand, if the output voltage is measured between two collectors, it is referred to as a balanced output because both the collectors are at the same dc potential w.r.t. ground. If the output is measured at one of the collectors w.r.t. ground, the configuration is called an unbalanced output.

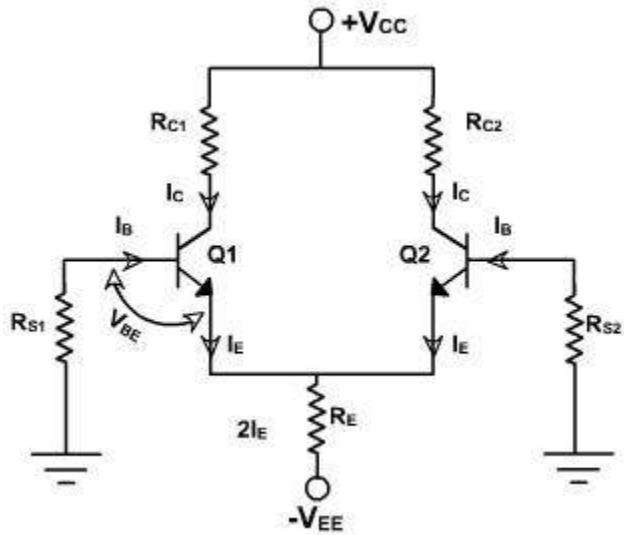
A multistage amplifier with a desired gain can be obtained using direct connection between successive stages of differential amplifiers. The advantage of direct coupling is that it removes the lower cut off frequency imposed by the coupling capacitors, and they are therefore, capable of amplifying dc as well as ac input signals.

### Dual Input, Balanced Output Differential Amplifier:

The circuit is shown in **fig. 1**,  $v_1$  and  $v_2$  are the two inputs, applied to the bases of  $Q_1$  and  $Q_2$  transistors. The output voltage is measured between the two collectors  $C_1$  and  $C_2$ , which are at same dc potentials.

### D.C. Analysis:

To obtain the operating point ( $I_{CC}$  and  $V_{CEQ}$ ) for differential amplifier dc equivalent circuit is drawn by reducing the input voltages  $v_1$  and  $v_2$  to zero as shown in **fig. 3**.



**Fig. 3**

The internal resistances of the input signals are denoted by  $R_S$  because  $R_{S1} = R_{S2}$ . Since both emitter biased sections of the differential amplifier are symmetrical in all respects, therefore, the operating point for only one section need to be determined. The same values of  $I_{CQ}$  and  $V_{CEQ}$  can be used for second transistor  $Q_2$ .

Applying KVL to the base emitter loop of the transistor  $Q_1$ .

$$R_S I_B + V_{BE} + 2 I_E R_E = V_{EE}$$

$$\text{But } I_B = \frac{I_E}{\beta_{dc}} \text{ and } I_C \approx I_E$$

$$\therefore I_E = I_C = \frac{V_{EE} - V_{BE}}{2R_E + R_S / \beta_{dc}} \quad (E-1)$$

$$V_{BE} = 0.6V \text{ for } S_i \text{ and } 0.2V \text{ for } G_e.$$

$$\text{Generally } \frac{R_S}{\beta_{dc}} \ll 2R_E \text{ because } R_S \text{ is the internal resistance of input signal.}$$

$$\therefore I_E = I_C = \frac{V_{EE} - V_{BE}}{2R_E}$$

The value of  $R_E$  sets up the emitter current in transistors  $Q_1$  and  $Q_2$  for a given value of  $V_{EE}$ . The emitter current in  $Q_1$  and  $Q_2$  are independent of collector resistance  $R_C$ .

The voltage at the emitter of  $Q_1$  is approximately equal to  $-V_{BE}$  if the voltage drop across  $R$  is negligible. Knowing the value of  $I_C$  the voltage at the collector  $V_C$  is given by

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

and  $V_{CE} = V_C - V_E$

$$= V_{CC} - I_C R_C + V_{BE}$$

$$V_{CE} = V_{CC} + V_{BE} - I_C R_C \quad (E-2)$$

From the two equations  $V_{CEQ}$  and  $I_{CQ}$  can be determined. This dc analysis applicable for all types of differential amplifier.

### Example - 1

The following specifications are given for the dual input, balanced-output differential amplifier of **fig.1**:

$R_C = 2.2 \text{ k}\Omega$ ,  $R_B = 4.7 \text{ k}\Omega$ ,  $R_{in1} = R_{in2} = 50 \text{ }\Omega$ ,  $+V_{CC} = 10\text{V}$ ,  $-V_{EE} = -10 \text{ V}$ ,  $\beta_{dc} = 100$  and  $V_{BE} = 0.715\text{V}$ .

Determine the operating points ( $I_{CQ}$  and  $V_{CEQ}$ ) of the two transistors.

### Solution:

The value of  $I_{CQ}$  can be obtained from equation (E-1).

$$I_{CQ} = I_E = \frac{V_{EE} - V_{BE}}{2R_E + R_{in} / \beta_{dc}}$$

$$= \frac{10 - 0.715}{9.4\text{k}\Omega + 50 / 100} = 0.988\text{mA}$$

The voltage  $V_{CEQ}$  can be obtained from equation (E-2).

$$V_{CEQ} = V_{CC} + V_{BE} - R_C I_{CQ}$$

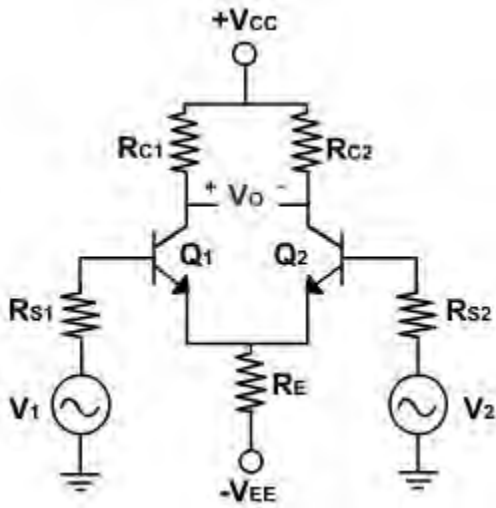
$$= 10 + 0.715 - (2.2\text{k}\Omega)(0.988\text{mA})$$

$$= 8.54\text{V}$$

The values of  $I_{CQ}$  and  $V_{CEQ}$  are same for both the transistors.

### Dual Input, Balanced Output Difference Amplifier:

The circuit is shown in [fig. 1](#)  $v_1$  and  $v_2$  are the two inputs, applied to the bases of  $Q_1$  and  $Q_2$  transistors. The output voltage is measured between the two collectors  $C_1$  and  $C_2$ , which are at same dc potentials.

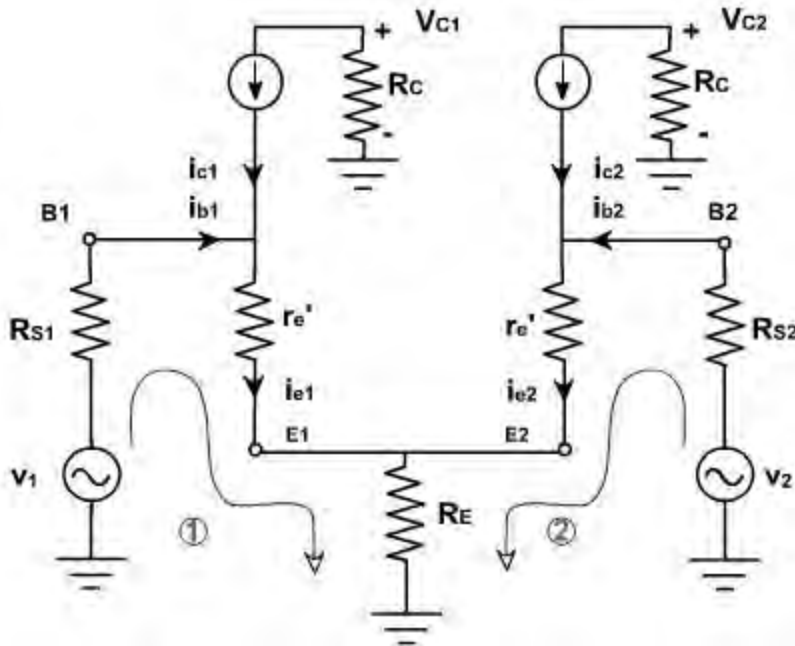


**Fig. 1**

#### A.C. Analysis :

In previous lecture dc analysis has been done to obtain the operating point of the two transistors.

To find the voltage gain  $A_d$  and the input resistance  $R_i$  of the differential amplifier, the ac equivalent circuit is drawn using r-parameters as shown in [fig. 2](#). The dc voltages are reduced to zero and the ac equivalent of CE configuration is used.



**Fig. 2**

Since the two dc emitter currents are equal. Therefore, resistance  $r'_{e1}$  and  $r'_{e2}$  are also equal and designated by  $r'_e$ . This voltage across each collector resistance is shown  $180^\circ$  out of phase with respect to the input voltages  $v_1$  and  $v_2$ . This is same as in CE configuration. The polarity of the output voltage is shown in Figure. The collector  $C_2$  is assumed to be more positive with respect to collector  $C_1$  even though both are negative with respect to ground.

Applying KVL in two loops 1 & 2.

$$v_1 = R_{S1} i_{b1} + i_{e1} r'_e + (i_{e1} + i_{e2}) R_E$$

$$v_2 = R_{S2} i_{b2} + i_{e2} r'_e + (i_{e1} + i_{e2}) R_E$$

Substituting current relations,

$$i_{b1} = \frac{i_{e1}}{\beta}, \quad i_{b2} = \frac{i_{e2}}{\beta}$$

$$V_1 = \frac{R_{S1}}{\beta} i_{e1} + r'_e i_{e1} + R_E (i_{e1} + i_{e2})$$

$$V_2 = \frac{R_{S2}}{\beta} i_{e2} + r'_e i_{e2} + R_E (i_{e1} + i_{e2})$$

Again, assuming  $R_{S1} / \beta$  and  $R_{S2} / \beta$  are very small in comparison with  $R_E$  and  $r'_e$  and therefore neglecting these terms,

$$(r'_e + R_E) i_{e1} + R_E i_{e2} = v_1$$

$$R_E i_{e1} + (r'_e + R_E) i_{e2} = v_2$$

Solving these two equations,  $i_{e1}$  and  $i_{e2}$  can be calculated.

$$i_{e1} = \frac{(r'_e + R_E) v_1 - R_E v_2}{(r'_e + R_E)^2 - R_E^2}$$

$$i_{e2} = \frac{(r'_e + R_E) v_2 - R_E v_1}{(r'_e + R_E)^2 - R_E^2}$$

The output voltage  $V_O$  is given by

$$V_O = V_{C2} - V_{C1}$$

$$= -R_C i_{C2} - (-R_C i_{C1})$$

$$= R_C (i_{C1} - i_{C2})$$

$$= R_C (i_{e1} - i_{e2})$$

Substituting  $i_{e1}$ , &  $i_{e2}$  in the above expression

$$v_o = R_C \left\{ \frac{(r'_e + R_E) v_1 - R_E v_2}{(r'_e + R_E)^2 - R_E^2} - \frac{(r'_e + R_E) v_2 - R_E v_1}{(r'_e + R_E)^2 - R_E^2} \right\}$$

$$= \frac{R_C (v_1 - v_2) (r'_e - 2R_E)}{r'_e (r'_e + 2R_E)}$$

$$\text{Therefore, } v_o = \frac{R_C}{r'_e} (v_1 - v_2) \quad (\text{E-1})$$

Thus a differential amplifier amplifies the difference between two input signals. Defining the difference of input signals as  $v_d = v_1 - v_2$  the voltage gain of the dual input balanced output differential amplifier can be given by

$$A_d = \frac{v_o}{v_d} = \frac{R_C}{r'_e} \quad (\text{E-2})$$

Differential Input Resistance:

Differential input resistance is defined as the equivalent resistance that would be measured at either input terminal with the other terminal grounded. This means that the input resistance  $R_{i1}$  seen from the input signal source  $v_1$  is determined with the signal source  $v_2$  set at zero. Similarly, the input signal  $v_1$  is set at zero to determine the input resistance  $R_{i2}$  seen from the input signal source  $v_2$ . Resistance  $R_{S1}$  and  $R_{S2}$  are ignored because they are very small.

$$R_{i1} = \left. \frac{v_1}{i_{b1}} \right|_{v_2=0}$$

$$= \left. \frac{v_1}{i_{e1}/\beta} \right|_{v_2=0}$$

Substituting  $i_{e1}$ ,

$$R_{i1} = \frac{\beta r'_e (r'_e + 2R_E)}{r'_e + R_E}$$

Since  $R_E \gg r'_e$

$$\therefore r'_e + 2R_E \gg 2R_E$$

$$\text{or } r'_e + R_E \gg R_E$$

$$\therefore R_{i1} = 2\beta r'_e \quad (\text{E-3})$$

Similarly,

$$R_{i2} = \left. \frac{V_2}{i_{b2}} \right|_{v_1=0}$$

$$= \left. \frac{V_2}{i_{e2}/\beta} \right|_{v_1=0}$$

$$R_{i2} = 2\beta r'_e \quad (\text{E-4})$$

The factor of 2 arises because the  $r'_e$  of each transistor is in series.

To get very high input impedance with differential amplifier is to use Darlington transistors. Another way is to use FET.

Output Resistance:

Output resistance is defined as the equivalent resistance that would be measured at output terminal with respect to ground. Therefore, the output resistance  $R_{O1}$  measured between collector  $C_1$  and ground is equal to that of the collector resistance  $R_C$ . Similarly the output resistance  $R_{O2}$  measured at  $C_2$  with respect to ground is equal to that of the collector resistor  $R_C$ .

$$R_{O1} = R_{O2} = R_C \quad (\text{E-5})$$

The current gain of the differential amplifier is undefined. Like CE amplifier the differential amplifier is a small signal amplifier. It is generally used as a voltage amplifier and not as current or power amplifier.

## Difference Amplifiers

A dual input, balanced output difference amplifier circuit is shown in [fig. 1](#).

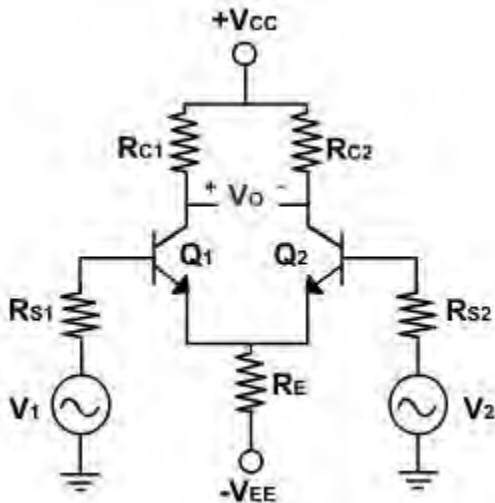


Fig. 1

Inverting & Non – inverting Inputs:

In differential amplifier the output voltage  $v_O$  is given by

$$\text{When } v_2 = 0, v_O = A_d v_1$$

$$\text{\& when } v_1 = 0, v_O = -A_d v_2$$

Therefore the input voltage  $v_1$  is called the non inverting input because a positive voltage  $v_1$  acting alone produces a positive output voltage  $v_O$ . Similarly, the positive voltage  $v_2$  acting alone produces a negative output voltage hence  $v_2$  is called inverting input. Consequently  $B_1$  is called noninverting input terminal and  $B_2$  is called inverting input terminal.

Common mode Gain:

A common mode signal is one that drives both inputs of a differential amplifier equally. The common mode signal is interference, static and other kinds of undesirable pickup etc.

The connecting wires on the input bases act like small antennas. If a differential amplifier is operating in an environment with lot of electromagnetic interference, each base picks up an unwanted interference voltage. If both the transistors were matched in all respects then the balanced output would be theoretically zero. This is the important characteristic of a differential amplifier. It discriminates against common mode input signals. In other words, it refuses to amplify the common mode signals.

The practical effectiveness of rejecting the common signal depends on the degree of matching between the two CE stages forming the differential amplifier. In other words, more closely are the currents in the input transistors, the better is the common mode signal rejection e.g. If  $v_1$  and  $v_2$  are the two input signals, then the output of a practical op-amp cannot be described by simply

$$v_0 = A_d (v_1 - v_2)$$

In practical differential amplifier, the output depends not only on difference signal but also upon the common mode signal (average).

$$v_d = (v_1 - v_2)$$

$$\text{and } v_c = \frac{1}{2} (v_1 + v_2)$$

The output voltage, therefore can be expressed as

$$v_0 = A_1 v_1 + A_2 v_2$$

Where  $A_1$  &  $A_2$  are the voltage amplification from input 1(2) to output under the condition that input 2 (1) is grounded.

$$\therefore v_1 = v_c + \frac{1}{2} v_d, \quad v_2 = v_c - \frac{1}{2} v_d$$

Substituting  $v_1$  &  $v_2$  in output voltage equation

$$\begin{aligned} v_0 &= A_1 \left( v_c + \frac{1}{2} v_d \right) + A_2 \left( v_c - \frac{1}{2} v_d \right) \\ &= \frac{1}{2} (A_1 - A_2) v_d + (A_1 + A_2) v_c \\ &= A_d v_d + A_c v_c \end{aligned}$$

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

The ability of a differential amplifier to reject a common mode signal is expressed by its common mode rejection ratio (CMRR). It is the ratio of differential gain  $A_d$  to the common mode gain  $A_c$ .

$$\text{CMRR} = \frac{A_d}{A_c} = \rho$$

$$\therefore v_0 = A_d v_d \left( 1 + \frac{1}{\rho} \frac{v_c}{v_d} \right)$$

Date sheet always specify CMRR in decibels  $\text{CMRR} = 20 \log \text{CMRR}$ .

Therefore, the differential amplifier should be designed so that  $\rho$  is large compared with the ratio of the common mode signal to the difference signal. If  $\rho = 1000$ ,  $v_c = 1 \text{ mV}$ ,  $v_d = 1 \text{ mV}$ , then

$$\frac{1 V_c}{\rho V_d} = \frac{1}{1000} \times \frac{1000 \mu V}{1 \mu V} = 1$$

It is equal to first term. Hence for an amplifier with  $r = 1000$ , a 1m V difference of potential between two inputs gives the same output as 1mV signal applied with the same polarity to both inputs.

#### Dual Input, Unbalanced Output Differential Amplifier:

In this case, two input signals are given however the output is measured at only one of the two-collector w.r.t. ground as shown in [fig. 2](#). The output is referred to as an unbalanced output because the collector at which the output voltage is measured is at some finite dc potential with respect to ground..

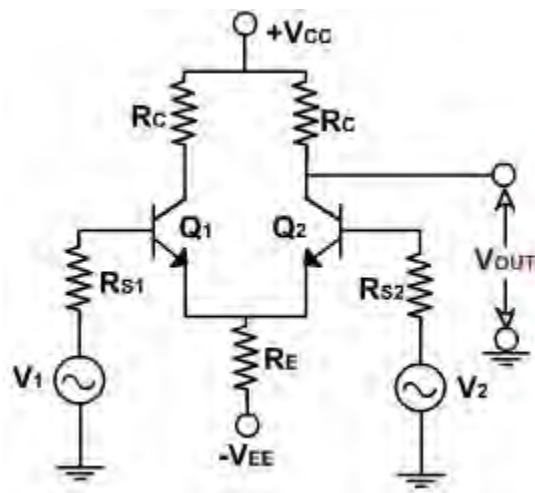


Fig. 2

In other words, there is some dc voltage at the output terminal without any input signal applied. DC analysis is exactly same as that of first case.

$$I_E = I_{CQ} = \frac{V_{EE} - V_{BE}}{2R_E + R_2 / \beta_{dc}}$$

$$V_{CEQ} = V_{CC} + V_{BE} - I_{CQ}R_C$$

AC Analysis:

The output voltage gain in this case is given by

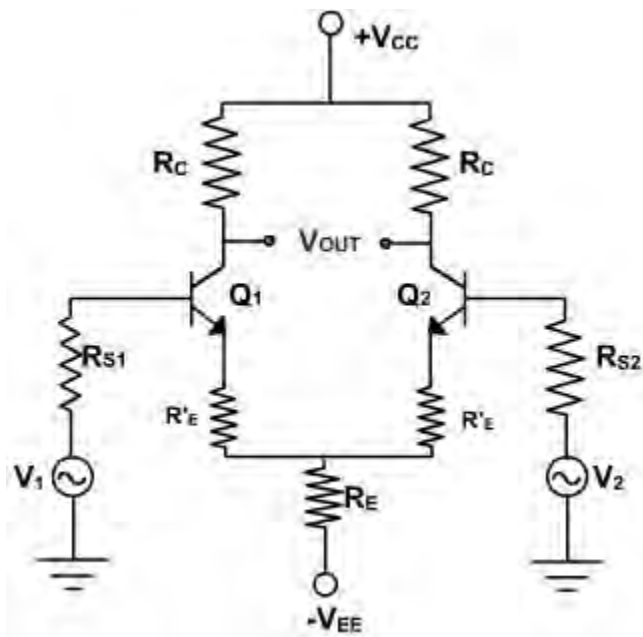
$$A_d = \frac{V_o}{V_d} = \frac{R_C}{2r_e}$$

The voltage gain is half the gain of the dual input, balanced output differential amplifier. Since at the output there is a dc error voltage, therefore, to reduce the voltage to zero, this configuration is normally followed by a level translator circuit.

Differential amplifier with swamping resistors:

By using external resistors  $R'_E$  in series with each emitter, the dependence of voltage gain on variations of  $r'_e$  can be reduced. It also increases the linearity range of the differential amplifier.

**Fig. 3**, shows the differential amplifier with swamping resistor  $R'_E$ . The value of  $R'_E$  is usually large enough to swamp the effect of  $r'_e$ .



**Fig. 3**

$$R_1 I_B + V_{BE} + R'_E I_E + 2 R_E I_E = V_{EE}$$

$$R_1 I_E / \beta_{dc} + V_{BE} + R'_E I_E + 2 R_E I_E = V_{EE}$$

From the equation,  $I_E$  can be obtained as

$$I_E = \frac{V_{EE} - V_{BE}}{R'_E + 2R_E + R_1 / \beta_{dc}}$$

$$V_{CEQ} = V_{CC} + V_{BE} - I_{CQ} R_C$$

The new voltage gain is given by  $A_d = \frac{R_C}{r_e + R'_E}$

The input resistance is given by  $R_{i1} = R_{i2} = 2\beta (r'_e + R'_E)$

The output resistance with or without  $R'_E$  is the same i.e.

$$R_{O1} = R_{O2} = R_C$$

## Biasing of Differential Amplifiers

### Constant Current Bias:

In the dc analysis of differential amplifier, we have seen that the emitter current  $I_E$  depends upon the value of  $b_{dc}$ . To make operating point stable  $I_E$  current should be constant irrespective value of  $b_{dc}$ .

For constant  $I_E$ ,  $R_E$  should be very large. This also increases the value of CMRR but if  $R_E$  value is increased to very large value,  $I_E$  (quiescent operating current) decreases. To maintain same value of  $I_E$ , the emitter supply  $V_{EE}$  must be increased. To get very high value of resistance  $R_E$  and constant  $I_E$ , current, current bias is used.

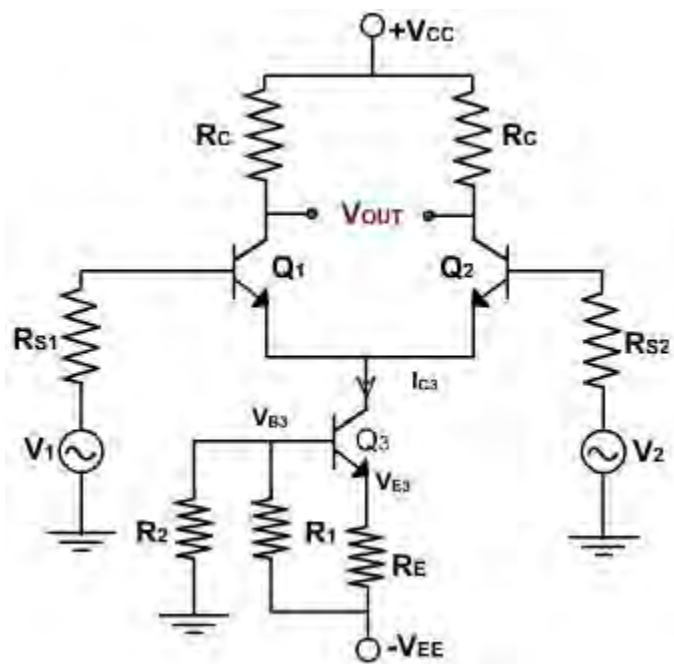


Figure 5.1

**Fig. 1**, shows the dual input balanced output differential amplifier using a constant current bias. The resistance  $R_E$  is replaced by constant current transistor  $Q_3$ . The dc collector current in  $Q_3$  is established by  $R_1$ ,  $R_2$ , &  $R_E$ .

Applying the voltage divider rule, the voltage at the base of  $Q_3$  is

$$\begin{aligned}
 V_{B3} &= \frac{R_2}{R_1 + R_2} (-V_{EE}) \\
 V_{E3} &= V_{B3} - V_{BE3} \\
 &= -\frac{R_2}{R_1 + R_2} V_{EE} - V_{BE3} \\
 I_{B3} &= I_{C3} = \frac{V_{E3} - (-V_{EE})}{R_E} \\
 &= \frac{V_{EE} - \left( \frac{R_2}{R_1 + R_2} \right) V_{EE} - V_{BE3}}{R_E}
 \end{aligned}$$

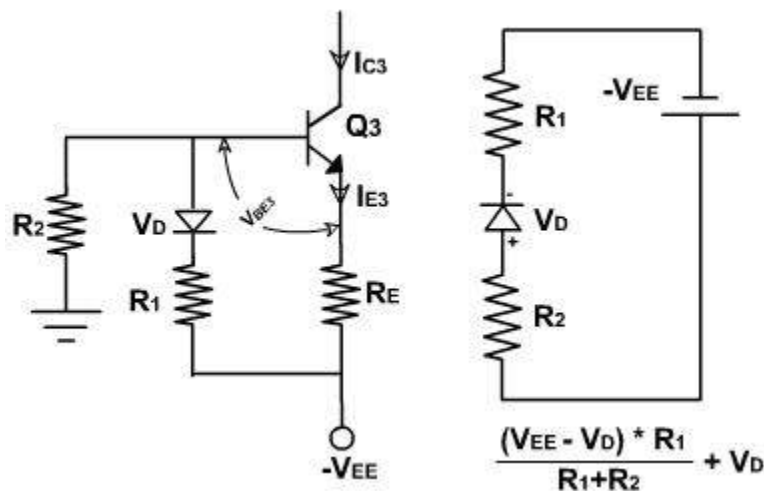
Because the two halves of the differential amplifiers are symmetrical, each has half of the current  $I_{C3}$ .

$$I_{E1} = I_{E2} = \frac{I_{C3}}{2} = \frac{V_{EE} - \left[ \frac{R_2}{R_1 + R_2} V_{EE} \right] - V_{BE3}}{2R_E}$$

The collector current,  $I_{C3}$  in transistor  $Q_3$  is fixed because no signal is injected into either the emitter or the base of  $Q_3$ .

Besides supplying constant emitter current, the constant current bias also provides a very high source resistance since the ac equivalent of the dc source is ideally an open circuit. Therefore, all the performance equations obtained for differential amplifier using emitter bias are also valid.

As seen in  $I_E$  expressions, the current depends upon  $V_{BE3}$ . If temperature changes,  $V_{BE}$  changes and current  $I_E$  also changes. To improve thermal stability, a diode is placed in series with resistance  $R_1$  as shown in [fig. 2](#).



**Fig. 2**

This helps to hold the current  $I_{E3}$  constant even though the temperature changes. Applying KVL to the base circuit of  $Q_3$ .

$$(V_{EE} - V_D) \frac{R_1}{R_1 + R_2} + V_D = V_{BE3} + I_{E3} R_E$$

where  $V_D$  is the diode voltage. Thus,

$$I_{E3} = \frac{1}{R_E} \left\{ V_{EE} \frac{R_1}{R_1 + R_2} + V_D \frac{R_1}{R_1 + R_2} - V_{BE3} \right\}$$

If  $R_1$  and  $R_2$  are so chosen that

$$\frac{R_2}{R_1 + R_2} V_D = V_{BE3}$$

then,

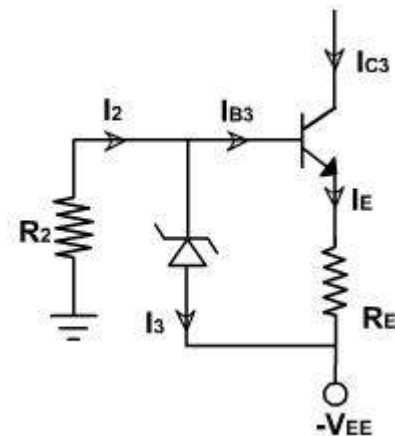
$$I_{E3} = \frac{1}{R_3} \cdot \frac{V_{EE} R_1}{R_1 + R_2}$$

Therefore, the current  $I_{E3}$  is constant and independent of temperature because of the added diode  $D$ . Without  $D$  the current would vary with temperature because  $V_{BE3}$  decreases approximately by  $2\text{mV}/^\circ\text{C}$ . The diode has same temperature dependence and hence the two variations cancel each other and  $I_{E3}$  does not vary appreciably with temperature. Since the cut – in voltage  $V_D$  of diode approximately the same value as the base to emitter voltage  $V_{BE3}$  of a transistor the above condition cannot be satisfied with one diode. Hence two diodes are used in series for  $V_D$ . In this case the common mode gain reduces to zero.

