by N positive ions hitting the cathode, obtain an expression for the current and hence deduce a relationship between β and γ.

Suggest experiments for determining the values of β and γ and estimating their relative importance. (*C.U., M.S.T. Part II*)

8. A vacuum tube 3 cm in diameter contains nitrogen at room temperature and at a pressure of 10^{-5} mm Hg. Assuming that the diameter of a nitrogen molecule is 4×10^{-8} cm, show that the molecular paths are usually terminated by collision with the wall of the tube. (*C.U., M.S.T. Part I*)

Bibliography

1. LINDSAY, R. B. (1941) *Introduction to Physical Statistics*. Wiley, New York and London.
2. CHAPMAN, S. and COWLING, T. G. (1952) *The Mathematical Theory of Non-uniform Gases*, 2nd edn. University Press, Cambridge.
3. VON ENGEL, A. (1965) *Ionized Gases*, 2nd edn. Oxford University Press, London.
4. MORGAN, C. GREY (1965) Fundamentals of electronic discharge in gases. In *Handbook of Vacuum Physics*, vol. 2, part 1, Ed. A. H. BECK. Pergamon, Oxford.
5. LLEWELLYN-JONES, F. (1966) *The Glow Discharge*. Methuen, London.
6. ACTON, J. R. and SWIFT, J. D. (1963) *Cold Cathode Discharge Tubes*. Heywood, London.
7. BECK, A. H. and JACKSON, T. M. (1959) *Electl Commun.*, 35, 251.
8. GREGORY, B. C. (1966) *Int. J. Electron.*, 20, 495.
9. BECK, A. H. (1963) *J. Instn elect. Engrs*, 9, 420.

65/70

10/92

92

3

Semiconductor physics

The name *semiconductor* is given to a group of materials whose resistivities lie between those of insulators and good (i.e. metallic) conductors. The range of experimentally observed resistivities for insulators, semiconductors and metals is shown in Fig. 3.1. The most striking difference between the physical behaviour of semiconductors and that of metals is that the *resistivity* of semiconductors *decreases* with *increasing* temperature instead of increasing in the manner of all metals.

The phenomena associated with semiconductors were first clearly recognized in the late 1920's and early 1930's by physicists investigating the physics of electrical rectification in copper oxide and selenium barrier layers. These materials are relatively complex and a full understanding of the subject was only obtained after very pure semiconducting elements, of which germanium and silicon are the most important examples, were investigated. This work was done during and after the Second World War.

The properties of semiconducting materials are very greatly influenced by the introduction of very small traces of impurity elements, even a proportion of one impurity atom to millions of atoms of the parent material being able to change the properties significantly. The physical constants, resistivity, thermal conductivity, etc., depend strongly on the temperature according to laws of the general form

$$\rho = \rho_0 \exp\left(\frac{eV}{kT}\right)$$

where, by way of example, ρ and ρ_0 are the resistivities at T K and 0 °K respectively. eV is a so-called activation energy, characteristic of the material, k is Boltzmann's constant, which merely serves to measure temperature on an energy basis.

A study of semiconductor physics has two major aims. First, it should allow one to understand the details of the process of electrical conduction in the semiconducting material of known, i.e. measured, properties; secondly, it should enable one to design special semiconducting materials with desired properties. Today, we have a rather adequate knowledge of the first part of this programme but there is still a great deal to be done before the second part

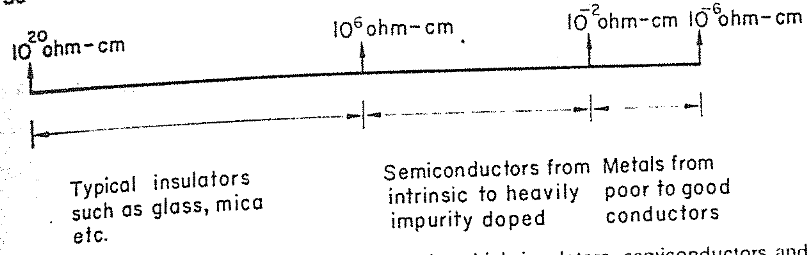


Fig. 3.1 The approximate resistivity ranges in which insulators, semiconductors and metallic conductors may be found

can be implemented. In this book we deal only with the analysis of the behaviour of semiconductors and with the characteristics and properties of semiconductor devices.

The properties of semiconductors are best understood by studying their lattice structure and by using the energy band concept of electrons in solids.

3.1. Band structure of semiconductors

The energy band structures of diamond, which is an insulator at all realizable temperatures, a semiconductor (germanium) and good metallic conductors are compared in Fig. 3.2. The insulator (Fig. 3.2a) is seen to have a

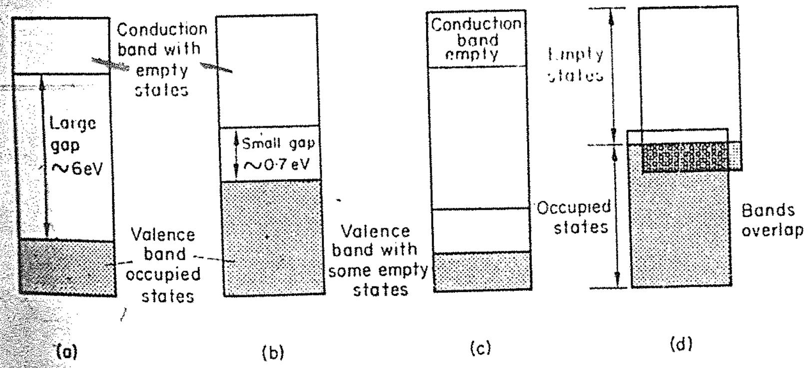


Fig. 3.2 Energy level diagrams for: (a) an insulator; (b) a semiconductor; (c) and (d) good metallic conductors

full valence band separated by a large energy gap from the empty conduction band. Since an energy difference of 1 eV is roughly equivalent to a temperature rise of 11,500 °K, it would be necessary to heat diamond to a temperature of 69,600 °K before more than a few electrons could gain sufficient energy to cross the gap. The semiconductor (Fig. 3.2b) also has a full valence band but the separating energy gap is much smaller, though expressed as a temperature it still corresponds to about 8,000 °K. The energy band pictures of Fig. 3.2 apply strictly only at a temperature of 0 °K. At any other temperature some electrons in the valence band will have enough energy to be excited into conduction band energy states, but their number would be very small unless the gap were very narrow or the temperature very high. In the case of the conductor (Fig. 3.2c and d), either one of the bands is only partially occupied or separated valence and conduction bands do not exist; the bands of allowed states overlap and at least one of them is only partially filled.

3.2. Free electrons and holes

The lattice structure of the common semiconductors is often the same as that of diamond and is therefore called the diamond type lattice. (Refs. 1 and 2 give details of lattice structures.) This lattice is shown in Fig. 3.3 and an ex-

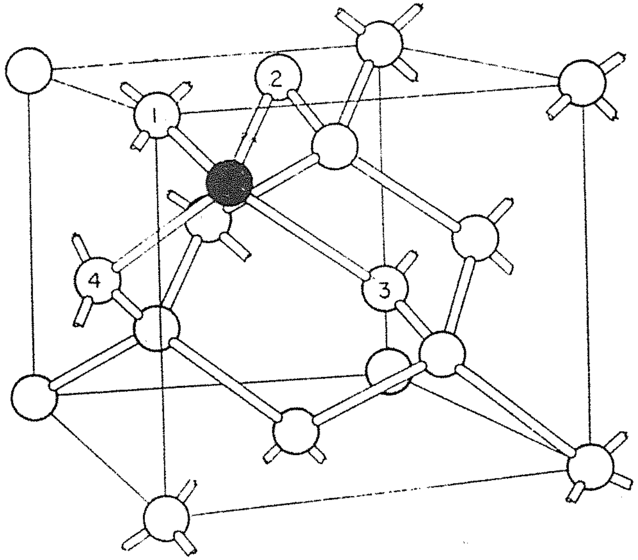


Fig. 3.3 Diamond lattice. Each atom is surrounded tetrahedrally by four others. (From Shockley, W. (1950) *Electrons and Holes in Semiconductors*. © D. Van Nostrand, Princeton, N.J.)

amination shows that each atom is surrounded tetrahedrally by four others. Each diamond atom has four electrons in the outer shell and each of these four electrons forms a covalent bond with an electron from an adjoining atom, as illustrated in a two-dimensional schematic representation in Fig. 3.4 In this idealized picture, all the electrons are firmly bound in the lattice but i either thermal energy is supplied to the crystal or light (photons) is incident or the crystal, the covalent bonds can be broken and the bound electrons re leased. They then move into the conduction band leaving behind vacancies called *holes*, in the valence band. These conduction band electrons are known as *free electrons*. It follows from this description that associated with the creation of a free electron is the necessary creation of an electron vacancy in

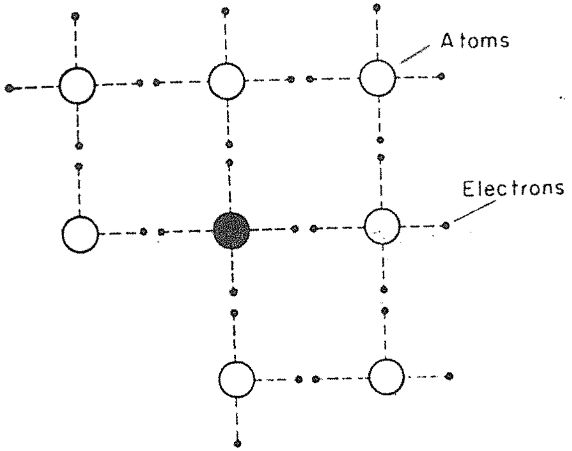


Fig. 3.4 A two-dimensional representation of the diamond type lattice

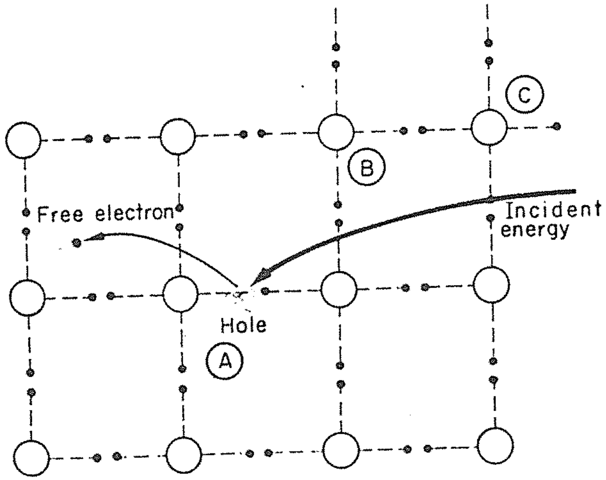


Fig. 3.5 Creation of a hole-electron pair by incident energy

our previously full structure. Figure 3.5 shows the creation of a free electron-hole pair within the lattice array. The hole, which is an unsaturated bond in the lattice, simulates a positive charge and is also free to move. The concept of hole movement is best understood by imagining a vacancy at (A) initially in

Fig. 3.5. An electron from (B) then fills the vacancy at (A), leaving a hole at (B) and the process is repeated by an electron from (C), etc. Thus the hole appears to move about in the crystal. The whole process described in this section, that of creating a free electron in the conduction band and thereby a hole in the valence band, is known as *electron-hole pair generation.*

3.3. Types of semiconductors

Semiconductors are classified into the broad categories indicated in Fig. 3.6. They may be either intrinsic or extrinsic, and the latter category is further subdivided into *n*-type and *p*-type semiconductors. The meanings of all these terms are explained in the next few sections.

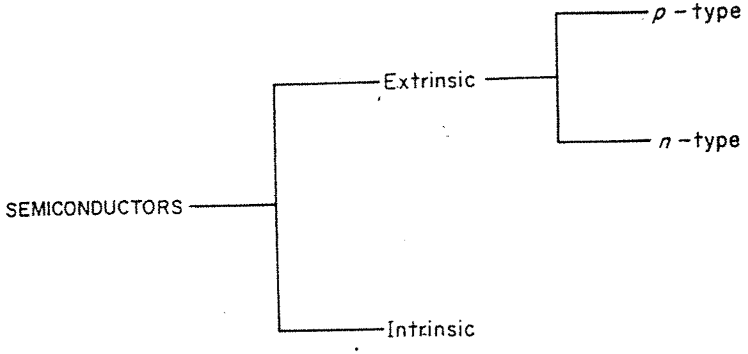


Fig. 3.6 Broad classification of semiconductors

3.3.1. The intrinsic semiconductor

A semiconductor crystal with equal numbers of electrons and holes is known as an *intrinsic* semiconductor. A very pure crystal without any im-purities and with only electron-hole pair generation producing free electrons and holes fulfils this condition. More realistically, a semiconductor with a very small impurity content and with a large number of thermally generatec electron-hole pairs is also known as intrinsic.

The addition of certain types of impurity atoms to the perfect crystallin structure we have so far considered, can produce either a large excess of fre electrons or a large excess of holes, and the properties of the semiconducto can be changed very considerably. If, however, two types of impurities ar present in the crystal, such that the contribution of free electrons due to on type is equally balanced by the contribution of holes due to the other, we hav a condition known as complete *compensation* and the semiconductor is agai considered to be intrinsic.

3.3.2. The *n*-type extrinsic semiconductor

The two best-known semiconducting elements, germanium and silicon crystallize in the diamond type lattice structure and, as described earlier, eac atom has four electrons in its outer shell and these form covalent bonds, one with each of the four nearest atoms. If an impurity atom of an element, such as arsenic or phosphorus, with five electrons in its outer shell is introduced into this ordered array, then four of its electrons form bonds with germanium atoms but the fifth remains only very loosely bound to its parent atom and moves around it with a large orbit. Figure 3.7 illustrates this. The fifth electron may very easily be separated from the impurity atom and moved into the conduction band as a free electron, leaving a net positive charge localized in the lattice. This separation of the fifth electron from its parent impurity atom is known as *ionization*. Any type of impurity which contributes an extra electron to the lattice is known as a *donor impurity*. A material with a majority of donor impurities and thus with an excess of free *negative* charges is known as an *n*-type semiconductor. The impurity atoms themselves are called donors and are said to be ionized when they have given up an electron.

