

UNIT 1: SEMICONDUCTOR PHYSICS

Syllabus:

- I.** *Formation of energy bands and classification of solids into conductors, Semiconductors and insulators.*
- II.** *Intrinsic and extrinsic semiconductors.*
- III.** *Fermi levels in semiconductor, energy gap and its temperature dependence and the physics of semiconductor junction.*
- IV.** *Hall Effect and its application.*

[Following are the reading materials for the students. Relevant references are given. Students are strongly encouraged to read the main books for more detailed discussion.]

I. Formation of energy bands

[Reference: Material science of Thin Films by Milton Ohring, Chapter 1.4]

A solid contains large number of atoms packed together. A single isolated atom has discrete energy levels. Widely spaced isolated atoms condense to form solids due to the energy reduction accompanying bond formation. If N atoms of type A in the gas phase (g) combine to form a solid (s), the binding energy E_b is released according to the equation



Energy E_b must be supplied to reverse the equation and decompose the solid. Let us picture the process of bonding by considering the energetics within and between atoms as the inter-atomic distance progressively shrinks. In each isolated atom, the electron energy levels are discrete, as shown in the Fig. 1. As the atoms approach one another, the wave functions start to overlap and the individual levels split, as a consequence of an extension of the Pauli's exclusion principle (that says that to a collective solid no two electrons can exist in the same quantum state). Level splitting and broadening occur first for the valence or outer electrons, since their electron clouds are the first to overlap. During atomic attraction, electrons populate these lower energy levels, reducing the overall energy of the solid. With further dimensional shrinkage, the overlap increases and the inner charge clouds begin to interact. Ion-core overlap now results in strong repulsive forces between atoms, raising the system energy. A compromise is reached at the equilibrium inter atomic distance in the solid where the system energy is minimized. At equilibrium, some of the levels have broadened into bands of energy levels. The bands span different ranges of energy, depending on the atoms and specific electron levels involved. Sometimes as in metals, bands of high energy overlap. Insulators and semiconductors have energy gaps of varying width

between bands where electron states are not allowed.

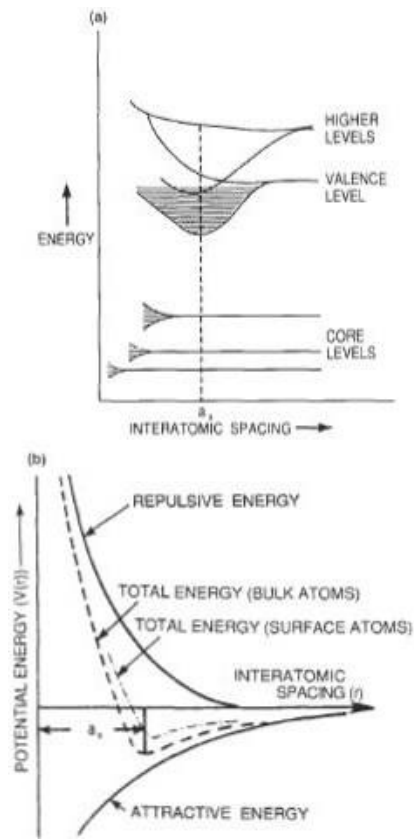


Fig. 1. (a) Splitting of electron levels and (b) energy interaction between atoms as a function of inter-atomic spacing.

The most important bands are:

Valence band – the last filled energy level at $T = 0$ K.

Conduction band – the first unfilled energy level at $T = 0$ K.

A common graphic means of distinguishing between different classes of solids involves the use of energy-band diagrams. Fig. 1 shows how individual energy levels broaden into bands when atoms are brought together to form solids. What is of interest here are the energies of electrons at the equilibrium atomic spacing in the crystal. For metals, insulators, and semiconductors the energy-band structures at the equilibrium spacing are schematically shown in Fig. 2 a, b, c. In each case the horizontal axis can be loosely interpreted as some macroscopic distance within the solid with much larger than atomic dimensions. This distance spans a region within the homogeneous bulk interior where the band energies are uniform from point to point. The uppermost band shown is called the conduction band because once electrons access its levels, they are essentially free to conduct electricity.

Metals have high conductivity because the conduction band contains electrons from the outset. One

has to imagine that there are a mind-boggling electrons per cubic centimeter (~ one per atom) in the conduction band, all of which occupy different quantum states. Furthermore, there are enormous numbers of states all at the same energy level, a phenomenon known as degeneracy. Lastly, the energy levels are extremely closely spaced and compressed within a typical 5-eV conduction-band energy width. The available electrons occupy states within the band up to a certain level known as the Fermi energy E_f (discussed in detail later). Above E_f are densely spaced excited levels, but they are all vacant. If electrons are excited sufficiently (e.g., by photons or through heating), they can gain enough energy to populate these states or even leave the metal altogether (e.g. by photo- and thermionic emission) and enter the vacuum. As indicated in Fig. 2a, the energy difference between the vacuum level and E_f is equal to $q\phi_M$, where ϕ_M is the work function and q is the electronic charge. Even under very tiny electric fields, the electrons in states at E_f can easily move into the unoccupied levels above it, resulting in a net current flow. For this reason, metals have high conductivities.

At the other extreme are insulators, in which the conduction band normally has no electrons. The valence electrons used in bonding completely fill the valence band. A large energy gap E_g ranging from 5 to 10 eV separates the filled valence band from the empty conduction band. There are normally no states and therefore no electrons within the energy gap. In order to conduct electricity, electrons must acquire sufficient energy to span the energy gap, but for all practical cases this energy barrier is all but insurmountable.

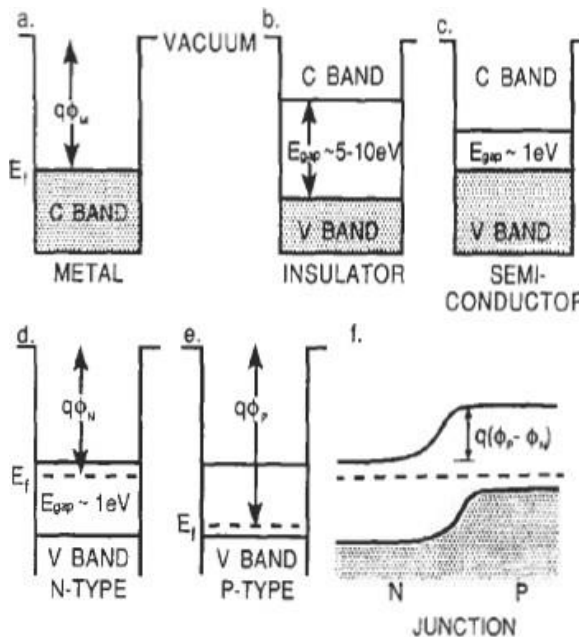


Fig. 2. Schematic band structure for (a) metal, (b) insulator, (c) semiconductors – (d) n-type semiconductor, (e) p-type semiconductor, (f) p-n semiconductor junction.