Some times zener diode may be used in place of diodes and resistance as shown in [fig. 3](#). Zeners are available over a wide range of voltages and can have matching temperature coefficient

The voltage at the base of transistor  $Q_B$  is

$$\begin{aligned} V_{B3} &= V_Z - V_{EE} \\ V_{E3} &= V_{B3} - V_{BE3} \\ &= V_Z - V_{EE} - V_{BE3} \\ \therefore I_{E3} &= \frac{V_{E3} - (-V_{EE})}{R_E} \\ &= \frac{V_Z - V_{BE3}}{R_E} \end{aligned}$$



**Fig. 3**

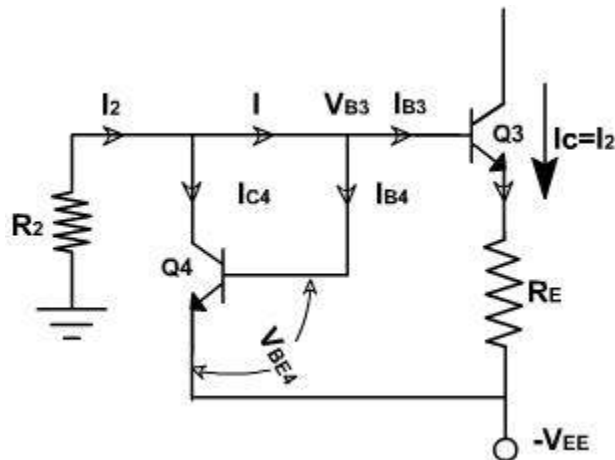
The value of  $R_2$  is selected so that  $I_2 \gg 1.2 I_{Z(\text{min})}$  where  $I_Z$  is the minimum current required to cause the zener diode to conduct in the reverse region, that is to block the rated voltage  $V_Z$ .

$$R_2 = \frac{V_{EE} - V_Z}{I_2}$$

Where  $I_2 = 1.2 I_{Z(\min)}$

### Current Mirror:

The circuit in which the output current is forced to equal the input current is said to be a current mirror circuit. Thus in a current mirror circuit, the output current is a mirror image of the input current. The current mirror circuit is shown in [fig. 4](#).



**Fig. 4**

Once the current  $I_2$  is set up, the current  $I_{C3}$  is automatically established to be nearly equal to  $I_2$ . The current mirror is a special case of constant current bias and the current mirror bias requires of constant current bias and therefore can be used to set up currents in differential amplifier stages. The current mirror bias requires fewer components than constant current bias circuits.

Since Q3 and Q4 are identical transistors the current and voltage are approximately same

$$V_{BE3} = V_{BE4}$$

$$I_{B3} = I_{B4}$$

$$I_{C3} = I_{C4}$$

Summing currents at node  $V_{B3}$

$$\begin{aligned} I_2 &= I_{C4} + I_{B4} \\ &= I_{C4} + 2I_{B4} = I_{C3} + 2I_{B3} \\ &= I_{C3} + 2\left(\frac{I_{C3}}{\beta_{dc}}\right) \\ &= I_{C3} \left(1 + \frac{2}{\beta_{dc}}\right) \end{aligned}$$

Generally  $\beta_{dc}$  is large enough, therefore  $\frac{2}{\beta_{dc}}$  is small.

$$\begin{aligned} \therefore I_2 &\approx I_{C3} \\ I_2 &= \frac{V_{EE} + V_{BE3}}{R_2} \end{aligned}$$

For satisfactory operation two identical transistors are necessary.

## Oscillators

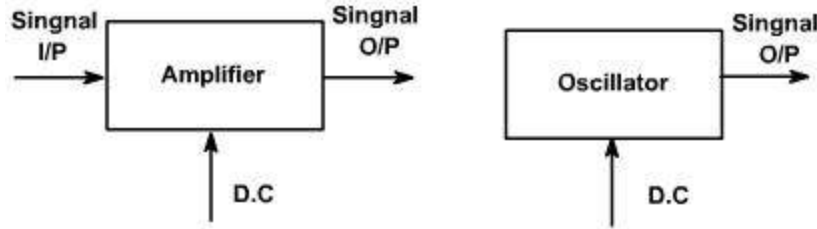
Oscillators:

An oscillator may be described as a source of alternating voltage. It is different than amplifier.

An amplifier delivers an output signal whose waveform corresponds to the input signal but whose power level is higher. The additional power content in the output signal is supplied by the DC power source used to bias the active device.

The amplifier can therefore be described as an energy converter, it accepts energy from the DC power supply and converts it to energy at the signal frequency. The process of energy conversion is controlled by the input signal, Thus if there is no input signal, no energy conversion takes place and there is no output signal.

The oscillator, on the other hand, requires no external signal to initiate or maintain the energy conversion process. Instead an output signals is produced as long as source of DC power is connected. [Fig. 1](#), shows the block diagram of an amplifier and an oscillator.



**Fig. 1**

Oscillators may be classified in terms of their output waveform, frequency range, components, or circuit configuration.

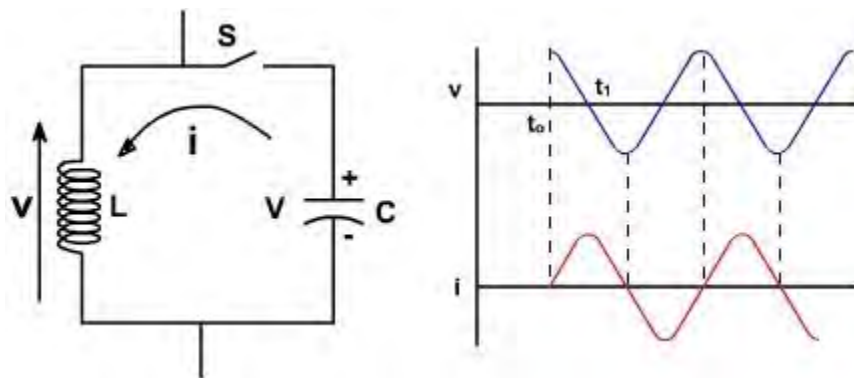
If the output waveform is sinusoidal, it is called harmonic oscillator otherwise it is called relaxation oscillator, which include square, triangular and saw tooth waveforms.

Oscillators employ both active and passive components. The active components provide energy conversion mechanism. Typical active devices are transistor, FET etc.

Passive components normally determine the frequency of oscillation. They also influence stability, which is a measure of the change in output frequency (drift) with time, temperature or other factors. Passive devices may include resistors, inductors, capacitors, transformers, and resonant crystals.

Capacitors used in oscillators circuits should be of high quality. Because of low losses and excellent stability, silver mica or ceramic capacitors are generally preferred.

An elementary sinusoidal oscillator is shown in [fig. 2](#). The inductor and capacitors are reactive elements i.e. they are capable of storing energy. The capacitor stores energy in its electric field. Whenever there is voltage across its plates, and the inductor stores energy in its magnetic field whenever current flows through it. Both C and L are assumed to be loss less. Energy can be introduced into the circuit by charging the capacitor with a voltage V as shown in [fig. 2](#). As long as the switch S is open, C cannot discharge and so  $i=0$  and  $V=0$ .



**Fig. 2**

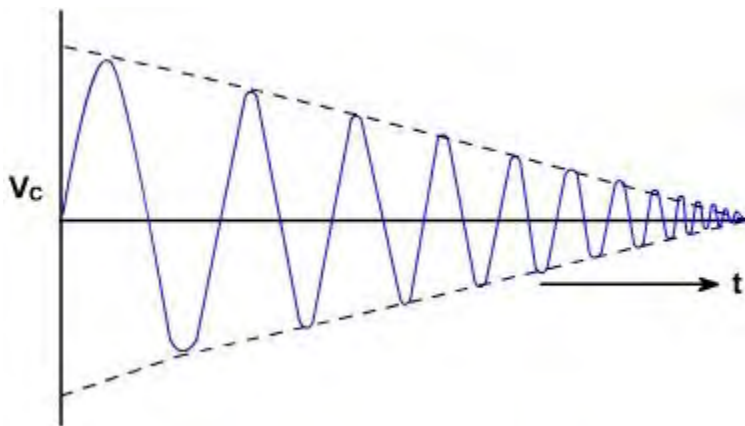
Now S is closed at  $t = t_0$ , This means V rises from 0 to V, Just before closing inductor current was zero and inductor current cannot be changed instantaneously. Current increases from zero value sinusoidally and is given by

$$i = V \sqrt{\frac{C}{L}} \sin \frac{t}{\sqrt{LC}}$$

The capacitor loses its charge and energy is simply transferred from capacitor to inductor magnetic field. The total energy is still same. At  $t = t_1$ , all the charge has been removed from the capacitor plates and voltage reduces to zero and at current reaches to its maximum value. The current for  $t > t_1$  charges C in the opposite direction and current decreases. Thus LC oscillation takes places. Both voltage and current are sinusoidal though no sinusoidal input was applied. The

frequency of oscillation is  $f = \frac{1}{2\pi\sqrt{LC}}$

The circuit discussed is not a practical oscillator because even if loss less components were available, one could not extract energy with out introducing an equivalent resistance. This would result in damped oscillations as shown in [fig. 3](#).



**Fig. 3**

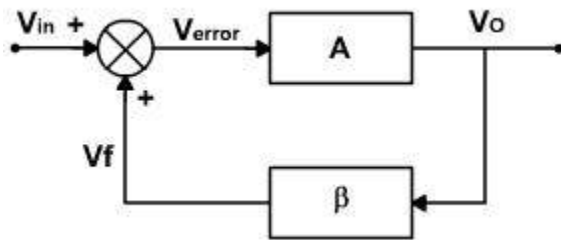
These oscillations decay to zero as soon as the energy in the tank is consumed. If we remove too much power from the circuit, the energy may be completely consumed before the first cycle of oscillations can take place yielding the over damped response.

It is possible to supply energy to the tank to make up for all losses (coil losses plus energy removed), thereby maintaining oscillations of constant amplitude.

Since energy lost may be related to a positive resistance, it follows that the circuit would gain energy if an equivalent negative resistance were available. The negative resistance, supplies

whatever energy the circuit lose due to positive resistance. Certain devices exhibit negative resistance characteristics, an increasing current for a decreasing voltage. The energy supplied by the negative resistance to the circuit, actually comes from DC source that is necessary to bias the device in its negative resistance region.

Another technique for producing oscillation is to use positive feedback considers an amplifier with an input signal  $v_{in}$  and output  $v_o$  as shown in [fig. 4](#).



**Fig. 4**

The amplifier is inverting amplifier and may be transistorized, or FET or OPAMP. The output is  $180^\circ$  out of phase with input signal  $v_o = -A v_{in}$ . ( $A$  is negative)

Now a feedback circuit is added. The output voltage is fed to the feedback circuit. The output of the feedback circuit is again  $180^\circ$  phase shifted and also gets attenuated. Thus the output from the feedback network is in phase with input signal  $v_{in}$  and it can also be made equal to input signal.

If this is so,  $V_f$  can be connected directly and externally applied signal can be removed and the circuit will continue to generate an output signal. The amplifier still has an input but the input is derived from the output amplifier. The output essentially feeds on itself and is continuously regenerated. This is positive feedback. The over all amplification from  $v_{in}$  to  $v_f$  is 1 and the total phase shift is zero. Thus the loop gain  $A \beta$  is equal to unity.

$$A\beta = 1 \angle 0^\circ$$

When this criterion is satisfied then the closed loop gain is infinite. i.e. an output is produced without any external input.

$$v_o = A v_{error}$$

$$= A (v_{in} + v_f)$$

$$= A (v_{in} + \beta v_o)$$

$$\text{or } (1 - A \beta) v_o = A v_{in}$$

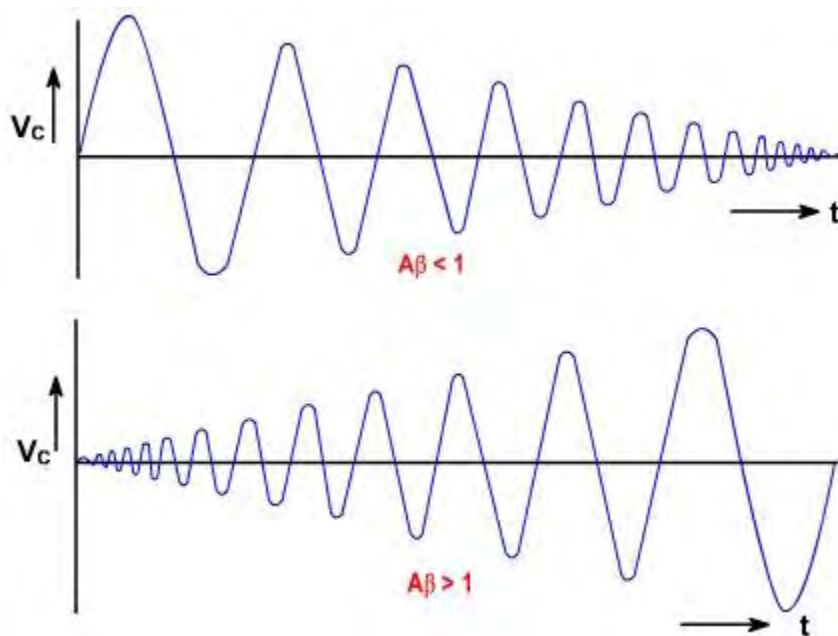
$$\text{or } \frac{v_o}{v_{in}} = \frac{A}{(1 - A\beta)}$$

When  $A\beta = 1$ ,  $v_o / v_{in} = \infty$

The criterion  $A\beta = 1$  is satisfied only at one frequency. This is known as **backhausen criterion**.

The frequency at which a sinusoidal oscillator will operate is the frequency for which the total phase shift introduced, as the signal proceeds from the input terminals, through the amplifier and feedback network and back again to the input is precisely zero or an integral multiple of  $2\pi$ . Thus the frequency of oscillation is determined by the condition that the loop phase shift is zero.

Oscillation will not be sustained, if at the oscillator frequency,  $A\beta < 1$  or  $A\beta > 1$ . [Fig. 5](#), show the output for two different conditions  $A\beta < 1$  and  $A\beta > 1$ .



**Fig. 5**

If  $A\beta$  is less than unity then  $A\beta v_{in}$  is less than  $v_{in}$ , and the output signal will die out, when the externally applied source is removed. If  $A\beta > 1$  then  $A\beta v_{in}$  is greater than  $v_{in}$  and the output voltage builds up gradually. If  $A\beta = 1$ , only then output voltage is sine wave under steady state conditions.

In a practical oscillator, it is not necessary to supply a signal to start the oscillations. Instead, oscillations are self-starting and begin as soon as power is applied. This is possible because of electrical noise present in all passive components.

Therefore, as soon as the power is applied, there is already some energy in the circuit at  $f_o$ , the frequency for which the circuit is designed to oscillate. This energy is very small and is mixed

with all the other frequency components also present, but it is there. Only at this frequency the loop gain is slightly greater than unity and the loop phase shift is zero. At all other frequency the Barkhausen criterion is not satisfied. The magnitude of the frequency component  $f_o$  is made slightly higher each time it goes around the loop. Soon the  $f_o$  component is much larger than all other components and ultimately its amplitude is limited by the circuits own non-linearities (reduction of gain at high current levels, saturation or cut off). Thus the loop gain reduces to unity and steady state is reached. If it does not, then the clipping may occur.

Practically,  $A\beta$  is made slightly greater than unity. So that due to disturbance the output does not change but if  $A\beta = 1$  and due to some reasons if  $A\beta$  decreases slightly than the oscillation may die out and oscillator stop functioning. In conclusion, all practical oscillations involve:

- An active device to supply loop gain or negative resistance.
- A frequency selective network to determine the frequency of oscillation.
- Some type of non-linearity to limit amplitude of oscillations.

### Example - 1

The gain of certain amplifier as a function of frequency is  $A(j\omega) = -16 \times 10^6 / j\omega$ . A feedback path connected around it has  $\beta(j\omega) = 10^3 / (2 \times 10^3 + j\omega)^2$ . Will the system oscillate? If so, at what frequency ?

#### Solution:

$$A\beta = \left( \frac{-16 \times 10^6}{j\omega} \right) \left[ \frac{10^3}{(2 \times 10^3 + j\omega)^2} \right] = \frac{-16 \times 10^9}{j\omega(2 \times 10^3 + j\omega)^2}$$

The loop gain is

To determine, if the system will oscillate, we will first determine the frequency, if any, at which the phase angle of  $A\beta$  ( $\theta = \angle A\beta$ ) equals to  $0^\circ$  or a multiple of  $360^\circ$ . Using phasor algebra, we have

$$\begin{aligned} \theta = \angle A\beta &= \angle \left( \frac{-16 \times 10^9}{j\omega(2 \times 10^3 + j\omega)^2} \right) = \angle 16 \times 10^9 + \angle(1/j\omega) + \angle \left( \frac{1}{(2 \times 10^3 + j\omega)^2} \right) \\ &= -180^\circ - 90^\circ - 2 \arctan \left( \frac{\omega}{2 \times 10^3} \right) \end{aligned}$$

This expression will equal  $-360^\circ$  if  $2 \arctan\left(\frac{\omega}{2 \times 10^3}\right) = 90^\circ$ ,

$$\arctan\left(\frac{\omega}{2 \times 10^3}\right) = 45^\circ$$

or  $\left(\frac{\omega}{2 \times 10^3}\right) = 1$

or  $\omega = 2 \times 10^3 \text{ rad/s}$

Thus, the phase shift around the loop is  $-360^\circ$  at  $\omega = 2000 \text{ rad/s}$ . We must now check to see if the gain magnitude  $|A\beta| = 1$  at  $\omega = 2 \times 10^3$ . The gain magnitude is

$$|A\beta| = \left| \frac{-16 \times 10^9}{j\omega(2 \times 10^3 + j\omega)^2} \right| = \frac{|-16 \times 10^9|}{|j\omega|(2 \times 10^3 + j\omega)^2}$$

$$= \frac{16 \times 10^9}{\omega \left[ (2 \times 10^3)^2 + \omega^2 \right]}$$

Substituting  $\omega = 2 \times 10^3$ , we find

$$|A\beta| = \frac{16 \times 10^9}{2 \times 10^3(4 \times 10^6 + 4 \times 10^6)} = 1$$

Thus, the Barkhausen criterion is satisfied at  $\omega = 2 \times 10^3 \text{ rad/s}$  and oscillation occurs at that frequency ( $2 \times 10^3 / 2\pi = 318.3 \text{ Hz}$ ).

## Harmonic Oscillators

According to Barkhausen criterion, a feedback type oscillator, having  $A\beta$  as loop gain, works if  $A\beta$  is made slightly greater than unity. As discussed in previous lectures, all practical oscillations involve:

- An active device to supply loop gain or negative resistance.
- A frequency selective network to determine the frequency of oscillation.
- Some type of non-linearity to limit amplitude of oscillations.

Harmonic Oscillator:

One feedback type harmonic oscillator circuit is shown in [fig. 1](#).

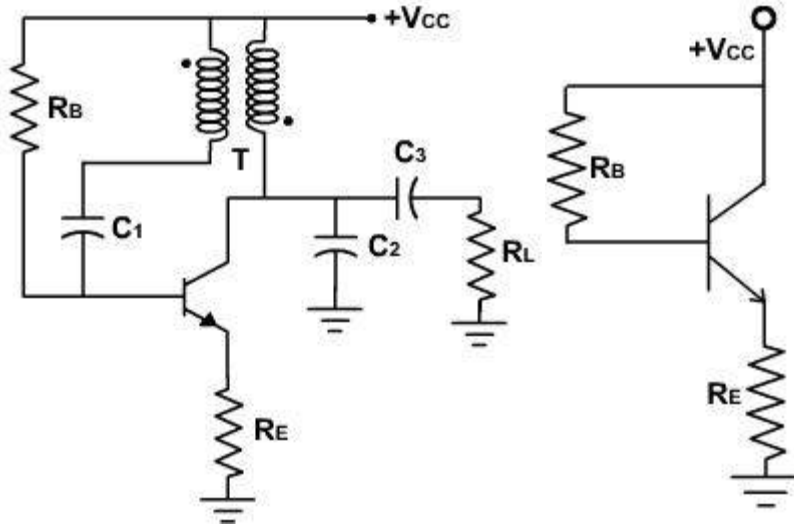


Fig. 1

Fig. 2

The dc equivalent circuit is shown in [fig. 2](#). The dc operating point is set by selecting  $V_{CC}$ ,  $R_B$  and  $R_E$ . The ac equivalent circuit is also shown in [fig. 3](#).

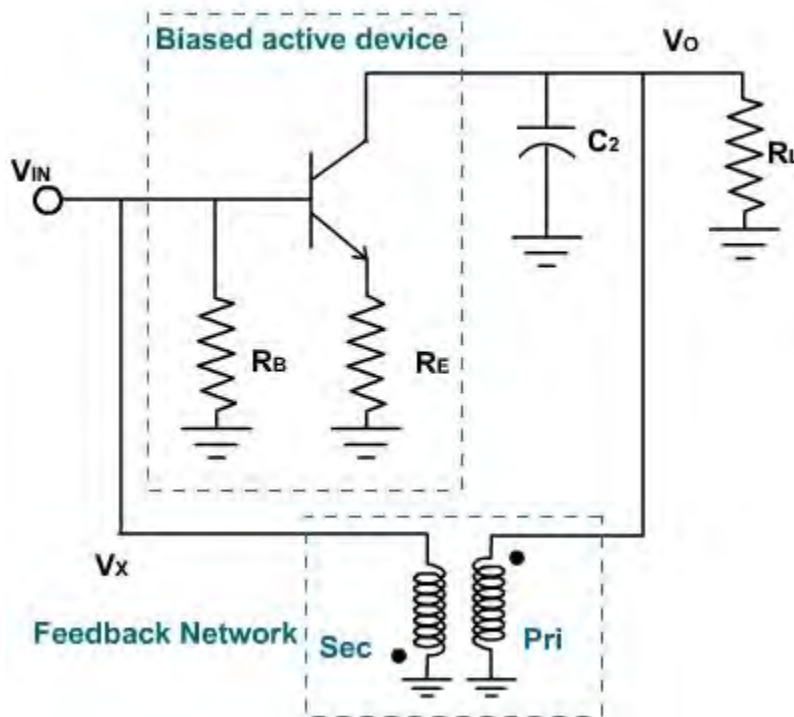


Fig. 3

The transformer provides  $180^\circ$  phase shift to ensure positive feedback so that at the desired frequency of oscillation, the total phase shift from  $v_{in}$  to  $v_x$  is made equal to  $0^\circ$  and magnitudes

are made equal by properly selecting the turns ratio.  $R_E$  also controls and stabilizes the gain through negative feedback.  $C_2$  and the transformer's equivalent inductance make up a resonant circuit that determines the frequency of oscillation.

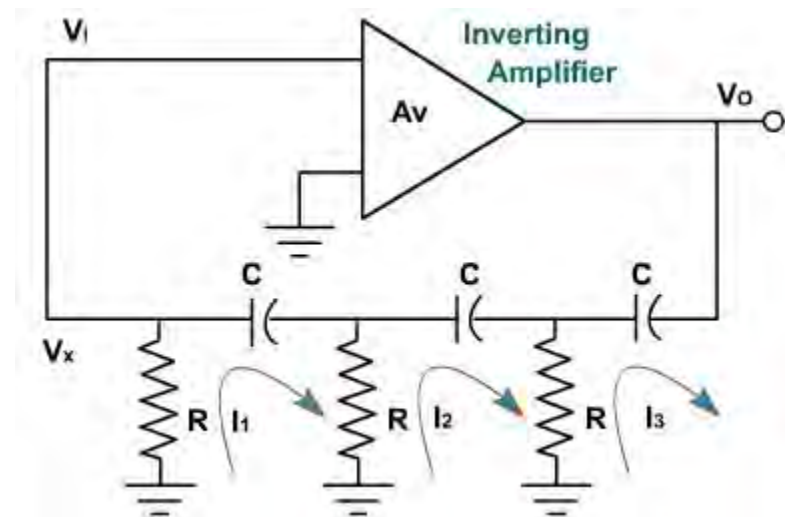
$C_1$  is used to block dc (Otherwise the base would be directly tied to  $V_{CC}$ ) through the low dc resistance of transformer primary.  $C_1$  has negligible reactance at the frequency of oscillation, therefore it is not apart of the frequency-determining network, the same applies to  $C_2$ .

In this circuit, there is an active device suitably biased to provide necessary gain. Since the active device produces loop phase shift  $180^\circ$  (from base to collector), a transformer in the feedback loop provides an additional  $180^\circ$  to yield to a loop phase shift of  $0^\circ$ . The feedback factor is equivalent to the transformer's turns ratio. There is also a tuned circuit, to determine the frequency of oscillation.

The load is in parallel with  $C_2$  and the transformer. If the load is resistive, which is usually the case, the  $Q$  of the tuned circuit and the loop gain are both affected, this must be taken into account when determining the minimum gain required for oscillation. If the load has a capacitive component, then the value of  $C_2$  should be reduced accordingly.

The RC Phase Shift Oscillator:

At low frequencies (around 100 KHz or less), resistors are usually employed to determine the frequency oscillation. Various circuits are used in the feedback circuit including ladder network.



**Fig. 4**

A block diagram of a ladder type RC phase shift oscillation is shown in [fig. 4](#). It consists of three resistor  $R$  and  $C$  capacitors. If the phase shift through the amplifier is  $180^\circ$ , then oscillation may occur at the frequency where the RC network produces an additional  $180^\circ$  phase shift.

To find the frequency of oscillation, let us neglect the loading of the phase shift network. Writing the KV equations,

$$V_x = -I_1 R$$

$$X_c = \frac{1}{\omega C}$$

$$I_1 R - j I_1 X_c + (I_1 - I_2) R = 0$$

$$(I_2 - I_1) R - j I_2 X_c + (I_2 - I_3) R = 0$$

$$(I_3 - I_1) R - j I_3 X_c + V_0 = 0$$

$$I_1 = -\frac{V_x}{R}$$

$$I_2 = \frac{I_1}{R} (2R - jX_c) = -\frac{V_x}{R^2} (2R - jX_c)$$

$$I_3 = \frac{1}{R} \left\{ (2R - jX_c)^2 \left( -\frac{V_x}{R^2} \right) + V_x \right\}$$

$$= -\left\{ \frac{3R^2 - X_c^2 - j4R X_c}{R^2} \right\} V_x$$

$$-V_0 = -(R - jX_c) \left\{ \frac{3R^2 - X_c^2 - j4R X_c}{R^2} \right\} V_x + \frac{V_x}{R} (2R - jX_c)$$

$$-V_0 = V_x \left\{ \frac{(3R^2 - R X_c^2 - j4R^2 X_c - j3R^2 X_c^3 - 4R X_c^2) + 2R^3 - jR^2 X_c}{R^3} \right\}$$

$$\frac{V_x}{V_0} = \frac{R^3}{R^3 - 5R X_c^2 - 6jR^2 X_c + jX_c^3}$$

putting  $X_c = \frac{1}{\omega C}$ , we get

$$\frac{V_x}{V_0} = \frac{R^3}{R^3 - \frac{5R}{\omega^2 C^2} - \frac{6jR^2}{\omega C} + \frac{j}{\omega^3 C^3}}$$

For phase shift equal to  $180^\circ$  between  $V_x$  and  $V_o$ , imaginary term of  $V_x / V_o$  must be zero.

$$\frac{j}{\omega^3 C^3} - \frac{6jR^2}{\omega C} = 0$$

$$\omega^2 C^2 = \frac{1}{6R^2}$$

$$\omega = \frac{1}{RC\sqrt{6}}$$

$$f = \frac{1}{2\pi RC\sqrt{6}}$$

Therefore,

This is the frequency of oscillation. Substituting this frequency in  $V_x / V_o$  expression.

$$\frac{V_x}{V_o} = \frac{R^3}{R^3 - 5R \cdot 6R^2} = -\frac{1}{29} = \beta$$

This shows that  $180^\circ$  phase shift from  $V_o$  to  $V_x$  can be obtained if

$$f = \frac{1}{2\pi RC\sqrt{6}}$$

and the gain of its feedback circuit becomes  $1/29$ .