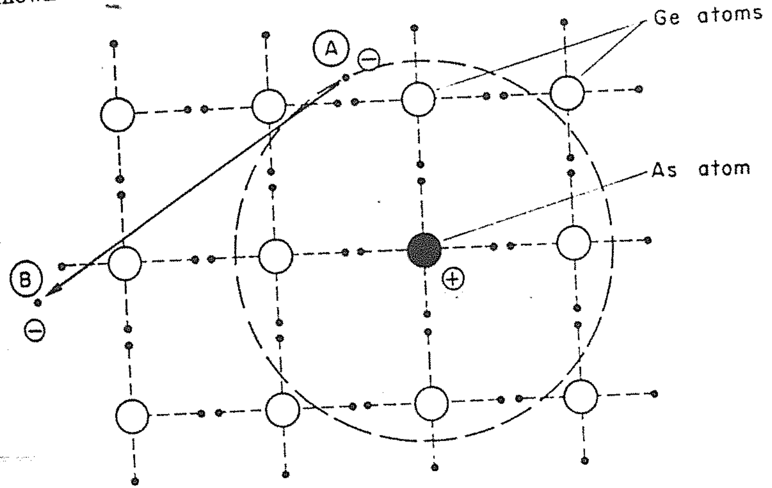


Fig. 3.7 Arsenic impurity atom in Ge crystal lattice. Loosely bound electron initially at A can become ionized and move to B

It is instructive to consider next an energy level diagram for an *n*-type semiconductor. The energy states corresponding to the electrons loosely bound to donor atoms lie just below the bottom of the conduction band level and in most cases an energy input of only ≑ 0·01 eV is sufficient to ionize the electrons so that they move into the conduction band leaving behind positively charged but immobile donor atoms in the lattice. The electrons in the conduction band are free to move while the positive charges are not. This contrasts with the creation of electron-hole pairs where the creation of a free electron in the conduction band is necessarily accompanied by the creation of a mobile hole in the valence band. The crystal remains electrically neutral in both cases, but flow in the case of the *n*-type semiconductor is primarily due to the movement of free electrons while in the intrinsic material it is due to the movement of both holes and electrons. The diagram of Fig. 3.8a illustrates the donor impurity level in the *n*-type crystal.

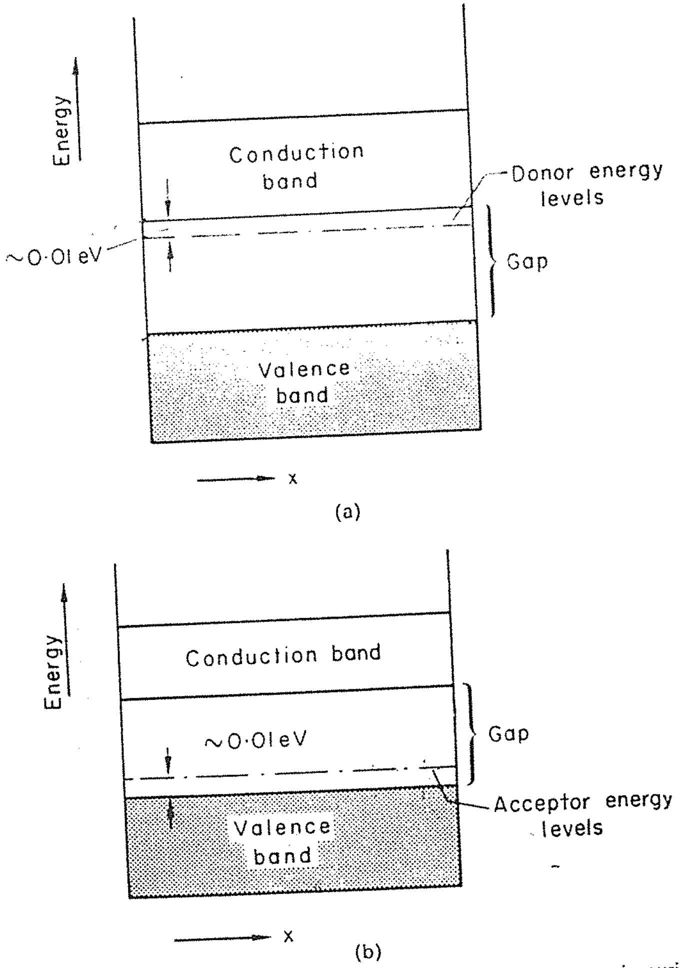


Fig. 3.8 (a) Donor impurity energy levels. (b) Acceptor impurity energy levels

3.3.3. The *p*-type extrinsic semiconductor

If impurity atoms with three valence electrons are introduced into the diamond type lattice, a *p*-type semiconductor is formed. Examples of elements with three electrons in the outer shell are boron, gallium, aluminium and indium. The three electrons of the impurity atom form covalent bonds with three of the electrons of the original array, and a space is left unoccupied. This absence of an electron, or hole, can be considered able to move in the same manner described earlier in section 3.2., i.e. an electron from a germanium atom is accepted by, say, an indium impurity atom, giving a negative local charge while the hole wanders about in the lattice. Figure 3.9 illustrates this process. The type of impurity which produces holes is called an *acceptor impurity*, and a material with an excess of acceptor impurities is known as a *p*-type semiconductor since the free charges are positive. The energy level diagram for a *p*-type semiconductor is shown in Fig. 3.8b. In this case the acceptor energy levels lie very close to the top of the valence band and electrons move from the valence band to fill acceptor energy levels, leaving behind them holes in the valence band. These holes are mobile while the negatively charged acceptor levels are bound to the lattice.

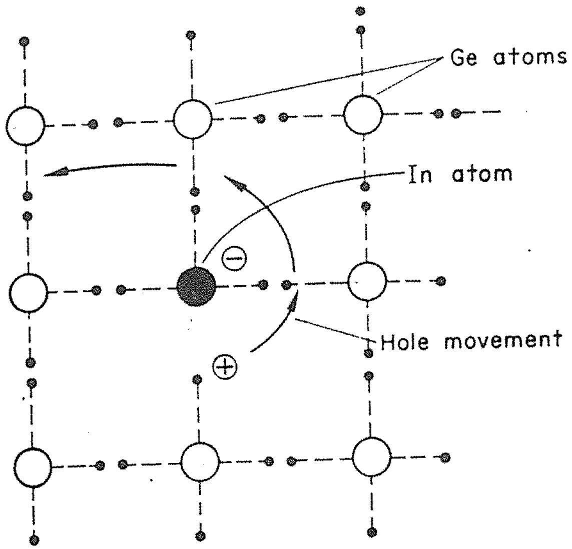


Fig. 3.9. Indium impurity atom in Ge crystal lattice. Hole associated with the In atom can move away

3.4. *Compensation*

In practice, semiconductors contain simultaneously both electrons and holes, the relative concentrations of which depend on the amounts of various intentional and unintentional impurities. Normally if donor impurities predominate the material is *n*-type, if acceptors predominate it is *p*-type. The effective density of the predominating impurity can be taken, on a reasonable approximation, as its total density less the density of impurity of the opposite type. The remaining impurity density is then described as compensated.

Summary

There are three types of semiconductor materials: (1) intrinsic material, in which the number of free electrons equals the number of holes; (2) *n*-type, in which free electrons predominate; and (3) *p*-type, in which holes predominate. Table 3.1 shows the properties of extrinsic germanium and silicon and lists some of the more common impurities.

Table 3.1 Extrinsic germanium and silicon

	n-Type or excess semiconductor	*p*-Type or defect semiconductor
	conduction by electrons	conduction by holes
Energy band with mobile carriers	Conduction band	Valence band
Sign of carrier	−ve	+ve
Valency of impurity atoms	5	3
General name for impurity	Donor	Acceptor
Common impurities	Phosphorus (P) Arsenic (As) Antimony (Sb)	Boron (B) Aluminum (Al) Indium (In)

3.5. *Semiconductor statistics*

In order to understand the electrical properties of materials, we need to know the number density of free charge carriers within the material as a function of their energy. To determine this density, we need to know two factors: (1) the density of energy states *available* in any particular band of energies. This density is designated $S(E)$ and is derived from quantum theory; (2) the probability that a particular state at energy E will actually be filled. The product of these two factors gives the number density of electrons which we require.

3.5.1. The density of states function, $S(E)$

The density of states is defined as $S(E)$ where $dS(E)$ is the number of states in an energy element dE located at E. From considerations based on quantum theory (Ref. 3), it can be shown that for electrons in the lower half of the band at least, $S(E)$ is given by:

$$S(E) = 4\pi(2m_e)^{3/2}h^{-3}(E - E_0)^{1/2}/\text{unit volume} \qquad (1)$$

where E_0 is the energy at the bottom of the band.

For electrons in the conduction band of a semiconductor

$$S(E) = 4\pi(2m_e)^{3/2}h^{-3}(E - E_c)^{1/2}$$

$(E - E_c)$ replaces $(E - E_0)$ in eqn. 1 because the electron with the lowest possible energy which can move from the valence band into the conduction band has an energy of at least E_c, the energy at the bottom of the conduction band. In this case $S(E) = 0$ at $E = E_c$.

For holes in the valence band we may consider $S(E)$ to increase as E falls below E_v, the energy at the top of the valence band, and $(E_v - E)$ replaces $(E - E_0)$ in eqn. 1. Hence

$$S(E) = 4\pi(2m_h)^{3/2}h^{-3}(E_v - E)^{1/2}$$

for states near the top of the valence band.

In all expressions for $S(E)$ the appropriate effective mass† has been used.

To gain some idea of the magnitudes involved, we find the number of available energy states/unit volume in the energy range 0–1 eV above Ec.

$$S(E)_{0-1\,eV} = \int_0^{1\,eV} \mathrm{d}S(E)\,\mathrm{d}E$$

$$= 4\pi\frac{(2m_e)^{3/2}}{h^3}\left(\tfrac{2}{3}E^{3/2}\right)/\text{unit volume}$$

$$= 4{\cdot}6 \times 10^{27}/\text{m}^3$$

and in a 1 mm cube there are $4{\cdot}6 \times 10^{18}$ energy states with energy less than 1 eV.

3.5.2. Fermi–Dirac probability function

We have implied earlier that not all the energy states specified by $S(E)$ are actually occupied by electrons. We may expect, intuitively, that there will be relatively few electrons in the higher energy states and that the numbers will depend upon the temperature of the crystal. In fact, statistical analysis shows that the probability of finding an electron at energy E is given by the expression (Ref. 4).

[illegible]

$$f(E) = \frac{1}{1 + \exp\left(\dfrac{E - \zeta}{kT}\right)} \qquad (2)$$

[illegible]

where ζ is a reference energy, known as the Fermi energy, to which we will give a meaning very shortly.

The function is obtained by considering the statistics of particles, indistinguishable from each other but with the restriction imposed by the Pauli exclusion principle that not more than one particle (here an electron) may exist in any particular quantum state. This differs from the Maxwell–Boltzmann function used in Chapter 2, in that the Pauli restriction does not apply in the M–B case and any number of particles may exist in any energy level.

---

† The concept of effective mass is discussed in Appendix A.

The F–D function is plotted in Fig. 3.10 for two values of temperature, $T = 0\,°$K and a higher temperature. It can be seen that for $T = 0\,°$K, $f(E) = 0$ for $E > \zeta$ and $f(E) = 1$ for $E < \zeta$. This defines ζ as the energy limit below which all electrons must be accommodated at $T = 0\,°$K. It should also be noted that $f(E) = \frac{1}{2}$ when $E = \zeta$ for all values of T greater than 0 °K, and this fact provides an alternative definition of ζ as the energy where the probability of finding a electron occupying a state is exactly 0·5.

$T = 0$

$T \sim \frac{\zeta}{10K}$

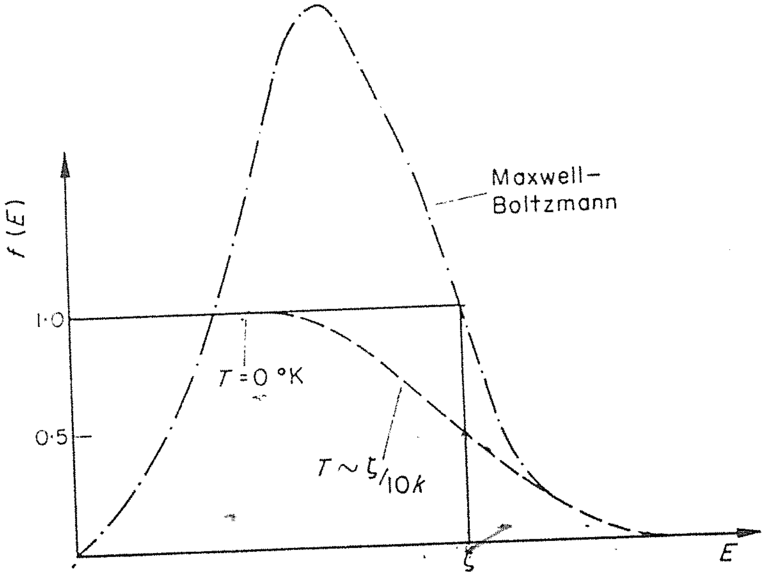


Fig. 3.10 The Fermi–Dirac probability function at $T = 0$ °K and at a higher temperature. The Maxwell–Boltzmann function is shown schematically for comparison

At room temperature (300 °K), $kT = 0{\cdot}026$ eV and for $E - \zeta > 3kT$, we have $\exp[(E - \zeta)/kT] \gg 1$ and the approximation

$$f(E) \doteqdot \frac{1}{\exp\left(\dfrac{E - \zeta}{kT}\right)} \qquad (3).$$

This approximation reduces the F–D function to the M–B function—an approximation which much simplifies the algebra of semiconductor statistics and is perfectly valid in all but some special cases in our study of semiconductors

3.5.3. The energy density of free electrons

Earlier in this section we defined the density of states in an energy range dE. So the number of electrons in an energy range dE is

$$\mathrm{d}N(E)\,\mathrm{d}E = S(E).f(E)\,\mathrm{d}E$$

therefore,

$$N(E) = \frac{4\pi(2m_e)^{3/2}}{h^3}\int \frac{E^{1/2}}{1+\exp\left(\dfrac{E-\zeta}{kT}\right)}\,\mathrm{d}E$$

assuming $E_0 = 0$. This number density is shown plotted as a function of E in Fig. 3.11 for two temperatures. At 0 °K it again shows all the energy states below ζ to be filled and all those above ζ to be empty, but at $T > 0$ °K, there is a definite number of electrons with energies $> \zeta$.

We shall use, in several parts of this book, the ideas outlined in this section to calculate the number densities of free charge carriers and their energy levels, with respect to the Fermi level.