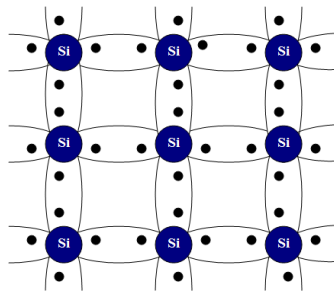
Pure (intrinsic) semiconductors at very low temperatures have a band structure like that of insulators, but E_g is smaller; e.g., $E_g = 1.1$ eV in Si and 0.68 eV in Ge. When a semiconductor is doped, new states are created within the energy gap. The electron (or hole) states associated with donors (or acceptors) are usually only a small fraction of an electron volt from the bottom of the conduction band (or top of the valence band). It now takes very little stimulation to excite electrons or holes to conduct electricity. The actual location of E_f with respect to the band diagram depends on the type and amount of doping atoms present. In an intrinsic semiconductor, E_f lies in the middle of the energy gap, because E_f is strictly defined as that energy level for which the probability of occupation is 1/2. If the semiconductor is doped with donor atoms to make it N-type, E_f lies above the midgap energy, as shown in Fig. 2d. If acceptor atoms are the predominant dopants, E_f lies below the midgap energy and a P-type semiconductor results (Fig. 2e).

In the following sections these things are elaborated.

[Reference: Solid State Electronic Devices by Ben G. Streetman, Chapter 3]

II. Intrinsic Semiconductors:

Semiconductors like germanium and silicon have crystalline structure. These are tetravalent materials. In order to acquire stability, four valance electrons form covalent bonds with neighboring atom's valance electrons. Following figure shows the two dimensional view of the Si crystal lattice covalent bonds.



A perfect semiconductor crystal with no impurities or lattice defects is called an *intrinsic* semiconductor. In such material there are no charge carriers at 0 K, since the valence band is filled with electrons and the conduction band is empty. At higher temperatures electron-hole pairs (EHP) are generated as valence band electrons are excited thermally across the band gap to the conduction band. These EHPs are the only charge carriers in intrinsic material. The generation of EHPs can be visualized in a qualitative way by considering the breaking of covalent bonds in the crystal lattice Fig. 3. At higher temperature some of the valance electrons acquire thermal energy to tear the covalent bond leaving behind an unsaturated bond - a vacancy called hole. They now act as free electrons and can move under the influence of small electric field. The energy required to break the bond is the band gap energy E_g .

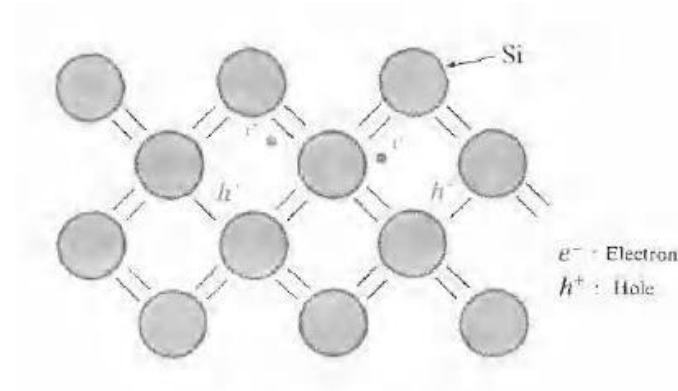


Fig. 3. Electron-hole pairs in the covalent bonding model of the Si crystal

This model helps in visualizing the physical mechanism of EHP creation, but the energy band model is more productive for purposes of quantitative calculation. One important difficulty in the "broken bond" model is that the free electron and the hole seem deceptively localized in the lattice. Actually, the positions of the free electron and the hole are spread out over several lattice spacing and should be considered quantum mechanically by probability distributions.

Since the electrons and holes are created in pairs, the conduction band electron concentration n (electrons per cm^3) is equal to the concentration of holes in the valence band p (holes per cm^3). Each of these intrinsic carrier concentrations is commonly referred to as n_i . Hence in intrinsic semiconductors

$$n = p = n_i$$

If a steady state carrier concentration is maintained, there must be *recombination* of EHPs at the same rate at which they are generated. Recombination occurs when an electron in the conduction band makes a transition to an empty state (hole) in the valence band, thus annihilating the pair.

Extrinsic Semiconductors:

Addition of controlled quantity of impurity to an intrinsic semiconductor to increase the conductivity is called doping and the resultant semiconductor is known as extrinsic semiconductor. The impurity atoms occupy lattice positions. Two types of extrinsic semiconductors are produced depending upon the group of impurity atom.

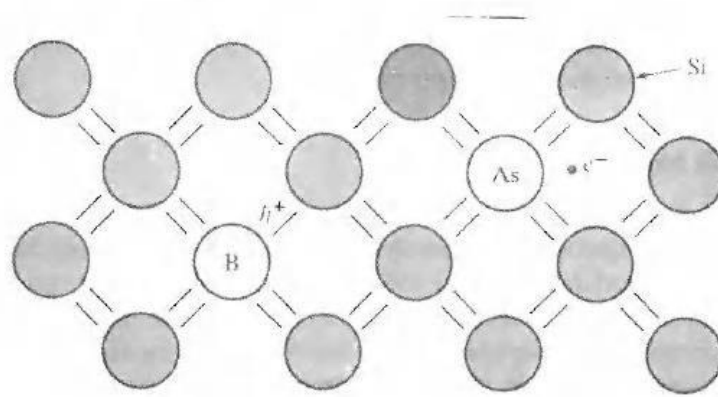


Fig. 4. Donor and acceptor atoms in the covalent bond model of Si-crystal.

n-type semiconductor:

If a pentavalent impurity (donor impurity) like arsenic (*As*) or phosphorous (*P*) is added to pure semiconductor like *Si* or *Ge*, n-type semiconductor is formed. Let's consider *As* atom (pentavalent) is doped in *Si* crystal. Each *As* atom forms covalent bonds with neighboring four *Si* atoms with four of its valance electrons. Fifth valance electron remains loosely bound to the parent impurity atom and cannot form covalent bond [Fig. 4].