In order to ensure the oscillation, initially  $|A\beta| > 1$  and under steady state  $A\beta = 1$ . This means the gain of the amplifier should be initially greater than 29 (so that  $A\beta > 1$ ) and under steady state conditions it reduces to 29.

This oscillator can be realized using FET amplifier as shown in [fig. 5](#). The feedback circuit is same as discussed above.

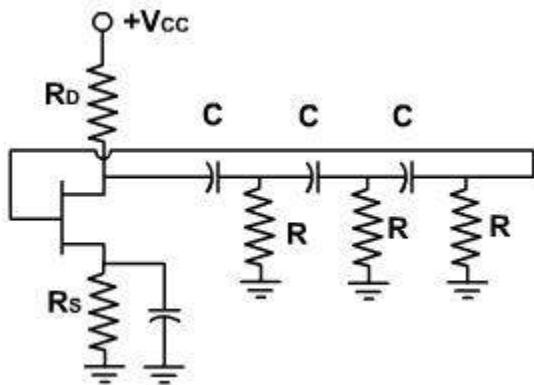


Fig. 5

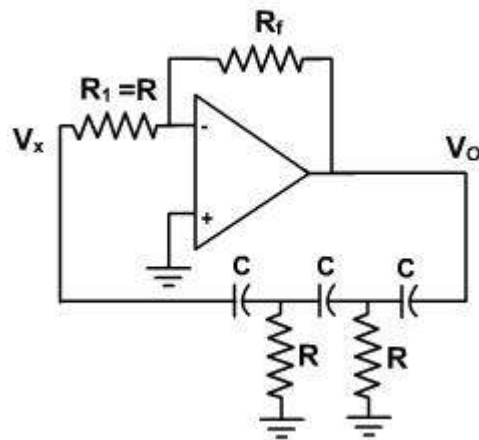


Fig. 6

The input impedance of FET is very high so that there is no loading of the feedback circuit. In this circuit, the feedback is voltage series feedback.

$$V_x = V_{GS} + V_s \quad \text{or} \quad V_{GS} = V_x - V$$

The same circuit can be realized using OPAMP. The circuit is shown in [fig. 6](#). The input impedance is very high and there is no overloading of feedback circuit. The OPAMP is connected in an inverting configuration and drives three cascaded RC sections. The inverting amplifier causes a  $180^\circ$  phase shift in the signal passing through it. RC network is used in the feedback to provide additional  $180^\circ$  phase shift. Therefore, the total phase shift in the signal, of a particular frequency, around the loop will equal  $360^\circ$  and oscillation will occur at that frequency. The gain necessary to overcome the loss in the RC network and bring the loop gain up to 1 is supplied by the amplifier. The gain is given by

$$\frac{v_o}{v_x} = - \frac{R_f}{R}$$

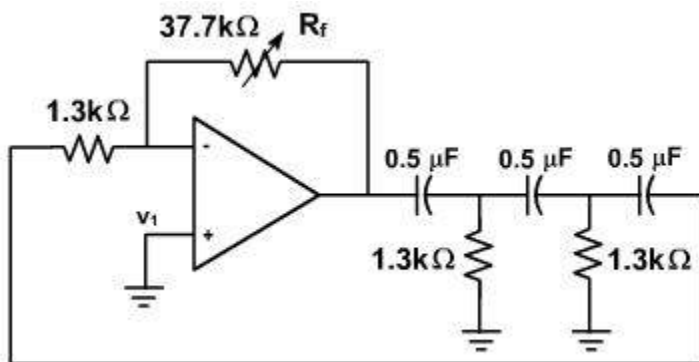
Note that input resistor to the inverting amplifier is also the last resistor of the RC feedback network.

Example -1:

Design a RC phase shift oscillator that will oscillate at 100 Hz.

**Solution:**

An RC phase shift oscillator using OPAMP is shown in [fig. 7](#). OPAMP is used as an inverting amplifier and provides  $180^\circ$  phase shift. RC network is used in the feedback to provide additional  $180^\circ$  phase shift.



**Fig. 7**

For an RC phase shift oscillator the frequency is given by

$$f = 100\text{Hz} = \frac{1}{2\pi\sqrt{6} RC}$$

Let  $C = 0.5 \mu\text{F}$ . Then

$$R = \frac{1}{100 (2\pi) \sqrt{6} (0.5 \times 10^{-6})} = 1300\Omega$$

Therefore,  $R_f = 29 R = 29 (1300\Omega) = 37.7 \text{ k}\Omega$ .

The completed circuit is shown in [fig. 7](#).  $R_f$  is made adjustable so the loop gain can be set precisely to 1.

Example - 2

For the network shown in [fig. 8](#) prove that

$$\frac{v_f}{v_o} = \frac{1}{3 + j(\omega RC - 1/\omega RC)}$$

This network is used with an OPAMP to form an oscillator. Show that the frequency of oscillation is  $f = 1/2\pi RC$  and the gain must exceed 3.

**Solution:**

To find the frequency of oscillation, let us neglect the loading of the phase shift network. Writing the KV equations,

$$V_f = I_2 R \quad (\text{E-1})$$

$$X_C = \frac{1}{\omega C} \quad (\text{E-2})$$

$$v_o = I_1 R - j(I_1 - I_2) X_C \quad (\text{E-3})$$

$$-j(I_2 - I_1) X_C - jI_2 X_C + I_2 R = 0 \quad (\text{E-4})$$

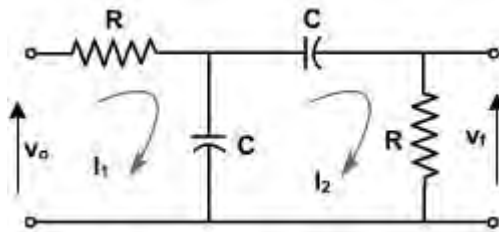
From equation (E-3),

$$I_1 = \frac{(v_o - jI_2 X_C)}{(R - jX_C)}$$

Substituting  $I_1$  in equation (E-4),

$$I_2 = \frac{-1}{(R - 2jX_C)} \left\{ \frac{v_o - jI_2 X_C}{R - jX_C} \right\}$$

Solving this equation we get,



**Fig. 8**

$$I_2 = - \left\{ \frac{-jX_c}{R^2 - 3RjX_c - X_c^2} \right\} V_o$$

Therefore, from equation (E-1),

$$\frac{V_f}{V_o} = \frac{-jRX_c}{R^2 - 3jRX_c - X_c^2}$$

Putting  $X_c = \frac{1}{\omega C}$ , we get,

$$\frac{V_f}{V_o} = \frac{1}{3 + j(R\omega C - 1/R\omega C)}$$

For phase shift equal to  $180^\circ$  between  $V_f$  and  $V_o$ , imaginary term of  $V_f / V_o$  must be zero. Therefore,

$$\begin{aligned} \omega RC - \frac{1}{\omega RC} &= 0 \\ \omega^2 C^2 &= \frac{1}{R^2} \\ \omega &= \frac{1}{RC} \\ f &= \frac{1}{2\pi RC} \end{aligned}$$

This is the frequency of oscillation. Substituting this frequency in  $V_f / V_o$  expression, we get,

$$\frac{V_f}{V_o} = \frac{1}{3} = \beta$$

This shows that  $0^\circ$  phase shift from  $V_o$  to  $V_f$  can be obtained if

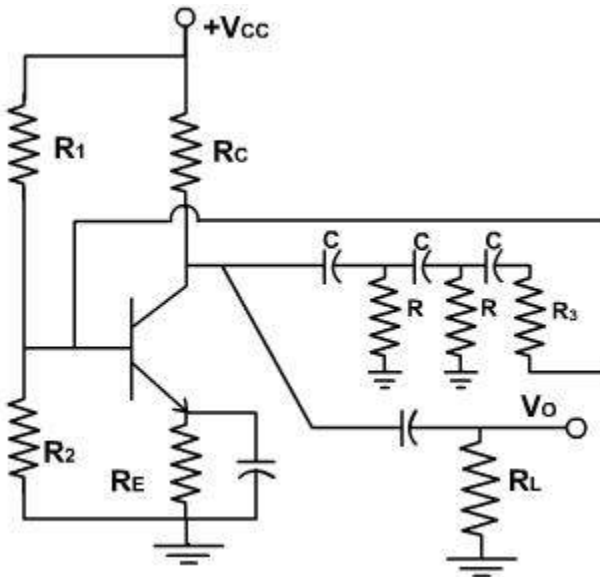
$$f = \frac{1}{2\pi RC}$$

and the gain of the feedback circuit becomes  $1/3$ . Therefore, oscillation takes place if the gain of the amplifier exceeds 3.

Transistor Phase Shift Oscillator:

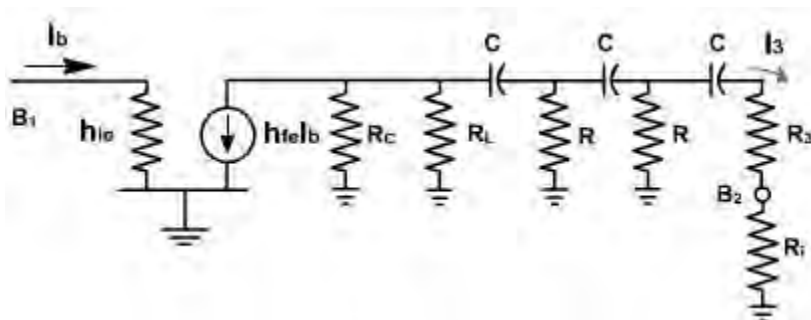
At low frequencies (around 100 kHz or less), resistors and capacitors are usually employed to determine the frequency of oscillation. [Fig. 1](#) shows transistorized phase shift oscillator circuit employing RC network. If the phase shift through the common emitter amplifier is  $180^\circ$ , then the oscillation may occur at the frequency where the RC network produces an additional  $180^\circ$  phase shift.

Since a transistor is used as the active element, the output across R of the feedback network is shunted by the relatively low input resistance of the transistor, because input diode is a forward biased diode



**Fig. 1**

Hence, instead of employing voltage series feedback, voltage shunt feedback is used for a transistor phase shift oscillator. The load resistance  $R_L$  is also connected via coupling capacitor. The equivalent circuit using h-parameter is shown in [fig. 2](#).



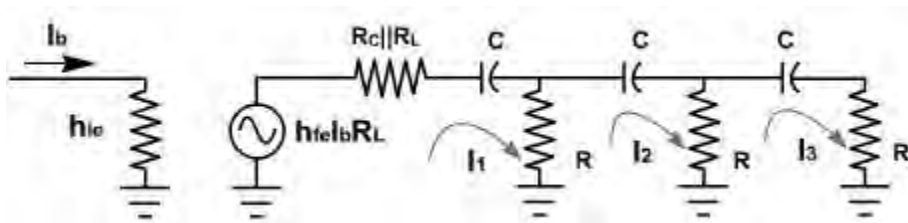
**Fig. 2**

For the circuit, the load resistance  $R_L$  may be lumped with  $R_C$  and the effective load resistance becomes  $R'_L (= R_C \parallel R_L)$ . The two h-parameters of the CE transistor amplifier,  $h_{oe}$  and  $h_{re}$  are neglected.

The capacitor  $C$  offers some impedance at the frequency of oscillation and, therefore, it is kept as it is, while the coupling capacitor behaves like ac short. The input resistance of the transistor is  $R_i \approx h_{ie}$ . Therefore the resistance  $R_3$  is selected such that  $R = R_3 + R_i = R_3 + h_{ie}$ . This choice makes the three RC selections alike and simplifies the calculation. The effect of biasing resistor  $R_1$ ,  $R_2$ , &  $R_E$  on the circuit operation is neglected.

Since this is a voltage shunt feedback, therefore instead of finding  $V_R / V_O$ , we should find the current gain of the feedback loop.

The simplified equivalent circuit is shown in [fig. 3](#).



**Fig. 3**

Applying KVL,

$$h_{fe} I_b R'_L + R'_L I_1 + X_C I_1 + R(I_1 - I_2) = 0$$

$$R(I_1 - I_2) + X_C I_2 + R(I_2 - I_3) = 0$$

$$R(I_3 - I_2) + X_C I_3 + R I_3 = 0$$

$$\therefore I_2 = \frac{1}{R} (2R + X_C) I_3$$

$$I_1 = \frac{1}{R} \left\{ \frac{(2R + X_C)^2}{R} - R \right\} I_3$$

$$\therefore h_{fe} I_b R'_L + (R'_L + R + X_C) \left\{ \frac{3R^2 + X_C^2 + 4R X_C}{R^2} \right\} I_3 - (2R + X_C) I_3 = 0$$

$$\begin{aligned} \frac{I_3}{I_b} &= \frac{h_{fe} R'_L R^2}{2R^3 + R^2 X_C - 3R^2 R'_L - X_C^2 R'_L - 4R R'_L X_C - 3R^3 - R X_C^3 - 4R^2 X_C - 3R^2 X_C - X_C^3 - 4R X_C^2} \\ &= \frac{h_{fe} R'_L R^2}{-R^3 - 6R^2 X_C - 5R X_C^2 - 3R'_L R^2 - 4R R'_L X_C - X_C^3 - R'_L X_C^2} \\ &= \frac{h_{fe} R'_L R^2}{(-R^3 - 5R X_C^2 - 3R'_L R^2 - R'_L X_C^2)(\text{Real}) + (-6R^2 X_C - 4R R'_L X_C - X_C^3)(\text{Imaginary})} \end{aligned}$$

Since  $I_3$  and  $I_b$  must be in phase to satisfy Barkhausen criterion, therefore

$$\begin{aligned} I_{\text{mag}} &= 0 \\ -6R^2 X_C - 4R R'_L X_C - X_C^3 &= 0 \\ -6R^2 - 4R R'_L &= \frac{-1}{\omega^2 C^2} \\ f &= \frac{1}{2\pi C \sqrt{6R^2 + 4R R'_L}} \\ &= \frac{1}{2\pi C R \sqrt{6 + 4\left(\frac{R'_L}{R}\right)}} \end{aligned}$$

Also initially  $I_3 > I_b$ , therefore, for oscillation to start,

$$\begin{aligned} h_{fe} R'_L R^2 &> (-R^3 - 5R X_C^2 - 3R'_L R^2 - R'_L X_C^2) \\ &> \{-R^3 - 5R(-6R^2 - 4R R'_L) - 3R'_L R^2 - R'_L(-6R^2 - 4R R'_L)\} \\ h_{fe} R'_L R^2 &> \{29R^3 + 23R^2 R'_L + 4R R'^2_L\} \\ h_{fe} &> \left\{ 29 \frac{R}{R'_L} + 23 + \frac{4R'_L}{R} \right\} \end{aligned}$$

Therefore, the two conditions must be satisfied for oscillation to start and sustain.

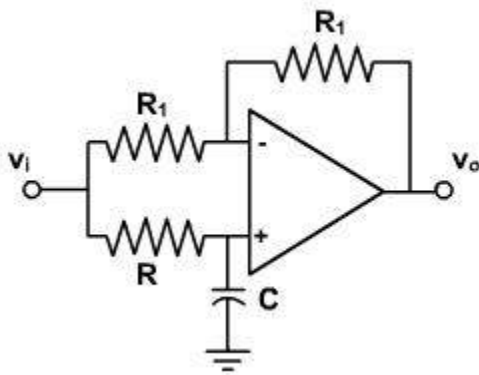
Example - 1

(a). Show that the OPAMP phase shifter shown in [fig. 4](#)

$$\frac{V_o}{V_i} = \frac{1 - j\omega RC}{1 + j\omega RC}$$

(b) Cascade two identical phase shifters of the type shown in [fig. 4](#). Complete the loop with an inverting amplifier. Show that the system will oscillate at the frequency  $f = 1 / 2\pi RC$  provided that the amplifier gain exceeds unity.

(c) Show that the circuit produces two quadrature sinusoids (sine wave differing in phase by  $90^\circ$ ).



**Fig. 4**

**Solution:**

(a) The voltage at the non-inverting terminal input of the OPAMP is given by

$$v_1 = \frac{-jX_C}{R - jX_C} V_i \quad \text{where } X_C = \frac{1}{\omega C}$$

Since the differential input voltage of OPAMP is negligible small, therefore, the voltage at the inverting terminal is also given by

$$v_2 = \frac{-jX_C}{R - jX_C} V_i$$

The input impedance of the OPAMP is very large, therefore,

$$\frac{V_i - v_2}{R_1} = \frac{v_2 - V_o}{R_1}$$

or  $V_i - 2V_2 = -V_o$

Substituting  $v_2$  in above equation, we get,

$$V_i - 2 \frac{-jX_c}{R - jX_c} V_i = -V_o$$

or,  $\frac{R + jX_c}{R - jX_c} V_i = -V_o$

Substituting  $X_c$ , we get

$$\frac{V_o}{V_i} = \frac{1 - j\omega RC}{1 + j\omega RC}$$

or,  $\frac{V_o}{V_i} = \frac{(1 - j\omega RC)(1 - j\omega RC)}{1 + (\omega RC)^2}$

or,  $\frac{V_o}{V_i} = \frac{1 - 2j\omega RC - (\omega RC)^2}{1 + (\omega RC)^2}$

$$\left| \frac{V_o}{V_i} \right| = \frac{\sqrt{(1 - (\omega RC)^2)^2 + (-2\omega RC)^2}}{1 + (\omega RC)^2}$$

Therefore,  $= 1$

and the phase angle between  $V_o$  and  $V_i$  is given by

$$\tan \phi = \frac{-2\omega RC}{1 - (\omega RC)^2}$$

or  $\phi = -2 \tan^{-1}(\omega RC)$

The magnitude of  $V_o / V_i$  is unity for all frequencies and the phase shift provided by this circuit is  $0^\circ$  for  $R = 0$  and  $-180^\circ$  for  $R \rightarrow \infty$ .

(b). If two such phase shifters are connected in cascade and an inverting amplifier with gain  $\beta$  is connected in the feedback loop, then the net loop gain becomes

$$\begin{aligned} \text{Loop gain} &= \text{Gain of phase shifter 1} \times \text{Gain of phase shifter 2} \times \beta \\ &= 1 \times 1 \times \beta \\ &= \beta \end{aligned}$$

Therefore, the oscillation takes place if gain  $\beta = 1$ , but it is kept  $> 1$  so that the losses taking place in the amplifier can be compensated.

The total phase shift around the loop is given by

$$\text{total phase shift} = -2 \tan^{-1}(\omega RC) - 2 \tan^{-1}(\omega RC) + 180^\circ$$

Further, for oscillation to take place the net phase shift around the loop would be  $0^\circ$ . Therefore,

$$-2 \tan^{-1}(\omega RC) - 2 \tan^{-1}(\omega RC) + 180^\circ = 0$$

$$\text{or } \omega RC = 1$$

$$\text{or } f = 1 / 2\pi R C$$

(c). The phase shift provided by amplifier in the feedback path is  $180^\circ$ , therefore, the phase shift provided by the phase shifters should also be  $180^\circ$  to have  $360^\circ$  or  $0^\circ$  phase shift. Thus, the phase shift provided by the individual shifter will be  $90^\circ$  as both are identical. Therefore, the sine wave produced by two phase shifters are  $90^\circ$  apart and the circuit produces two quadrature sinusoids.

### Wien Bridge Oscillator:

The Wien Bridge oscillator is a standard oscillator circuit for low to moderate frequencies, in the range 5Hz to about 1MHz. It is mainly used in audio frequency generators.

The Wien Bridge oscillator uses a feedback circuit called a lead lag network as shown in [fig. 1](#).

At very low frequencies, the series capacitor looks open to the input signal and there is no output signal. At very high frequencies the shunt capacitor looks shorted, and there is no output. In between these extremes, the output voltage reaches a maximum value. The frequency at which the output is maximized is called the resonant frequency. At this frequency, the feedback fraction reaches a maximum value of  $1/3$ .

At very low frequencies, the phase angle is positive, and the circuit acts like a lead network. On the other hand, at very high frequencies, the phase angle is negative, and the circuit acts like a lag network. In between, there is a resonant frequency  $f_r$  at which the phase angle equals  $0^\circ$ .

The output of the lag lead network is

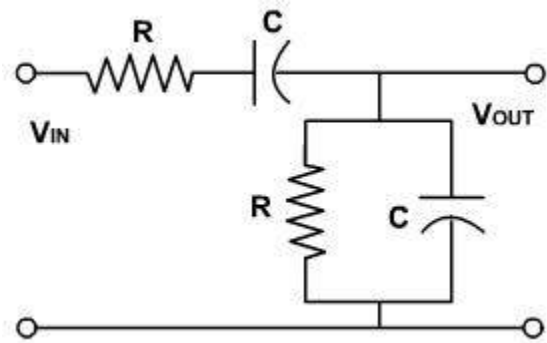


Fig. 1

$$\begin{aligned}
V_{out} &= \frac{R \parallel (-jX_c)}{R - jX_c + R \parallel (-jX_c)} V_{in} \\
&= \frac{\frac{-jR X_c}{R - jX_c}}{R - jX_c + \frac{-jR X_c}{R - jX_c}} V_{in} \\
V_{out} &= \frac{-jR X_c}{R^2 - 2jR X_c - X_c^2 - jR X_c} V_{in} \\
&= \frac{-jR X_c}{(R^2 - X_c^2) - j(3R X_c)} V_{in} \\
\frac{V_{out}}{V_{in}} &= \frac{-jR X_c \{ (R^2 - X_c^2) + j(3R X_c) \}}{(R^2 - X_c^2)^2 - j(3R X_c)^2} \\
\frac{V_{out}}{V_{in}} &= \frac{R X_c \{ 3R X_c - j(R^2 - X_c^2) \}}{(R^2 - X_c^2)^2 - j(3R X_c)^2}
\end{aligned}$$

The gain of the feedback circuit is given by

$$\begin{aligned}
\frac{V_{out}}{V_{in}} &= \frac{R X_c}{\sqrt{(R^2 - X_c^2)^2 + (3R X_c)^2}} \\
&= \frac{R X_c}{\sqrt{R^4 + X_c^4 - 2R^2 X_c^2 + 9R^2 X_c^2}} \\
&= \frac{R X_c}{R X_c \sqrt{9 + \left(\frac{R}{X_c} - \frac{X_c}{R}\right)^2}} \\
&= \frac{1}{\sqrt{9 + \left(\frac{R}{X_c} - \frac{X_c}{R}\right)^2}}
\end{aligned}$$

The phase angle between  $V_{out}$  and  $V_{in}$  is given by

$$\begin{aligned}
\varphi &= \tan^{-1} \left( \frac{-(R^2 - X_c^2)}{3R X_c} \right) \\
&= \tan^{-1} \left( \frac{X_c/R - R/X_c}{3} \right)
\end{aligned}$$

These equations show that maximum value of gain occurs at  $X_C = R$ , and phase angle also becomes  $0^\circ$ . This represents the resonant frequency of lead lag network. [Fig. 2](#), shows the gain and phase vs frequency.

$$R = \frac{1}{2\pi f_r C}$$

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

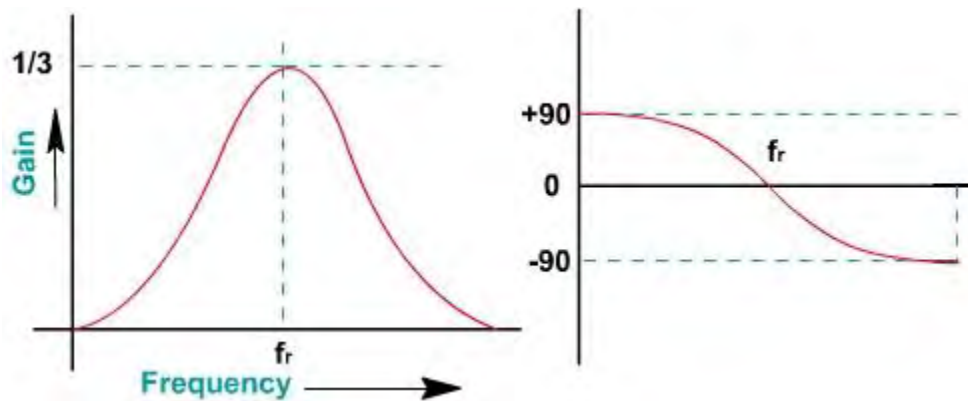
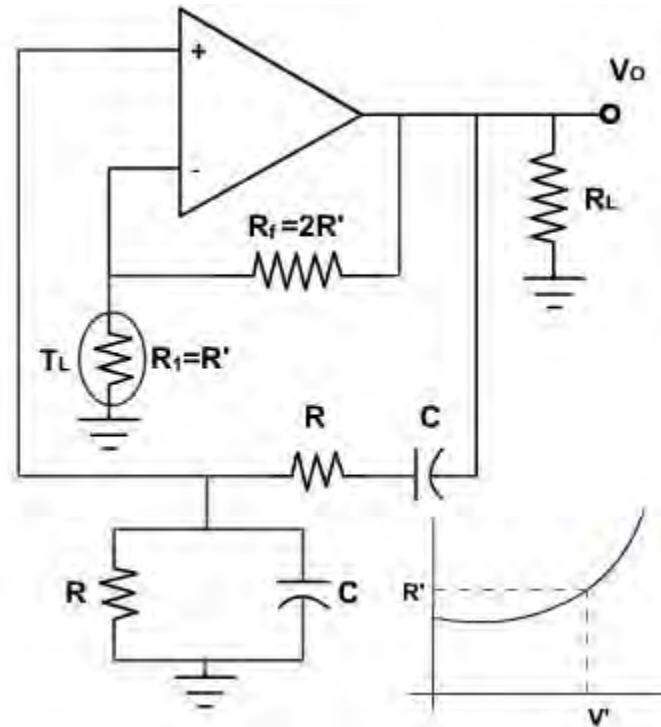


Fig. 2

How Wien Bridge Oscillator Works:

[Fig. 3](#), shows a Wien Bridge oscillator. The operational amplifier is used in a non-inverting configuration, and the lead-lag network provides the feedback. Resistors  $R_f$  and  $R_1$  determine the amplifier gain and are selected to make the loop gain equal to 1. If the feedback circuit parameters are chosen properly, there will be some frequency at which there is zero phase shift in the signal fed back to non-inverting terminal. Because the amplifier is non-inverting, it also contributes zero phase shift, so the total phase shift around the loop is 0 at that frequency, as required for oscillation.

The oscillator uses positive and negative feedback. The positive feedback helps the oscillations to build up when the power is turned on. After the output signal reaches the desired level the negative feedback reduces the loop gain to 1. The positive feedback is through the lead-lag network to the non-inverting input. Negative feedback is through the voltage divider to the inverting input.



**Fig. 3**

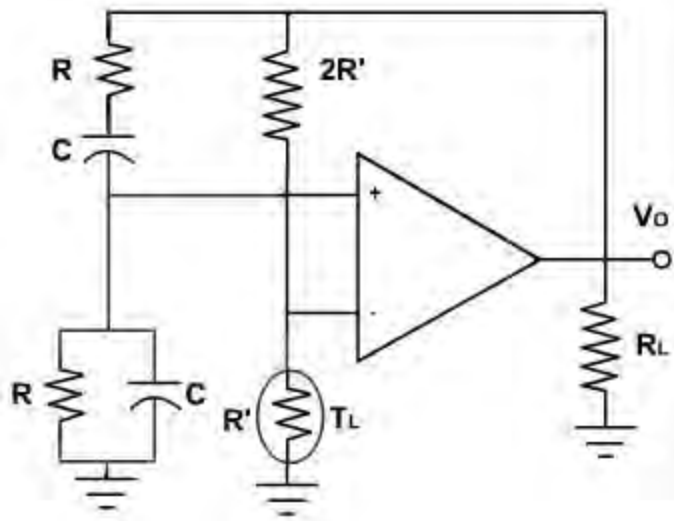
At power up, the tungsten lamp has a low resistance, and therefore, negative feedback is less. For this, reason, the loop gain AB is greater than 1, and oscillations can build up at the resonant frequency  $f_r$ . As the oscillations build up, the tungsten lamp heats up slightly and its resistance increases. At the desired output level the tungsten lamp has a resistance  $R'$ . At this point

$$A_{CL} = 1 + \frac{R_1}{R_2} = 1 + \frac{2R'}{R} = 3$$

Since the lead lag network has a gain (=B) of 1/3, the loop gain AB equals unity and then the output amplitude levels off and becomes constant. The frequency of oscillation can be adjusted by selecting R and C as

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

The amplifier must have a closed loop cut off frequency well above the resonant frequency,  $f_r$ .



**Fig. 4**

[Fig. 4](#), shows another way to represent Wein Bridge oscillator. The lead lag network is the left side of the bridge and the voltage divider is the right side. This ac bridge is called a Wein Bridge. The error voltage is the output of the Wein Bridge. When the bridge approaches balance, the error voltage approaches zero.

Example -1:

Design a Wien-bridge oscillator that oscillates at 25 kHz.