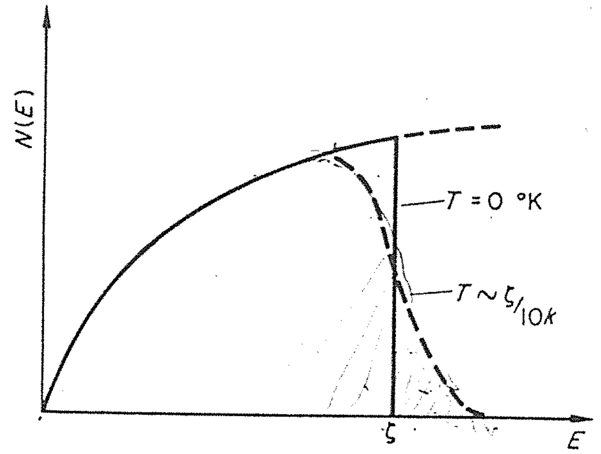


Fig. 3.11 The density of free electrons at $T = 0$ °K and at a higher temperature, $T \sim \zeta/10\mathrm{k}$

3.5.4. Intrinsic semiconductors

We consider first the relatively simple case of the intrinsic semiconductor in which the numbers of holes (p) and electrons (n) are equal:

$$n = p \tag{4}$$

Figure 3.12a shows the energy level diagram again, and Fig. 3.12b shows the density of energy states $S(E)$ which are available in the conduction and valence bands, plotted as a function of energy. The probability of these states being occupied by an electron is given by the Fermi–Dirac probability function $f(E)$, shown again in Fig. 3.12c. The total number of energy states occupied by electrons in the energy range E to $E + \mathrm{d}E$ is, as before, the density of available energy states times the probability of the state being occupied,

$$\mathrm{d}N(E)\,\mathrm{d}E = S(E).f(E)\,\mathrm{d}E$$

The density of positive holes in the valence band can be considered in a similar manner. A hole is an energy state in the valence band which is not filled by an electron although, from energy considerations, this is possible. Hence, the probability of a hole at an energy level less than E_v, the top of the valence band, is $1 - f(E)$. Therefore the number of holes in the energy range E to $E + \mathrm{d}E$ is

$$\mathrm{d}P(E)\,\mathrm{d}E = S(E)(1 - f(E))\,\mathrm{d}E \tag{5}$$

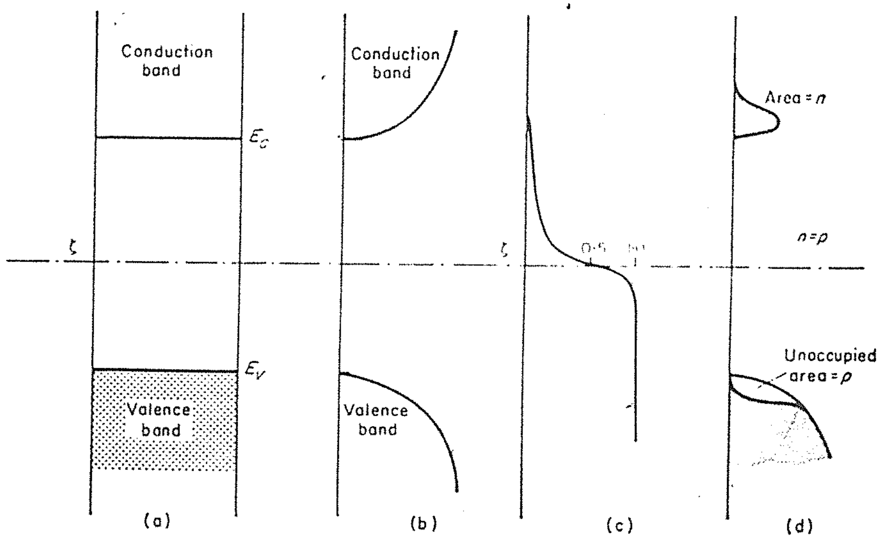


Fig. 3.12 The intrinsic semiconductor without impurities: (a) energy level diagram; (b) density of energy states $S(E)$; (c) the Fermi function; (d) density of electrons and holes

Using the approximate expression of eqn. 3, we write

$$f(E) \doteq \frac{1}{\exp\left(\dfrac{E-\zeta}{kT}\right)} = \exp\left(\frac{\zeta - E}{kT}\right)$$

For holes we have the probability $1 - f(E)$ giving

$$1 - f(E) \simeq 1 - \frac{1}{1+\exp\left(\dfrac{E-\zeta}{kT}\right)} \tag{6}$$

Here again we assume $\zeta - E >$ say $3kT$, therefore the term $\exp(E-\zeta)/kT \ll 1$. Therefore, expanding eqn. 6 using the binomial expansion we have

$$1 - f(E) \doteq \exp\left(\frac{E-\zeta}{kT}\right) \tag{7}$$

We calculate the total number of electrons in the conduction band at any temperature.

Using

$$S(E) = 4\pi(2m_e)^{3/2}h^{-3}(E - E_c)^{1/2}$$

therefore

$$dN(E)\,dE = 4\pi(2m_e)^{3/2}h^{-3}(E - E_c)^{1/2}\exp\left(\frac{\zeta - E}{kT}\right)dE \qquad (8)$$

and the total number of electrons in the conduction band is

$$n = \int_{E_c}^{\infty} dN(E)\,dE = 4\pi(2m_e)^{3/2}h^{-3}\int_{E_c}^{\infty}(E - E_c)^{1/2}\exp\left(\frac{\zeta - E}{kT}\right)dE$$

$$= 4\pi(2m_e)^{3/2}h^{-3}\exp\left(\frac{\zeta - E_c}{kT}\right)\int_{E_c}^{\infty}(E - E_c)^{1/2}\exp-\left(\frac{E - E_c}{kT}\right)dE \qquad (9)$$

Using a standard integral

$$n = 2\left(\frac{2\pi m_e kT}{h^2}\right)^{3/2}\exp-\left(\frac{E_c - \zeta}{kT}\right) \qquad (10)$$

This can be written as

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

where

$$N_c = 2\left(\frac{2\pi m_e kT}{h^2}\right)^{3/2}$$

and N_c may be regarded as a fictitious number of states, all lying at the energy E_c and giving the same answer as integration over the whole band of energies.

Similarly for holes in the valence band

$$p = 2\left(\frac{2\pi m_h kT}{h^2}\right)^{3/2}\exp-\left(\frac{\zeta - E_v}{kT}\right) \qquad (11)$$

The product pn is obtained by multiplying eqns. 10 and 11 to give

$$p \times n = 4\left(\frac{2\pi kT}{h^2}\right)^3(m_e m_h)^{3/2}\exp-\left(\frac{E_g}{kT}\right) \qquad (12)$$

where $E_g = E_c - E_v$ and is called the forbidden energy gap for the semiconductor. Numerically,

$$pn = 2{\cdot}33 \times 10^{43}T^3\left(\frac{m_e m_h}{h^2}\right)^{3/2}\exp\left(-\frac{E_g}{kT}\right)$$

where m is the free-space mass of the electron.

The pn product depends on the energy gap and not on the position of the Fermi level, therefore it is the same whether the semiconductor is intrinsic, p-type or n-type.

For an intrinsic semiconductor we use p_i and n_i to denote the hole and electron densities and use eqn. 1 to give

$$pn = p_i . n_i = n_i^2 = p_i^2 \qquad (13)$$

and hence

$$n_i = 2\left(\frac{2\pi kT}{h^2}\right)^{3/2}.(m_e m_h)^{3/4}\exp\left(-\frac{E_g}{2kT}\right) \qquad (14)$$

This equation gives the total number density of free electrons in an intrinsic semiconductor, and for Ge and Si at room temperature of 300 °K the values of n_i are $2.5.10^{13}/\text{cm}^3$ and $1.6.10^{10}/\text{cm}^3$ respectively.

The position of the Fermi level may be determined by equating the r.h.s's of eqns. 10 and 11 which give

$$m_e^{3/2}\exp-\left(\frac{E_c - \zeta}{kT}\right) = m_h^{3/2}\exp-\left(\frac{\zeta - E_v}{kT}\right) \qquad (15)$$

$$\zeta = \tfrac{1}{2}(E_c + E_v) + \tfrac{3}{4}kT\ln\left(\frac{m_h}{m_e}\right) \qquad (16)$$

If m_e and m_h are the same, then E_F lies in the middle of the gap at $(E_c + E_v)/2$, even at $T > 0$ °K.

3.5.5. Extrinsic semiconductors

The most general representation of the energy levels in a doped semiconductor containing both group 3 and group 5 impurities is that of Fig. 3.13 where both donor and acceptor levels are shown in addition to the conduction and valence bands. At low temperatures electrons are at low-energy levels and almost all of them are located in the valence band or in the donor level. This state of affairs is altered significantly as the temperature is raised. The first change is that electrons in the valence band are excited into the acceptor level and simultaneously electrons from the donor level move into the conduction band. This situation is illustrated by the lines in Fig. 3.13. At higher temperatures, more and more electrons can go all the way from the valence band to the conduction band. An electron leaving the valence band for an acceptor level leaves behind a *free hole* and becomes a bound electron. Similarly, electrons leaving the donor level leave behind *bound holes*. The direct movement of an electron from the valence band into the conduction band leaves a free hole in the valence band and introduces a free electron in the conduction band, as expected from previous consideration, but this process is relatively improbable because of the width of the gap. All these processes leading to the production of free and bound holes and electrons are also illustrated in Fig. 3.13.

The numbers of free electrons and free holes available determine the electrical characteristics of the semiconductor material, and the next step is to calculate these numbers.

Let N_a be the density of acceptor atoms
N_d be the density of donor atoms
p be the density of holes in the material
n be the density of electrons in the material
n_d be the number of unionized donor atoms, which retain their fifth electron
p_a be the number of acceptor atoms which do not trap a fourth electron.

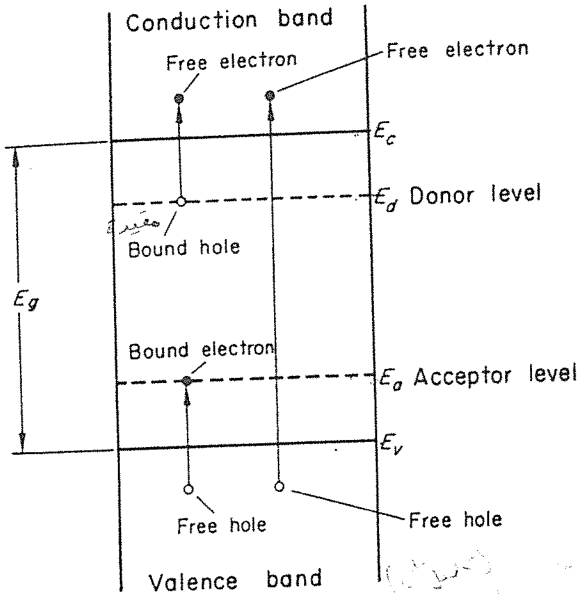


Fig. 3.13 The energy levels in an extrinsic semiconductor and the mechanisms of charge carrier generation

Now we know that any donor atom which ionizes becomes positively charged, and any acceptor atom which gains a fourth electron becomes negatively charged. However, the crystal as a whole cannot be charged up and must remain electrically neutral and therefore we may equate the total number of positive charges to the total of negative charges:

donor acceptor

$$p + (N_d' - n_d) = n + (N_a' - p_a) \qquad (17)$$

We have previously shown that n, the number of electrons in the conduction band, is given by

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

and p, the number of holes in the valence band is

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

Now N_d donor atoms exist, therefore there are N_d energy states at E_d and the number of electrons trapped at donor levels is N_d times the probability of an electron having the energy E_d.

$$n_d = N_d . f(E_d) = \frac{N_d}{1 + \exp\left(\frac{E_d - \zeta}{kT}\right)} \qquad (18)$$

and if $E_d - \zeta \gg kT$

$$n_d = N_d \exp -\left(\frac{E_d - \zeta}{kT}\right) \qquad (19)$$

This approximation is valid at most normal temperatures and doping levels, but invalid at very high doping levels and very low temperatures.

For holes, we have p_a as the number of unionized acceptor atoms and $N_a - p_a$ as the number of electrons at level E_a.

Then,

$$N_a - p_a = N_a . f(E_a)$$

or

$$p_a = N_a\{1 - f(E_a)\} \qquad (20)$$

If $E_a - \zeta \gg kT$, we may use reasoning similar to section 3.5.4 and obtain

$$p_a = N_a \exp -\left(\frac{\zeta - E_a}{kT}\right) \qquad (21)$$

This expression again is invalid when the doping level is very high and when the temperature is very low.

3.5.6 Position of the Fermi level

In an extrinsic semiconductor the Fermi level is no longer approximately at the centre of the energy gap as in intrinsic semiconductors, but moves nearer to either the conduction band or valence band levels depending on the nature of the impurity doping. Let us consider a semiconductor in which $n > p$, then

$$N_c \exp -\left(\frac{E_c - \zeta}{kT}\right) > N_v \exp -\left(\frac{\zeta - E_v}{kT}\right)$$

but

$$N_c \doteqdot N_v$$

therefore

$$(E_c - \zeta) < (\zeta - E_v) \qquad (22)$$

i.e. ζ has moved from near the middle of the gap to be nearer the bottom of the conduction band. In the same way, when $p > n$, the level of ζ shifts to be nearer the valence band.

3.5.6.1. *n-Type semiconductors*

If $N_d \gg N_a$, we can neglect the term $N_a - p_a$ in eqn. 17. If we also assume complete ionization of the donor atoms, $n_d = 0$ and assume that holes in the valence band are few compared with electrons in the conduction band, we may write eqn. 17 as

$$n \doteqdot N_d \qquad (23)$$

and the semiconductor is called an n-type semiconductor.

The Fermi level is determined by using

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

and

$$\zeta = E_c - kT \ln\left(\frac{N_c}{N_d}\right) \qquad (24)$$

3.5.6.2 *p-Type semiconductor*

Here we assume $N_a \gg N_d$ and neglect $N_d - n_d$. All the acceptor atoms are assumed to have acquired a fourth electron. Hence $p_a = 0$ and p is assumed much greater than n. Then

$$p \doteq N_a$$

and the semiconductor is called p-type, and

$$\zeta = E_v + kT \ln \left(\frac{N_v}{N_a}\right) \tag{25}$$

3.6. *The general equations for extrinsic semiconductors*

For any semiconductor, we are given E_g, E_c and E_v. The type of doping determines E_d and E_a. The doping levels then fix N_d and N_a. We then have five equations which have been stated earlier in the text and are now summarized for convenience:

$$n = N_c \exp -\left(\frac{E_c - \zeta}{kT}\right) \tag{10}$$

$$p = N_v \exp -\left(\frac{\zeta - E_v}{kT}\right) \tag{11}$$

$$n_d = N_d \exp -\left(\frac{E_d - \zeta}{kT}\right) \tag{19}$$

$$p_a = N_a \exp -\left(\frac{\zeta - E_a}{kT}\right) \tag{21}$$

$$p + (N_d - n_d) = n + (N_a - p_a) \tag{17}$$

The unknown quantities we need to know at any temperature, T, are n, p (the densities of electrons and holes), n_d (the number of electrons unionized from donor atoms) and p_a (the number of holes trapped at acceptor levels). Since there are five unknowns and five equations, a solution is possible but a general solution is very laborious and rarely necessary. Some perfectly valid simplifying conditions may be applied to ease the algebra and, indeed, to provide a clearer physical picture.