When impurities or lattice defects are introduced into an otherwise perfect crystal, additional levels are created in the energy band structure, usually within the band gap. For example, an impurity from column V of the periodic table (*P*, *As*, and *Sb*) introduces an energy level very near the conduction band in *Ge* or *Si*. This level is filled with electrons at 0 K, and very little thermal energy is required to excite these electrons to the conduction band. The energy required to remove the fifth valance electron is very small (0.04 eV). Hence the addition of donor impurity generates new energy levels in the band picture. This is donor energy level (E_D) which lies just below the bottom of the conduction band, as shown in Fig. 2d.

The material doped with donor impurities can have a considerable concentration of electrons in the conduction band, even when the temperature is too low for the intrinsic EHP concentration to be appreciable. Thus semiconductors doped with a significant number of donor atoms will have $n_0 \gg \{n_i, p_0\}$ at room temperature. This is n-type material.

p-type semiconductor:

If a trivalent impurity (acceptor impurity), like boron (*B*) is added to a pure semiconductor like *Si* or *Ge*, p-type semiconductor is formed. Consider Boron (*B*) atom is doped to *Si* crystal. Each *B* atom forms covalent bonds with neighboring three *Si* atoms with three valance electrons and falls short of one electron for completing fourth covalent bonds. As a result, a vacancy is left in the bonding [Fig. 4]. With a small amount of thermal energy, this incomplete bond can be transferred to other atoms as the bonding electrons exchange positions. The idea of an electron "hopping" from an adjacent bond into the incomplete bond at the B site provides some physical insight into the

behavior of an acceptor. The impurity atoms thus supply holes which are ready to accept electrons. Hence the impurity is known as acceptor impurity.

Atoms from column III (*B, Al, Ga, and In*) introduce impurity levels in Ge or Si near the valence band (Fig. 1e). These levels are empty of electrons at 0 K. At low temperatures, enough thermal energy is available to excite electrons from the valence band into the impurity level, leaving behind holes in the valence band. Since this type of impurity level "accepts" electrons from the valence band, it is called an *acceptor* level, and the column III impurities are acceptor impurities in *Ge* and *Si*. Doping with acceptor impurities can create a semiconductor with a hole concentration p_0 much greater than the conduction band electron concentration n_0 , this type of materials are called p-type semiconductor.

Direct and Indirect Gap Semiconductors:

[OPTIONAL]

[Reference: Solid State Electronic Devices by Ben G. Streetman, Chapter 3]

The "thought experiment" in which isolated atoms were brought together to form a solid, is useful in pointing out the existence of energy bands and some of their properties. However, when quantitative calculations are made of band structures, a single electron is assumed to travel through a perfectly periodic lattice. The wave function of the electron is assumed to be in the form of a plane wave moving, for example, in the x -direction with propagation constant k , also called a wavevector. The space-dependent wave function for the electron is

$$\psi_k(x) = U(k_x, x)e^{ik_x x}$$

Where the function $U(k_x, x)$ modulates the wave function according to the periodicity of the lattice.

In such a calculation, allowed values of energy can be plotted vs. the propagation constant k . Since the periodicity of most lattices is different in various directions, the (E, k) diagram must be plotted for the various crystal directions and the full relationship between E and k is a complex surface which should be visualized in three dimensions.

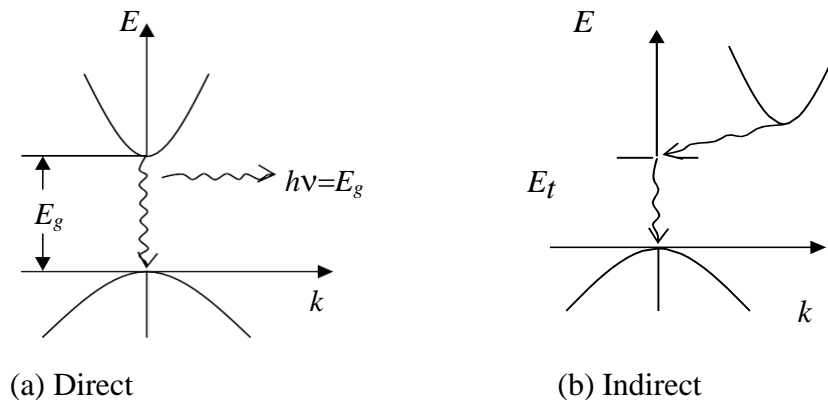


Fig. 5. (a) Direct and (b) Indirect transitions in semiconductors.

The band structure of *GaAs* has a minimum in the conduction band and a maximum in the valence band for the same k value ($k = 0$). On the other hand, *Si* has its valence band maximum at a different value of k than its conduction band minimum. Thus an electron making a smallest-energy transition from the conduction band to the valence band in *GaAs* can do so without a change in k value; on the other hand, a transition from the minimum point in the *Si* conduction band to the maximum point of the valence band requires some change in k . Thus there are two classes of semiconductor energy bands **direct** (the first case, as in *GaAs*) and **indirect** (as in *Si*).

In a direct gap semiconductor such as *GaAs*, an electron in the conduction band can fall to an empty state in the valence band, giving off the energy difference E_g as a photon of light. On the other hand, an electron in the conduction band minimum of an indirect semiconductor such as *Si* cannot fall directly to the valence band maximum but must undergo a momentum change as well as changing its energy. For example, it may go through some defect state (E_t) within the band gap. In an indirect transition which involves a change in k , the energy is generally given up as heat to the lattice rather than as an emitted photon.

This difference between direct and indirect band structures is very important for deciding which semiconductors can be used in devices requiring light output. For example, semiconductor light emitters and lasers generally must be made of materials capable of direct band-to-band transitions or of indirect materials with vertical transitions between defect states.