**Solution:**

Let  $C_1 = C_2 = 0.001 \mu\text{F}$ . Then, the frequency of oscillation is given by,

$$f = 25 \times 10^3 \text{ Hz} = \frac{1}{2\pi R (10^{-9}\text{F})}$$

$$\text{or, } R = \frac{1}{2\pi(25 \times 10^3 \text{ Hz})(10^{-9}\text{F})} = 6366\Omega$$

Let  $R_1 = 10 \text{ K}\Omega$ . Then,

$$\frac{R_f}{R_1} = 2$$

or,  $R_f = 20\text{K}\Omega$

## Tuned Oscillator:

A variety of oscillator circuits can be built using LC tuned circuits. A general form of tuned oscillator circuit is shown in [fig. 1](#). It is assumed that the active device used in the oscillator has very high input resistance such as FET, or an operational amplifier.

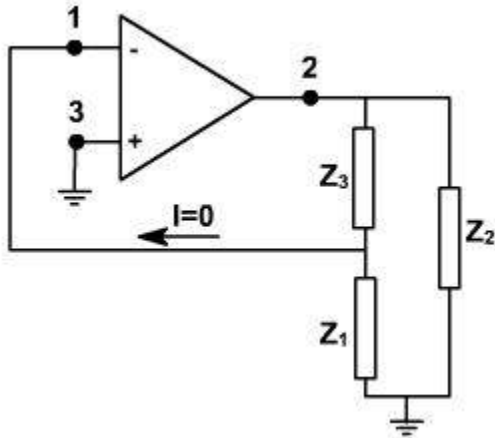


Fig. 1

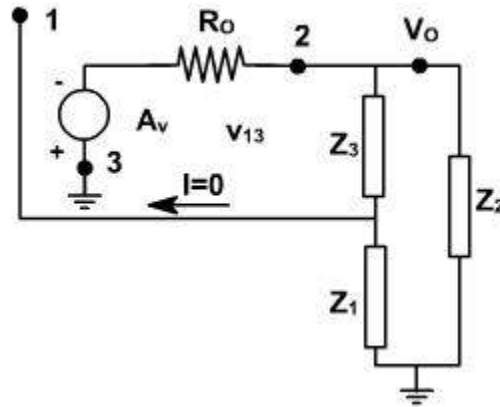


Fig. 2

[Fig. 2](#) shows linear equivalent circuit of [fig. 1](#) using an amplifier with an open circuit gain  $-A_v$  and output resistance  $R_o$ . It is clear from the topology of the circuit that it is voltage series feedback type circuit.

The loop gain of the circuit  $-A\beta$  can be obtained by considering the circuit to be a feedback amplifier with output taken from terminals 2 and 3 and with input terminals 1 and 3. The load impedance  $Z_L$  consists of  $Z_2$  in parallel with the series combination of  $Z_1$  and  $Z_3$ . The gain of the the amplifier without feedback will be given by

$$A = - \frac{A_v Z_L}{Z_L + R_o}$$

The feedback circuit gain is given by

$$\begin{aligned} \beta &= - \frac{V_f}{V_o} \\ &= - \frac{Z_1}{Z_1 + Z_3} \end{aligned}$$

Therefore, the loop gain is given by

$$\begin{aligned}
-A\beta &= -\frac{A_v Z_L}{Z_L + R_o} \cdot \frac{Z_1}{Z_1 + Z_3} \\
&= -\frac{A_v}{1 + \frac{R_o}{Z_L}} \cdot \frac{Z_1}{Z_1 + Z_3} \\
&= -\frac{A_v}{1 + \frac{R_o(Z_2 + Z_1 + Z_3)}{Z_2(Z_1 + Z_3)}} \cdot \frac{Z_1}{Z_1 + Z_3} \\
&= -\frac{A_v Z_1 Z_2}{Z_2(Z_1 + Z_3) + R_o(Z_2 + Z_1 + Z_3)}
\end{aligned}$$

If the impedances are pure reactances (either inductive or capacitive), then  $Z_1 = jX_1$ ,  $Z_2 = jX_2$  and  $Z_3 = jX_3$ . Then

$$-A\beta = \frac{A_v X_1 X_2}{jR_o(X_2 + X_1 + X_3) - X_2(X_1 + X_3)}$$

For the loop gain to be real (zero phase shift around the loop),

$$X_1 + X_2 + X_3 = 0$$

$$\begin{aligned}
-A\beta &= \frac{A_v X_1 X_2}{-X_2(X_1 + X_3)} \\
&= \frac{-A_v X_1}{(X_1 + X_3)} \\
&= \frac{-A_v X_1}{X_2}
\end{aligned}$$

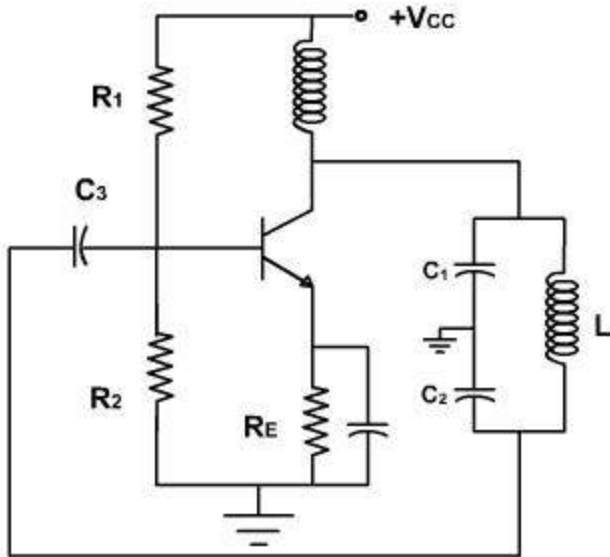
and

Therefore, the circuit will oscillate at the resonant frequency of the series combination of  $X_1$ ,  $X_2$  and  $X_3$ . Since  $-A\beta$  must be positive and at least unity in magnitude, then  $X_1$  and  $X_2$  must have the same sign ( $A_v$  is positive). In other words, they must be the same kind of reactance, either both inductive or both capacitive.

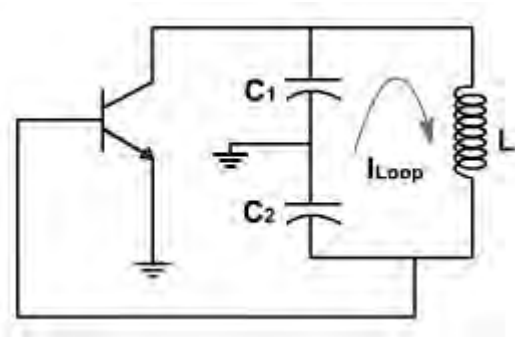
### The Colpitts Oscillator:

Wein bridge oscillator is not suited to high frequencies (above 1MHz). The main problem is the phase shift through the amplifier.

The alternative is an LC oscillator, a circuit that can be used for frequencies between 1MHz and 500MHz. The frequency range is beyond the frequency limit of most OPAMPs. With an amplifier and LC tank circuit, we can feedback a signal with the right amplitude and phase is feedback to sustain oscillations. [Fig. 3](#), shows the circuit of colpitts oscillator.



**Fig. 3**



**Fig. 4**

The voltage divider bias sets up a quiescent operating point. The circuit then has a low frequency voltage gain of  $r_c / r'_e$  where  $r_c$  is the ac resistance seen by the selector. Because of the base and collector lag networks, the high frequency voltage gain is less than  $r_c / r'_e$ .

**Fig. 4**, shows a simplified ac equivalent circuit. The circulating or loop current in the tank flows through  $C_1$  in series with  $C_2$ . The voltage output equals the voltage across  $C_1$ . The feedback voltage  $v_f$  appears across  $C_2$ . This feedback voltage drives the base and sustains the oscillations developed across the tank circuit provided there is enough voltage gain at the oscillation frequency. Since the emitter is at ac ground the circuit is a CE connection.

Most LC oscillators use tank circuit with a Q greater than 10. The Q of the feedback circuit is given by

$$Q = R_1 \sqrt{\frac{C}{L}}$$

Because of this, the approximate resonant frequency is

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

This is accurate and better than 1% when Q is greater than 10. The capacitance C is the equivalent capacitance the circulation current passes through. In the Colpitts tank the circulating current flows through  $C_1$  in series with  $C_2$ .

Therefore  $C = C_1 C_2 / (C_1 + C_2)$

The required starting condition for any oscillator is  $A\beta > 1$  at the resonant frequency or  $A > 1/\beta$ . The voltage gain  $A$  in the expression is the gain at the oscillation frequency. The feedback gain  $\beta$  is given by

$$\beta = v_f / v_{out} \approx X_{C1} / X_{C2}$$

Because same current flow through  $C_1$  and  $C_2$ , therefore

$$\beta = C_1 / C_2; \quad A > 1/\beta; \quad A > C_1 / C_2$$

This is a crude approximation because it ignores the impedance looking into the base. An exact analysis would take the base impedance into account because it is in parallel with  $C_2$ .

With small  $\beta$ , the value of  $A$  is only slightly larger than  $1/\beta$ . and the operation is approximately close  $A$ . When the power is switched on, the oscillations build up, and the signal swings over more and more of ac load line. With this increased signal swing, the operation changes from small signal to large signal. As this happen, the voltage gain decreases slightly. With light feedback the value of  $A\beta$  can decrease to 1 without excessive clapping.

With heavy feedback, the large feedback signal drives the base into saturation and cut off. This charges capacitor  $C_3$  producing negative dc clamping at the base and changing the operation from class A to class C. The negative damping automatically adjusts the value of  $A\beta$  to 1.

### Example - 1

Design a Colpitts oscillator that will oscillate at 100 kHz.

#### Solution:

Let us choose  $R_1 = R_f = 5 \text{ k}\Omega$  and  $C = 0.001 \text{ }\mu\text{F}$ . From the frequency expression,

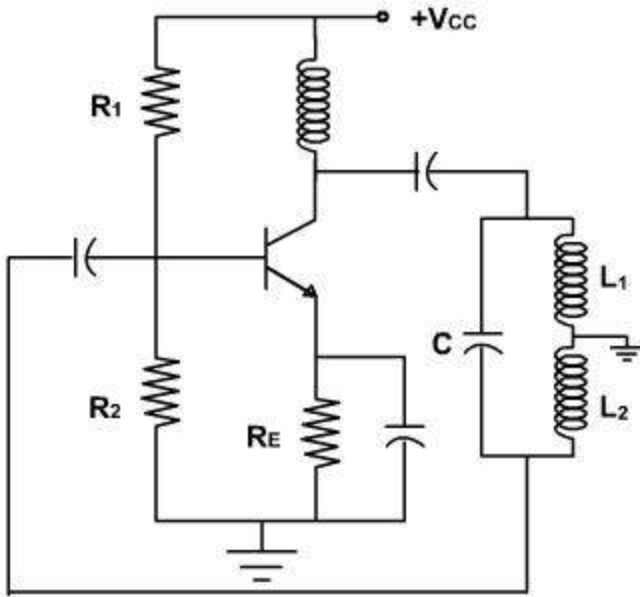
$$L = \frac{2}{(2\pi)^2 f_0^2 C} = \frac{2}{(2\pi)^2 (100 \times 10^3 \text{ Hz})^2 (0.001 \times 10^{-6} \text{ F})} \approx 5 \text{ mH}$$

The quality factor (Q) of the LC circuit is given by:

$$Q = 5 \text{ k} \sqrt{\frac{1 \text{ nF}}{5 \text{ mH}}} \approx 2.2$$

#### Hartley Oscillator:

[Fig. 5](#), shows Hartley oscillator when the LC tank is resonant, the circulating current flows through  $L_1$  in series with  $L_2$ . Thus, the equivalent inductance is  $L = L_1 + L_2$ .



**Fig. 5**

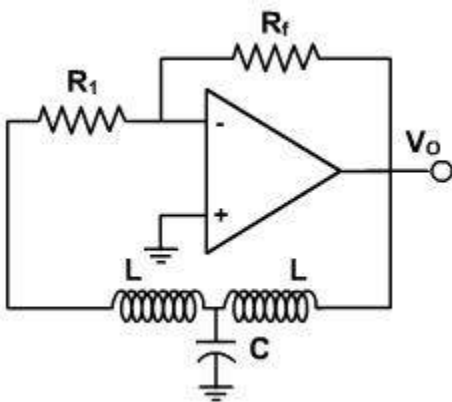
In the oscillator, the feedback voltage is developed by the inductive voltage divider,  $L_1$  &  $L_2$ . Since the output voltage appears across  $L_1$  and the feedback voltage across  $L_2$ , the feedback fraction is

$$\beta = V / V_{out} = X_{L2} / X_{L1} = L_2 / L_1$$

As usual, the loading effect of the base is ignored. For oscillations to start, the voltage gain must be greater than  $1/\beta$ . The frequency of oscillation is given by

$$f_r = \frac{1}{2\pi\sqrt{2LC}}$$

Similarly, an opamp based Hartley oscillator circuit is shown in [fig. 6](#).



**Fig. 6**

**Crystal Oscillator:**

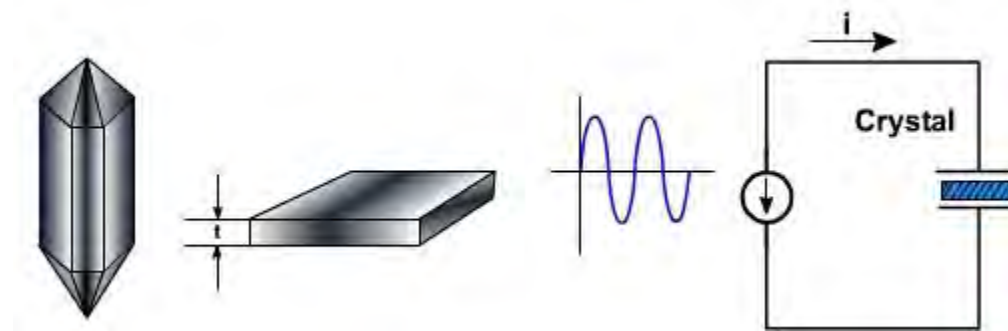
Some crystals found in nature exhibit the piezoelectric effect i.e. when an ac voltage is applied across them, they vibrate at the frequency of the applied voltage. Conversely, if they are mechanically pressed, they generate an ac voltage. The main substances that produce this piezoelectric effect are Quartz, Rochelle salts, and Tourmaline.

Rochelle salts have greatest piezoelectric activity, for a given ac voltage, they vibrate more than quartz or tourmaline. Mechanically, they are the weakest they break easily. They are used in microphones, phonograph pickups, headsets and loudspeakers.

Tourmaline shows the least piezoelectric activity but is a strongest of the three. It is also the most expensive and used at very high frequencies.

Quartz is a compromise between the piezoelectric activity of Rochelle salts and the strength of tourmaline. It is inexpensive and easily available in nature. It is most widely used for RF oscillators and filters.

The natural shape of a quartz crystal is a hexagonal prism with pyramids at the ends. To get a useable crystal out of this it is sliced in a rectangular slab form of thickness  $t$ . The number of slabs we can get from a natural crystal depends on the size of the slabs and the angle of cut.



**Fig. 1**

For use in electronic circuits, the slab is mounted between two metal plates, as shown in [fig. 1](#). In this circuit the amount of crystal vibration depends upon the frequency of applied voltage. By changing the frequency, one can find resonant frequencies at which the crystal vibrations reach a maximum. Since the energy for the vibrations must be supplied by the ac source, the ac current is maximized at each resonant frequency. Most of the time, the crystal is cut and mounted to vibrate best at one of its resonant frequencies, usually the fundamental or lowest frequency. Higher resonant frequencies, called overtones, are almost exact multiplies of the fundamental

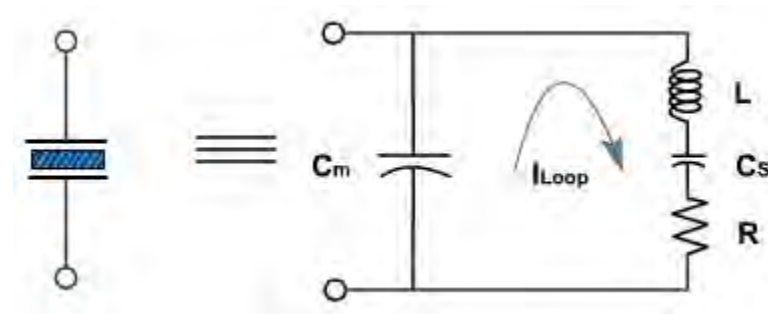
frequency e.g. a crystal with a fundamental frequency of 1 MHz has a overtones of 2 MHz, 3 MHz and so on. The formula for the fundamental frequency of a crystal is

$$f = K / t.$$

where K is a constant that depends on the cut and other factors, t is the thickness of crystal, f is inversely proportional to thickness t. The thinner the crystal, the more fragile it becomes and the more likely it is to break because of vibrations. Quartz crystals may have fundamental frequency up to 10 MHz. To get higher frequencies, a crystal is mounted to vibrate on overtones; we can reach frequencies up to 100 MHz.

### AC Equivalent Circuit:

When the mounted crystal is not vibrating, it is equivalent to a capacitance  $C_m$ , because it has two metal plates separated by dielectric,  $C_m$  is known as mounting capacitance.



**Fig. 2**

When the crystal is vibrating, it acts like a tuned circuit. [Fig. 2](#), shows the ac equivalent circuit of a crystal vibrating at or near its fundamental frequency. Typical values are L is henrys, C in fractions of a Pico farad, R in hundreds of ohms and  $C_m$  in Pico farads

$$L_s = 3\text{Hz}, \quad C_s = 0.05 \text{ pf}, \quad R_s = 2\text{K}, \quad C_m = 10 \text{ pf}.$$

The Q of the circuit is very very high. Compared with L-C tank circuit. For the given values, Q comes out to be 3000. Because of very high Q, a crystal leads to oscillators with very stable frequency values.

The series resonant frequency  $f_s$  of a crystal is the sonant frequency of the LCR branch. At this frequency, the branch current reaches a maximum value because  $L_s$  resonant with  $C_s$ .

$$f_s = \frac{1}{2\pi\sqrt{L_s C_s}}$$

Above  $f_s$ , the crystal behaves inductively. The parallel resonant frequency is the frequency at which the circulating or loop current reaches a maximum value. Since this loop current must flow through the series combination of  $C_s$  and  $C_m$ , the equivalent  $C_{loop}$  is

$$C_{loop} = \frac{C_m C_s}{C_m + C_s}$$

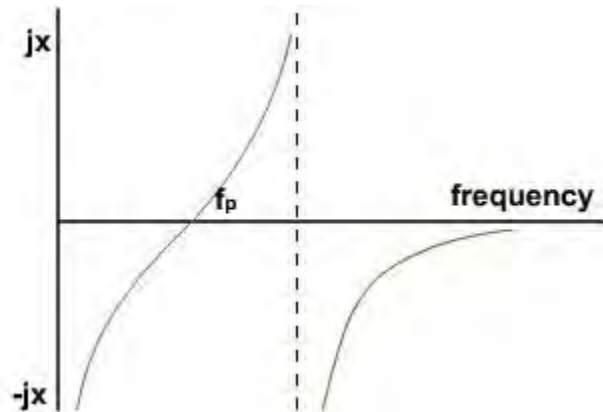
and  $f_p = \frac{1}{2\pi\sqrt{L_s C_{loop}}}$

Since  $C_{loop} > C_s$ , therefore,  $f_p > f_s$ .

Since  $C_m > C_s$ , therefore,  $C_m \parallel C_s$  is slightly lesser than  $C_s$ . Therefore  $f_p$  is slightly greater than  $f_s$ . Because of the other circuit capacitances that appear across  $C_m$  the actual frequency will lie between  $f_s$  and  $f_p$ .  $f_s$  and  $f_p$  are the upper and lower limits of frequency. The impedance of the crystal oscillator can be plotted as a function of frequency as shown in [fig. 3](#).

At frequency  $f_s$ , the circuit behaves like resistive circuit. At  $f_p$  the impedance reaches to maximum, beyond  $f_p$ , the circuit is highly capacitive.

The frequency of an oscillator tends to change slightly with time. The drift is produced by temperature, aging and other causes. In a crystal oscillator the frequency drift with time is very small, typically less than 1 part in  $10^6$  per day. They can be used in electronic wristwatches. If the drift is 1 part in  $10^{10}$ , a clock with this drift will take 30 years to gain or lose 1 sec.



**Fig. 3**

Crystals can be manufactured with values of  $f_s$  as low as 10 kHz; at these frequencies the crystal is relatively thick. On the high frequency side,  $f_s$  can be as high as 1- MHz; here the crystal is very thin.

The temperature coefficient of crystals is usually small and can be made zero. When extreme temperature stability is required, the crystal may be housed in an oven to maintain it at a constant temperature. The high Q of the crystal also contributes to the relatively drift free oscillation of crystal oscillators.

### Example - 1

The parameters of the equivalent circuit of a crystal are given below:

$L = 0.4 \text{ H}$ ,  $C_s = 0.06 \text{ pF}$ ,  $R = 5 \text{ k}\Omega$ ,  $C_m = 1.0 \text{ pF}$ .

Determine the series and parallel resonant frequencies of the crystal.

**Solution:**

With reference to **fig. 2**, the admittance of the crystal  $Y$  is given by

$$\begin{aligned}
 Y &= j\omega C_m + \frac{1}{(j\omega L + R) + 1/(j\omega C_s)} \\
 &= j\omega C_m + \frac{j\omega C_s [(1 - \omega^2 LC_s) - j\omega C_s R]}{(1 - \omega^2 LC_s)^2 + (\omega^2 C_s^2 R^2)} \\
 &= G + jB
 \end{aligned}$$

$$G = \frac{\omega^2 C_s^2 R}{(1 - \omega^2 LC_s)^2 + (\omega^2 C_s^2 R^2)}$$

where,

and 
$$B = \omega C_m + \frac{\omega C_s [1 - \omega^2 LC_s]}{(1 - \omega^2 LC_s)^2 + (\omega^2 C_s^2 R^2)}$$

The resonant frequencies are obtained by putting  $B = 0$ . Thus,

$$\begin{aligned}
 \omega^4 L^2 C_s^2 + \omega^2 [(-2LC_s) + (C_s^2 R^2) - (LC_s^2/C_m)] + [1 + (C_s/C_m)] &= 0 \\
 \omega^4 + [\omega^2/(LC_s)] [(-2) + (C_s R^2/L) - (C_s/C_m)] + [1/L^2 C_s^2][1 + (C_s/C_m)] &= 0
 \end{aligned}$$

Consider the term  $C_s R^2 / L_s = C_s R / [L R]$ . In a crystal, the time constant  $(L / R)$  is very much greater than  $C_s R$ . Thus the ratio is very much less than 1. For the values given, this ratio is of the order of  $10^{-6}$ . Neglecting this term in comparison with 2, we get two roots as

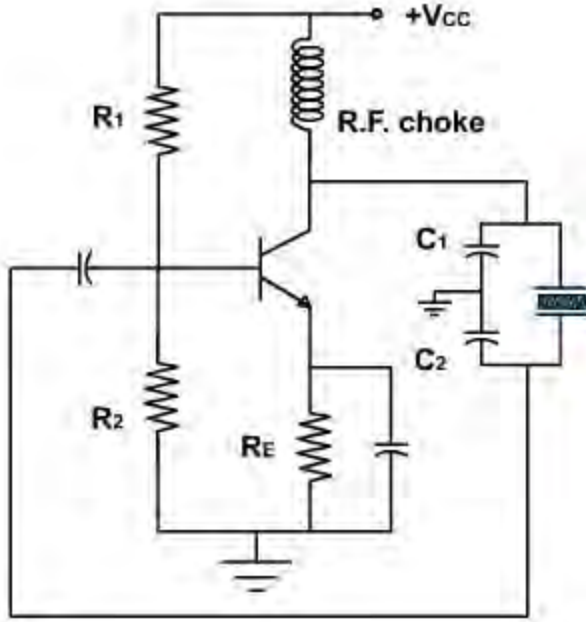
$$\begin{aligned}
 \omega_s^2 &= 1/(LC_s) \\
 \text{and } \omega_p^2 &= 1/[L\{(C_s C_m)/(C_s + C_m)\}]
 \end{aligned}$$

where,  $\omega_s$  and  $\omega_p$  are the series and parallel resonant frequencies respectively. Substituting the values, we get

$$\omega_s = 6.45 \text{ M Hz. and } \omega_p = 6.64 \text{ MHz}$$

Crystal Oscillators:

**Fig. 4**, shows a colpitts crystal oscillator.

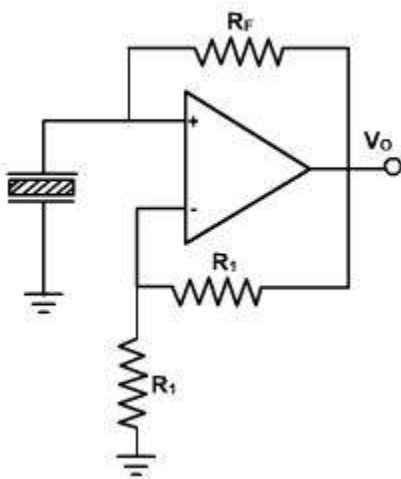


**Fi. 4**

The capacitive voltage divider produces the feedback voltage for the base of transistor. The crystal acts like an inductor that resonates with  $C_1$  and  $C_2$ . The oscillation frequency is between the series and parallel resonant frequencies.

Example-2:

If the crystal of example-1 is used in the oscillator circuit as shown [fig. 5](#), determine the values of  $R$  for the circuit to oscillate.



**Fig. 5**

**Solution:**

The equivalent circuit of crystal (discussed earlier) shows that it has a parallel resonant frequency ( $\omega_p$ ) at which the impedance becomes maximum. The amplified signal output of the circuit is applied across the potential divider consisting of R and the crystal circuit. At the resonant frequency the impedance of crystal becomes maximum (magnitude R) and thus the loop gain will be greater than or equal to unity. At frequencies away from  $\omega_p$  the loop gain becomes less than unity. The loop phase shift is also zero around  $\omega_p$ . Thus both the conditions required for sustained oscillations are satisfied and the circuit oscillates.

The value of G of the crystal at  $\omega = \omega_p$  is given by

$$G = \frac{1}{R_j \left[ \frac{(L_s/R_s)(C_s/C_m)}{(C_s + C_m)R_s} + 1 \right]} \approx \frac{1}{4.3 \times 10^6} \Omega$$

Thus the resistor R should be less than  $5 \times 10^6 \Omega$ .

## UNIT V

### PULSE CIRCUITS

#### Clipper Circuits

Clippers:

Clipping circuits are used to select that portion of the input wave which lies above or below some reference level. Some of the clipper circuits are discussed here. The transfer characteristic ( $v_o$  vs  $v_i$ ) and the output voltage waveform for a given input voltage are also discussed.

#### Clipper Circuit 1:

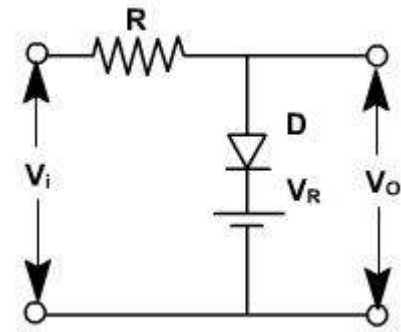
The circuit shown in [fig. 3](#), clips the input signal above a reference voltage ( $V_R$ ).

In this clipper circuit,

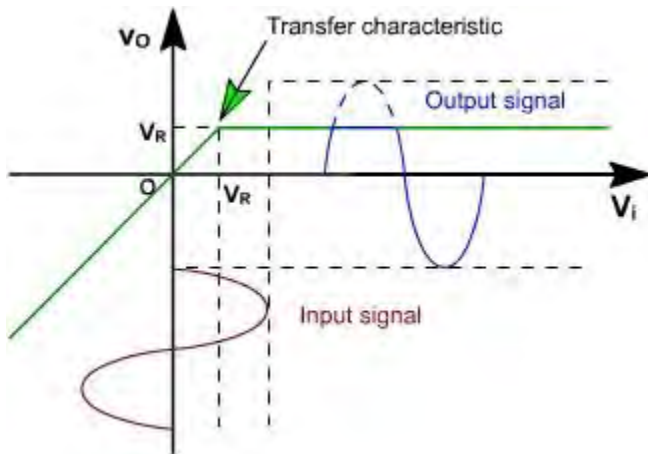
If  $v_i < V_R$ , diode is reversed biased and does not conduct. Therefore,  $v_o = v_i$

and, if  $v_i > V_R$ , diode is forward biased and thus,  $v_o = V_R$ .

The transfer characteristic of the clippers is shown in [fig. 4](#).



**Fig. 3**



**Fig. 4**

**Clipper Circuit 2:**

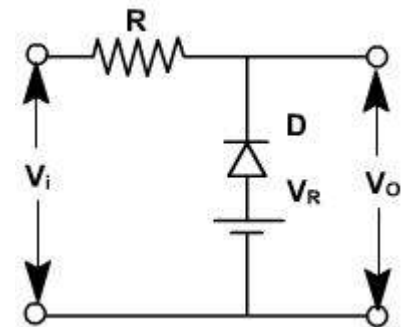
The clipper circuit shown in [fig. 5](#) clips the input signal below reference voltage  $V_R$ .