One simplifying assumption which may be used is the concept of compensation which was introduced earlier in the discussion of intrinsic materials. This can be applied to a semiconductor containing both donor and acceptor impurities by taking the difference between N_d and N_a.

If $N_d - N_a \doteq 0$, the material is considered intrinsic. If $N_d \gg N_a$, the material is n-type and the apparent donor concentration is $N_d - N_a$ and the equation involving N_a may be ignored. Hence, for n-type material we assume $N_a = 0$ and $p_a = 0$. A similar reasoning is applied to p-type materials where it is assumed $N_d = 0$ and $n_d = 0$.

Another simplification is to consider the semiconductor in various temperature ranges, rather than generally for all values of T. This enables a considerable simplification of the expressions in eqns. 10, 11, 19 and 21.

Finally, approximations are possible based on the level of doping and on the width of the forbidden energy gap. Some typical cases are now considered using these approximations.

3.6.1. Low temperatures

It is difficult to be precise about the temperature range because the behaviour of different semiconductors at any fixed temperature can be quite different. For example, extrinsic Ge crystals behave like intrinsic material at a temperature of about 100 °C, while Si continues to exhibit extrinsic semiconductor properties. The low temperature range would be at temperatures well below 0 °C for Ge and Si.

We consider an n-type material in the following examples, but the considerations apply equally to p-type material. At this low temperature, there is a large number of unionized donor atoms, the $n.p$ product is small and, in the case of n-type material, $n \gg p$. Therefore the neutrality equation is written as

$$N_d - n_d = n \quad = N_d - N_d \exp -\left(\frac{E_d - \xi}{kT}\right) \tag{26}$$

where N_d is the compensated density of impurity atoms.

Using eqns. 10 and 19

$$\frac{(19)}{(10)}$$

$$26 \quad n_d \doteq n \frac{N_d}{N_c} \exp -\left(\frac{E_d - E_c}{kT}\right)$$

substituting in eqn. 20

$$n = \frac{N_c N_d}{N_c + N_d \exp\left(\frac{E_c - E_d}{kT}\right)} \tag{27}$$

and, from eqn. 10

$$\zeta = E_c + kT \ln \frac{N_d}{N_c + N_d \exp\left(\frac{E_c - E_d}{kT}\right)} \tag{28}$$

Any increase in temperature causes a rapid increase in the number of ionized donor atoms and the conductivity increases rapidly with temperature.

3.6.2. Moderate temperatures

This implies temperatures around room temperature for Ge and Si. We can still assume that there are very few thermally produced holes, and $n \gg p$. In this temperature range it may also be assumed that *almost all the donors are ionized* even if the number of thermally produced electron-hole pairs is small. In eqn. 27 we have $N_c \gg n_d$, therefore,

$$N_c \gg N_d \exp\left(\frac{E_c - E_d}{kT}\right)$$

and we may write

$$n \doteqdot \frac{N_c N_d}{N_c} = N_d \tag{29}$$

and

$$\zeta = E_c + kT \ln \frac{n}{N_c} = E_c + kT \ln \frac{N_d}{N_c} \tag{30}$$

Most semiconductor devices are normally operated in this temperature range.

3.6.3. High temperatures

At still higher temperatures we begin to have a contribution from electron-hole pair generation, so that the number of holes may no longer be considered negligible, i.e. $p \neq 0$. Again, all the donor atoms are certainly ionized, therefore $n_d = 0$. Hence, eqn. 17 becomes

$$N_d + p = n$$

but

$$p.n = n_i^2$$

therefore

$$N_d + \frac{n_i^2}{n} = n \quad \text{or} \quad n = \frac{N_d \pm (N_d^2 + 4n_i^2)^{1/2}}{2} \tag{31}$$

The negative sign is ignored since n cannot be negative and

$$n = \frac{N_d + (N_d^2 + 4n_i^2)^{1/2}}{2} \tag{32}$$

Using $n_i^2 = N_c^2 \exp -(E_g/kT)$ we may obtain n and ζ.

$n_i = N_c$ [illegible] $-($[illegible]$)$

3.6.4. Very high temperatures

At very high temperatures, n_i increases rapidly so that

$$n_i \gg N_d$$

and eqn. 17 gives $n = n_i = p$ and the material becomes intrinsic. The changes in the semiconductor as the temperature increases are illustrated in Fig. 3.14. The number of carriers, the shift in the Fermi level and the resistivity are all shown as functions of temperature and the appropriate temperature ranges are marked. The moderate temperature range is the most important because most devices operate in this range. For an n-type material, in this range, $n = N_d$ and for Ge and Si there are somewhere between 10^{14} to 10^{17} impurity atoms/cm^3 and almost all these are ionized. The contribution from electron-hole pairs is usually negligible.

An interesting comparison of semiconductor materials may be made with reference to Fig. 3.14. Si transistors may be satisfactorily operated at a temperature of 100 °C because Si has an energy gap of 1·1 eV and very few electron-hole pairs exist at this temperature. Ge, on the other hand, has an energy gap of only 0·72 eV and at a temperature of 100 °C its behaviour would correspond to the high temperature range or even the intrinsic range. At another extreme, InSb devices would have to operate at liquid nitrogen temperature to remain in the moderate temperature range, and at room temperature the material is intrinsic. Si and Ge devices at liquid nitrogen temperature are operating in the low temperature range and transistor action is unsatisfactory. (Refs. 5 and 6 discuss semiconductor statistics in greater detail.)

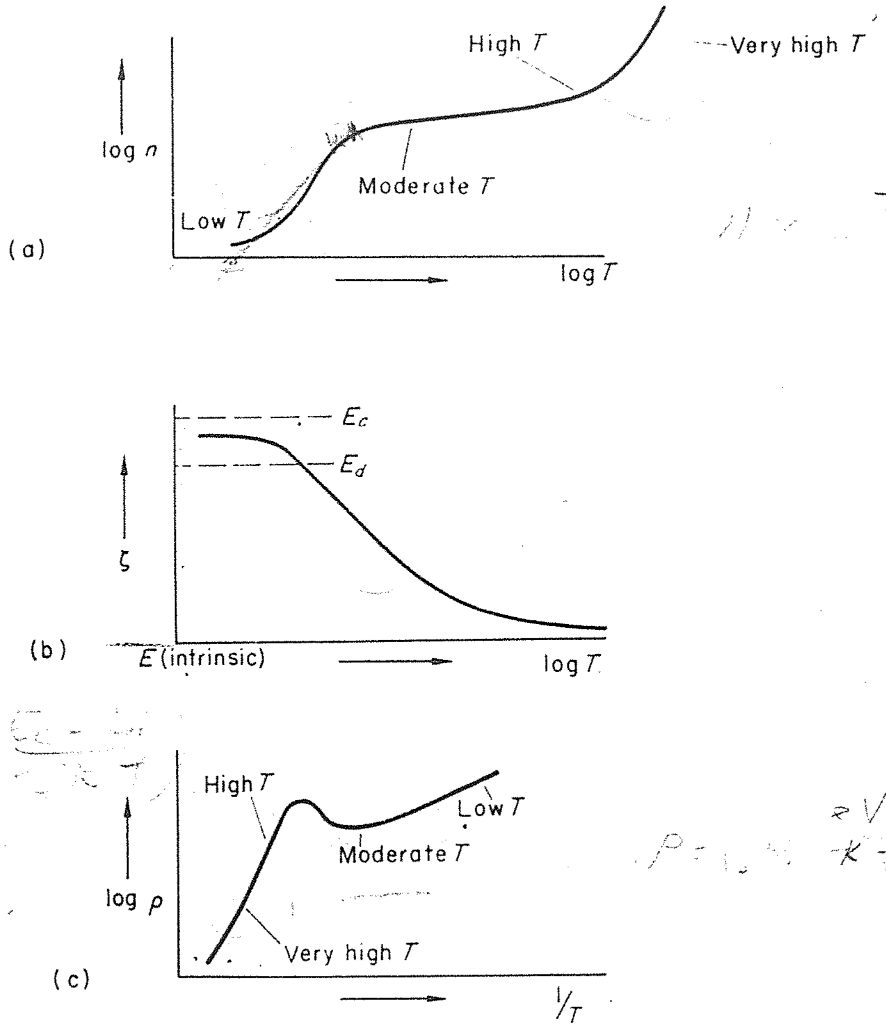


Fig. 3.14 (a) The change in carrier concentration with increasing temperature. (b) The shift in the Fermi level. (c) The change in resistivity

3.7. *Movement of charge carriers in a semiconductor*

In this section we describe the movement of carriers in a semiconductor both with and without an applied electric field. If the movement is under the

اتيار

influence of a field, it is known as *drift*, and when there is no applied field, the movement is by *diffusion.*

3.7.1. Drift mobility and conductivity in a semiconductor

Free electrons in vacuum accelerate continuously under the influence of an electric field, but/in a semiconductor free electrons move with an *average velocity* which is proportional to the applied field provided the temperature is constant.

The detailed motion of electrons under the influence of both an external applied field and the surrounding lattice is very complicated. An electron accelerates under the influence of the field for a time known as the mean free time and traverses a mean free path. It then undergoes a collision such that the velocity in the direction of the field is destroyed. Immediately after the collision, the electron starts to accelerate again until it makes another collision after traversing another mean free path. The sequence of events repeats itself and the electron gradually drifts in the direction of the field in between collisions. It must be noted that the velocity of the electron is not in the direction of the field at all times.

The experimental evidence is that carriers move with a velocity proportional to the applied electric field, and the constant of proportionality is known as the *mobility*. Thus

قابلية الحركة

$$v_n = -\mu_n E \tag{33a}$$

$$v_p = +\mu_p E \tag{33b}$$

where μ_n and μ_p are the electron and hole mobilities respectively, with dimensions of (length)²/volt sec. v_n and v_p are the average drift velocities; E is the applied electric field.

In general, the hole and electron mobilities are different; the former is usually much lower in the case of germanium and silicon. It can be shown that mobility is directly proportional to the mean free time of the carrier between collisions (t_n and t_p for electrons and holes respectively) and inversely proportional to the effective mass. Thus,

$$\mu_n = \frac{et_n}{m_e} \tag{34a}$$

and

$$\mu_p = \frac{et_p}{m_h} \tag{34b}$$

Numerical values of mobilities at 300 °K are (Ref. 7):

germanium	$\mu_n \sim 3{,}900$ cm²/volt sec
	$\mu_p \sim 1{,}900$ cm²/volt sec
silicon	$\mu_n \sim 1{,}500$ cm²/volt sec
	$\mu_p \sim 500$ cm²/volt sec

Making the simplest possible assumption that all electrons move with the

average velocity v_n and all holes with the velocity v_p, the current densities due to the motion of the carriers are

$$J_n = -env_n = +en\mu_n E \tag{35a}$$

and

$$J_p = epv_p = ep\mu_p E \tag{35b}$$

where e is the electronic charge. The current density with both carriers present is

$$J = J_n + J_p = e(pv_p - nv_n)$$

and using eqns. 33a and b we have

$$J = eE(n\mu_n + p\mu_p) \tag{36}$$

Now conductivity is defined, as usual, by

[illegible]

$$\sigma = J/E \tag{37}$$

hence

$$\sigma = e(n\mu_n + p\mu_p) \tag{38}$$

Thus conductivity is a function both of mobilities and of carrier densities.

We now consider the variation of conductivity with temperature for an intrinsic semiconductor.

Putting $m_e = m_h = m$ in eqn. 14 and using eqn. 38,

we have

$$\sigma = N_c e(\mu_n + \mu_p) \exp -\frac{E_g}{2kT} \tag{39}$$

which is of the same general form as eqn. 1.

The conductivity of an intrinsic semiconductor thus increases exponentially as the temperature is increased, a situation that may be contrasted with the behaviour of metals when they are heated.

Equation 39 has the important practical consequence that the width of the energy gap for any semiconductor can be determined by measuring the variation of conductivity with temperature. To elaborate, we have the equation

$$\ln \sigma = \ln [N_c e(\mu_n + \mu_p)] - \frac{E_g}{2kT} \tag{40}$$

and a plot of $\ln \sigma$ versus $1/T$ gives a straight line because the term $\ln [N_c e(\mu_n + \mu_p)]$ does not vary appreciably with temperature, although the mobilities are functions of temperature. The slope of the line may be used to calculate E_g.

For n- and p-type materials, the expressions derived in section 3.6 can be inserted in eqn. 38 provided due attention is paid to the choice of the appropriate temperature range.

3.7.2. Majority and minority carriers

Semiconductor crystals are usually doped in such a way that the movement of one type of carrier constitutes the bulk of the current. For an n-type semiconductor nearly all the electrons lie in the conduction band and both the number of electrons bound to donor atoms and the number of holes are small fractions of the number of free electrons in the conduction band. The move- ment of these conduction band electrons dominates the properties of the *n*-type semiconductor because they are responsible for the bulk of the current carried in the crystal and they are known as the *majority carriers*. The holes in such a crystal carry a relatively small current and are called *minority carriers*. In a *p*-type crystal the roles of holes and electrons are reversed and holes be- come the majority carriers, while the minority carriers are electrons. In some semiconductors the electron mobility may be much greater than hole mobility and these behave as *n*-type in their current-carrying properties even when they are intrinsic.

3.7.3. Diffusion in semiconductors

In addition to drift there is another mechanism of carrier flow which takes place in the absence of an electric field and plays an important part in the operation of many semiconductor devices. This carrier flow is called a *diffusion* *current*. It occurs whenever carriers are injected into, or produced in, a semi- conductor in such a way as to make the spatial distribution non-uniform, or produce localized concentrations. Under these conditions carriers diffuse away from the concentration regions in an attempt to restore the uniform carrier density all over the crystal. The situation is analogous to the diffusion of gas molecules between two interconnected vessels at different pressures.

It is found, as is usual in all physical diffusion examples, that diffusion current density is proportional to the negative gradient of the carrier density, and the constant of proportionality is known as the *diffusion constant*.

Thus the diffusion current density due to holes is given by

$$J_p = -eD_p \frac{dp}{dx} \tag{41a}$$

and due to electrons by

$$J_n = eD_n \frac{dn}{dx} \tag{41b}$$

The values of D_n and D_p for Si and Ge are given in Table 3.2. A relation is de- rived between the diffusion constant and mobility in section 3.7.8 and is known as the Einstein relation.