III. Fermi Level and Fermi Function:

[Reference: *Solid State Electronic Devices* by Ben G. Streetman, Chapter 3]

Electrons in solids obey Fermi-Dirac statistics. The Fermi function states that the distribution of electrons over a range of allowed energy levels at thermal equilibrium is

$$f(E) = \frac{1}{1 + e^{(E - E_F)/kT}}$$

Where k is Boltzmann's constant ($= 8.62 \times 10^{-5} \text{ eV/K} = 1.38 \times 10^{-23} \text{ J/K}$), T is the absolute temperature. The function $f(E)$, the Fermi-Dirac distribution function, gives the probability that an available energy state at E will be occupied by an electron at absolute temperature T .

The quantity E_F is called the Fermi level. For an energy E equal to the Fermi level energy E_F , the occupation probability is

$$f(E_F) = \frac{1}{1 + e^{(E_F - E_F)/kT}} = \frac{1}{2}$$

Thus an energy state at Fermi level has a probability of occupancy by an electron equal to $\frac{1}{2}$. This distribution function is shown in Fig. 6.

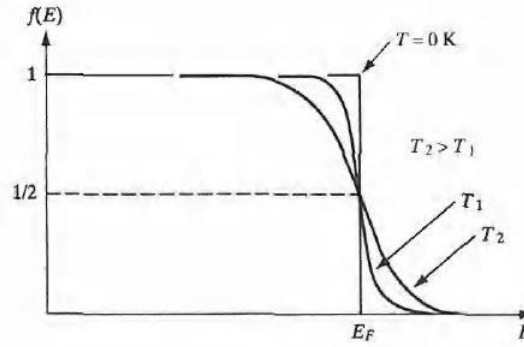


Fig. 6. The Fermi-Dirac distribution function.

Thus at 0K, every available energy state up to E_F is filled, and all states above E_F are empty. Therefore, in other words, Fermi level is the highest energy state at $T = 0$ K up to which electron can occupy. At temperatures higher than 0 K, there exists some probability for states above the Fermi level to be filled by electrons.

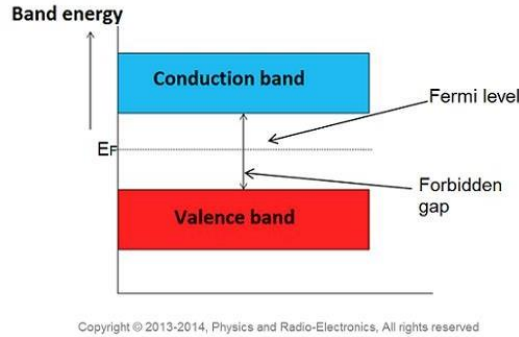
The Fermi distribution function is used to calculate the concentration of electrons and holes in a semiconductor, if the densities of available states in the valence and conduction bands are known. The concentration of electrons in the conduction band is given by

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

Where $N(E)dE$ is the density of states (cm^{-3}) in the energy range dE . The subscript 0 used with the electron and hole concentration symbols (n_0 , p_0) indicates equilibrium conditions. The number of electrons per unit volume in the energy range dE is the product of the density of states and the probability of occupancy $f(E)$. Thus the total electron concentration is the integral over the entire conduction band. It can be shown that $N(E)$ is proportional to $E^{1/2}$, so the density of states in the conduction band increases with electron energy. On the other hand, Fermi function becomes extremely small for large energies. The result is that the product $f(E)N(E)$ decreases rapidly above E_c , and very few electrons occupy energy states far above the conduction band edge. Similarly, the probability of finding an empty state (hole) in the valence band $[1 - f(E)]$ decreases rapidly below E_v , and most holes occupy states near the top of the valence bands.

Fermi Level in Intrinsic Semiconductors:

In pure or intrinsic semiconductor, the concentration of free electrons in the conduction band and the concentration of holes in the valance band are same. For an intrinsic semiconductor the Fermi level lies half way between valence and conduction band.



Consider an intrinsic semiconductor at T K.

The concentration of electrons in the conduction band is

$$n = N_c \exp\left[-\left(\frac{E_c - E_F}{kT}\right)\right]$$

Where, $N_c = 2(2\pi m_n kT / h^2)^{3/2}$ is the effective density of electrons at the conduction band edge (E_c).

m_n = effective mass of the e, k = Boltzmann's constant, h = Planck's constant

Similarly, the concentration of holes in the valence band is given by

$$p = N_v \exp\left[-\left(\frac{E_F - E_v}{kT}\right)\right]$$

Where, $N_v = 2(2\pi m_p kT / h^2)^{3/2}$ represents the effective density of holes at the valence band edge, E_v .

m_p = effective mass of the hole.

For an intrinsic semiconductor,

$$n = p = n_i$$

Therefore,

$$N_c \exp\left[-\left(\frac{E_c - E_F}{kT}\right)\right] = N_v \exp\left[-\left(\frac{E_F - E_v}{kT}\right)\right]$$

$$\text{Or} \quad \exp\left(\frac{2E_F - E_c - E_v}{kT}\right) = \frac{N_v}{N_c} \quad \text{or} \quad \frac{2E_F - E_c - E_v}{kT} = \ln\left(\frac{N_v}{N_c}\right)$$

Or
$$E_F = E_i = \frac{E_c + E_v}{2} + \frac{kT}{2} \ln\left(\frac{N_v}{N_c}\right)$$

Putting the values of N_v and N_c we get

$$E_F = \frac{E_c + E_v}{2} + \frac{3}{4} kT \ln\left(\frac{m_p}{m_n}\right)$$

At $T = 0$ K

$$E_F = \frac{E_c + E_v}{2}$$

i.e. the Fermi level lies in the middle of the conduction and the valence band. This is also true for other temperatures provided $m_p = m_n$. However, in general $m_p > m_n$ and the Fermi level is raised slightly as T exceeds 0 K.