In this clipper circuit,

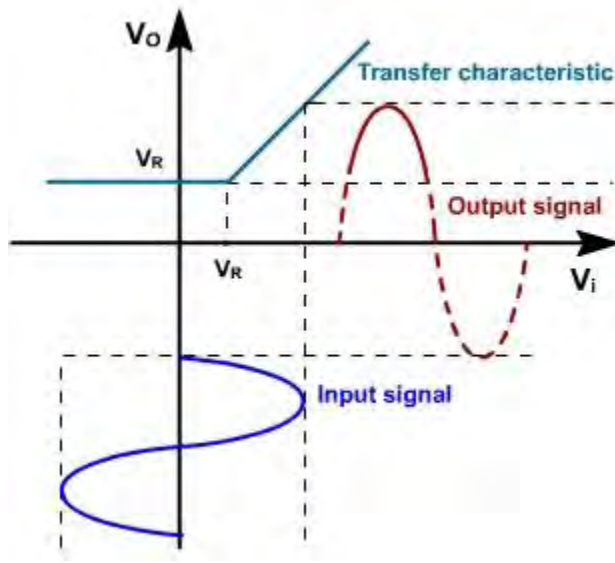
If  $v_i > V_R$ , diode is reverse biased.  $v_o = v_i$

and, If  $v_i < V_R$ , diode is forward biased.  $v_o = V_R$

The transfer characteristic of the circuit is shown in [fig. 6](#).



**Fig. 5**



**Fig. 6**

**Clipper Circuit 3:**

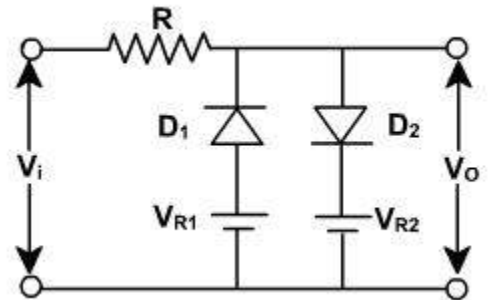
To clip the input signal between two independent levels ( $V_{R1} < V_{R2}$ ), the clipper circuit is shown in [fig. 7](#).

The diodes  $D_1$  &  $D_2$  are assumed ideal diodes.

For this clipper circuit, when  $v_i \leq V_{R1}$ ,  $v_o = V_{R1}$

and,  $v_i \geq V_{R2}$ ,  $v_o = V_{R2}$

and,  $V_{R1} < v_i < V_{R2}$   $v_o = v_i$



**Fig. 7**

The transfer characteristic of the clipper is shown in [fig. 8](#).

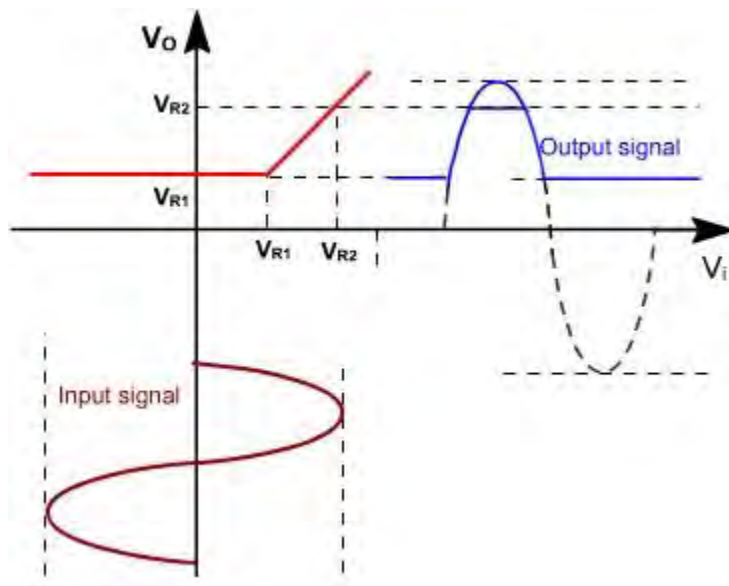


Fig. 8

## Clipper Circuits

### Example – 1:

Draw the transfer characteristic of the circuit shown in [fig. 9](#).

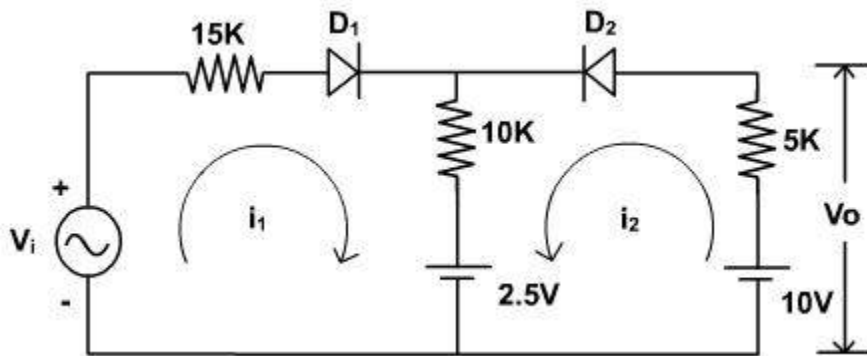


Fig. 9

### Solution:

When diode  $D_1$  is off,  $i_1 = 0$ ,  $D_2$  must be ON.

$$i_2 = \frac{10 - 2.5}{10k + 5k} = \frac{7.5}{15k} = 0.5 \text{ mA}$$

and  $v_o = 10 - 5 \times 0.25 = 7.5 \text{ V}$

$$v_p = v_o = 7.5 \text{ V}$$

Therefore,  $D_1$  is reverse biased only if  $v_i < 7.5 \text{ V}$

If  $D_2$  is off and  $D_1$  is ON,  $i_2 = 0$

$$i_1 = \frac{v_i - 2.5}{15k + 10k} = 0.04v_i - 0.1$$

and  $v_p = 10 ( 0.04 v_i - 0.1 ) + 2.5 = 0.4 v_i + 1.5$

For  $D_2$  to be reverse biased,

$$v_p > 10 \text{ V}$$

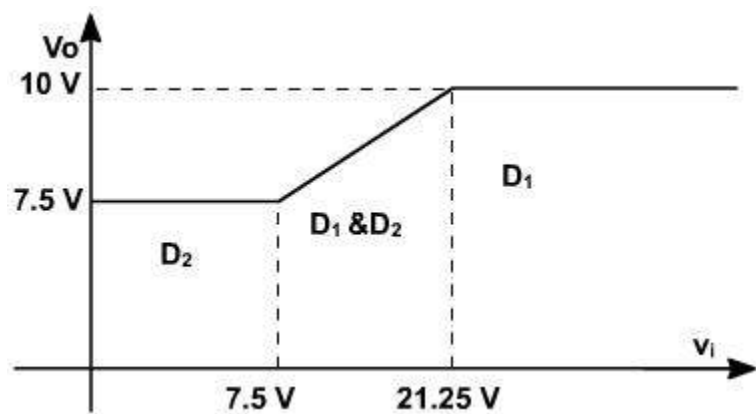
$$0.04v_i + 1.5 > 10 \text{ V}$$

$$0.4v_i > 8.5 \text{ V}$$

$$v_i > 21.25 \text{ V}$$

Between 7.5 V and 21.25 V both the diodes are ON.

$$\begin{aligned} v_i &= 15i_1 + 10(i_1 + i_2) + 2.5 \\ 10 &= 5i_2 + 10(i_1 + i_2) + 2.5 \\ i_2 &= \frac{10v_i - 250 - 25 + 62.5}{100 - 375} \\ v_o &= 10 - 5i_2 \\ &= 10 - 5 \left( \frac{10v_i - 275 + 62.5}{-275} \right) \\ &= 10 + \frac{50v_i - 1375 + 312.5}{275} \\ v_o &= \frac{2v_i + 67.5}{11} \end{aligned}$$



**Fig. 10**

The transfer characteristic of the circuit is shown in [fig. 10](#).

## Clipper and Clamper Circuits

Clippers:

In the clipper circuits, discussed so far, diodes are assumed to be ideal device. If third approximation circuit of diode is used, the transfer characteristics of the clipper circuits will be modified.

Clipper Circuit 4:

Consider the clipper circuit shown in [fig. 1](#) to clip the input signal above reference voltage

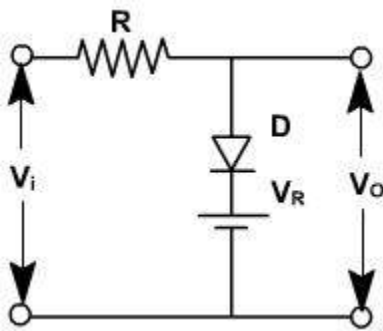


Fig. 1

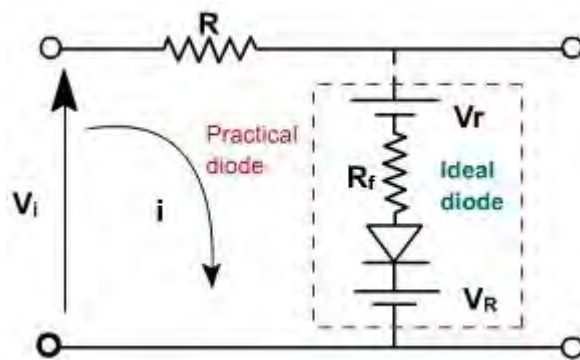


Fig. 2

When  $v_i < (V_R + V_r)$ , diode D is reverse biased and therefore,  $v_o = v_i$ .

and when  $v_i > (V_R + V_r)$ , diode D is forward biased and conducts. The equivalent circuit, in this case is shown in [fig. 2](#).

The current  $i$  in the circuit is given by

$$i = \frac{v_i - V_r - V_R}{R + R_f}$$

$$v_o = V_r + V_R + iR_f$$

$$= V_r + V_R + \frac{(V_i - V_r - V_R)R_f}{R + R_f}$$

$$v_o = (V_r + V_R) \frac{R}{R + R_f} + \frac{R_f}{R + R_f} v_i$$

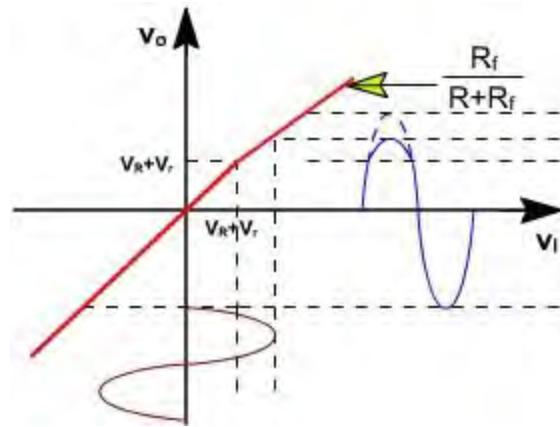


Fig. 3

The transfer characteristic of the circuit is shown in [fig. 3](#).

### Clipper Circuit 5:

Consider the clipper circuit shown in [fig. 4](#), which clips the input signal below the reference level ( $V_R$ ).

If  $v_i > (V_R - V_r)$ , diode  $D$  is reverse biased, thus  $v_o = v_i$  and when  $v_i < (V_R - V_r)$ ,  $D$  conducts and the equivalent circuit becomes as shown in [fig. 5](#).

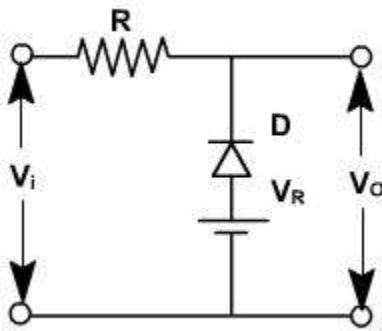


Fig. 4

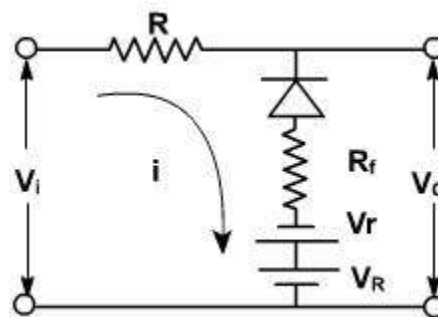


Fig. 5

Therefore,

$$i = \frac{v_i + V_r + V_R}{R + R_f}$$

$$v_o = iR_f + V_R - V_r$$

$$= \frac{R_f}{R + R_f} (v_i - (V_R - V_r)) + (V_R + V_r)$$

$$= \frac{R}{R + R_f} (V_R - V_r) + \frac{R_f v_i}{R + R_f}$$

The transfer characteristic of the circuit is shown in [fig. 6](#).

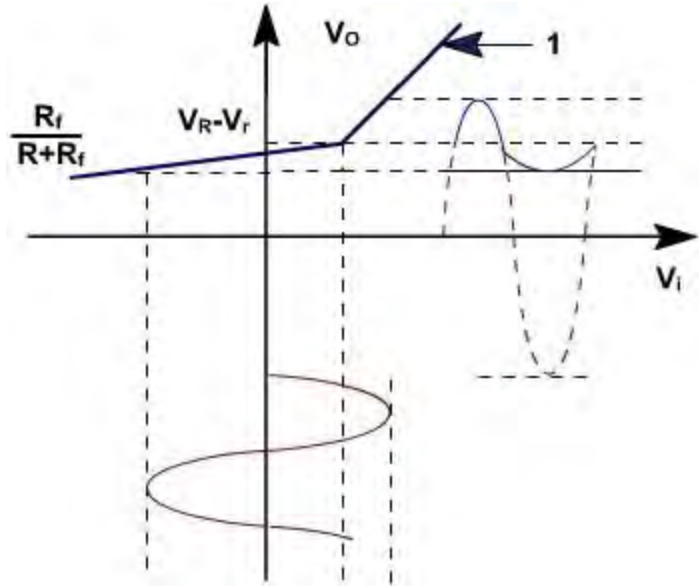


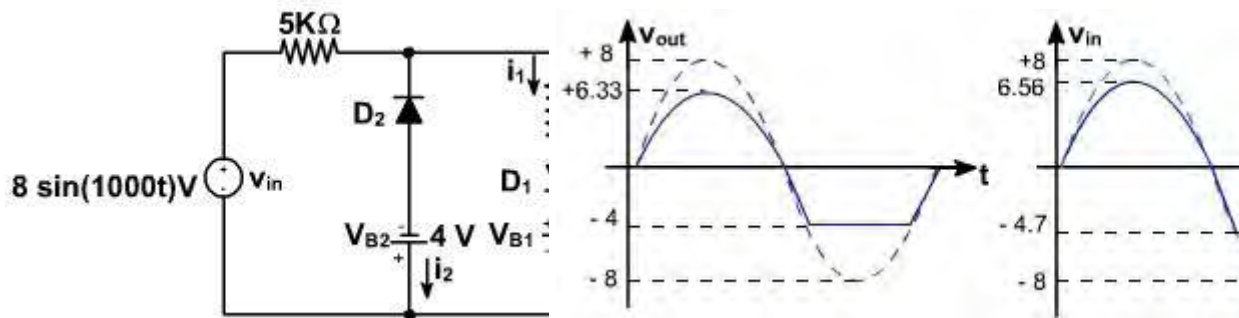
Fig. 6

## Clipper and Clamper Circuits

Example - 1:

Find the output voltage  $v$  out of the clipper circuit of [fig. 7\(a\)](#) assuming that the diodes are

- ideal.
- $V_{on} = 0.7$  V. For both cases, assume  $R_F$  is zero.



**Fig. 7(a)**

**Fig. 7(b)**

**Solution:**

(a). When  $v_{in}$  is positive and  $v_{in} < 3$ , then  $v_{out} = v_{in}$

and when  $v_{in}$  is positive and  $v_{in} > 3$ , then

$$i_1 = \frac{v_{in} - 3}{1.5 \times 10^4}$$

$$v_{out} = 10^4 i_1 + 3 = \frac{2}{3} v_{in} + 1$$

At  $v_{in} = 8$  V(peak),  $v_{out} = 6.33$  V.

When  $v_{in}$  is negative and  $v_{in} > -4$ , then  $v_{out} = v_{in}$

When  $v_{in}$  is negative and  $v_{in} < -4$ , then  $v_{out} = -4$  V

The resulting output wave shape is shown in [fig. 7\(b\)](#).

(b). When  $V_{ON} = 0.7$  V,  $v_{in}$  is positive and  $v_{in} < 3.7$  V, then  $v_{out} = v_{in}$

When  $v_{in} > 3.7$  V, then

$$i_1 = \frac{v_{in} - 3.7}{1.5 \times 10^4}$$

$$v_{out} = 10^4 i_1 + 3.7 = \frac{2}{3} v_{in} + 1.23$$

When  $v_{in} = 8$  V,  $v_{out} = 6.56$  V.

When  $v_{in}$  is negative and  $v_{in} > -4.7$  V, then  $v_{out} = v_{in}$

When  $v_{in} < -4.7$  V, then  $v_{out} = -4.7$  V

The resulting output wave form is shown in [fig. 7\(b\)](#).

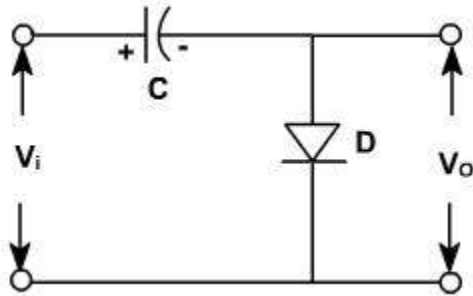
**Clamper Circuits:**

Clamping is a process of introducing a dc level into a signal. For example, if the input voltage swings from -10 V and +10 V, a positive dc clamper, which introduces +10 V in the input will produce the output that swings ideally from 0 V to +20 V. The complete waveform is lifted up

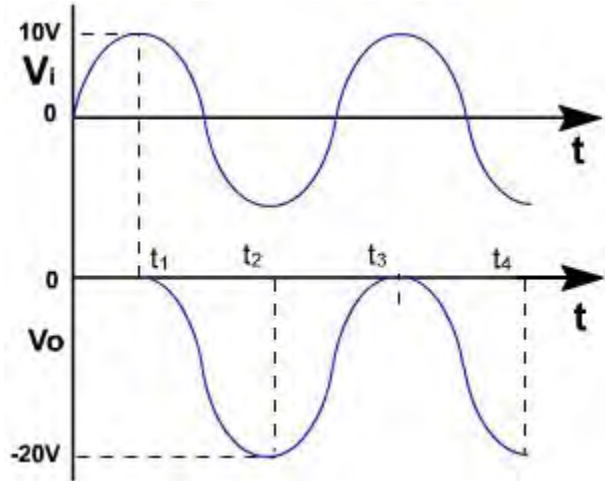
by +10 V.

Negative Diode clamper:

A negative diode clamper is shown in [fig. 8](#), which introduces a negative dc voltage equal to peak value of input in the input signal.



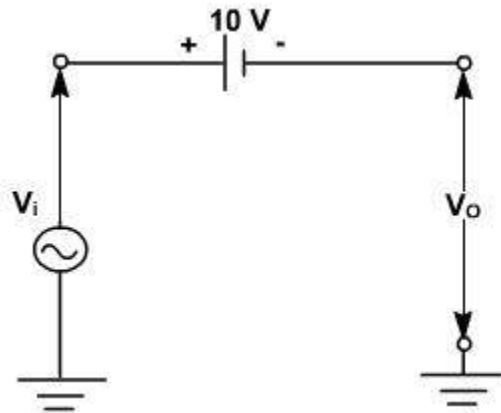
**Fig. 8**



**Fig. 9**

Let the input signal swings form +10 V to -10 V. During first positive half cycle as  $V_i$  rises from 0 to 10 V, the diode conducts. Assuming an ideal diode, its voltage, which is also the output must be zero during the time from 0 to  $t_1$ . The capacitor charges during this period to 10 V, with the polarity shown.

At that  $V_i$  starts to drop which means the anode of D is negative relative to cathode, ( $V_D = v_i - v_c$ ) thus reverse biasing the diode and preventing the capacitor from discharging. [Fig. 9](#). Since the capacitor is holding its charge it behaves as a DC voltage source while the diode appears as an open circuit, therefore the equivalent circuit becomes an input supply in series with -10 V dc voltage as shown in [fig. 10](#), and the resultant output voltage is the sum



**Fig. 10**

of instantaneous input voltage and dc voltage

(-10 V).

## Clamper Circuits

### Positive Clamper:

The positive clamper circuit is shown in [fig. 1](#), which introduces positive dc voltage equal to the peak of input signal. The operation of the circuit is same as of negative clamper.

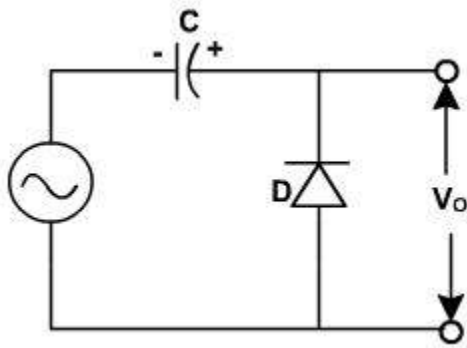


Fig. 1

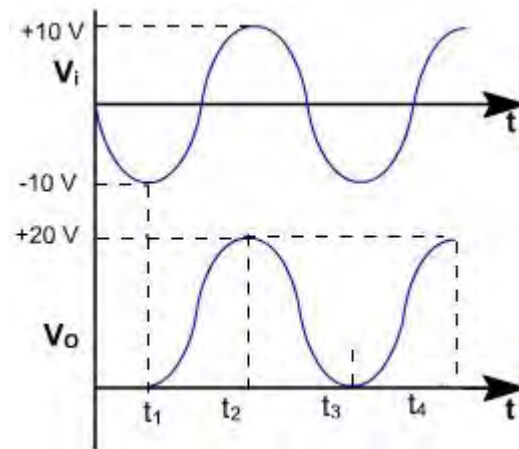


Fig. 2

Let the input signal swings from +10 V to -10 V. During first negative half cycle as  $V_i$  rises from 0 to -10 V, the diode conducts. Assuming an ideal diode, its voltage, which is also the output must be zero during the time from 0 to  $t_1$ . The capacitor charges during this period to 10 V, with the polarity shown.

After that  $V_i$  starts to drop which means the anode of D is negative relative to cathode, ( $V_D = v_i - v_C$ ) thus reverse biasing the diode and preventing the capacitor from discharging. [Fig. 2](#). Since the capacitor is holding its charge it behaves as a DC voltage source while the diode appears as an open circuit, therefore the equivalent circuit becomes an input supply in series with +10 V dc voltage and the resultant output voltage is the sum of instantaneous input voltage and dc voltage (+10 V).

To clamp the input signal by a voltage other than peak value, a dc source is required. As shown in [fig. 3](#), the dc source is reverse biasing the diode.

The input voltage swings from +10 V to -10 V. In the negative half cycle when the voltage exceed 5V then D conduct. During input voltage variation from -5 V to -10 V, the capacitor charges to 5 V with the polarity shown in [fig. 3](#). After that D becomes reverse biased and open circuited. Then complete ac signal is shifted upward by 5 V. The output waveform is shown in [fig. 4](#).

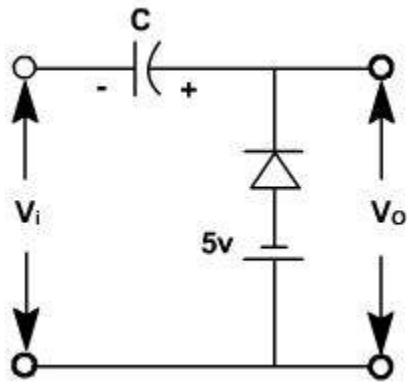


Fig. 3

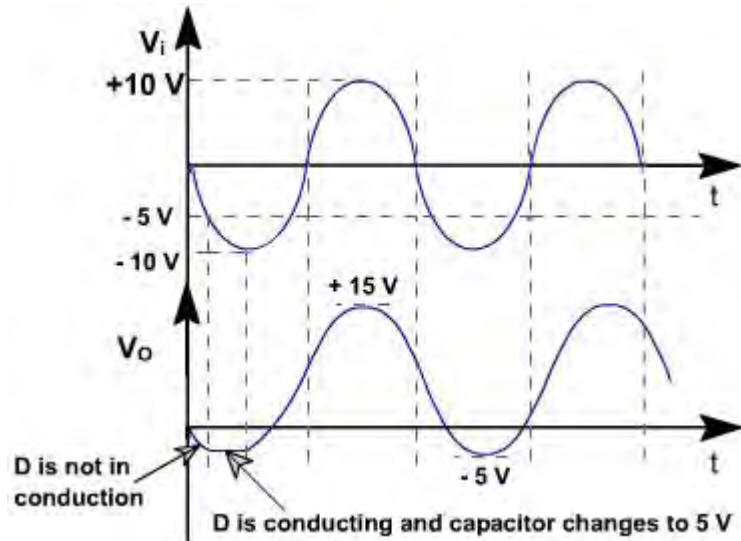
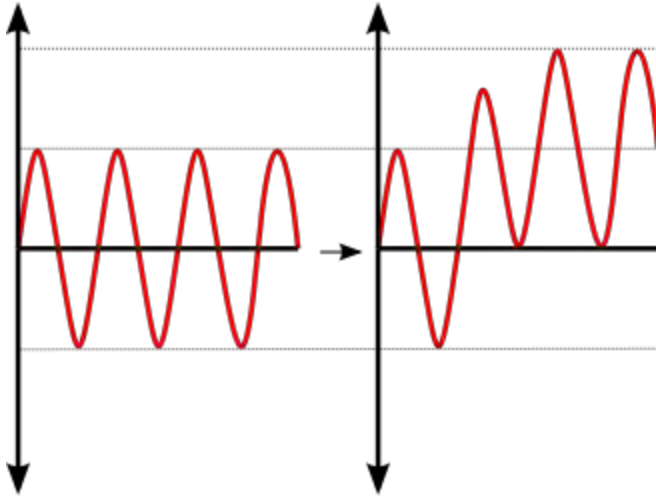


Fig. 4

### Clamper (electronics)

From Wikipedia, the free encyclopedia  
 Jump to: [navigation](#), [search](#)



Positive unbiased voltage clamping shifts the amplitude of the input waveform so that all parts of it are greater than 0V

A **clammer** is an [electronic circuit](#) that prevents a signal from exceeding a certain defined magnitude by shifting its DC value. The clamper does not restrict the peak-to-peak excursion of the signal, but moves it up or down by a fixed value. A **diode clamp** (a simple, common type) relies on a [diode](#), which conducts electric current in only one direction; [resistors](#) and [capacitors](#) in the circuit are used to maintain an altered dc level at the clamper output.

### General function

A clamping circuit (also known as a clamper) will bind the upper or lower extreme of a waveform to a fixed DC voltage level. These circuits are also known as DC voltage restorers. Clampers can be constructed in both positive and negative polarities. When unbiased, clamping circuits will fix the voltage lower limit (or upper limit, in the case of negative clampers) to 0 Volts. These circuits clamp a peak of a waveform to a specific DC level compared with a capacitively coupled signal which swings about its average DC level (usually 0V).

### Types

Clamp circuits are categorised by their operation; negative or positive and biased and unbiased. A positive clamp circuit outputs a purely positive waveform from an input signal; it offsets the input signal so that all of the waveform is greater than 0V. A negative clamp is the opposite of this - this clamp outputs a purely negative waveform from an input signal.

A bias voltage between the diode and ground offsets the output voltage by that amount.