3.7.4. Generation and recombination of carriers

In a semiconductor, thermal agitation produces electron-hole pairs and both sorts of carrier produced are mobile within the crystal lattice. Impurity atoms, too, may be ionized and this results in the production of a mobile majority carrier together with an unionized impurity atom which is immobile. Thus, at any temperature there exists an equilibrium concentration of nega- tive and positive charge carriers. If this situation is disturbed by applying an external energy source to the semiconductor, e.g. a light source supplying photons of energy hf, then some electrons are raised to conduction band energy levels and extra carriers, known as *excess carriers*, are generated. The

Table 3.2 Some properties of germanium and silicon (Ref. 7 gives much data on semi- conductor properties)

Property	Ge	Si
Atomic no.	32	14
Rel. permittivity, ϵ_r	16	12
Energy gap, E_g	0·72	1·1 eV
n_i at 300 °K	$2·5 \times 10^{13}$	$\sim 1·6 \times 10^{10}/\text{cm}^3$
ρ intrinsic at 300 °K	47	~ 230,000 Ωcm
Mobility μ_n at 300 °K	3,900	1,500 cm²/volt sec
μ_p	1,900	500 cm²/volt sec
D_n	99	39 cm²/sec
D_p	49	13 cm²/sec

case where the number of carriers is reduced below the equilibrium value is known as *carrier depletion*.

As soon as the light source is turned off, carrier generation ceases and the excess concentration decays back to the original equilibrium condition by means of a process known as *recombination*. Ideally, this is the transition of electrons from the conduction band back into the valence band energy levels, so that they no longer remain mobile. The energy released in such a transition must equal the forbidden energy gap which implies the release of a photon or the generation of many phonons.† These are relatively improbable occurrences and if recombination were entirely dependent on their occurrence, the lifetime of the carriers would range from 0·1 to 1·0 sec. It is observed experimentally that carrier lifetimes are two orders of magnitude shorter than these predic- tions and it is assumed that some more probable recombination process must occur.

A recombination process which gives lifetimes nearer the true value arises if we assume the existence of extra energy levels lying within the energy gap. Such levels may arise from a variety of causes, including the existence of un- wanted impurities in the crystal or the presence of lattice imperfections. The probability of a hole or electron being captured and subsequently recombining at the intermediate levels is much greater than the probability of a band-to- band transition. Sites giving rise to extra energy levels are known as *re- combination centres* because they facilitate the recombination process. The probability that a localized energy level can capture an excess carrier is expressed in terms of its *capture cross-section*. The greater the probability, the greater this cross-section. If the probability of capturing an electron is greater than the probability of capturing a hole, then the recombination centre is known as an electron trap. Similarly, energy levels where the proba- bility of capturing a hole is greater than the probability of capturing an electron are known as hole traps. Any real semiconductor has a large number

---

† Phonons are quantized lattice vibrations in a crystal and have a much lower energy than photons.

of electron and hole traps. Figure 3.15 shows, in a schematic manner, the generation of carriers, band-to-band recombination and the trapping process just described.

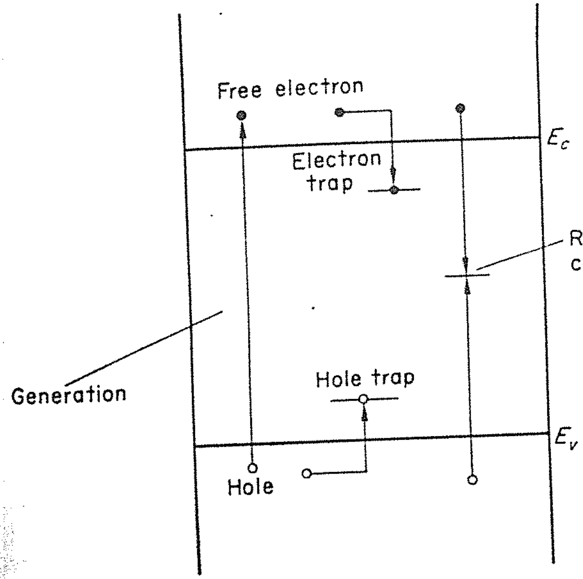


Fig. 3.15 Generation, trapping and recombination of holes and electrons

3.7.5. Carrier lifetime

We have stated earlier that whenever a semiconductor in thermal equilibrium is irradiated by an energy source, excess carriers are generated; when the source is turned off, recombination causes the carriers to decay in number to the original state. This decay process takes a finite time which is called the *carrier lifetime*.

In the extrinsic semiconductor it is usually the concentration of minority carriers which is markedly affected by the presence of an external energy source. The concentration of majority carriers is already so great that it is hardly altered. The total numbers of electrons and holes created are, of course, always the same and the crystal remains neutral.

In an n-type crystal the minority carriers are holes and the rate of change of excess hole density will be

$$\nearrow \quad \frac{dp}{dt} = \overset{\text{generation}}{g} - r\text{ - recombination} \qquad (42)$$

where g is the rate of generation of holes, say, as a result of incident photons or light, and r is the rate of recombination. If dt/τ is the probability that a hole

will combine in a time dt, then the rate at which holes combine is p/τ and

$$\frac{dp}{dt} = g - \frac{p}{\tau} \qquad (43)$$

If the light is now turned off, $g = 0$ and

$$\frac{dp}{dt} + \frac{p}{\tau} = 0$$

and

$$p = p_0 \exp\left(-\frac{t}{\tau}\right) \qquad (44)$$

where p_0 is the excess hole density at the end of generation time, i.e. $t = 0$ for convenience.

Equation 44 shows that carrier densities decay exponentially with time after the light is removed, with the rate of decay governed by τ, the minority carrier lifetime.

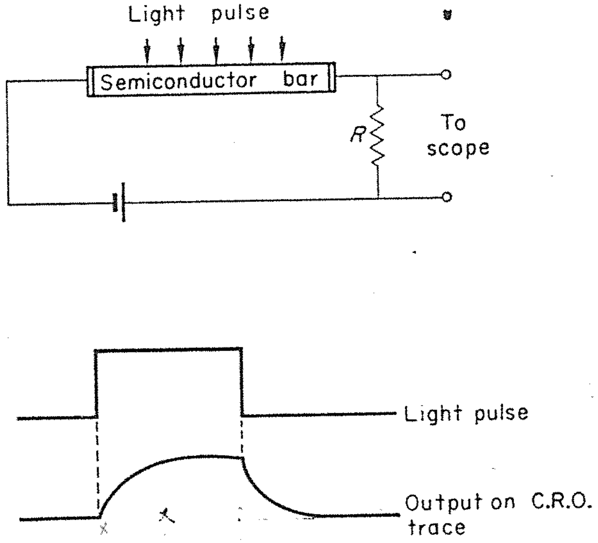


Fig. 3.16 Principle of minority carrier lifetime measurement

The lifetime may be measured experimentally using the arrangement shown in Fig. 3.16. The light source is turned on suddenly, using a fast-acting shutter, and irradiates the semiconductor. Excess minority carriers are immediately generated and their number increases until the rate of generation equals the rate of recombination. The current flow in the circuit increases according to

$$\Delta I = \text{const.}\left[1 - \exp\left(-\frac{t}{\tau}\right)\right]$$

On switching off,

$$\Delta I = \text{const.}\left[\exp\left(-\frac{t}{\tau}\right)\right]$$

The shape of the pulse in Fig. 3.16 expresses this increase and decrease, and the value of τ may be deduced from an estimate of the rise or fall times. The applied voltage must be kept small and only the centre of the semiconducting rod illuminated. This ensures that nearly all the excess minority carriers recombine before reaching the end terminal. The light source must be sufficiently weak to ensure that minority carrier concentration is much less than the majority carrier concentration. In semiconductors with very short lifetimes the light is switched on and off with fast-acting shutters such as Kerr cells.

Before leaving the subject of lifetime, it is worth while mentioning that the surfaces of semiconducting crystals contain many more imperfections than the bulk, so that recombination is far more likely at the surface. In fact, many of the carriers have such long lifetimes in the bulk of the crystal that they almost always recombine only after reaching the surface. The two lifetimes, i.e. the bulk and the surface, are combined to give the true or observed lifetime

$$\frac{1}{\tau_{\text{obs}}} = \frac{1}{\tau_{\text{bulk}}} + \frac{1}{\tau_{\text{surface}}} \tag{45}$$

In many cases where lifetime must be kept as long as possible, special chemical etching processes are used to ensure that the surface is as smooth as possible because this increases the surface lifetime.

3.7.6. Diffusion length

Another parameter used to describe the behaviour of semiconductor devices is the average distance a carrier may travel before it is forced to recombine. This distance is known as the *diffusion length.*

Let us postulate the generation of a number of excess carriers at a point in a bar-shaped crystal so that all the currents flow along the length of the bar in the x direction. Let p be the number of holes/unit volume at a section dx along the bar. Then

$$\frac{\mathrm{d}p}{\mathrm{d}t} = g - r - \frac{1}{e}\frac{\mathrm{d}J_p}{\mathrm{d}x} \tag{46}$$

The last term arises because, if the current has increased, some holes must have been removed from the volume each carrying a positive charge $+e$, and

$$J_{p\,\text{diff}} = -eD_p\frac{\mathrm{d}p}{\mathrm{d}x}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = g - r + D_p\frac{\mathrm{d}^2p}{\mathrm{d}x^2} \tag{47}$$

This is a general equation for holes, and a sign change enables a similar equation for electrons to be written down.

P. [illegible]

Let the bar be n-type with light incident on it at $x = 0$. Hence, hole-electron pairs are produced at $x = 0$; $r = p/\tau_p$ where τ_p is the lifetime of holes. At some time after the light is switched on, an equilibrium state in time will be reached. The hole concentration will vary with distance but not with time (assuming that there is no generation of holes at any other point in the bar except at $x = 0$).

Therefore

$$\frac{\mathrm{d}p}{\mathrm{d}t} = 0$$

$$g = 0$$

therefore

$$D_p\frac{\mathrm{d}^2p}{\mathrm{d}x^2} = r = \frac{p}{\tau_p}$$

or

$$\frac{\mathrm{d}^2p}{\mathrm{d}x^2} = \frac{p}{D_p\tau_p}$$

Solving for the condition at $x = 0$, $p = p_0$

$$p = p_0\exp\left[-\frac{x}{(D_p\tau_p)^{1/2}}\right] \tag{48}$$

From this expression it is seen that holes will travel a distance

$$L_p = (D_p\tau_p)^{1/2} \tag{49}$$

before the concentration falls to $1/e$ of its value at the start. The concentration gradient at $x = 0$ is given by

$$\left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)_{x=0} = \frac{-p_0}{L_p} \tag{50}$$

L_p is the average distance travelled by a hole before recombination and is known as the diffusion length.

3.7.6.1. *Measurement of the diffusion length*

A line of light is focused on a semiconductor bar at right angles to its length The concentration of, say, holes increases from an equilibrium value of p_0 to p_1. The additional holes recombine as they diffuse away, giving a concentration p at x along the bar.

$$(p - p_0) = (p_1 - p_0)\exp\left(-\frac{x}{(D_p\tau_p)^{1/2}}\right)$$

A negatively biased collector at a distance x from the light collects a current $I \propto p - p_0$. Hence a plot of ln I versus x gives a straight line of slope equal to the diffusion length, L_p.

3.7.7. The continuity equation

The separate effects of carrier generation, recombination, diffusion and drift can be conveniently summarized in an equation known as the *continuity equation.* This equation is derived by equating the rate of change of charge density in a volume of semiconductor to the algebraic sum of the charge entering or being created, and the charge leaving or being lost in the volume element. Thus, generally, for holes

$$\frac{\mathrm{d}p}{\mathrm{d}t} = g - \frac{p - p_0}{\tau_p} + D_p\nabla^2 p - \mu_p E \operatorname{grad} p \tag{51}$$

where $D_p\nabla^2 p$ expresses the diffusion of particles, and $\mu_p E \operatorname{grad} p$ expresses the drift of particles.

In most cases the one-dimensional form of this equation is sufficient, and we write

$$\frac{\mathrm{d}p}{\mathrm{d}t} = g - \frac{p - p_0}{\tau_p} + D_p \frac{\mathrm{d}^2 p}{\mathrm{d}x^2} - \mu_p \frac{\mathrm{d}}{\mathrm{d}x}(pE) \tag{52}$$

and for electrons

$$\frac{\mathrm{d}n}{\mathrm{d}t} = g - \frac{n - n_0}{\tau_n} - D_n \frac{\mathrm{d}^2 n}{\mathrm{d}x^2} + \mu_n \frac{\mathrm{d}}{\mathrm{d}x}(nE) \tag{53}$$

3.7.8. Einstein relation

Consider a bar of p-type semiconducting material which is inhomogeneously doped in such a way that the concentration of holes at $x = 0$ is p_0 and the concentration at a point x away is p_x where $p_x < p_0$. We can write the continuity equation as

$$J_p = -eD_p \frac{\mathrm{d}p}{\mathrm{d}x} + pe\mu_p E_x \tag{54}$$

at equilibrium $J_p = 0$

$$\frac{\mathrm{d}p}{\mathrm{d}x} = \frac{\mu_p}{D_p} \cdot pE_x \tag{55}$$

The holes diffuse from the region of high concentration into the region of low concentration leaving immobile negative charges in the heavily doped region. The result is that a field E_x, and hence a potential barrier V_x, is set up to oppose the further diffusion of holes.

In equilibrium, the number diffusing over the potential barrier equals the number returning under the influence of the favourable potential. Therefore

$$p_x = p_0 \exp\left(-\frac{eV_x}{kT}\right) \tag{56}$$

and

$$\frac{\mathrm{d}p_x}{\mathrm{d}x} = -\frac{p_0 e}{kT} \exp\left(-\frac{eV_x}{kT}\right)\frac{\mathrm{d}V_x}{\mathrm{d}x} \tag{57}$$

Using eqns. 55 and 57 and $E = -\dfrac{\mathrm{d}V_x}{\mathrm{d}x}$

$$\mu_p = \frac{e}{kT} D_p \tag{58}$$

This is known as the Einstein relation and by a similar analysis we find, for electrons

$$\mu_n = \frac{e}{kT} D_n \tag{59}$$

3.7.9. Current flow in semiconductors

From the continuity equation we may consider the case where g and r are negligible and derive expressions for the current flow in terms of the diffusion and drift components

$$J_n = eD_n \frac{\mathrm{d}n}{\mathrm{d}x} + en\mu_n E \tag{60}$$

$$J_p = -eD_p \frac{\mathrm{d}p}{\mathrm{d}x} + ep\mu_p E \tag{61}$$

Using the Einstein relations,

$$J_n = eD_n\left[\frac{\mathrm{d}n}{\mathrm{d}x} + \frac{eE}{kT}n\right] \tag{62}$$

$$J_p = eD_p\left[-\frac{\mathrm{d}p}{\mathrm{d}x} + \frac{eE}{kT}p\right] \tag{63}$$

These two equations are important and extensively used to describe the operation of semiconductor devices.