Fermi level in an extrinsic semiconductor

Let N_D and N_A be the concentration of ionized donors and ionized acceptor, respectively, in an extrinsic semiconductor. Since the semiconductor is electrically neutral, the sum of the concentration of electrons (n) and negatively charged acceptor ions must be equal to the sum of the concentrations of holes (p) and the positively charged donor ions. Thus

$$n + N_A = p + N_D$$

For an n -type material, we can put $p = 0$ and $N_A = 0$. Therefore, $n = N_D$ and we can write

$$n = N_D = N_c \exp\left[-\left(\frac{E_c - E_F}{kT}\right)\right]$$

Solving for E_F , we get

$$E_F = E_c - kT \ln\left(\frac{N_c}{N_D}\right)$$

Generally, $N_c > N_D$, so that $E_F < E_c$. Thus, the Fermi level lies in the forbidden gap slightly below the conduction band in the n -type semiconductor (Fig. 7b). As the donor concentration (N_D) increases, the Fermi level moves up and enters the conduction band. The semiconductor then becomes degenerate and behaves as a metal. As the temperature increases, the Fermi level moves down, and the semiconductor tends to become intrinsic.

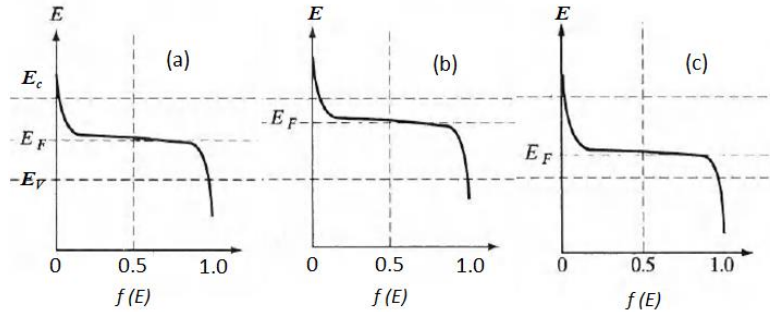


Fig. 7. Fermi-Dirac distribution for (a) intrinsic, (b) n -type and (c) p -type semiconductors.

Similarly, **for a p-type semiconductor**, we can put $n = 0$ and $N_D = 0$, so that

$$p = N_A = N_v \exp\left[-\left(\frac{E_F - E_v}{kT}\right)\right]$$

Or

$$E_F = E_v + kT \ln\left(\frac{N_v}{N_A}\right)$$

Usually, $N_v > N_A$, consequently $E_F > E_v$. Thus, for a p-type semiconductor the Fermi level lies in the forbidden gap slightly above the top of the valence band (Fig. 7c). If N_A is increased, the Fermi level moves down and enters the valence band. Also with increasing temperature, the Fermi level moves towards the center of the band gap, and the material becomes intrinsic.

If $N_A = N_D$, i.e. the semiconductor is fully compensated, then $n = p$ and the Fermi level is determined by the expression for intrinsic semiconductor.

Invariance of Fermi level at equilibrium

[Reference: Solid State Electronic Devices by Ben G. Streetman, Chapter 3]

All the system, in general, and the semiconductor devices, in particular, work on the principle of the Invariance of the Fermi level at equilibrium. It states that no discontinuity or gradient can arise in the equilibrium Fermi level E_f .

Let us consider two materials (with density of states $N_1(E)$ and $N_2(E)$, respectively, and the Fermi distribution $f_1(E)$ and $f_2(E)$, respectively) in intimate contact such that electrons can move between the two. These may be, for example, dissimilar semiconductors, n- and p-type regions, a metal and a semiconductor, or simply two adjacent regions of a non-uniformly doped semiconductor. Each material is described by a Fermi-Dirac distribution function and some distribution of available energy states that electrons can occupy.

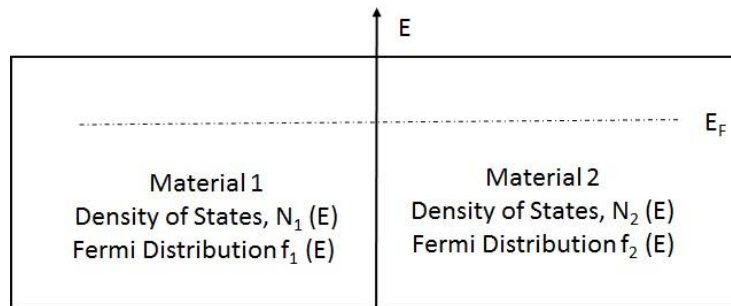


Fig. 7. Two materials in intimate contact at equilibrium.

At thermal equilibrium, there is no current, and therefore no net charge transport and no transfer of energy. So for each energy E in the above figure any transfer of electron from material 1 to material 2 must be balanced by equal and opposite transfer of electron from 2 to 1.

At energy E the rate of transfer of electrons from 1 to 2 is proportional to the number of filled states at E in material 1 times the number of empty states at E in material 2.

Thus,

$$\text{Rate from 1 to 2} \propto N_1(E)f_1(E) \cdot N_2(E)[1-f_2(E)]$$

Similarly,

$$\text{Rate from 2 to 1} \propto N_2(E)f_2(E) \cdot N_1(E)[1-f_1(E)]$$

At equilibrium this must be equal. Thus we have

$$N_1(E)f_1(E) \cdot N_2(E)[1-f_2(E)] = N_2(E)f_2(E) \cdot N_1(E)[1-f_1(E)]$$

Rearranging,

$$f_1(E) = f_2(E)$$

$$\text{i.e.} \quad \left[1 + e^{\frac{E-E_{F1}}{kT}}\right]^{-1} = \left[1 + e^{\frac{E-E_{F2}}{kT}}\right]^{-1}$$

$$\text{or} \quad E_{F1} = E_{F2}$$

Therefore, there is no discontinuity in the equilibrium Fermi level.

Law of mass action for a semiconductor

Since for an intrinsic semiconductor

$$n = p = n_i$$

Then,

$$\begin{aligned} np &= n_i^2 = N_c N_v \exp \left(-\frac{E_c - E_v}{kT} \right) \\ &= N_c N_v \exp \left(-\frac{E_g}{kT} \right) \\ &= 4 \left(\frac{2\pi kT}{h^2} \right)^3 (m_n m_p)^{\frac{3}{2}} \exp \left(-\frac{E_g}{kT} \right) \end{aligned}$$

This shows that for a given semiconductor, the product of electron and hole concentrations is a constant at a given temperature and is equal to the square of the intrinsic carrier concentration. This is known as the law of mass action and holds for both intrinsic and extrinsic semiconductors. If impurity atoms are added to a semiconductor to increase n , there will be corresponding decrease in p such that the product np remains constant. Thus we always have

$$np = n_i^2$$

Drift velocity, mobility and conductivity of intrinsic semiconductor

The charge carriers in a solid are in constant random motion, even at thermal equilibrium. At room temperature the thermal motion of an individual electron may be visualized as random scattering from lattice vibrations, impurities, other electrons and defects as shown in the figure below.