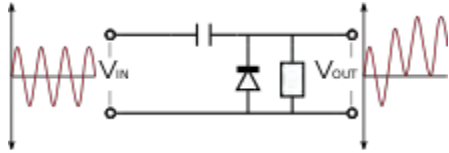
For example, an input signal of peak value 5V ( $V_{IN} = 5V$ ) is applied to a positive clamp with a bias of 3V ( $V_{BIAS} = 3V$ ), the peak output voltage will be

$$V_{OUT} = 2V_{IN} + V_{BIAS}$$

$$V_{OUT} = 2 * 5 + 3$$

$$V_{OUT} = 13V$$

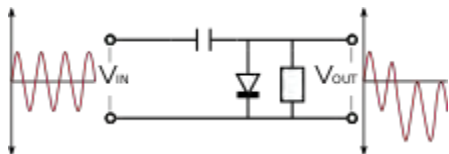
### Positive unbiased



A positive unbiased clamp

In the negative cycle of the input AC signal, the diode is forward biased and conducts, charging the capacitor to the peak positive value of  $V_{IN}$ . During the positive cycle, the diode is reverse biased and thus does not conduct. The output voltage is therefore equal to the voltage stored in the capacitor plus the input voltage again, so  $V_{OUT} = 2V_{IN}$

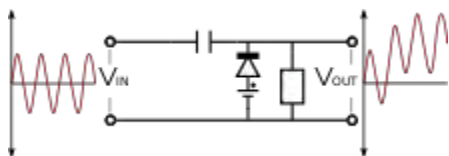
### ] Negative unbiased



A negative unbiased clamp

A negative unbiased clamp is the opposite of the equivalent positive clamp. In the positive cycle of the input AC signal, the diode is forward biased and conducts, charging the capacitor to the peak value of  $V_{IN}$ . During the negative cycle, the diode is reverse biased and thus does not conduct. The output voltage is therefore equal to the voltage stored in the capacitor plus the input voltage again, so  $V_{OUT} = -2V_{IN}$

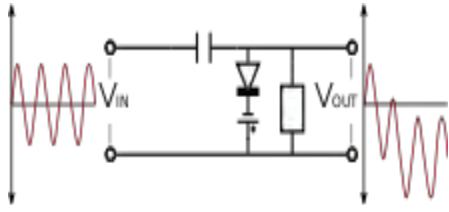
### Positive biased



A positive biased clamp

A positive biased voltage clamp is identical to an equivalent unbiased clamp but with the output voltage offset by the bias amount  $V_{BIAS}$ . Thus,  $V_{OUT} = 2V_{IN} + V_{BIAS}$

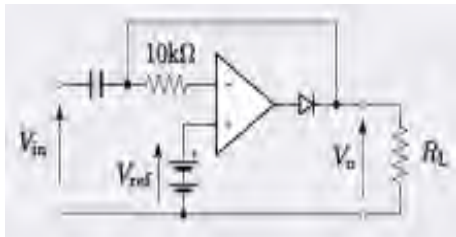
## Negative biased



A negative biased clamp

A negative biased voltage clamp is likewise identical to an equivalent unbiased clamp but with the output voltage offset in the negative direction by the bias amount  $V_{BIAS}$ . Thus,  $V_{OUT} = -2V_{IN} - V_{BIAS}$

## Op-amp circuit



Precision op-amp clamp circuit<sup>[1]</sup>

The figure shows an op-amp clamp circuit with a non-zero reference clamping voltage. The advantage here is that the clamping level is at precisely the reference voltage. There is no need to take into account the forward volt drop of the diode (which is necessary in the preceding simple circuits as this adds to the reference voltage). The effect of the diode volt drop on the circuit output will be divided down by the gain of the amplifier, resulting in an insignificant error.

## Clamping for input protection

Clamping can be used to adapt an input signal to a device that cannot make use of or may be damaged by the signal range of the original input.

## Principles of operation

The schematic of a clamper reveals that it is a relatively simple device. The two components creating the clamping effect are a capacitor, followed by a diode in parallel with the load. The clamper circuit relies on a change in the capacitor's time constant; this is the result of the diode changing current path with the changing input voltage. The magnitude of  $R$  and  $C$  are chosen so

that  $\tau=RC$  is large enough to ensure that the voltage across the capacitor does not discharge significantly during the diode's "Non conducting" interval. During the first negative phase of the AC input voltage, the capacitor in the positive clamper charges rapidly. As  $V_{in}$  becomes positive, the capacitor serves as a voltage doubler; since it has stored the equivalent of  $V_{in}$  during the negative cycle, it provides nearly that voltage during the positive cycle; this essentially doubles the voltage seen by the load. As  $V_{in}$  becomes negative, the capacitor acts as a battery of the same voltage of  $V_{in}$ . The voltage source and the capacitor counteract each other, resulting in a net voltage of zero as seen by the load.

### **Biased versus non-biased**

By using a voltage source and resistor, the clamper can be biased to bind the output voltage to a different value. The voltage supplied to the potentiometer will be equal to the offset from zero (assuming an ideal diode) in the case of either a positive or negative clamper (the clamper type will determine the direction of the offset. If a negative voltage is supplied to either positive or negative, the waveform will cross the x-axis and be bound to a value of this magnitude on the opposite side. Zener diodes can also be used in place of a voltage source and potentiometer, hence setting the offset at the Zener voltage.

### **Examples**

One common such clamping circuit is the DC restorer circuit in [analog television](#) receiver, which returns the voltage of the signal during the back porch of the line blanking period to 0V. Since the back porch is required to be at 0V on transmission, any DC or low frequency hum that has been induced onto the signal can be effectively removed via this method.

### **[Diode Clippers](#)**

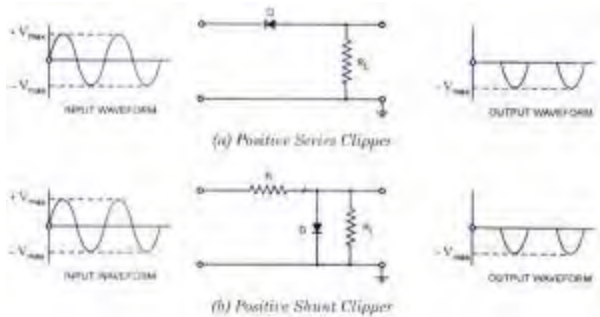
For a clipping circuit at least two components—an ideal diode and resistor are required and sometimes a dc battery is also employed for fixing the clipping level. The diode acts as a closed switch when forward biased and an open switch when reverse biased. The input waveform can be clipped at different levels by simply changing the voltage of the battery and by interchanging the positions of the various elements.

Depending on the orientation of the diode, the positive or negative region of the input signal is “clipped” off and accordingly the diode clippers may be positive or negative clippers.

There are two general categories of clippers: series and parallel (or shunt). The series configuration is defined as one where diode is in series with the load, while the shunt clipper has the diode in a branch parallel to the load.

#### **1. Positive Clipper**

The clipper which removes the positive half cycles of the input voltage is called the positive clipper. The circuit arrangements for a positive clipper are illustrated in the figure given below.



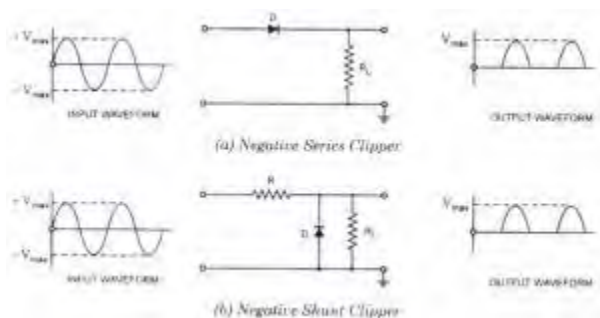
Positive Series Clipper and Positive Shunt Clipper

The figure illustrates the positive series clipper circuit (that is, diode in series with the load). From the figure (a) it is seen that while the input is positive, diode D is reverse biased and so the output remains at zero that is, positive half cycle is clipped off. During the negative half cycle of the input, the diode is forward biased and so the negative half cycle appears across the output.

Figure (b) illustrates the positive shunt clipper circuit (that is, diode in parallel with the load). From the figure (b) it is seen that while input side is positive, the diode D is forward biased and conducts heavily (that is, diode acts as a closed switch). So the voltage drop across the diode or across the load resistance  $R_L$  is zero. Thus output voltage during the positive half cycles is zero, as shown in the output waveform. During the negative half cycles of the input signal voltage, the diode D is reverse biased and behaves as an open switch. Consequently the entire input voltage appears across the diode or across the load resistance  $R_L$  if R is much smaller than  $R_L$

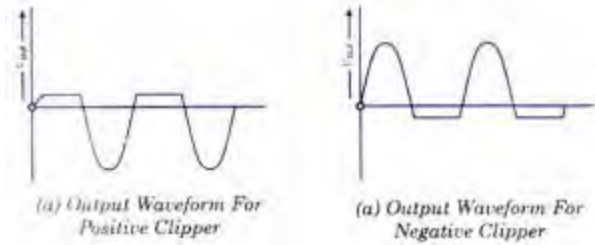
Actually the circuit behaves as a voltage divider with an output voltage of  $[R_L / R + R_L] V_{max} = -V_{max}$  when  $R_L \gg R$

Note: If the diode in figures (a) and (b) is reconnected with reversed polarity, the circuits will become for a negative series clipper and negative shunt clipper respectively. The negative series and negative shunt clippers are shown in figures (a) and (b) as given below.



Negative Series Clipper and Negative Shunt Clipper

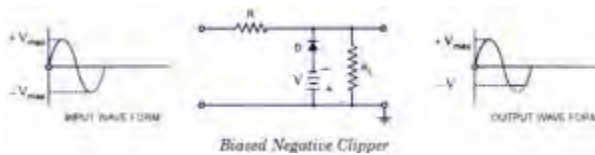
In the above discussion, the diode is considered to be ideal one. If second approximation for diode is considered the barrier potential (0.7 V for silicon and 0.3 V for Germanium) of diode, will be taken into account. Then the output waveforms for positive and negative clippers will be of the shape shown in the figure below.



Output Waveform – Positive Clipper and Negative Clipper

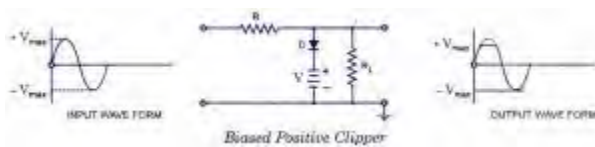
## 2. Biased Clipper

Sometimes it is desired to remove a small portion of positive or negative half cycles of the signal voltage. Biased clippers are employed for this purpose. The circuit diagram for a biased negative clipper (that is for removing a small portion of each negative half cycle) is illustrated in figure (a).



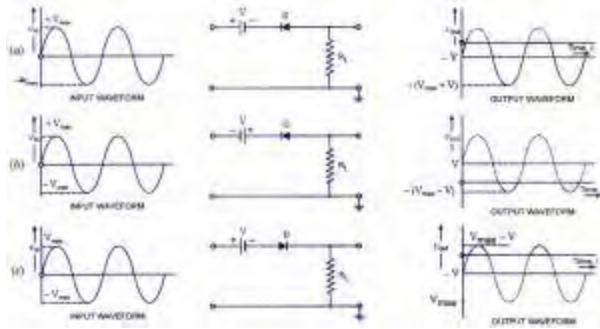
**Biased Negative Clipper**

The action of the circuit is explained below. When the input signal voltage is positive, the diode D is reverse-biased and behaves as an open-switch, the entire positive half cycle appears across the load, as illustrated by output waveform [figure (a)]. When the input signal voltage is negative but does not exceed battery voltage V, the diode D remains reverse-biased and most of the input voltage appears across the output. When during the negative half cycle of input signal, the signal voltage exceeds the battery voltage V, the diode D is forward biased i.e conducts heavily. The output voltage is equal to  $-V$  and stays at  $-V$  as long as the input signal voltage is greater than battery voltage V in magnitude. Thus a biased negative clipper removes input voltage when the input signal voltage exceeds the battery voltage. Clipping can be changed by reversing the battery and diode connections, as illustrated in figure (b).

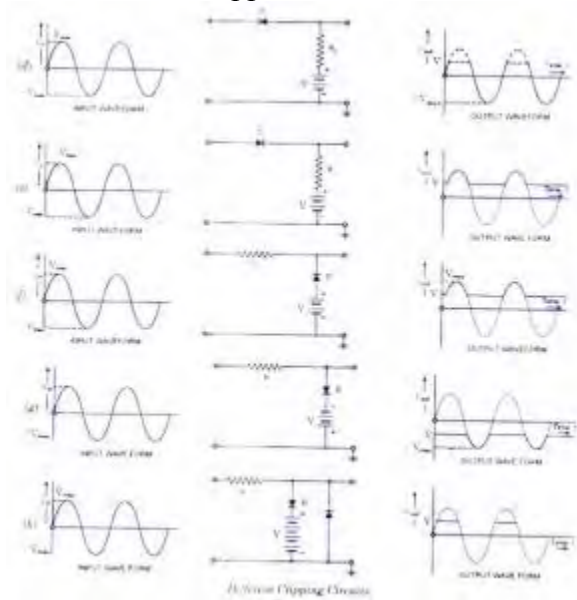


**Biased Positive Clipper**

Some of other biased clipper circuits are given below in the figure. While drawing the wave-shape of the output basic principle discussed above are followed. The diode has been considered as an ideal one.



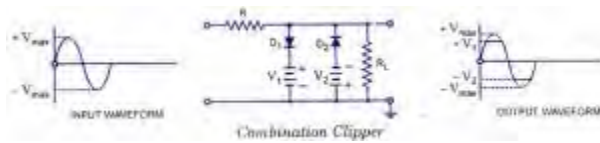
Biased Clipper Circuits



Different Clipping Circuits

### 3. Combination Clipper

When a portion of both positive and negative of each half cycle of the input voltage is to be clipped (or removed), combination clipper is employed. The circuit for such a clipper is given in the figure below.



#### Combination Clipper

The action of the circuit is summarized below. For positive input voltage signal when input voltage exceeds battery voltage  $+V_1$  diode  $D_1$  conducts heavily while diode  $D_2$  is reverse biased and so voltage  $+V_1$  appears across the output. This output voltage  $+V_1$  stays as long as the input signal voltage exceeds  $+V_1$ . On the other hand for the negative input voltage signal, the diode  $D_1$  remains reverse biased and diode  $D_2$  conducts heavily only when input voltage exceeds

battery voltage  $V_2$  in magnitude. Thus during the negative half cycle the output stays at  $-V_2$  so long as the input signal voltage is greater than  $-V_2$ .

### Drawbacks of Series and Shunt Diode Clippers

In series clippers, when diode is in off position, there should be no transmission of input signal to output. But in case of high frequency signals transmission occurs through diode capacitance which is undesirable. This is the drawback of using diode as a series element in such clippers.

In shunt clippers, when diode is in the off condition, transmission of input signal should take place to output. But in case of high frequency input signals, diode capacitance affects the circuit operation adversely and the signal gets attenuated (that is, it passes through diode capacitance to ground).

### Multivibrator

A **multivibrator** is an [electronic circuit](#) used to implement a variety of simple two-state systems such as [oscillators](#), [timers](#) and [flip-flops](#). It is characterized by two amplifying devices (transistors, electron tubes or other devices) cross-coupled by resistors and capacitors.

There are three types of multivibrator circuit:

- **astable**, in which the circuit is not stable in either state—it continuously oscillates from one state to the other. Due to this, it does not require an input (Clock pulse or other).
- **monostable**, in which one of the states is stable, but the other is not—the circuit will flip into the unstable state for a determined period, but will eventually return to the stable state. Such a circuit is useful for creating a timing period of fixed duration in response to some external event. This circuit is also known as a **one shot**. A common application is in eliminating [switch bounce](#).
- **bistable**, in which the circuit will remain in either state indefinitely. The circuit can be flipped from one state to the other by an external event or trigger. Such a circuit is important as the fundamental building block of a [register](#) or [memory](#) device. This circuit is also known as a [latch](#) or a [flip-flop](#).

In its simplest form the multivibrator circuit consists of two cross-coupled [transistors](#). Using [resistor-capacitor](#) networks within the circuit to define the time periods of the unstable states, the various types may be implemented. Multivibrators find applications in a variety of systems where square waves or timed intervals are required. Simple circuits tend to be inaccurate since many factors affect their timing, so they are rarely used where very high precision is required.

Before the advent of low-cost integrated circuits, chains of multivibrators found use as [frequency dividers](#). A free-running multivibrator with a frequency of one-half to one-tenth of the reference frequency would accurately lock to the reference frequency. This technique was used in early electronic organs, to keep notes of different octaves accurately in tune. Other applications included early [television](#) systems, where the various line and frame frequencies were kept synchronized by pulses included in the video signal.

## Astable multivibrator circuit

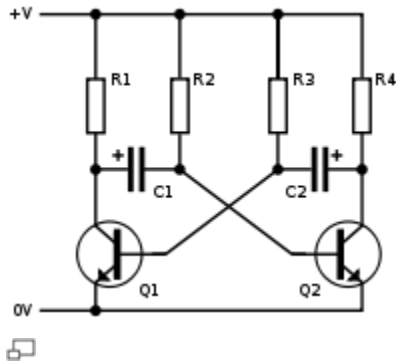


Figure 1: Basic [BJT](#) astable multivibrator

This circuit shows a typical simple astable circuit, with an output from the collector of Q1, and an inverted output from the collector of Q2.

### Basic mode of operation

The circuit keeps one transistor switched on and the other switched off. Suppose that initially, Q1 is switched on and Q2 is switched off.

#### State 1:

- Q1 holds the bottom of R1 (and the left side of C1) near ground (0 V).
- The right side of C1 (and the base of Q2) is being charged by R2 from below ground to 0.6 V.
- R3 is pulling the base of Q1 up, but its base-emitter diode prevents the voltage from rising above 0.6 V.
- R4 is charging the right side of C2 up to the power supply voltage (+V). Because R4 is less than R2, C2 charges faster than C1.

When the base of Q2 reaches 0.6 V, Q2 turns on, and the following [positive feedback](#) loop occurs:

- Q2 abruptly pulls the right side of C2 down to near 0 V.
- Because the voltage across a capacitor cannot suddenly change, this causes the left side of C2 to suddenly fall to almost  $-V$ , well below 0 V.
- Q1 switches off due to the sudden disappearance of its base voltage.

- R1 and R2 work to pull both ends of C1 toward +V, completing Q2's turn on. The process is stopped by the B-E diode of Q2, which will not let the right side of C1 rise very far.

This now takes us to **State 2**, the mirror image of the initial state, where Q1 is switched off and Q2 is switched on. Then R1 rapidly pulls C1's left side toward +V, while R3 more slowly pulls C2's left side toward +0.6 V. When C2's left side reaches 0.6 V, the cycle repeats.

### Multivibrator frequency

The period of each *half* of the multivibrator is given by  $t = \ln(2)RC$ . The total period of oscillation is given by:

$$T = t_1 + t_2 = \ln(2)R_2 C_1 + \ln(2)R_3 C_2$$

$$f = \frac{1}{T} = \frac{1}{\ln(2) \cdot (R_2 C_1 + R_3 C_2)} \approx \frac{1}{0.693 \cdot (R_2 C_1 + R_3 C_2)}$$

where...

- $f$  is [frequency](#) in [hertz](#).
- $R_2$  and  $R_3$  are [resistor](#) values in [ohms](#).
- $C_1$  and  $C_2$  are [capacitor](#) values in [farads](#).
- $T$  is [period time](#) (In this case, the sum of two period durations).

**For the special case** where

- $t_1 = t_2$  (50% duty cycle)
- $R_2 = R_3$
- $C_1 = C_2$

$$f = \frac{1}{T} = \frac{1}{\ln(2) \cdot 2RC} \approx \frac{0.721}{RC}$$

### Initial power-up

When the circuit is first powered up, neither transistor will be switched on. However, this means that at this stage they will both have high base voltages and therefore a tendency to switch on, and inevitable slight asymmetries will mean that one of the transistors is first to switch on. This will quickly put the circuit into one of the above states, and oscillation will ensue. In practice, oscillation always occurs for practical values of  $R$  and  $C$ .

However, if the circuit is temporarily held with both bases high, for longer than it takes for both capacitors to charge fully, then the circuit will remain in this stable state, with both bases at 0.6

V, both collectors at 0 V, and both capacitors charged backwards to  $-0.6$  V. This can occur at startup without external intervention, if  $R$  and  $C$  are both very small. For example, a 10 MHz oscillator of this type will often be unreliable. (Different oscillator designs, such as [relaxation oscillators](#), are required at high frequencies.)

### Period of oscillation

Very roughly, the duration of state 1 (low output) will be related to the time constant  $R_2C_1$  as it depends on the charging of  $C_1$ , and the duration of state 2 (high output) will be related to the time constant  $R_3C_2$  as it depends on the charging of  $C_2$ . Because they do not need to be the same, an asymmetric [duty cycle](#) is easily achieved.

However, the duration of each state also depends on the initial state of charge of the capacitor in question, and this in turn will depend on the amount of discharge during the previous state, which will also depend on the resistors used during discharge ( $R_1$  and  $R_4$ ) and also on the duration of the previous state, *etc.* The result is that when first powered up, the period will be quite long as the capacitors are initially fully discharged, but the period will quickly shorten and stabilise.

The period will also depend on any current drawn from the output and on the supply voltage.

### Protective components

While not fundamental to circuit operation, [diodes](#) connected in series with the base or emitter of the transistors are required to prevent the base-emitter junction being driven into reverse breakdown when the supply voltage is in excess of the  $V_{eb}$  breakdown voltage, typically around 5-10 volts for general purpose silicon transistors. In the monostable configuration, only one of the transistors requires protection.

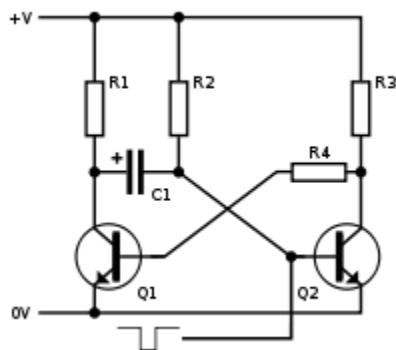


Figure 2: Basic [BJT](#) monostable multivibrator.

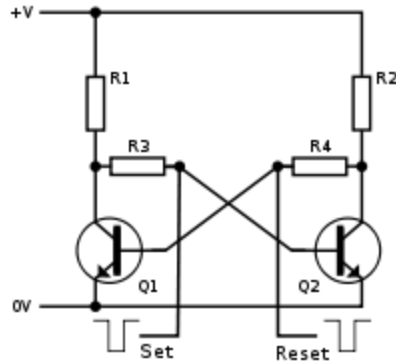


Figure 3: Basic [BJT](#) bistable multivibrator.

### Non-electronic astables

Astable oscillators are usually thought of as [electronic circuits](#), but need not be. [Bimetallic strips](#) are used which switch an electric current on as they cool and off as they heat—flashing Christmas and car [indicator lights](#) may use this mechanism.

### Monostable multivibrator circuit

When triggered by an input pulse, a monostable multivibrator will switch to its unstable position for a period of time, and then return to its stable state. The time period monostable multivibrator remains in unstable state is given by  $t = \ln(2)R_2C_1$ . If repeated application of the input pulse maintains the circuit in the unstable state, it is called a *retriggerable* monostable. If further trigger pulses do not affect the period, the circuit is a *non-retriggerable* multivibrator.

### Bistable multivibrator circuit

Suggested values:

- **R1, R2** = 10 k $\Omega$
- **R3, R4** = 10 k $\Omega$

This [latch](#) circuit is similar to an astable multivibrator, except that there is no charge or discharge time, due to the absence of capacitors. Hence, when the circuit is switched on, if Q1 is on, its collector is at 0 V. As a result, Q2 gets switched off. This results in more than half +V volts being applied to R4 causing current into the base of Q1, thus keeping it on. Thus, the circuit remains stable in a single state continuously. Similarly, Q2 remains on continuously, if it happens to get switched on first.

Switching of state can be done via Set and Reset terminals connected to the bases. For example, if Q2 is on and Set is grounded momentarily, this switches Q2 off, and makes Q1 on. Thus, Set is used to "set" Q1 on, and Reset is used to "reset" it to off state.

## Astable Multivibrator

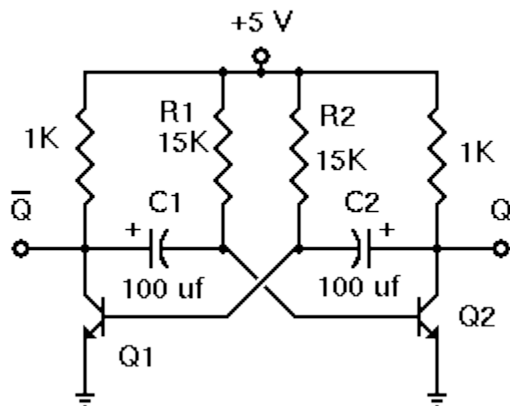
### Introduction

Now that we have seen the bistable multivibrator and then modified it to form a monostable multivibrator, the next question is, Can we modify it further to use capacitor coupling on both sides? And what would happen if we did?

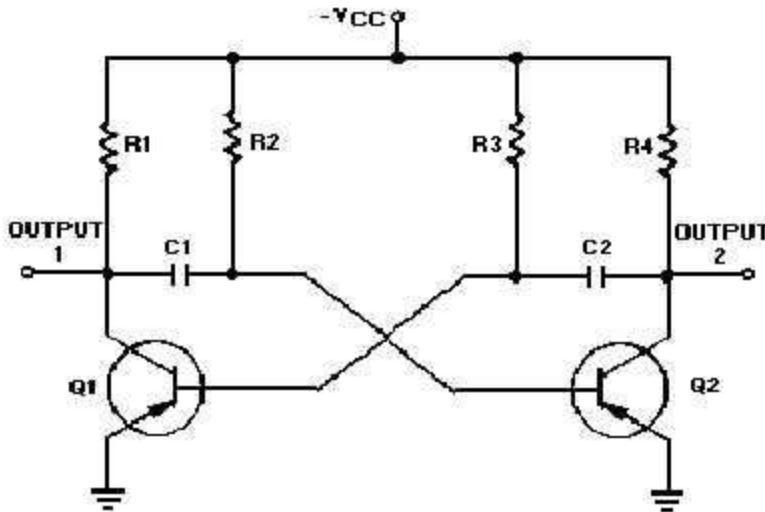
Well, of course we *can* do this; the further question would be, Do we *want* to? And that depends on what happens when we build the circuit this way.

Since the use of one capacitor prevents the circuit from remaining stable in one of its two possible states, it seems likely that with both sides coupled this way the circuit will be unable to remain stable in either state. That is in fact the case, and in this experiment we will construct and demonstrate this circuit.

### Schematic Diagram



As shown in the schematic diagram here, the astable multivibrator simply extends the modification that converted the bistable multivibrator to a monostable version of the circuit. Now, both transistors are coupled to each other through capacitors. Whichever transistor is off at any moment cannot remain off indefinitely; its base will become forward biased as that capacitor charges towards +5 volts. **Once that happens, that transistor will turn on, thereby turning the other one off.**



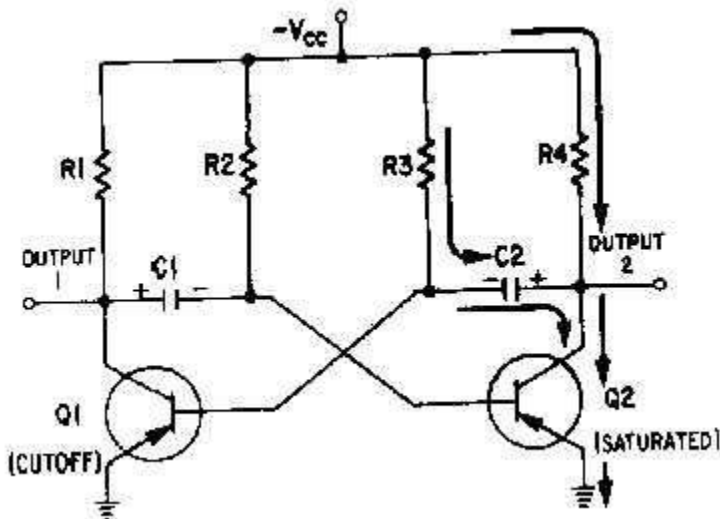
3-6 Figure 3-3.—Astable Multivibrator.

When an input signal to one amplifier is large enough, the transistor can be driven into cutoff, and its collector voltage will be almost  $V_{CC}$ . However, when the transistor is driven into saturation, its collector voltage will be about 0 volts. A circuit that is designed to go quickly from cutoff to saturation will produce a square or rectangular wave at its output. This principle is used in multivibrators. Multivibrators are classified according to the number of steady (stable) states of the circuit. A steady state exists when circuit operation is essentially constant; that is, one transistor remains in conduction and the other remains cut off until an external signal is applied. The three types of multivibrators are the ASTABLE, MONOSTABLE, and BISTABLE. The astable circuit has no stable state. With no external signal applied, the transistors alternately switch from cutoff to saturation at a frequency determined by the RC time constants of the coupling circuits. The monostable circuit has one stable state; one transistor conducts while the other is cut off. A signal must be applied to change this condition. After a period of time, determined by the internal RC components, the circuit will return to its original condition where it remains until the next signal arrives. The bistable multivibrator has two stable states. It remains in one of the stable states until a trigger is applied. It then FLIPS to the other stable condition and remains there until another trigger is applied. The multivibrator then changes back (FLOPS) to its first stable state.