3.8. *The Hall effect*

The effect depends on the fact that moving charge carriers are deflected by a magnetic field applied at right angles to the direction of current flow. The redistribution of charge carriers creates an electric field which is orthogonal to both the magnetic field and the charge motion. The direction of the electric field depends on whether the charge carriers are electrons or holes, and the magnitude of the field is related to the mobility and concentration of the carriers. It was first discovered in thin metal foils by Hall in 1897.

Consider a semiconductor crystal with a battery connected between contacts on opposite faces. Carriers move along its length in the x-direction with a velocity v and a current I_x flows. If a magnetic field, B, is applied at right angles to the direction in which current is flowing, as shown in Fig. 3.17, there is a force $ev \times B$ on the carriers and they are deflected in a direction which is mutually perpendicular both to the original direction of motion and to the magnetic field. The initial deflection of these carriers builds up an electric field across the semiconductor of such magnitude that the force which it exerts on the carriers just balances the force exerted by the magnetic field.

In Fig. 3.17b an n-type semiconductor is shown in which the concentration of minority carriers is assumed to be negligible. The current flow is in the positive x-direction and the magnetic field in the positive z-direction. The

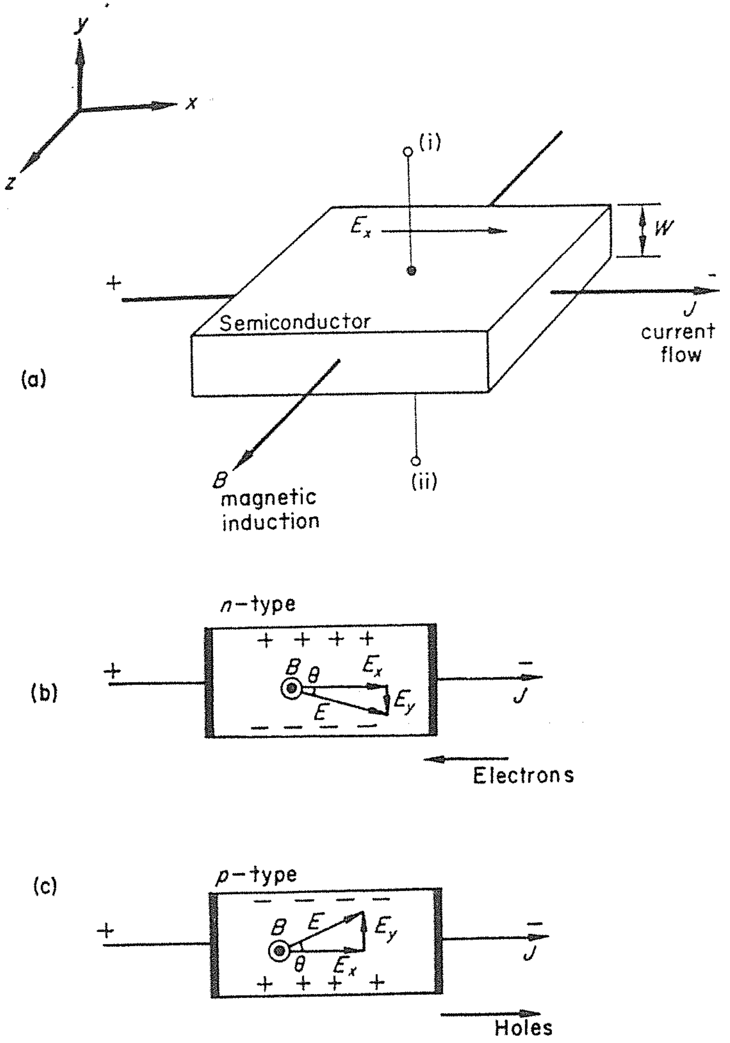


Fig. 3.17 The Hall effect: (a) the applied fields; (b) *n*-type semiconductor; (c) *p*-type semiconductor

electrons flowing from right to left are deflected in the negative y-direction and a transverse field, E_y, acts as shown. The total electric field, the resultant of the applied field E_x and the transverse field E_y, makes an angle θ with the direction of current flow as shown in the inset vector diagram.

If the drift velocity of the electrons is v_n and their concentration n_n, we have

$$v_n = -\frac{J_x}{n_n e} \tag{64}$$

Equating the transverse forces on the electron

$$eE_y = eBv_n = -\frac{BJ_x}{n_n} \tag{65}$$

The component of the electric field parallel to the direction of current flow is defined by

$$J_x = n_n e\mu_n E_x \tag{66}$$

whence

$$\tan\theta = \frac{E_y}{E_x} = -B\mu_n \tag{67}$$

θ is known as the *Hall angle* and the transverse electric field E_y as the *Hall field*.

It is usual to write the Hall field as

$$E_H = RJB \tag{68}$$

where R is known as the *Hall constant*, which, from eqn. 65 is given by

$$R_n = -\frac{1}{en_n} \tag{69}$$

For p-type material the charge carriers are again deflected in the negative y-direction as shown in Fig. 3.17c, but the direction of the Hall field is now in the positive y-direction because the holes are positively charged and

$$R_p = \frac{1}{pe} \tag{70}$$

If the width of the semiconductor is w, a voltage appears between the terminals i and ii shown in Fig. 3.17a and is known as the *Hall voltage*, V_H. It is given by

$$V_H = E_H . w \tag{71}$$

and using eqn. 68, R can be found, and from the sign of the Hall voltage the predominant carrier in the crystal may be identified.

In deriving the expression for the Hall constant a steady drift velocity has been assumed, but when various refinements are taken into account and the average speed is calculated, it is found that

$$R_n = -\frac{3\pi}{8n_n e} = -\frac{1{\cdot}18}{n_n e} \tag{72}$$

and

$$R_p = \frac{1{\cdot}18}{pe} \tag{73}$$

We now derive an expression for the Hall constant when both electrons and holes must be considered.

Consider a transverse current component

$$J_y = ne\mu_n E_y + pe\mu_p E_y \tag{74}$$

or, from eqn. 67 and the corresponding eqn. for holes,

$$J_y = eBE_x(p\mu_p^2 - n\mu_n^2) \tag{75}$$

To reduce the net transverse current to zero, a field must exist

$$E_y = \frac{eBE_x}{\sigma}(p\mu_p^2 - n\mu_n^2)$$

$$= \frac{eBJ_x}{\sigma^2}(p\mu_p^2 - n\mu_n^2)$$

$$\therefore\ R = \frac{E_H}{BJ_x} = \frac{p\mu_p^2 - n\mu_n^2}{e(n\mu_n + p\mu_p)^2} \tag{76}$$

Using $\sigma = e(n\mu_n + p\mu_p)$. If the material is intrinsic and $n = p = n_i$

$$R_i = \frac{1}{en_i}\left(\frac{\mu_p - \mu_n}{\mu_n + \mu_p}\right) \tag{77}$$

3.8.1. The Hall effect in semiconductor measurements

The Hall constant can be used to determine a number of semiconductor parameters when the current is predominantly due to one type of carrier only.

(a) As we have already mentioned, the type of carrier (i.e. hole or electron) can be identified from the sign of R.
(b) The concentration of the majority carriers can be determined using eqns. 69 and 70.
(c) The mobility of the majority carriers can be found using eqn. 66, and mobilities measured in this way are known as *Hall mobilities*.
(d) The energy gap of an intrinsic semiconductor may be determined by the use of eqn. 77 in which we can write

$$n_i = AT^{3/2}\exp\left(-\frac{E_g}{2kT}\right)$$

(14) [illegible] ارجع

where A is a constant. Thus,

$$R_i = \frac{1}{eAT^{3/2}\exp\left(-\frac{E_g}{2kT}\right)}\left(\frac{\mu_n - \mu_p}{\mu_n + \mu_p}\right) \tag{78}$$

whence a plot of R_i versus T gives E_g, provided the values of E_g and of the mobilities remain constant with temperature.

3.9. *Metal–semiconductor contacts*

We describe, qualitatively, the behaviour of metal–semiconductor contacts which are used extensively in semiconductor devices. The contact can either be rectifying or it can be ohmic (i.e. with a low resistance and a linear I–V characteristic). Rectifying contacts are used in the metal rectifiers such as selenium or copper-oxide rectifiers whose use predates the advent of germanium and silicon junction devices. Ohmic contacts are used to attach connecting terminals to all types of semiconductor devices.

✗3.9.1. Rectifying contacts

The contact between a metal and a semiconductor will rectify when the work function of the metal is greater than that of the semiconductor.

We consider a metal of work function ϕ_m. In such a metal, all the energy states are occupied up to the Fermi level, ζ, and the application of a further energy input equal to ϕ_m is necessary before electrons can be emitted from the surface. This statement defines the term 'work function' until we consider its meaning in more detail in Chapter 6. An n-type semiconductor in which the Fermi level lies above the centre of the band-gap is placed adjacent to, but not in contact with, the metal (Fig. 3.18a). The semiconductor work function is defined as ϕ_s, the energy measured from the Fermi level to the vacuum level. The energy marked χ in Fig. 3.18a is known as the electron affinity and is the

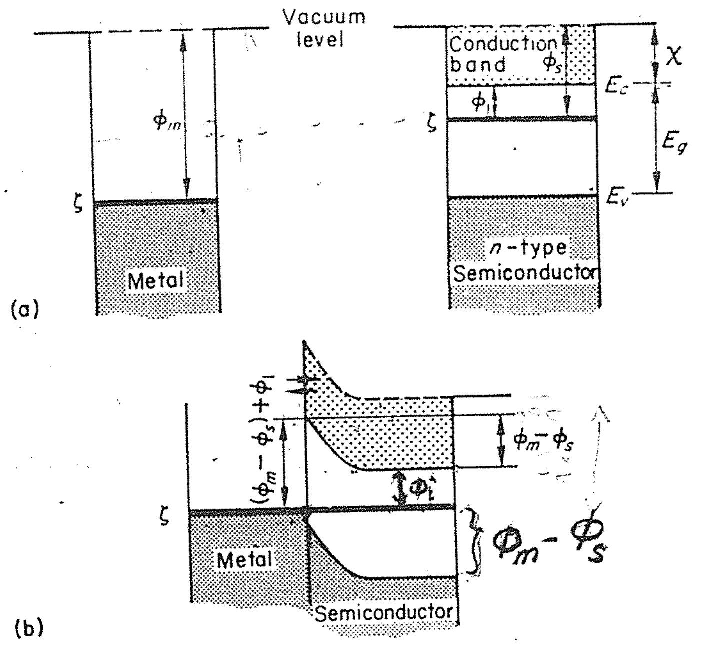


Fig. 3.18 (a) Energy bands in a metal and n-type semiconductor before contact. (b) Rectifying contact between metal and semiconductor, $\phi_m > \phi_s$

energy required to emit an electron from the bottom of the conduction band. Finally, ϕ_1 is the energy difference from the level ζ to E_c.

If the metal and the semiconductor are now brought into contact, the energy bands are modified as shown in Fig. 3.18b. Electrons occupying conduction band energy states move into the metal and leave behind positive charges associated with their immobile donor atoms. The movement continues until the Fermi levels on the two sides of the contact are aligned and the levels E_c and E_v bend upwards at the junction, so that ζ now lies at an energy $\phi_1 + \phi_m - \phi_s$ below E_c.

تعلق

The metallic surface becomes negatively charged while the semiconductor becomes positively charged as a result of the initial charge movement, and further movement of electrons from semiconductor to metal is inhibited by a

يعيق ، تمنع

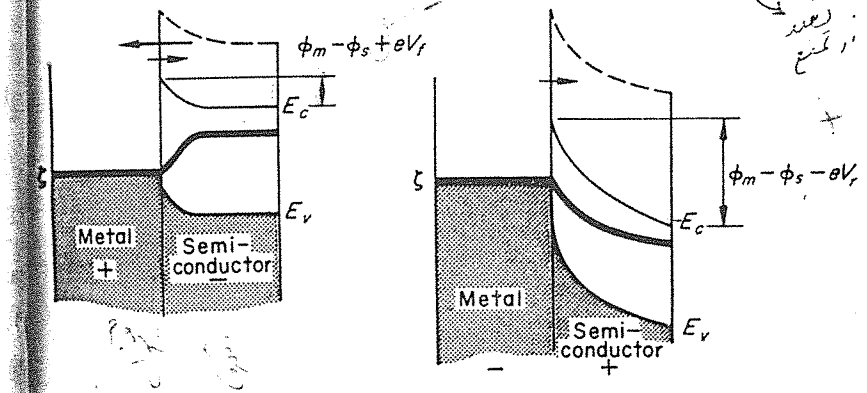


Fig. 3.19 Forward bias and reverse bias applied to a metal–semiconductor rectifying contact

barrier of height $\phi_m - \phi_s$. Electron flow from metal to semiconductor is inhibited by a barrier of height $(\phi_m - \phi_s) + \phi_1$. In equilibrium, small, equal and opposite currents flow from semiconductor to metal and vice versa and there is no net current.

Suppose the situation is now altered by the application of a voltage, V, such that the metal is positive and the semiconductor negative. The barrier from semiconductor to metal is *reduced* by an amount eV and a large number of electrons flow from semiconductor to metal while the number from metal to semiconductor remains unchanged. This situation is known as forward bias. Reversal of the bias polarity causes the barrier from semiconductor to metal to be *increased* by eV and the current flow is even further reduced than in the unbiased condition. Figure 3.19 illustrates the two cases.

3.9.2. Ohmic contacts

If $\phi_m < \phi_s$, the energy bands are modified on contact as shown in Fig. 3.20. The Fermi levels again align, electrons flow from metal to semiconductor and a negative charge accumulates on the semiconductor which causes the levels E_c and E_v to bend downwards. There is no rectifying potential barrier and none exists on the application of either forward or reverse bias. Thus current flow is not impeded by the existence of a barrier and is considered ohmic. In practice, the condition $\phi_m < \phi_s$ does not appear to be sufficient to make ohmic contacts.

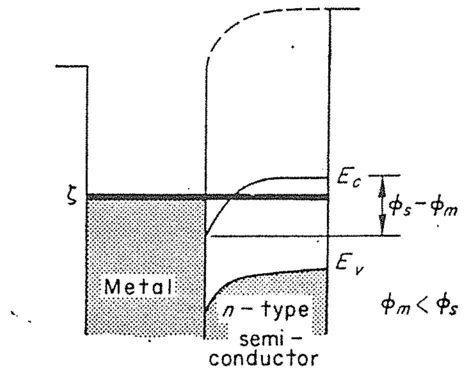


Fig. 3.20 Ohmic contact between metal and n-type semiconductor

Finally, it may be pointed out that for the p-type semiconductor metal contact, it is required that $\phi_m < \phi_s$ to make a rectifying contact, and $\phi_m > \phi_s$ to make an ohmic contact. The energy band diagrams for this case may be drawn using the principles outlined for the n-type semiconductor.