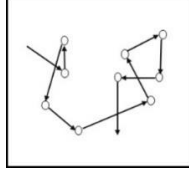


Fig. 8. Thermal motion of electron in solids

Since the scattering is random, there is no net motion of electrons (or holes in valence band, as at ordinary temperature a semiconductor has both free electrons in conduction band and equal number of holes in valence band) over a period of time to a particular direction. Hence there will be no net current flow.

If an electric field \mathbf{E} is applied, each electron experiences a net force $-q\mathbf{E}$ from the field. Though this force may be insufficient to alter appreciably the random path of an individual electron; the effect when averaged over all the electrons, however, is a net motion of the group in the direction opposite to the field direction. Same logic also holds good for holes, though, in this case, the motion of holes will be in the direction of the field.

Therefore, when an electric field is applied the velocity of charge carriers begin to increase. However, the velocity acceleration is not continuous as the net acceleration is just balanced in steady state by deceleration of the collision process.

The extra (i.e. other than thermal diffusion) velocity acquired by the carriers in the presence of applied electric field is called **drift velocity** (v_d). It is proportional to the strength of the applied electric field, i.e.

$$v_d \propto \mathbf{E} \text{ or } v_d = \mu \mathbf{E}$$

where the constant of proportionality, μ , is called the mobility of the charge carrier. It is defined as the drift velocity per unit electric field. S.I. unit for μ is $m^2/V\text{-sec}$.

In an intrinsic semiconductor, since the electrons move in a nearly empty conduction band and holes move in a nearly full valence band, we have to consider both of their motion.

Let v_{dn} , μ_n and n denote the drift velocity, mobility and concentration of electrons, respectively. Then current density due to electrons is given by

$$J_n = -nqv_{dn}$$

Now,

$$v_{dn} = -\mu_n \mathbf{E}$$

[negative sign indicates that the electron moves against the electric field]

Combining the above two equations, we get,

$$J_n = nq\mu_n \mathbf{E}$$

Comparing with the Ohm's law: $J_n = \sigma_n E$

where, σ_n = electrical conductivity of the material

Thus we get,

$$\sigma_n = nq\mu_n$$

For hole conduction we change n to p , $-q$ to $+q$, and μ_n to μ_p ($= v_{dp}/E$) and the conductivity due to holes in the valence band is

$$\sigma_p = pq\mu_p \quad \text{where, } p = \text{concentration of holes}$$

Thus the total conductivity of the material is

$$\sigma = \sigma_n + \sigma_p = q(n\mu_n + p\mu_p)$$

For an intrinsic semiconductor,

$$n = p = n_i$$

Therefore,

$$\sigma = qn_i(\mu_n + \mu_p)$$

N.B.

1. The drift speed is usually much smaller than the random speed due to thermal motion.
2. In semiconductors, the movement of carriers or flow of current is consequence of two processes;
 - a) drift current due to the drift of carriers under the effect of an electric field.
 - b) diffusion current due to diffusion of carriers under the effect of concentration gradient of dopants present in the semiconductor. It is generally negligible for uniform distribution of impurities.

***p-n* junction:**

When a *p*-type semiconductor is joined together with an *n*-type semiconductor a *p-n* junction is formed.

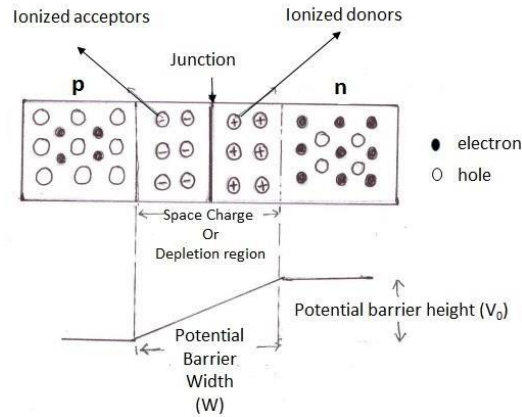


Fig. 9. Properties of an equilibrium *p-n* junction

When the junction is formed, due to concentration gradient holes diffuse from *p*-side to *n*-side and electrons diffuse from *n* to *p* side. The diffusing majority carriers from the regions recombine near the junction and disappear, leaving behind uncompensated acceptor and donor ions. Since this region of width [*W* (say)] near the junction is devoid of mobile charge carriers, this is known as **depletion region**. This is also called the **space charge region**, since space charge exists within *W*, while neutrality is maintained outside this region.

Because of the development of the '+ve' space charge region near the *n*-side of the junction and '-ve' space charge region near the *p*-side, an electric field (*E*) is created from '+ve' charge towards '-ve' charge. This electric field appears in the region of width '*W*' about the junction (known as **potential barrier width**) and is directed opposite to that of diffusion current for each of the carriers. Therefore, the field creates a drift component of current from *n* to *p*, opposing diffusion current. The electric field *E* builds up to the point where net current is zero at equilibrium.

There is a gradient of potential in the direction opposite to *E*, in accordance to the relation $E = -dV/dx$. In the neutral region outside *W*, a constant potential is maintained. The potential difference, V_0 , between the two ends of the barrier width is known as the contact or junction potential or the **built-in potential barrier**, in that it is necessary to the maintenance of the equilibrium at junction and any electron or hole must surmount this potential to disturb the equilibrium.

It is worthwhile to note here that the contact or barrier potential cannot be measured by placing a

Volt-meter across the device, because new contact potentials are formed at each probe, just cancelling each V_0 . By definition, V_0 is an equilibrium quantity and no net current can result from it.

The energy-band diagram of a p - n junction at equilibrium is shown in Fig. 10. It is quite evident that because of the invariance of Fermi level in the system, the bands in both the p and n sides bend.

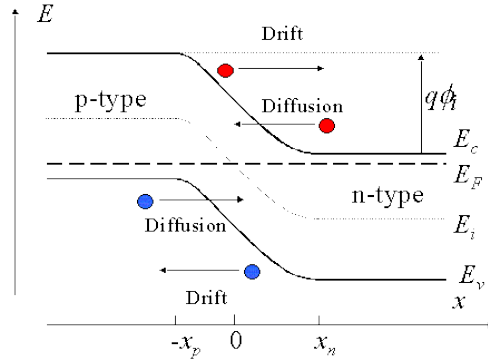


Fig. 10. The energy band diagram of a p - n junction at equilibrium.