- Q1. What type circuit is used to produce square or rectangular waves? Q2. What type of multivibrator does not have a stable state? Q3. What type of multivibrator has one stable state? Q4. What type of multivibrator has two stable states?*

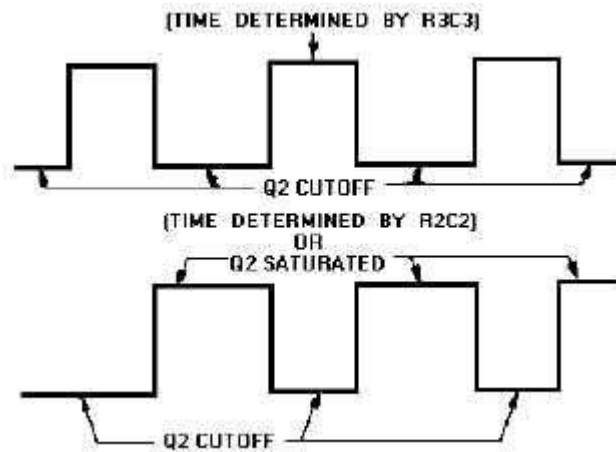
**Astable Multivibrator** An astable multivibrator is also known as a FREE-RUNNING MULTIVIBRATOR. It is called free-running because it alternates between two different output voltage levels during the time it is on. The output remains at each voltage level for a definite period of time. If you looked at this output on an oscilloscope, you would see continuous square or rectangular waveforms. The astable multivibrator has two outputs, but NO

inputs. Let's look at the multivibrator in figure 3-3 again. This is an astable multivibrator. The astable multivibrator is said to oscillate. To understand why the astable multivibrator oscillates, assume that transistor Q1 saturates and transistor Q2 cuts off when the circuit is energized. This situation is shown in figure 3-4. We assume Q1 saturates and Q2 is in cutoff because the circuit is symmetrical; that is,  $R1 = R4$ ,  $R2 = R3$ ,  $C1 = C2$ , and  $Q1 = Q2$ . It is impossible to tell which transistor will actually conduct when the circuit is energized. For this reason, either of the transistors may be assumed to conduct for circuit analysis purposes. **Figure 3-4.—Astable multivibrator (Q1 saturated).** Essentially, all the current in the circuit flows through Q1; Q1 offers almost no resistance to current flow. Notice that capacitor C1 is charging. Since Q1 offers almost no resistance in its saturated state, the rate of charge of C1 depends only on the time constant of R2 and C1 (recall that  $TC = RC$ ). Notice that the right-hand side of capacitor C1 is connected to the base of transistor Q2, which is now at cutoff. Let's analyze what is happening. The right-hand side of capacitor C1 is becoming increasingly negative. If the base of Q2 becomes sufficiently negative, Q2 will conduct. After a certain period of time, the base of Q2 will become sufficiently negative to cause Q2 to change states from cutoff to conduction. The time necessary for Q2 to become saturated is determined by the time constant  $R2C1$ . The next state is shown in figure 3-5. The negative voltage accumulated on the right side on capacitor C1 has caused Q2 to conduct. Now the following sequence of events takes place almost instantaneously. Q2 starts conducting and quickly saturates, and the voltage at output 2 changes from  $V_{CC}$  approximately  $V_{CC}$  to approximately 0 volts. This change in voltage is coupled through C2 to the base of Q1, forcing Q1 to cutoff. Now Q1 is in cutoff and Q2 is in saturation. This is the circuit situation shown in figure 3-6.

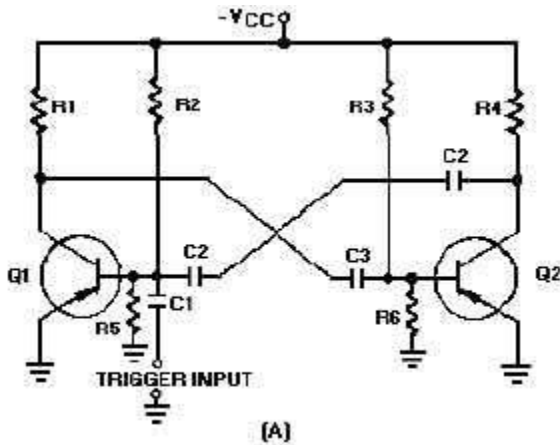


3-8 **Figure 3-5.—Astable multivibrator. Figure 3-6.—Astable multivibrator. (Q2 saturated).** Notice that figure 3-6 is the mirror image of figure 3-4. In figure 3-6 the left side of capacitor C2 becomes more negative at a rate determined by the time constant  $R3C2$ . As the left side of C2 becomes more negative, the base of Q1 also becomes more negative. When the base of Q1 becomes negative enough to allow Q1 to conduct, Q1 will again go into saturation. The

resulting change in voltage at output 1 will cause Q2 to return to the cutoff state. Look at the output waveform from transistor Q2, as shown in figure 3-7. The output voltage (from either output of the multivibrator) alternates from approximately 0 volts to  $V_{CC}$  approximately  $V_{CC}$ , remaining in each state for a definite period of time. The time may range from a microsecond to as much as a second or two.  $V_{CC}$  In some applications, the time period of higher voltage ( $V_{CC}$ ) and the time period of lower voltage (0 volts) will be equal. Other applications require differing higher- and lower-voltage times. For example, timing and gating circuits often have different

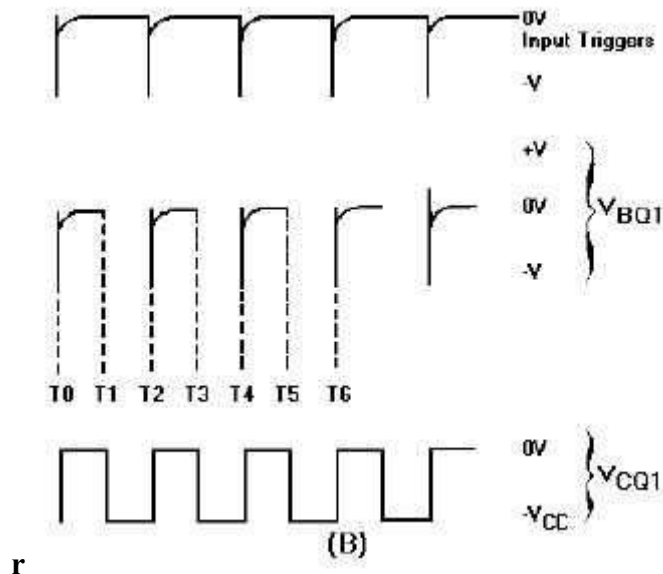


pulse widths as shown in figure 3-8.



**3-9 Figure 3-7.—Square wave output from Q2. Figure 3-8.—Rectangular waves. FREQUENCY STABILITY.—**Some astable multivibrators must have a high degree of frequency stability. One way to obtain a high degree of frequency stability is to apply triggers. Figure 3-9, view (A), shows the diagram of a triggered, astable multivibrator. At time  $T_0$ , a negative input trigger to the base of Q1 (through C1) causes Q1 to go into saturation, which drives Q2 to cutoff. The circuit will remain in this condition as long as the base voltage of Q2 is positive. The length of time the base of Q2 will remain positive is determined by C3, R3, and R6. Observe the parallel paths for C3 to discharge. **Figure 3-9A.—Triggered astable multivibrator and output.**

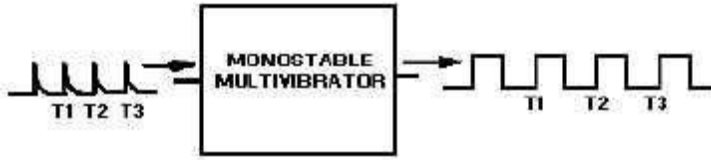
View (B) of figure 3-9 shows the waveforms associated with the circuit. At time T1, Q2 comes out of cutoff and goes into saturation. Also, Q1 is caused to come out of saturation and is cut off. The base voltage waveform of Q1 shows a positive potential that is holding Q1 at cutoff. This voltage would normally hold Q1 at cutoff until a point between T2 and T3. However, at time T2 another trigger is applied to the base of Q1, causing it to begin conducting. Q1 goes into saturation and Q2 is caused to cut off. This action repeats each time a trigger (T2, T4, T6) is applied.



**Figure 3-9B.—Triggered astable multivibrator and output.**

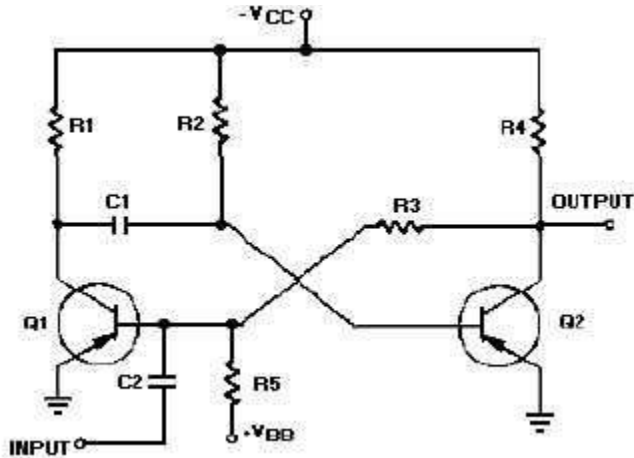
The prt of the input triggers must be shorter than the natural free-running prt of the astable multivibrator, or the trigger prf must be slightly higher than the free-running prf of the circuit. This is to make certain the triggers control the prt of the output.

**Monostable Multivibrator** The monostable multivibrator (sometimes called a ONE-SHOT MULTIVIBRATOR) is a square- or rectangular-wave generator with just one stable condition. With no input signal (quiescent condition) one amplifier conducts and the other is in cutoff. The monostable multivibrator is basically used for pulse stretching. It is used in computer logic systems and communication navigation equipment. The operation of the monostable multivibrator is relatively simple. The input is triggered with a pulse of voltage. The output changes from one voltage level to a different voltage level. The output remains at this new voltage level for a definite period of time. Then the circuit automatically reverts to its original condition and remains that way until another trigger pulse is applied to the input. The monostable multivibrator actually takes this series of input triggers and converts them to uniform square pulses, as shown in figure 3-10. All of the square output pulses are of the same amplitude and time duration.



**Figure 3-10.—Monostable multivibrator block diagram.**

The schematic for a monostable multivibrator is shown in figure 3-11. Like the astable multivibrator, one transistor conducts and the other cuts off when the circuit is energized.



**Figure 3-11.—Monostable multivibrator schematic.**

Recall that when the astable multivibrator was first energized, it was impossible to predict which transistor would initially go to cutoff because of circuit symmetry. The one-shot circuit is not symmetrical like the astable multivibrator. Positive voltage  $V_{BB}$  is applied through  $R_5$  to the base of  $Q_1$ . This positive voltage causes  $Q_1$  to cut off. Transistor  $Q_2$  saturates because of the negative  $V_{CC}$  voltage applied from  $CC$  to its base through  $R_2$ . Therefore,  $Q_1$  is cut off and  $Q_2$  is saturated before a trigger pulse is applied, as shown in figure 3-12. The circuit is shown in its stable state.

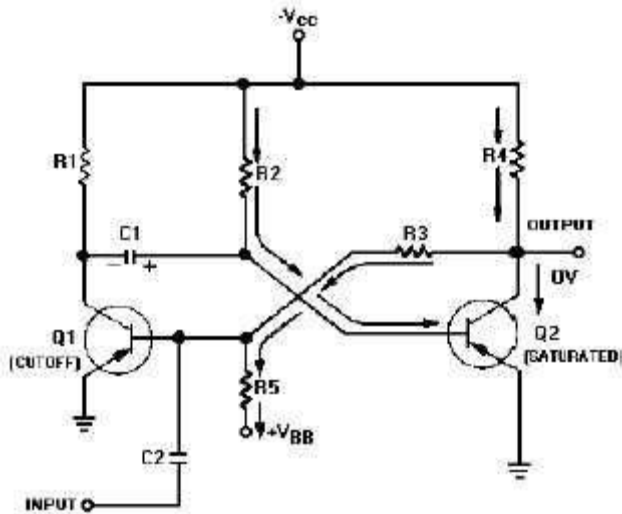
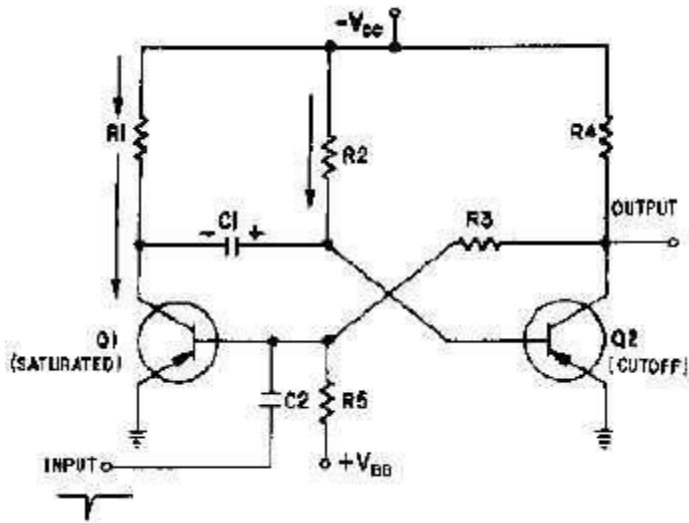


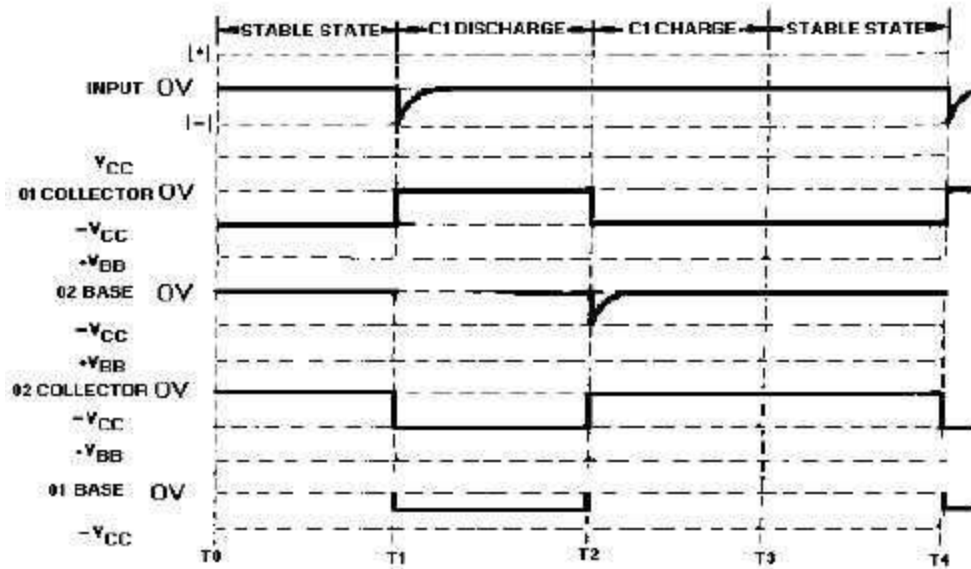
Figure 3-12.—Monostable multivibrator (stable state).

Let's take a more detailed look at the circuit conditions in this stable state (refer to figure 3-12). As stated above, Q1 is cut off, so no current flows through R1, and the collector of Q1  $V_{C1}$  is at  $V_{CC}$ . Q2 is saturated and has practically no voltage drop across it, so its collector is essentially at 0 volts. R5 and R3 form a voltage divider from  $V_{BB}$  to the ground potential at the collector of Q2. The tie point between these two resistors will be positive. Thus, the base of Q1 is held positive, ensuring that Q1 remains cutoff. Q2 will remain saturated because the base of Q2 is very slightly negative as a result of the  $V_{BE}$  voltage drop across R2. If the collector of Q1 is near  $V_{CC}$  and the base of Q2 is near ground, C1 must be charged to nearly  $V_{CC}$  volts with the polarity shown. Now that all the components and voltages have been described for the stable state, let us see how the circuit operates (see figure 3-13). Assume that a negative pulse is applied at the input terminal. C2 couples this voltage change to the base of Q1 and starts Q1 conducting. Q1 quickly saturates, and its collector voltage immediately rises to ground potential. This sharp voltage increase is coupled through C1 to the base of Q2, causing Q2 to cut off; the collector voltage of Q2 immediately drops to  $V_{CC}$ . The voltage divider formed by R5 and R3 then holds the base of Q1 negative, and Q1 is locked in saturation.



**Figure 3-13.—Monostable multivibrator (triggered).**

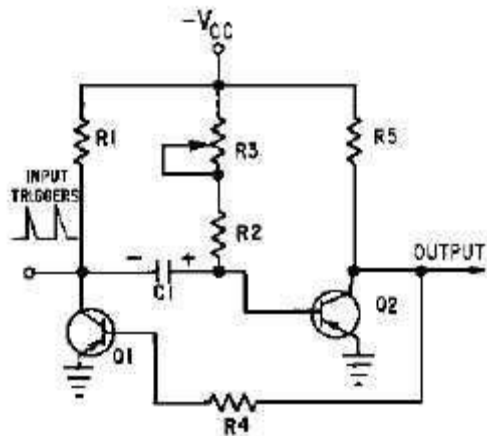
The one-shot multivibrator has now been turned on by applying a pulse at the input. It will turn itself off after a period of time. To see how it does this, look at figure 3-13 again. Q1 is held in saturation by the negative voltage applied through R3 to its base, so the circuit cannot be turned off here. Notice that the base of Q2 is connected to C1. The positive charge on C1 keeps Q2 cutoff. Remember that a positive voltage change (essentially a pulse) was coupled from the collector of Q1 when it began conducting to the base of Q2, placing Q2 in cutoff. When the collector of Q1 switches from  $-V_{CC}$  volts to 0 volts, the charge on C1 acts like a battery with its negative terminal on the collector of Q1, and its positive terminal connected to the base of Q2. This voltage is what cuts off Q2. C1 will now begin to discharge through Q1 to ground, back through  $V_{CC}$ , through R2 to the other side of C1. The time required for C1 to discharge depends on the RC time constant of C1 and R2. Figure 3-14 is a timing diagram that shows the negative input pulse and the resultant waveforms that you would expect to see for this circuit description.



**Figure 3-14.—Waveforms of a monostable multivibrator (triggered).**

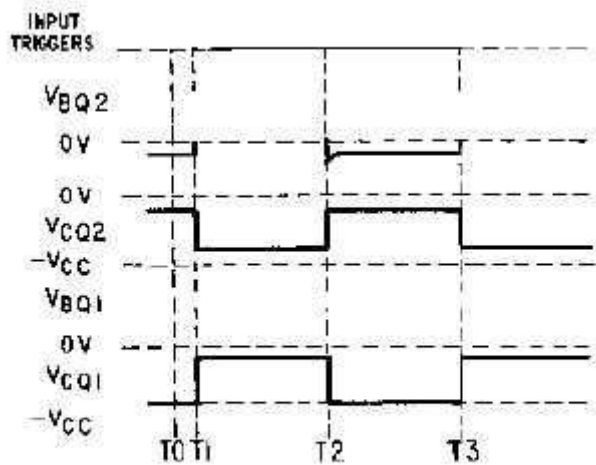
The only part of the operation not described so far is the short C1 charge time that occurs right after Q1 and Q2 return to their stable states. This is simply the time required for C1 to gain electrons on its left side. This charge time is determined by the  $R1C1$  time constant.

Another version of the monostable multivibrator is shown in figure 3-15. View (A) is the circuit and view (B) shows the associated waveforms. In its stable condition (T0), Q1 is cut off and Q2 is conducting. The input trigger (positive pulse at T1) is applied to the collector of Q1 and coupled by C1 to the base of Q2 causing Q2 to be cut off. The collector voltage of Q2 then goes to  $+V_{CC}$ . The more negative voltage at the collector of Q2 forward biases Q1 through R4. With the forward bias, Q1 conducts, and the collector voltage of Q1 goes to about 0 volts. C1 now discharges and keeps Q2 cut off. Q2 remains cut off until C1 discharges enough to allow Q2 to conduct again (T2). When Q2 conducts again, its collector voltage goes toward 0 volts and Q1 is cut off. The circuit returns to its quiescent state and has completed a cycle. The circuit remains in this stable state until the next trigger arrives (T3).



(A) SCHEMATIC

Figure 3-15A.—Monostable multivibrator and waveshapes. Schematic.

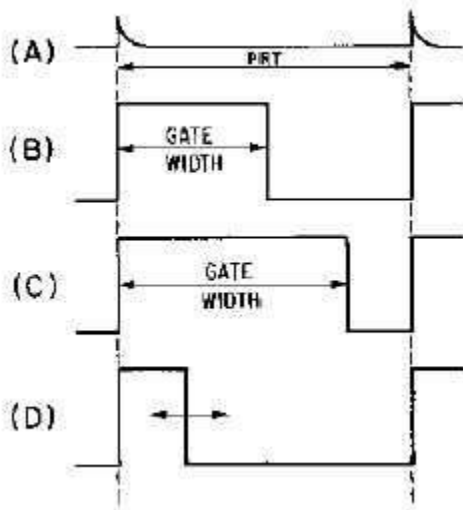


(B) WAVESHAPES

3-14 Figure 3-15B.—Monostable multivibrator and waveshapes. Waveshapes

Note that R3 is variable to allow adjustment of the gate width. Increasing R3 increases the discharge time for C1 which increases the cutoff time for Q2. Increasing the value of R3 widens the gate. To decrease the gate width, decrease the value of R3. Figure 3-16 shows the relationships between the trigger and the output signal. View (A) of the figure shows the input trigger; views (B) and (C) show the different gate widths made available by R3. Although the durations of the gates are different, the duration of the complete cycle remains the same as the

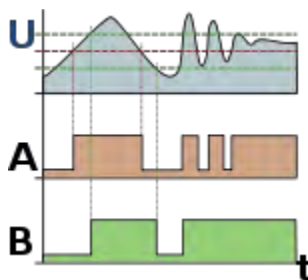
pulse repetition time of the triggers. View (D) of the figure illustrates that the trailing edge of the positive alternation is variable.



**Figure 3-16.—Monostable multivibrator waveforms with a variable gate.**

The reason the monostable multivibrator is also called a one-shot multivibrator can easily be seen. For every trigger pulse applied to the multivibrator, a complete cycle, or a positive and negative alternation of the output, is completed. *Q5. In an astable multivibrator, which components determine the pulse repetition frequency?*

### Schmitt trigger



The effect of using a Schmitt trigger (B) instead of a comparator (A).

In [electronics](#), a **Schmitt trigger** is a [comparator circuit](#) that incorporates [positive feedback](#). In the non-inverting configuration, when the input is higher than a certain chosen threshold, the output is high; when the input is below a different (lower) chosen threshold, the output is low; when the input is between the two, the output retains its value. The trigger is so named because the output retains its value until the input changes sufficiently to trigger a change. This dual

threshold action is called [hysteresis](#), and implies that the Schmitt trigger has some [memory](#). In fact, the Schmitt trigger is a [bistable multivibrator](#).

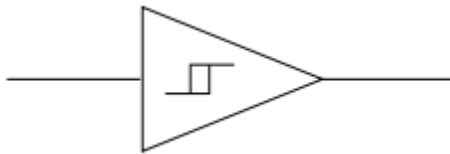
Schmitt trigger devices are typically used in [open loop](#) configurations for noise immunity and [closed loop positive feedback](#) configurations to implement [multivibrators](#).

### Invention

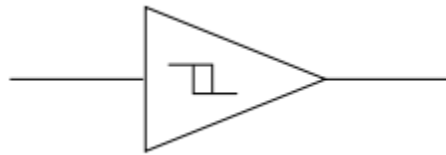
The Schmitt trigger was invented by [US](#) scientist [Otto H. Schmitt](#) in 1934 while he was still a graduate student,<sup>[1]</sup> later described in his doctoral dissertation (1937) as a "thermionic trigger".<sup>[2]</sup> It was a direct result of Schmitt's study of the neural impulse propagation in [squid](#) nerves.<sup>[2]</sup>

### Symbol

The symbol for Schmitt triggers in circuit diagrams is a triangle with an inverting or non-inverting [hysteresis](#) symbol. The symbol depicts the corresponding ideal hysteresis curve.



Standard Schmitt trigger



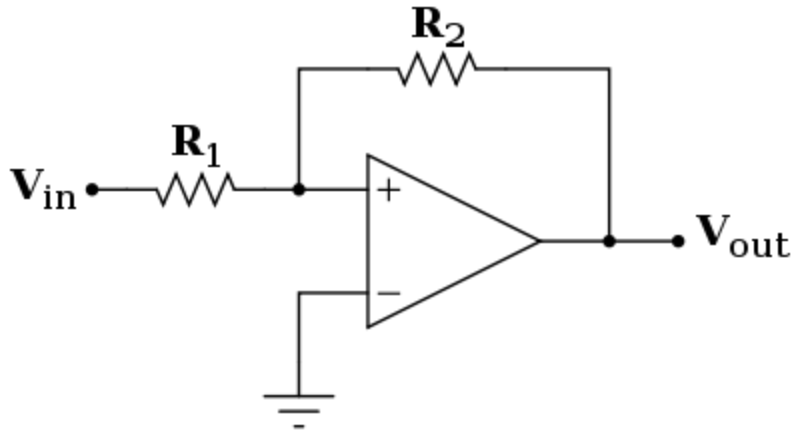
Inverting Schmitt trigger (WRONG IMAGE! Should have a circle at the right point of the triangle indicating inversion!)

### Implementation

A Schmitt trigger can be implemented with a simple [tunnel diode](#), a diode with an "N"-shaped current–voltage characteristic in the first [quadrant](#). An oscillating input will cause the diode to move from one rising leg of the "N" to the other and back again as the input crosses the rising and falling switching thresholds. However, the performance of this Schmitt trigger can be improved with transistor-based devices that make explicit use of [positive feedback](#) to implement the switching.

### Comparator implementation

Schmitt triggers are commonly implemented using a [comparator](#)<sup>[nb 1]</sup> connected to have positive feedback (i.e., instead of the usual negative feedback used in [operational amplifier](#) circuits). For this circuit, the switching occurs near ground, with the amount of [hysteresis](#) controlled by the [resistances](#) of  $R_1$  and  $R_2$ :

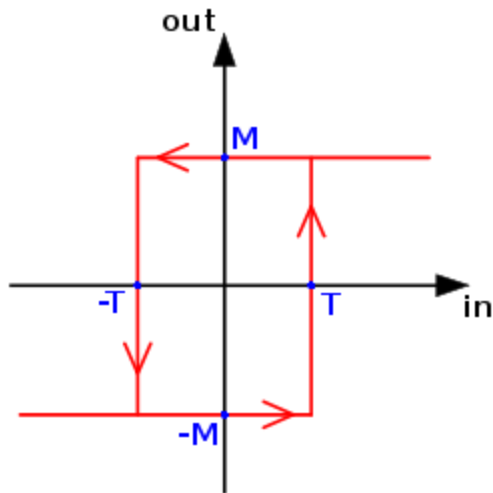


The comparator extracts the [sign](#) of the difference between its two inputs. When the non-inverting (+) input is at a higher voltage than the inverting (-) input, the comparator output switches to  $+V_S$ , which is its high supply voltage. When the non-inverting (+) input is at a lower voltage than the inverting (-) input, the comparator output switches to  $-V_S$ , which is its low supply voltage. In this case, the inverting (-) input is [grounded](#), and so the comparator implements the [sign function](#) – its 2-state output (i.e., either high or low) always has the same sign as the continuous input at its non-inverting (+) terminal.

Because of the resistor network connecting the Schmitt trigger input, the non-inverting (+) terminal of the comparator, and the comparator output, the Schmitt trigger acts like a comparator that switches at a different point depending on whether the output of the comparator is high or low. For very negative inputs, the output will be low, and for very positive inputs, the output will be high, and so this is an implementation of a "non-inverting" Schmitt trigger. However, for intermediate inputs, the state of the output depends on both the input and the output. For instance, if the Schmitt trigger is currently in the high state, the output will be at the positive power supply rail ( $+V_S$ ).  $V_+$  is then a voltage divider between  $V_{in}$  and  $+V_S$ . The comparator will switch when  $V_+=0$  (ground). Current conservation shows that this requires

$$\frac{V_{in}}{R_1} = -\frac{V_S}{R_2}$$

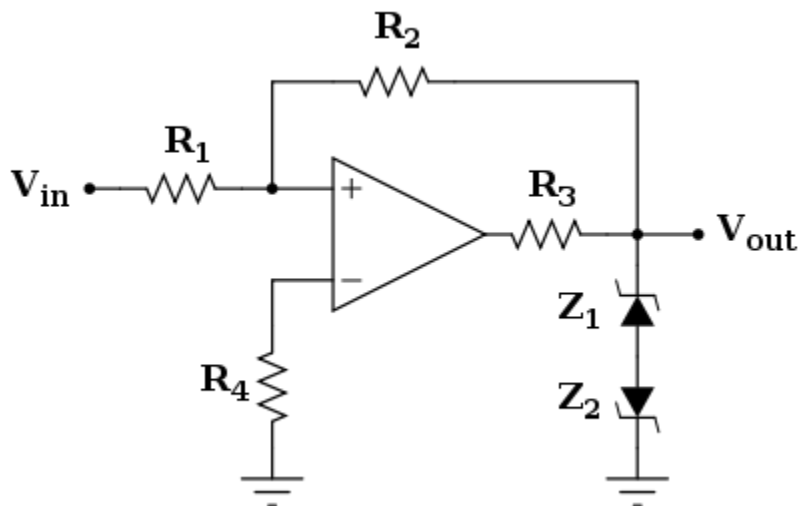
and so  $V_{in}$  must drop below  $-\frac{R_1}{R_2}V_S$  to get the output to switch. Once the comparator output has switched to  $-V_S$ , the threshold becomes  $+\frac{R_1}{R_2}V_S$  to switch back to high.