Appendix

The concept of effective mass

Solid-state physicists often find it convenient to use semi-classical treatments to explain the main features of electron motion, electrical and thermal conductivity and the like in metals and semiconductors. Such methods give a useful insight into the relevant physical processes but obviously they have to be used with discretion. In the theory of electrical conductivity the concept of effective mass is widely used. This concept arises because, in quantum mechanics, an electron travels with a velocity equal to the *group* velocity of the probability waves. For a *free* electron the probability waves are non-dispersive, that is their phase and group velocities are equal. However, this is an unusual special case and in general the waves are dispersive so that the phase velocity is a function of energy and no longer equals the group velocity. If we model the particulate electron, in engineering terms, it can be thought of as a signal propagated by the waves and therefore moving at their group velocity, as stated. If we wish to think of this signal in particulate terms, it is clearly useful if we can preserve the ordinary relationships, of Newtonian mechanics, between the force and the momentum and energy. This can only be done if we introduce an effective mass, which differs from the mass of a free particle and which depends on the group velocity, i.e. on the dispersion equation for the Schrödinger waves for the particular case under discussion.

The precise definition of effective mass is obtained as follows. For any wave of angular frequency ω and propagation constant β, the group velocity is given by

$$g = \frac{d\omega}{d\beta} \tag{A1}$$

Also, the energy is related to ω in quantum mechanics by ($\hbar = h/2\pi$)

$$E = \hbar\omega \tag{A2}$$

Therefore

$$g = \frac{1}{\hbar}\frac{dE}{d\beta} \tag{A3}$$

The momentum p of a particle is given by de Broglie's relation

$$p = \hbar\beta \tag{A4}$$

and the force is

$$dp/dt = \hbar\, d\beta/dt \tag{A5}$$

since $\hbar$ is a constant.

We now define the effective mass by

$$p = m^* g$$

Then

$$F = \frac{dp}{dt} = \frac{m^*}{\hbar}\cdot\frac{d^2E}{d\beta^2}\cdot\frac{d\beta}{dt} \tag{A6}$$

or, using eqn. A5

$$m^* = \frac{\hbar^2}{\left(\frac{d^2E}{d\beta^2}\right)} = \left(\frac{d^2E}{dp^2}\right)^{-1} \tag{A7}$$

Thus, theoretically we can calculate the effective mass if we know the relationship between E and β, which results from the solution of Schrödinger's equation. It is not often necessary to assign a numerical value to m^* but Hall effect and conductivity measurement yield values *averaged* over a range of energies, which when inserted in the relevant equations give numerically accurate results.

Questions

X **1.** Describe what is meant by mobile and immobile charges in an impurity semiconductor and explain the formation of a depletion layer at a *p-n* junction.

An *n*-type semiconductor has 10^{14} donor atoms/cm³ which are completely activated at 300 °K: the electrons in the conduction band can be represented by $N_c = 2{\cdot}5 \times 10^{19}$ available levels/cm³ at the bottom of the band E_c with a probability of occupation equal to $\exp[-(E_c - E_F)/kT]$. Determine the position of the Fermi level, E_F. At 300 °K, $kT = 0{\cdot}026$ eV. (*L.U., B.Sc. Eng. Elec.*)

✓ **2.** It can be shown that, in a semiconductor, the number of electrons in the conduction band is given by ✓

$$n = N_c e^{E_F/kT} \text{ per m}^3$$

where E_F is the Fermi energy measured with reference to the bottom of the conduction band and

$$N_c = 2(2\pi m_e kT/h^2)^{3/2}$$

An *n*-type semiconductor has a concentration N_D of donor levels at energy ΔE below the bottom of the conduction band, intrinsic conductivity is negligible and it may be assumed that the classical approximation to the Fermi function is valid.

✓ (a) Show that, if nearly all of the donors are ionized,

$$E_F \doteqdot kT \ln N_D/N_C$$

X (b) Show that, at sufficiently low temperatures, where relatively few of the donors are ionized,

$$E_F \doteqdot -\tfrac{1}{2}\Delta E \quad \text{and} \quad n = \sqrt{N_C N_D}\, e^{\Delta E/2kT}$$

✓ **3.** Sketch the form of the curve giving the log of resistivity as a function of $1/T$ for a semiconducting material which contains sufficient donor impurity atoms (N_d) for the material to be predominantly *n*-type at room temperatures, where the impurities are fully ionised. Explain fully the principal conduction processes responsible for different regions of this curve.

A sample of *n*-type germanium has a resistivity of 1·5 Ωcm at 300 °K. Evaluate the number of electrons, n, per cm³ in the conduction band using the values: $e = 1{\cdot}602 \times 10^{-19}$ C for the charge on the electron and $\mu_n = 3{,}600$ cm² V⁻¹ sec⁻¹ for the electron mobility. Check that the density of holes, p, is negligible compared with n, given that in intrinsic germanium at this temperature, $n = n_i = 2{\cdot}41 \times 10^{13}$ cm⁻³. (*L.U., B.Sc. Eng. Elec.*)

X **4.** The variation of the resistivity of intrinsic germanium with temperature is given by the following table:

T (°K)	385	455	556	714
ρ (Ωm)	0·028	0·0061	0·0013	0·000274

It may be assumed, as a rough approximation, that the hole and electron mobilities both vary as $T^{-3/2}$ and that the forbidden energy gap E_g is independent of temperature.

(a) Determine the value of E_g in electron volts.

(b) At about what wavelength would you expect the fundamental absorption edge to occur?

✓ **⑤** The resistivity of intrinsic germanium at room temperature (300 °K) is 0·47 Ωm and the mobilities at this temperature are

$$\mu_p = 0{\cdot}18 \text{ m}^2/\text{Vs} \qquad \mu_n = 0{\cdot}38 \text{ m}^2/\text{Vs}\ , e = 1{\cdot}602 \times 10^{-19}\,\text{C}$$

(a) If p and n are the concentrations of holes and electrons respectively, what is the value of the pn product? $5{\cdot}66 \times 10^{38}$

(b) If the germanium had contained 5×10^{19} donor atoms per cubic metre and if all of these had been ionized, what would then have been the resistivity of the germanium? $\rho = 0{\cdot}223\ \Omega\,\text{m}$

X **6.** Discuss the interdependence of electrical conductivity, carrier mobility, diffusivity, minority carrier lifetime and diffusion length in semiconductors.

The radiative recombination lifetime for germanium and indium antimonide may be taken as 1 s and 0·5 µs respectively. Experiments on the decay of photoconduction in monocrystalline InSb give decay times of order 0·35 µs. Experiments on the measurement of diffusion length of holes in a crystal of *n*-type germanium give values of 1 mm for this quantity. If the hole mobility may be taken as 1750 $cm^2 V^{-1} s^{-1}$ and kT/e as 0·025 V, what is the hole lifetime in the specimen of germanium?

Discuss the physical interpretation of the above observations and the probability or otherwise of obtaining lifetime values approaching 1 s for germanium.

7. Distinguish between *drift* current and *diffusion* current in a semiconductor and derive an expression for the diffusion current flowing across an infinite plane *p-n* junction across which a potential difference V is applied. State clearly all assumptions made. (*G. Inst. P.*)

Calculate, making reasonable approximations, the ratio of electron drift current to hole drift current for intrinsic, 5 Ωcm *p*-type, and 5 Ωcm *n*-type germanium, given that the electron mobility is 3000 cm^2 volt-s, the hole mobility is 1900 cm^2/volt-s and that the product of the electron density and the hole density is $6{\cdot}25 \times 10^{26}$ cm^{-6}. (*Grad. I.E.E.*)

8. Describe how the Hall coefficient is measured for a material such as germanium. In an experiment on a semiconductor specimen of resistivity 5 Ωcm the Hall coefficient was found to be 4,500 cm^3 C^{-1}. Assuming the electronic charge to be $1{\cdot}6 \times 10^{-19}$ C, what are the free charge carriers, what is their concentration, and what is their mobility?

Mention two practical applications of the Hall effect. (*G. Inst. P.*)

9. Derive the vector equation for the Hall effect and interpret its significance with the aid of a vector diagram.

A current of 2 mA flows along the axis of a rectangular specimen of *n*-type germanium of length 0·005 m and of square section of side 0·001 m. If the specimen is situated so that the axis and the plane of one side are perpendicular to a uniform magnetic field of flux density 0·15 Wb/m^2, the Hall voltage measured between the mid-points of the opposite pair of sides is found to be 2·5 mV. If the potential difference between the ends of the specimen is found to be 0·25 V, calculate :

(a) the Hall coefficient;
(b) the conductivity;
(c) the electron mobility;
(d) the current carrier density.

Indicate the importance of the use of Hall-effect measurements for the determination of semiconductor properties and describe two further practical devices utilizing the Hall effect. (*Grad. I.E.E.*)

10. Explain in *physical* terms the mechanism of the Hall effect in a semiconductor.

A thin *p*-type germanium specimen of dimensions 2 × 1 cm is located in a plane normal to the earth's magnetic field. Calculate the potential difference which must be applied between the faces 2 cm apart if a Hall voltage of 10 mV is to arise. The earth's magnetic field may be taken as 44 μWb/m^2. The mobility of positive holes in germanium is 0·18 m^2/Vs.

Indicate on a diagram where this Hall voltage will appear, showing its direction, the direction of the magnetic field and of the applied voltage.

Explain fully what would be the effect of increasing this applied voltage by a factor of 200. (*Grad. I.E.E.*)

Bibliography

1. DEKKER, A. J. (1957) *Solid State Physics*. Prentice-Hall, Englewood Cliff, N.J.; Macmillan, London.
2. KITTEL, C. (1966) *Introduction to Solid State Physics*, 3rd edn. Wiley, New York and London.

3. MOTT, N. F. and JONES, H. (1959) *The Theory and Properties of Metals and Alloys*. Oxford University Press, London.
4. MOLL, J. L. (1964) *Physics of Semiconductors*; chapter 5. McGraw-Hill, New York and Maidenhead.
5. BLAKEMORE, J. S. (1962) *Semiconductor Statistics*. Pergamon, Oxford.
6. SHIVE, J. N. (1959) *Semiconductor Devices*. Van Nostrand, Princeton, N.J. and London.
7. CONWELL, E. M. (1952) *Proc. Inst. Radio Engrs*, **40**, 1327; (1958) *ibid.*, **46**, 1281.
8. SHIVE, J. N. *op. cit.*; chapters 2 and 24.
9. *ibid.*; chapter 20.

General reference

MOLL, J. L. (1964) *Physics of Semiconductors*. McGraw-Hill, New York and Maidenhead.

4

Semiconductor diodes and junction transistors

Since 1948, when the first transistor was announced by Bardeen and Brattain (Ref. 1) at the Bell Telephone Laboratories, a large number of semiconductor devices have been invented, and in this chapter and the next we describe some of the more important ones.

The p–n junction is described first and its mechanism of operation is considered in some detail since it forms the basis for almost all the other semiconductor devices. The junction transistor and the drift transistor have had such a very great impact on post-war electronics that it is unnecessary to stress their importance any further. The more recent junction-field-effect devices are increasingly important because they extend the use of semiconductor devices to circuit applications which had hitherto been beyond the capabilities of the junction transistor. Finally, the metal-oxide-semiconductor device, which should play an important part in the development and production of microcircuits, is described.

Notation

n	electron concentration
n_p	electron concentration in p-type region (minority carriers)
n_{po}	electron concentration in bulk p-type crystal
n_n	electron concentration in n-type region (majority carriers)
p	hole concentration
p_n	hole concentration in n-type region (minority carriers)
p_{no}	hole concentration in bulk n-type crystal
p_p	hole concentration in p-type region (majority carriers)
V_a	barrier voltage at unbiased p–n junction
V_b	applied battery bias
L_p, L_n	hole and electron diffusion lengths
D_p, D_n	diffusion constants
τ_p, τ_n	lifetimes
J_p, J_n	current densities
A	junction area

4.1. The p–n junction

An exact theoretical analysis of the conduction process at a practical p–n junction is very difficult, but a close approximation to the experimental properties can be obtained by assuming a fairly simple model for the p–n junction itself. We assume that (a) the junction is abrupt, (b) recombination in the region of the junction is negligible and (c) there are no fields in either semiconductor material away from the junction region. Assuming this simple model we first describe the action at the junction qualitatively and then derive expressions for the current flow across the junction, the width of the junction region and the junction capacitance.

The action at the junction with no external applied voltage is described with reference to Fig. 4.1. A p-type crystal, with its high concentration of holes, is brought into close contact with an n-type crystal containing a high concentration of free electrons. The concentrations are shown in Fig. 4.1b. From the discussion of diffusion in section 3.7.3 we know that there is an immediate diffusion of holes from the p-region, where they are abundant, to the n-region, where they are scarce, and there is a corresponding movement of electrons from the n-region to the p-region. The movements are shown schematically by the firm arrows in Fig. 4.1a. These movements last for a very short while and mobile holes and electrons known as carriers subsequently recombine as they move away from the junction into the bulk of the crystal.

Before the junction was made, both crystals were electrically neutral so that the movement of holes from the p-region leaves a number of negatively charged acceptor atoms on that side, which remain bound to the lattice, and similarly the movement of electrons from the n-region leaves behind positively charged donor atoms. The result is the formation of a space-charge layer in the vicinity of the junction with a positive charge on the n-type region and a negative charge in the p-type region, and with the reversal of charge occurring at the junction. The immobile charged atoms are shown schematically in Fig. 4.1a and the magnitude and distribution of the space charge is shown in Fig. 4.1c. The extent of this space-charge layer between the two crystals is known as the *depletion region* or layer because mobile carriers have been removed from it. Another feature of these two space-charge regions is that they are equal in magnitude, though opposed in polarity. This arises because we have assumed that there is no field in the crystal far from the junction. The existence of the oppositely charged layers produces a field in the depletion region and all field lines from positive donor atoms must terminate on an equal number of negative acceptor atoms.