A p - n junction can act as a diode. Diode is a semiconductor device that essentially acts as a one way switch for a current. It allows current to flow easily in one direction, and severely restricts current from flowing in the other direction. Diode can be operated in two ways – as forward biased or as reversed biased. In the following section we discuss their features.

Built-in Potential Barrier

Assuming that no voltage is applied across the p - n junction and the junction is in thermal equilibrium, the Fermi energy level is constant throughout the entire system, as shown below.

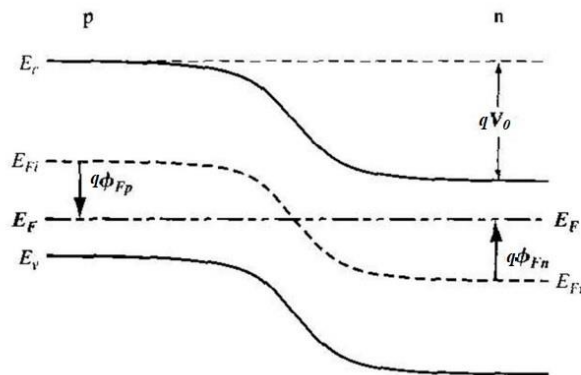


Fig. 11. Energy-band diagram of a p - n junction in thermal equilibrium.

The conduction and valence band energies must bend as we move through the space charge region, since the relative position of the conduction and valence bands with respect to the Fermi energy changes between p and n regions.

Electrons in the conduction band of the n region will see a potential barrier in trying to move into the conduction band of the p region. This potential is known as the built-in potential barrier and is denoted by V_0 . The role of potential barrier V_0 is to maintain the equilibrium so that no current is produced by this voltage.

The intrinsic Fermi level is equidistant from the conduction band edge through the junction, thus built-in potential barrier can be determined as the difference between the intrinsic Fermi levels in the p and n regions. Let us define the potentials ϕ_{Fn} and ϕ_{Fp} as shown in Fig. 11, so that we have

$$V_0 = \phi_{Fn} + \phi_{Fp} \quad (1)$$

In the n region, the electron concentration in the conduction band is given by

$$n = N_c \exp\left[-\left(\frac{E_c - E_F}{kT}\right)\right] \quad (2)$$

which can also be written as

$$n = n_i \exp\left(\frac{E_F - E_{Fi}}{kT}\right) \quad (3)$$

where n_i and E_{Fi} are the intrinsic carrier concentration and the intrinsic Fermi energy, respectively. The potential Φ_{Fn} in the n region may be defined as

$$q \cdot \Phi_{Fn} = E_{Fi} - E_F \quad (4)$$

Substituting Eqn. 4 in Eqn. 3 gives

$$n = n_i \exp\left[-\left(\frac{q \cdot \Phi_{Fn}}{kT}\right)\right] \quad (5)$$

Taking the *natural log* on both sides of Eqn. 5 and setting $n = N_D$ and solving for the potential gives

$$\Phi_{Fn} = \frac{-kT}{q} \ln\left(\frac{N_D}{n_i}\right) \quad (6)$$

Similarly, in the p region, the hole concentration is given by

$$p = N_A = n_i \exp\left(\frac{E_{Fi} - E_F}{kT}\right) \quad (7)$$

Where N_A is the acceptor concentration. We define the potential Φ_{Fp} in the p region as

$$q \cdot \Phi_{Fp} = E_{Fi} - E_F \quad (8)$$

Combining Eqns. 7 and 8, we get

$$\Phi_{Fp} = \frac{+kT}{q} \ln\left(\frac{N_A}{n_i}\right) \quad (9)$$

Therefore, the built-in potential barrier for the step function is found by substituting Eqns. 6 and 9 in Eqn. 1 as

$$V_0 = \frac{kT}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right) \quad (10)$$

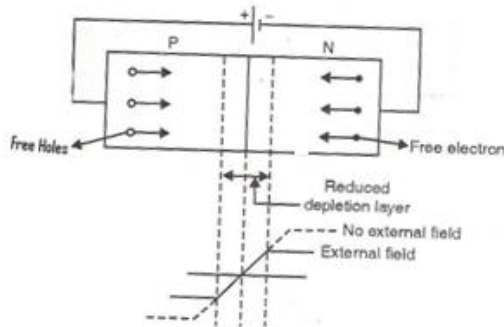
kT/q is defined as the thermal voltage.

Forward biased p - n junction

[OPTIONAL]

When positive terminal of the battery is connected to p -side and negative terminal of the battery is connected to the n -side of the p - n junction, then the junction is said to be forward biased. The electrons in the n -region are repelled by the negative terminal of the battery and the holes in the p -region are repelled by the positive terminal of the battery. As a result of which both the width of the depletion region and the barrier height decreases, as shown below. Now with a comparatively smaller applied potential the electrons and holes acquire energy to overcome the potential barrier and cross the junction.

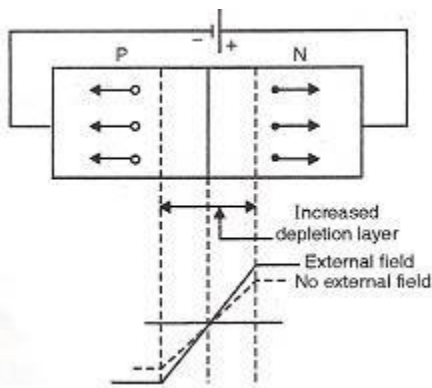
In this case we have a large increase in current for a small increase in voltage.



Reverse biased p - n junction

When the external potential applied to the p - n junction is such that the positive terminal of the battery is connected to the n -side and negative terminal to the p -side then the junction is said to be reverse biased. The reverse bias increases the potential barrier and also the width of the depletion layer (Fig. below). Under this condition, the majority charge carriers in both the regions move away from the junction. Hence there is no current due to the majority carriers.

However, a very small current flows under reverse bias due to minority charge carriers. Hence the reverse resistance is very high.