Typical hysteresis curve (which matches the curve shown on a Schmitt trigger symbol)

So this circuit creates a switching band centered around zero, with trigger levels  $\pm \frac{R_1}{R_2} V_s$ . The input voltage must rise above the top of the band, and then below the bottom of the band, for the output to switch on and then back off. If  $R_1$  is zero or  $R_2$  is infinity (i.e., an [open circuit](#)), the band collapses to zero width, and it behaves as a standard comparator. The output characteristic is shown in the picture on the right. The value of the threshold  $T$  is given by  $\frac{R_1}{R_2} V_s$  and the maximum value of the output  $M$  is the power supply rail.

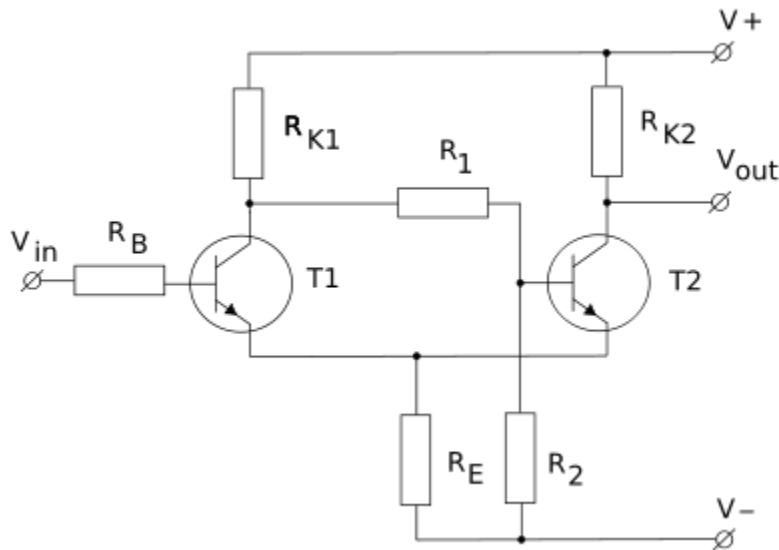
A practical Schmitt trigger configuration is shown below.



The output characteristic has exactly the same shape of the previous basic configuration, and the threshold values are the same as well. On the other hand, in the previous case, the output voltage was depending on the power supply, while now it is defined by the [Zener diodes](#) (which could also be replaced with a single [double-anode Zener diode](#)). In this configuration, the output levels can be modified by appropriate choice of Zener diode, and these levels are resistant to power supply fluctuations (i.e., they increase the [PSRR](#) of the comparator). The resistor  $R_3$  is there to limit the current through the diodes, and the resistor  $R_4$  minimizes the input voltage offset caused by the comparator's input leakage currents (see [Limitations of real op-amps](#)).

### Schmitt trigger with two transistors

In the positive-feedback configuration used in the implementation of a Schmitt trigger, most of the complexity of the comparator's own implementation is unused. Hence, it can be replaced with two cross-coupled transistors (i.e., the transistors that would otherwise implement the input stage of the comparator). An example of such a 2-transistor-based configuration is shown below. The chain  $R_{K1} R_1 R_2$  sets the base voltage for transistor T2. This divider, however, is affected by transistor T1, providing higher voltage if T1 is open. Hence the threshold voltage for switching between the states depends on the present state of the trigger.



For NPN transistors as shown, when the input voltage is well below the shared emitter voltage, T1 does not conduct. The base voltage of transistor T2 is determined by the mentioned divider. Due to negative feedback, the voltage at the shared emitters must be almost as high as that set by the divider so that T2 is conducting, and the trigger output is in the low state. T1 will conduct when the input voltage (T1 base voltage) rises slightly above the voltage across resistor  $R_E$  (emitter voltage). When T1 begins to conduct, T2 ceases to conduct, because the voltage divider now provides lower T2 base voltage while the emitter voltage does not drop because T1 is now drawing current across  $R_E$ . With T2 now not conducting the trigger has transitioned to the high state.

With the trigger now in the high state, if the input voltage lowers enough, the current through T1 reduces, lowering the shared emitter voltage and raising the base voltage for T2. As T2 begins to conduct, the voltage across  $R_E$  rises, further reducing the T1 base-emitter potential and T1 ceases to conduct.

In the high state, the output voltage is close to  $V_+$ , but in the low state it is still well above  $V_-$ . This may not be low enough to be a "logical zero" for digital circuits. This may require additional amplifiers following the trigger circuit.

The circuit can be simplified:  $R_1$  can be replaced with a short circuit connection, connecting the T2 base directly to the T1 collector, and  $R_2$  can be taken out and replaced with an open circuit. The key to its operation is that less current flows through  $R_E$  when T1 is switched on (as a result of input current into its base) than when T1 is off, because turning T1 on turns T2 off, and T2, when on, provides more current through  $R_E$  than does T1. With less current entering  $R_E$ , the voltage across it will be lower, and so once current gets going into T1, the input voltage must go lower to turn T1 back off as now its emitter voltage has been lowered. This Schmitt trigger buffer can also be turned into a Schmitt trigger inverter and another resistor saved in the process, by replacing  $R_{K2}$  with a short connection, and connecting  $V_{out}$  to the emitter of T2 instead of its collector. In this case however, a larger value of resistance should be used for  $R_E$  as it now serves as the pull-down resistor on the output, lowering the voltage on the output when the output should be low, instead of serving as only a small resistance which is only intended to develop a small voltage across it that actually adds to the output voltage when it should be at a digital low.

## Applications

Schmitt triggers are typically used in [open loop](#) configurations for noise immunity and [closed loop positive feedback](#) configurations to implement [multivibrators](#).

## Noise immunity

One application of a Schmitt trigger is to increase the noise immunity in a circuit with only a single input threshold. With only one input threshold, a [noisy](#) input signal near that threshold could cause the output to switch rapidly back and forth from noise alone. A noisy Schmitt Trigger input signal near one threshold can cause only one switch in output value, after which it would have to move beyond the other threshold in order to cause another switch.

For example, in [Fairchild Semiconductor's](#) QSE15x family of infrared photosensors<sup>[3]</sup>, an [amplified infrared photodiode](#) generates an electric signal that switches frequently between its absolute lowest value and its absolute highest value. This signal is then [low-pass filtered](#) to form a smooth signal that rises and falls corresponding to the relative amount of time the switching signal is on and off. That filtered output passes to the input of a Schmitt trigger. The net effect is that the output of the Schmitt trigger only passes from low to high after a received infrared signal excites the photodiode for longer than some known delay, and once the Schmitt trigger is high, it only moves low after the infrared signal ceases to excite the photodiode for longer than a similar known delay. Whereas the photodiode is prone to spurious switching due to noise from the

environment, the delay added by the filter and Schmitt trigger ensures that the output only switches when there is certainly an input stimulating the device.

### **Devices that include a built-in Schmitt trigger**

As discussed in the example above, the Fairchild Semiconductor QSE15x family of photosensors use a Schmitt trigger internally for noise immunity. Schmitt triggers are common in many switching circuits for similar reasons (e.g., for switch [debouncing](#)).

The following [7400 series](#) devices include a Schmitt trigger on their input or on each of their inputs:

- 7413: Dual Schmitt trigger 4-input NAND Gate
- 7414: Hex Schmitt trigger Inverter
- 7418: Dual Schmitt trigger 4-input NAND Gate
- 7419: Hex Schmitt trigger Inverter
- 74121: Monostable Multivibrator with Schmitt Trigger Inputs
- 74132: Quad 2-input NAND Schmitt Trigger
- 74221: Dual Monostable Multivibrator with Schmitt Trigger Input
- 74232: Quad NOR Schmitt Trigger
- 74310: Octal Buffer with Schmitt Trigger Inputs
- 74340: Octal Buffer with Schmitt Trigger Inputs and three-state inverted outputs
- 74341: Octal Buffer with Schmitt Trigger Inputs and three-state noninverted outputs
- 74344: Octal Buffer with Schmitt Trigger Inputs and three-state noninverted outputs
- 74(HC/HCT)7541 Octal Buffer with Schmitt Trigger Inputs and Three-State Noninverted Outputs
- SN74LV8151 is a 10-bit universal Schmitt-trigger buffer with 3-state outputs

A number of [4000 series](#) devices include a Schmitt trigger on inputs, for example:

- 4093: Quad 2-Input NAND
- 40106: Hex Inverter
- 14538: Dual Monostable Multivibrator
- 4020: 14-Stage Binary Ripple Counter
- 4024: 7-Stage Binary Ripple Counter
- 4040: 12-Stage Binary Ripple Counter
- 4017: Decade Counter with Decoded Outputs
- 4022: Octal Counter with Decoded Outputs
- 4093: Quad Dual Input NAND gate

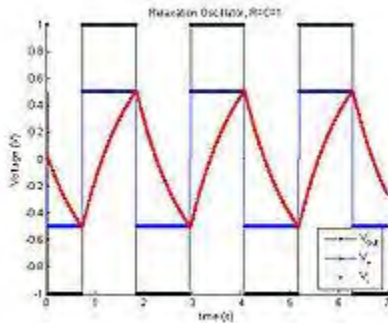
Dual Schmitt input configurable single-gate CMOS logic, AND, OR, XOR, NAND, NOR, XNOR

- NC7SZ57 Fairchild
- NC7SZ58 Fairchild
- SN74LVC1G57 Texas Instruments

- SN74LVC1G58 Texas Instruments

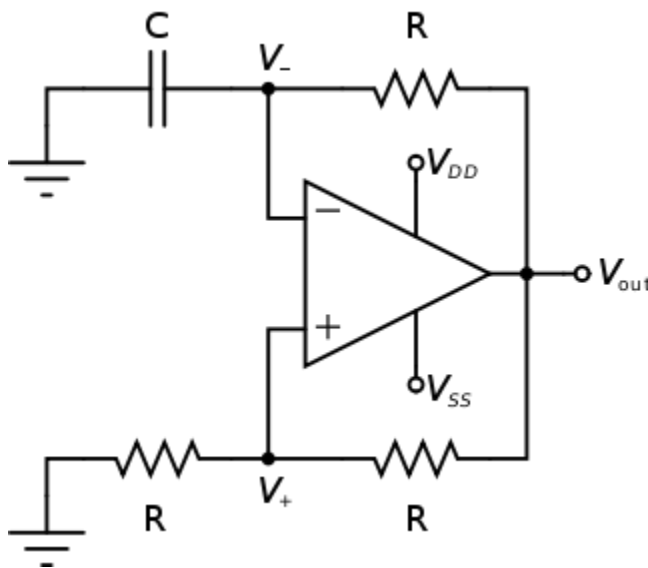
### Use as an oscillator

A Schmitt trigger is a [bistable multivibrator](#), and it can be used to implement another type of multivibrator, the [relaxation oscillator](#). This is achieved by connecting a single [resistor-capacitor](#) network to an inverting Schmitt trigger – the capacitor connects between the input and ground and the resistor connects between the output and the input. The output will be a continuous [square wave](#) whose [frequency](#) depends on the values of R and C, and the threshold points of the Schmitt trigger. Since multiple Schmitt trigger circuits can be provided by a single [integrated circuit](#) (e.g. the [4000 series CMOS](#) device type 40106 contains 6 of them), a spare section of the IC can be quickly pressed into service as a simple and reliable oscillator with only two external components.



Output and [capacitor](#) waveforms for [comparator-based relaxation oscillator](#).

For example, the comparator-based implementation of a [relaxation oscillator](#) is shown below.



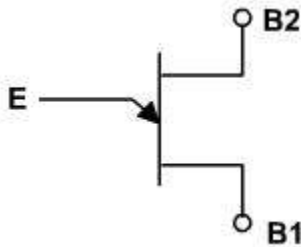
Here, a comparator-based Schmitt trigger is used in its inverting configuration. That is, the input and [ground](#) are swapped from the Schmitt trigger shown above, and so very negative signals

correspond to a positive output and very positive signals correspond to a negative output. Additionally, slow [negative feedback](#) is added with an [RC network](#). The result, which is shown on the right, is that the output automatically oscillates from  $V_{SS}$  to  $V_{DD}$  as the capacitor charges from one Schmitt trigger threshold to the other.

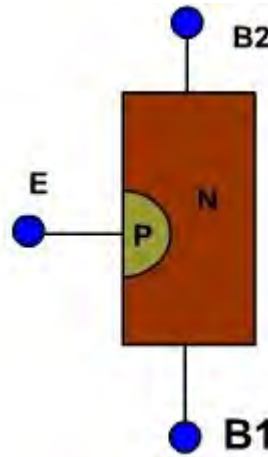
### Υνι-φονχτιον τρανσιιστορ

The UJT as the name implies, is characterized by a single pn junction. It exhibits negative resistance characteristic that makes it useful in oscillator circuits.

The symbol for UJT is shown in [fig. 1](#). The UJT is having three terminals base1 (B1), base2 (B2) and emitter (E). The UJT is made up of an N-type silicon bar which acts as the base as shown in [fig. 2](#). It is very lightly doped. A P-type impurity is introduced into the base, producing a single PN junction called emitter. The PN junction exhibits the properties of a conventional diode.



**Fig. 1**

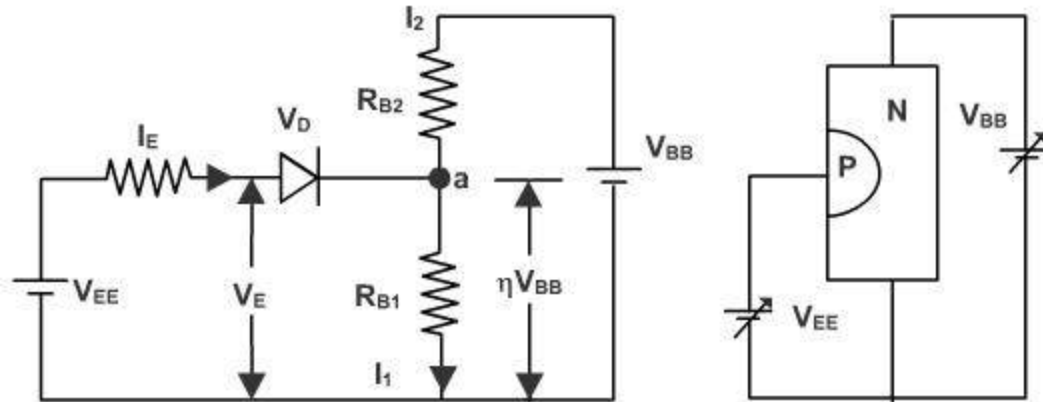


**Fig. 2**

A complementary UJT is formed by a P-type base and N-type emitter. Except for the polarity of voltage and current the characteristic is similar to those of a conventional UJT.

A simplified equivalent circuit for the UJT is shown in [fig. 3](#).  $V_{BB}$  is a source of biasing voltage connected between B2 and B1. When the emitter is open, the total resistance from B2 to B1 is simply the resistance of the silicon bar, this is known as the inter base resistance  $R_{BB}$ . Since the N-channel is lightly doped, therefore  $R_{BB}$  is relatively high, typically 5 to 10K ohm.  $R_{B2}$  is the resistance between B2 and point a', while  $R_{B1}$  is the resistance from point a' to B1, therefore the interbase resistance  $R_{BB}$  is

$$R_{BB} = R_{B1} + R_{B2}$$



**Fig. 3**

The diode accounts for the rectifying properties of the PN junction.  $V_D$  is the diode's threshold voltage. With the emitter open,  $I_E = 0$ , and  $I_1 = I_2$ . The interbase current is given by

$$I_1 = I_2 = V_{BB} / R_{BB} .$$

Part of  $V_{BB}$  is dropped across  $R_{B2}$  while the rest of voltage is dropped across  $R_{B1}$ . The voltage across  $R_{B1}$  is

$$V_a = V_{BB} * (R_{B1}) / (R_{B1} + R_{B2})$$

The ratio  $R_{B1} / (R_{B1} + R_{B2})$  is called intrinsic standoff ratio

$$\square = R_{B1} / (R_{B1} + R_{B2}) \text{ i.e. } V_a = \square V_{BB} .$$

The ratio  $\square$  is a property of UJT and it is always less than one and usually between 0.4 and 0.85. As long as  $I_B = 0$ , the circuit behaves as a voltage divider.

Assume now that  $v_E$  is gradually increased from zero using an emitter supply  $V_{EE}$ . The diode remains reverse biased till  $v_E$  voltage is less than  $\square V_{BB}$  and no emitter current flows except leakage current. The emitter diode will be reversed biased.

When  $v_E = V_D + \square V_{BB}$ , then appreciable emitter current begins to flow where  $V_D$  is the diode's threshold voltage. The value of  $v_E$  that causes, the diode to start conducting is called the peak point voltage and the current is called peak point current  $I_P$ .

$$V_P = V_D + \square V_{BB} .$$

## Uni-junction transistor

The graph of [fig. 4](#) shows the relationship between the emitter voltage and current.  $v_E$  is plotted on the vertical axis and  $I_E$  is plotted on the horizontal axis. The region from  $v_E = 0$  to  $v_E = V_P$  is called cut off region because no emitter current flows (except for leakage). Once  $v_E$  exceeds the peak point voltage,  $I_E$  increases, but  $v_E$  decreases. up to certain point called valley point ( $V_V$  and  $I_V$ ). This is called negative resistance region. Beyond this,  $I_E$  increases with  $v_E$  this is the saturation region, which exhibits a positive resistance characteristic.

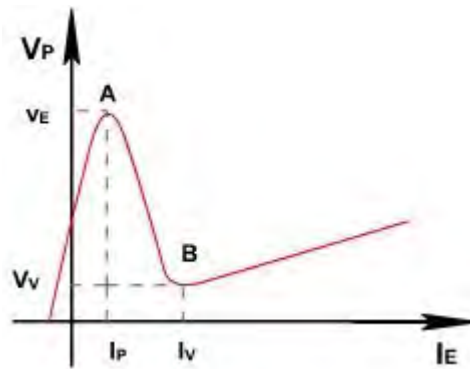


Fig. 4

The physical process responsible for the negative resistance characteristic is called conductivity modulation. When the  $v_E$  exceeds  $V_P$  voltage, holes from P emitter are injected into N base. Since the P region is heavily doped compared with the N-region, holes are injected to the lower half of the UJT.

The lightly doped N region gives these holes a long lifetime. These holes move towards B1 to complete their path by re-entering at the negative terminal of  $V_{EE}$ . The large holes create a conducting path between the emitter and the lower base. These increased charge carriers represent a decrease in resistance  $R_{B1}$ , therefore can be considered as variable resistance. It decreases up to 50 ohm.

Since  $\alpha$  is a function of  $R_{B1}$  it follows that the reduction of  $R_{B1}$  causes a corresponding reduction in intrinsic standoff ratio. Thus as  $I_E$  increases,  $R_{B1}$  decreases,  $\alpha$  decreases, and  $V_a$  decreases. The decrease in  $V_a$  causes more emitter current to flow which causes further reduction in  $R_{B1}$ ,  $\alpha$ , and  $V_a$ . This process is regenerative and therefore  $V_a$  as well as  $v_E$  quickly drops while  $I_E$  increases. Although  $R_B$  decreases in value, but it is always positive resistance. It is only the dynamic resistance between  $V_V$  and  $V_P$ . At point B, the entire base1 region will saturate with carriers and resistance  $R_{B1}$  will not decrease any more. A further increase in  $I_e$  will be followed by a voltage rise.

The diode threshold voltage decreases with temperature and  $R_{BB}$  resistance increases with temperature because Si has positive temperature coefficient.

## Uni-junction transistor

### UJT Relaxation Oscillator:

The characteristic of UJT was discussed in previous lecture. It is having negative resistance region. The negative dynamic resistance region of UJT can be used to realize an oscillator.

The circuit of UJT relaxation oscillator is shown in [fig. 1](#). It includes two resistors  $R_1$  and  $R_2$  for taking two outputs  $R_2$  may be a few hundred ohms and  $R_1$  should be less than 50 ohms. The dc source  $V_{CC}$  supplies the necessary bias. The interbase voltage  $V_{BB}$  is the difference between  $V_{CC}$  and the voltage drops across  $R_1$  and  $R_2$ . Usually  $R_{BB}$  is much larger than  $R_1$  and  $R_2$  so that  $V_{BB}$  approximately equal to  $V$ . Note,  $R_{B1}$  and  $R_{B2}$  are inter-resistance of UJT while  $R_1$  and  $R_2$  is the actual resistor.  $R_{B1}$  is in series with  $R_1$  and  $R_{B2}$  is in series with  $R_2$ .

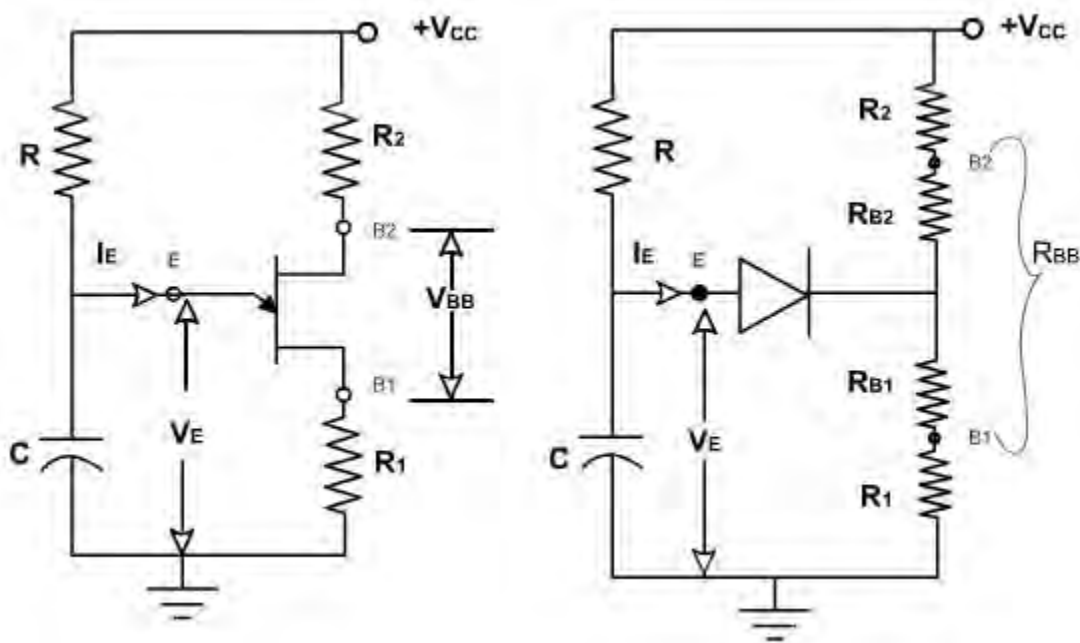


Fig. 1

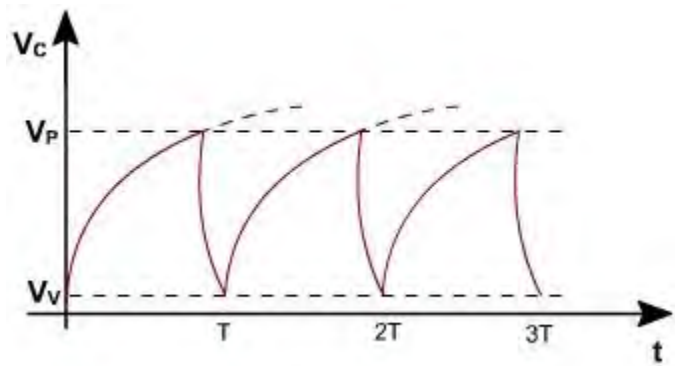
As soon as power is applied to the circuit capacitor begins to charge toward  $V$ . The voltage across  $C$ , which is also  $V_E$ , rises exponentially with a time constant

$$\tau = R C$$

As long as  $V_E < V_P$ ,  $I_E = 0$ . the diode remains reverse biased as long as  $V_E < V_P$ . When the capacitor charges up to  $V_P$ , the diode conducts and  $R_{B1}$  decreases and capacitor starts discharging. The reduction in  $R_{B1}$  causes capacitor  $C$  voltage to drop very quickly to the valley voltage  $V_V$  because of the fast time constant due to the low value of  $R_{B1}$  and  $R_1$ . As soon as  $V_E$

drops below  $V_a + V_D$  the diode is no longer forward biased and it stops conduction. It now reverts to the previous state and C begins to charge once more toward  $V_{CC}$ .

The emitter voltage is shown in [fig. 2](#),  $V_E$  rises exponentially toward  $V_{CC}$  but drops to a very low value after it reaches  $V_P$ . The time for the  $V_E$  to drop from  $V_P$  to  $V_V$  is relatively small and usually neglected. The period T can therefore be approximated as follows:



**Fig. 2**

Let  $T$  be the Time required for  $V_E$  to rise from 0 to  $V_P$ .

As long as  $R_{BB} \gg (R_1 + R_2)$ ;

$$V_{BB} \approx V_{CC} \text{ and } V_P = \eta V_{CC}$$

The capacitor charging voltage is given by

$$V_E = V_{CC} (1 - e^{-t/RC})$$

Where  $V_E$  is the instantaneous capacitor voltage.

Note that at  $t = T$ ,  $V_E = V_P = \eta V_{CC}$

$$\eta V = V_{CC} (1 - e^{-T/RC})$$

or  $\eta = (1 - e^{-T/RC})$

or  $e^{-T/RC} = 1 - \eta$

Thus

$$e^{-T/RC} = 1 - \eta$$

$$T = RC \ln\left(\frac{1}{1 - \eta}\right)$$

$$T = RC K$$

The frequency of oscillation is, therefore, given by

$$f_c = \frac{1}{T RC K}$$

The parameter  $K$  varies with  $\eta$

### Uni-junction transistor

There are two additional outputs possible for the UJT oscillation one of these is the voltage developed at B1 due to capacitor discharge while the other is voltage developed at B2 as shown in [fig. 3](#).

When UJT fires (at  $t = T$ )  $V_a$  drops, causing a corresponding voltage drop at B2. The duration of outputs at B1 and B2 are determined by C discharge time.

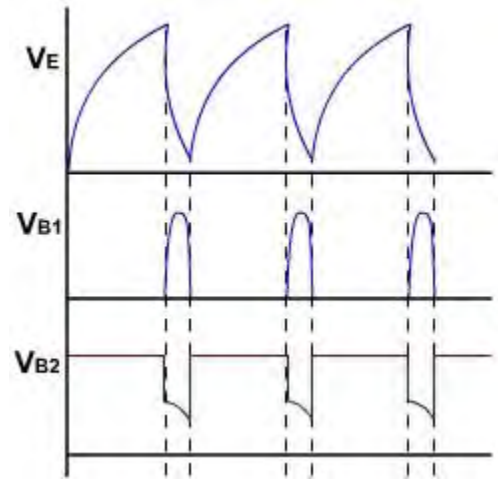
If  $R_1$  is very small, C discharges very quickly and very narrow pulse is produced at the output. If  $R_1 = 0$ , obviously no pulses appear at B1.

If  $R_2 = 0$ , no pulse can be generated at B2. If  $R_1$  is too large, its positive resistance may swamp the negative resistance and prevent the UJT from switching back after it has fired.

$R_2$ , in addition to providing a source of pulse at B2, is useful for temperature stabilization of the UJT's peak point voltage .

$$V_P = V_D + \alpha V_{BB}.$$

As the temperature increases,  $V_p$  decreases. The temperature coefficient of  $R_{BB}$  is positive.  $R_s$  is essentially independent of temperature. It is therefore possible to select  $R_2$  so that  $\alpha V_{BB}$  increases with temperature by the same amount as  $V_D$  decreases. This provides a constant  $V_P$  and, in turn, frequency of oscillation.



**Figure 31.3**

### Selection of R and C:

In the circuit, R is required to pass only the capacitor charging current. At the instant when  $V_P$  is reached; R must supply the peak current. It is therefore, necessary, that the current through R should be slightly greater than the peak point.

$$I_R > I_P$$

$$\frac{V_{CC} - V_P}{R} > I_P$$

$$R < \frac{V_{CC} - V_P}{I_P}$$

Once the UJT fires,  $V_E$  drops to the valley voltage  $V_V$  .  $I_E$  should not be allowed to increase

beyond the valley point  $I_V$ , otherwise the UJT is taken into saturation region and does not switch back,  $R$  therefore must be selected large enough to ensure that

$$I_E < I_V$$
$$\frac{V_{CC} - V_V}{R} < I_V$$
$$R > \frac{V_{CC} - V_V}{I_V}$$

Therefore,  $\frac{V_{CC} - V_V}{I_P} > R > \frac{V_{CC} - V_V}{I_V}$

As long as  $R$  is chosen between these extremes, reliable operation results.