IV. The Hall Effect and its applications

[Reference: Solid State Electronic Devices by Ben G. Streetman, Chapter 3]

In 1879, E.H. Hall observed that when a current carrying conductor is placed in a magnetic field, a potential proportional to the current and to the magnetic field is developed across the material in a direction perpendicular to both the current and to the magnetic field. The establishment of electric field is known as Hall effect and the resultant voltage is known as Hall voltage.

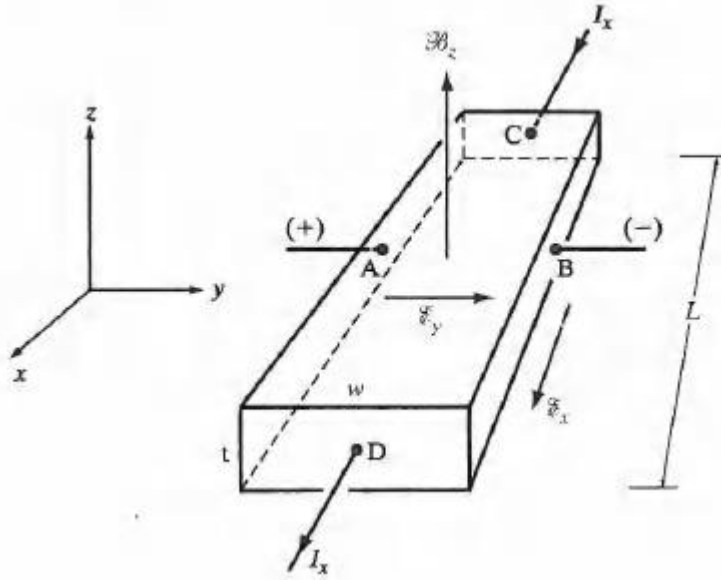


Fig. 12. The Hall effect.

Let us consider a p-type bar placed along the x -axis as shown in Fig. 12. A current I is flowing along x direction and the magnetic field is along the z -direction. Then the total force on a single hole due to electric and magnetic fields is given by Lorentz equation

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

In the y -direction the force is

$$F_y = q(E_y + v_x B_z)$$

Therefore, unless an electric field E_y is established along the width of the bar, each hole will experience a net force (and, therefore, an acceleration) in the negative y -direction due to the $qv_x B_z$ product. To maintain a steady state flow of holes down the length of the bar, the electric field E_y must just balance the product $v_x B_z$, i.e.

$$E_y = v_x B_z$$

So that the net force $F_y = 0$.

Actually, the electric field is set up when the magnetic field shifts the hole distribution slightly in the negative (-) y direction. Once E_y balances $v_x B_z$, no lateral force is experienced by the holes as they drift along the bar.

The establishment of the electric field E_y is known as the Hall effect, and the resulting voltage $V_{AB} = E_y w$ is called the Hall voltage (where w is the width of the bar).

Now, if J is the current density and p_0 is the concentration of holes, then

$$J = qp_0 v$$

Therefore, we can express E_y as

$$E_y = \frac{J_x}{qp_0} B_z = R_H J_x B_z \quad \text{with} \quad R_H = \frac{1}{qp_0}$$

Thus the Hall field is proportional to the product of the current density and the magnetic flux density.

The proportionality constant $R_H = (qp_0)^{-1}$ is called the Hall coefficient.

Measurement of carrier concentration

The measurement of Hall voltage for a known current and magnetic field can be used to find the hole concentration p_0 .

$$p_0 = \frac{1}{qR_H} = \frac{J_x B_z}{qE_y} = \frac{(\frac{I_x}{wt}) B_z}{q(\frac{V_{AB}}{w})} = \frac{I_x B_z}{qtV_{AB}}$$

Where w and t are the width and the thickness of the bar, respectively.

Since all the quantities of the right hand side of the above equation can be measured, Hall effect can be used to determine the carrier concentration accurately.

Measurement of mobility

If R is the resistance of the sample, then resistivity

$$\rho = \frac{R \times (\text{Area})}{(\text{Length})} = \frac{Rwt}{L} = \frac{V_{CD}/I_x}{L/wt}$$

Now, conductivity, $\sigma = 1/\rho = q\mu_p p_0$, where μ_p = hole mobility.

Hence,

$$\mu_p = \frac{\sigma}{qp_0} = \frac{1/\rho}{1/R_H} = \frac{R_H}{\rho}$$

Therefore, the measurements of the Hall coefficient and the resistivity over a range of temperature give plots of majority carrier concentration and mobility vs. temperature. Such measurements are extremely useful in the analysis of semiconductor materials.

For n -type semiconductor similar results can be obtained by replacing ' q ' by ' $-q$ ' and p_0 by n_0 . i.e. in case of n -type semiconductor,

$$R_H = -\frac{1}{qn_0} \quad \text{and} \quad E_y = -\frac{J_x}{qn_0} B_z$$

Measurement of the sign of the Hall voltage is a common technique for determining the type (i.e. p -type or n -type) of an unknown semiconductor.

Problem:

1. A Si sample is doped with 10^{17} phosphorous atoms/cm³. Mobility of the charge carrier is 700 cm²/V-s. Find the resistivity and Hall voltage in a sample 100 μ m thick, if $I_x = 1$ mA and $B_z = 1$ KG = 10^{-5} Wb/cm².

Solution: Conductivity

$$\sigma = q\mu_n n_0 = (1.6 \times 10^{-19} \text{C}) \left(700 \frac{\text{cm}^2}{\text{V}} \right) (10^{17} \text{cm}^{-3}) = 11.2 (\Omega - \text{cm})^{-1}$$

Since p_0 is negligible in n -type semiconductor.

Therefore, resistivity, $\rho = \sigma^{-1} = 0.0893 \Omega - \text{cm}$

The Hall coefficient is

$$R_H = -\frac{1}{qn_0} = -62.5 \frac{\text{cm}^3}{\text{C}}$$

The Hall voltage,

$$V = \frac{I_x}{qtn_0} B_z = \frac{I_x B_z}{t} R_H = \frac{(10^{-3} \text{A}) \left(\frac{10^{-5} \text{Wb}}{\text{cm}^2} \right)}{10^{-2} \text{cm}} \left(-62.5 \frac{\text{cm}^3}{\text{C}} \right) = -62.5 \mu\text{V}$$