The field intensity between the two field-free regions is shown as a function of distance in Fig. 4.1d. The maximum is exactly at the junction where the greatest number of lines cross. The direction of this field is such as to oppose the further movement of carriers and the diffusion process is therefore inhibited. A state of equilibrium is eventually reached when the field is strong enough to stop the net current flow across the junction completely. The field does cause some of the few holes in the n-type region to drift into the p-type

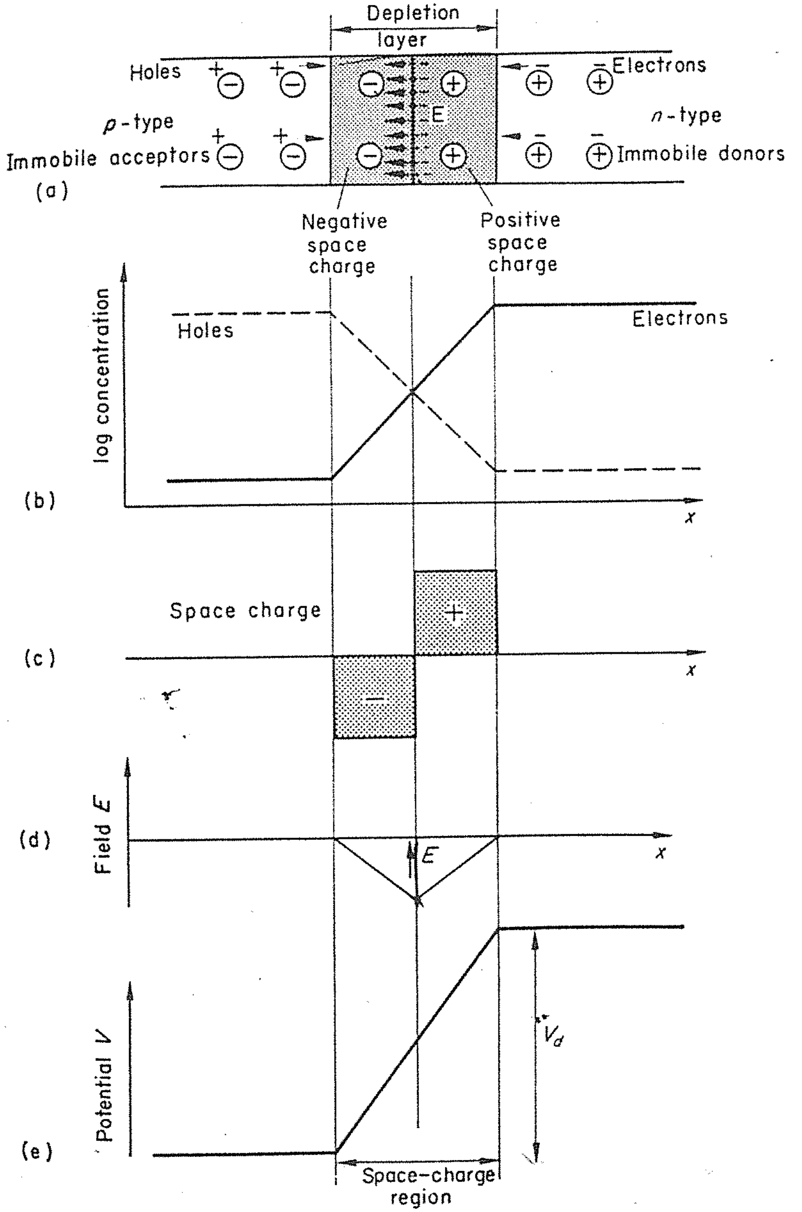


Fig. 4.1 (a) Schematic diagram of depletion layer formation. (b) Concentration of holes and electrons. (c) Space charge in the depletion layer. (d) Electric field in the depletion layer. (e) Potential barrier

region, but they are balanced by an equal number of holes which overcome the field and still diffuse into the *n*-type region from the *p*-type region; the net hole movement is zero. A similar argument applies in the case of electrons and the net current is zero.

A field at the junction implies that there is a potential difference between the two regions. It is shown plotted in Fig. 4.1e and is known as the potential barrier. The sense is such as to oppose the diffusion of carriers in either direction. To summarize, at a *p–n* junction in equilibrium with no applied bias, there is a potential barrier across a depletion layer and the net current flow is zero.

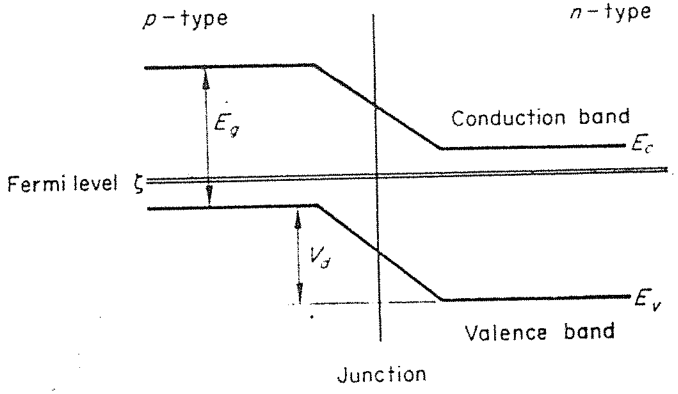


Fig. 4.2 Energy level diagram for *p–n* junction with no applied bias

The energy level diagram for the *p–n* junction in the equilibrium state is shown in Fig. 4.2. The energy gap for the semiconductor, E_g, is as shown. The Fermi level is continuous through the junction and lies near the valence band in the *p*-type material and near the conduction band in the *n*-type material. The potential barrier, V_d, is also shown in the same figure.

The conduction of charge across the *p–n* junction will be considered under three conditions: (1) with no applied bias voltage, (2) with a reverse bias applied to the junction, and finally (3) with a forward bias applied to the junction. The three bias conditions and the resulting potential barriers are shown in Fig. 4.3.

4.1.1. The unbiased *p–n* junction

The junction is in equilibrium with no applied voltage. There are four components of charge moving across the junction although the net current is zero.

(a) There are a few electrons on the *p*-side which move into the *n*-side with the aid of the field, causing a current which is proportional to the concentration of electrons on the *p*-side.

$$I_{n1} = C_1 n_{p0} \tag{1}$$

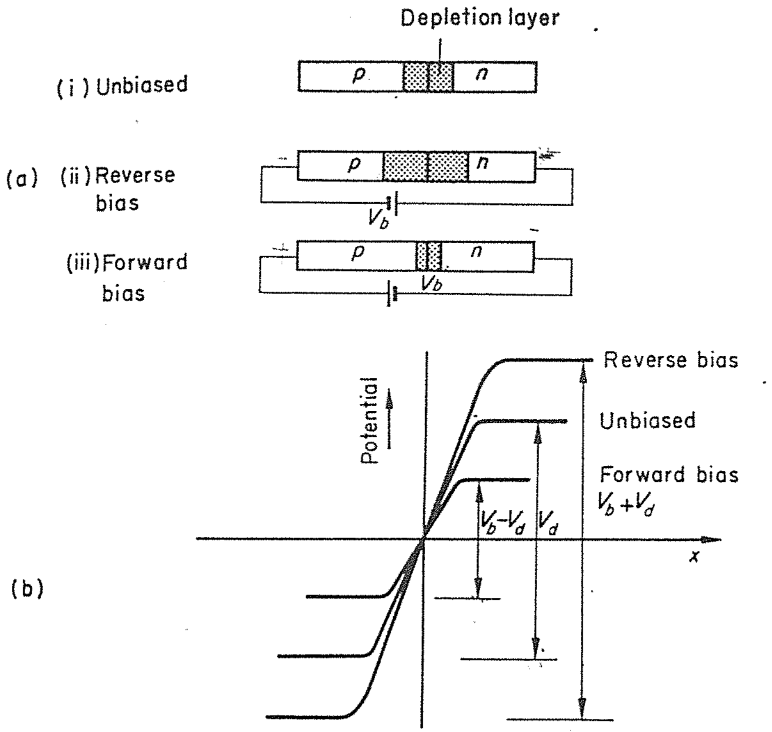


Fig. 4.3 (a) Bias conditions on a *p–n* junction. (b) Potential barrier formed with the three bias conditions

The constant C_1 depends upon various semiconductor parameters such as diffusion constant, diffusion length, lifetime and the junction area and is given a more precise meaning in section 4.1.4.

(b) There are a number of electrons which move from the *n*-side, overcome the potential barrier V_d and enter the *p*-side. This number is proportional to the concentration of electrons in the *n*-side multiplied by the probability of these electrons being able to surmount the potential barrier.

$$I_{n2} = C_1 n_n \exp\left(-\frac{eV_d}{kT}\right) \tag{2}$$

The usual Boltzmann relation is used to express the probability of electrons overcoming the barrier and the constant C_1 is the same as in eqn. 1, since it again depends on the same parameters as before which apply to the conduction of electrons in a crystal.

(c) In a manner similar to case (a), a few holes will move from the *n*-side to the *p*-side, giving

$$I_{p1} = C_2 p_{n0} \tag{3}$$

where C_2 is a constant which applies to the conduction of holes in a crystal.

(d) Some holes overcome the barrier and move from the *p*-side to the *n*-side and this current is given by

$$I_{p2} = C_2 p_p \exp\left(-\frac{eV_d}{kT}\right) \tag{4}$$

In equilibrium, for the unbiased junction, since the net current is zero,

$$I_{n1} = I_{n2} \quad \text{and} \quad I_{p1} = I_{p2}$$

Hence
$$n_{p0} = n_n \exp\left(-\frac{eV_d}{kT}\right). \tag{5}$$

and
$$p_{n0} = p_p \exp\left(-\frac{eV_d}{kT}\right) \tag{6}$$

4.1.2. The *p–n* junction with a reverse bias

Reverse bias is applied to a *p–n* junction as shown in Fig. 4.3a with the negative terminal of the battery connected to the *p*-type crystal and the positive terminal to the *n*-type crystal. The height of the potential barrier is increased by the magnitude of the applied bias as shown in Fig. 4.3b. The effect of this increase in barrier height is that the currents I_{n2} and I_{p2} are reduced, to negligible values when V_b is large enough, while the currents I_{n1} and I_{p1} remain unchanged. We have, for the reverse bias

$$I_{n2} = C_1 n_n \exp -\frac{e(V_d + V_b)}{kT} \tag{7}$$

and
$$I_{p2} = C_2 p_p \exp -\frac{e(V_d + V_b)}{kT} \tag{8}$$

The net current flow under these conditions is given by the algebraic sum of the four components

$$I_R = (C_1 n_n + C_2 p_p) \exp -e(V_d + V_b)/kT - C_1 n_{p0} - C_2 p_{n0}$$

Using eqns. 5 and 6 we have

$$I_R = (C_1 n_{p0} + C_2 p_{n0})\left(\exp -\frac{eV_b}{kT} - 1\right) \tag{9}$$

or
$$I_R = I_0\left(\exp -\frac{eV_b}{kT} - 1\right) \tag{10}$$

where $I_0 = (C_1 n_{p0} + C_2 p_{n0})$ is known as the saturation current. The presence of the negative exponent ensures that for bias voltages greater than ~ 1 volt I_R is equal to the negative of the saturation current which flows as a result of the small number of electrons in the *p*-type region and the small number of holes in the *n*-type region.

4.1.3. *p–n* Junction with forward bias

Forward bias is applied to the *p–n* junction as shown in Fig. 4.3a, and the potential barrier is reduced by the amount of the battery voltage as shown in Fig. 4.3b. The currents I_{n2} and I_{p2} are considerably increased because they have a much lower barrier to surmount while the currents I_{p1} and I_{n1} remain unchanged. The total current in the forward direction is

$$I_F = (C_1 n_n + C_2 p_p) \exp -\frac{e(V_d - V_b)}{kT} - C_1 n_{p0} - C_2 p_{n0}$$

Using eqns. 5 and 6 again, we have

$$I_F = I_0 \left(\exp \frac{eV_b}{kT} - 1\right) \tag{11}$$

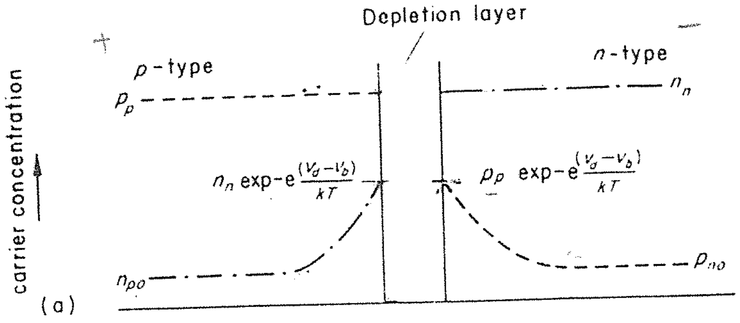
This is exactly the same as eqn. 10 except for the sign of V_b. I_F increases very fast for $V_b > 0{\cdot}02$V. Equations 10 and 11 are known as the diode equations and give the relation between current and applied bias. Experiments on real *p–n* junction diodes show that eqns. 10 and 11 give curves which agree very closely with experiment.

4.1.4. Interpretation of the saturation current I_0

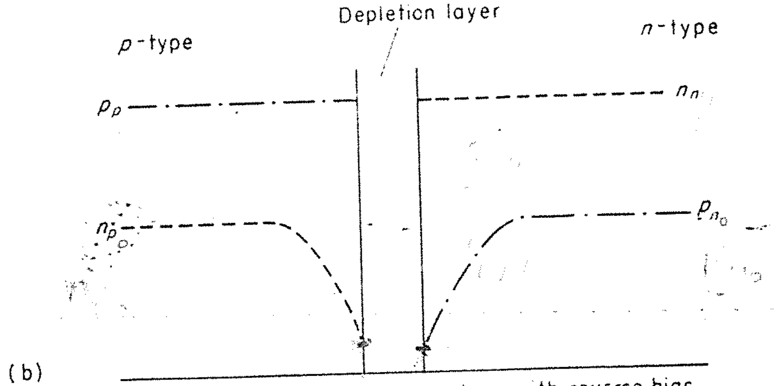
We have said earlier that the minority carrier flow across the junction depends on the parameters of the semiconductor material. To interpret the nature of the saturation current we must investigate the behaviour of these minority carriers more closely. عدل-تغير

Forward bias applied to the junction alters the concentration of minority carriers near the junction. Figure 4.4 shows a large increase in minority carriers near the junction for both *p*- and *n*-regions. This happens because there is a high concentration p_p, of holes on the *p*-side where they form the majority carriers and some of these cross the potential barrier so that the concentration just beyond the depletion layer on the *n*-side is $p_p \exp[-e(V_d - V_b)/kT]$. This high concentration decreases as one proceeds further into the *n*-side, to reach the equilibrium concentration of minority carriers (holes) in the *n*-side, p_{n0}. Similarly, there is a high concentration of majority carriers n_n on the *n*-side of the depletion layer and this concentration becomes $n_n \exp[-e(V_d - V_b)/kT]$ just on the *p*-side of the depletion layer and decays to the equilibrium value of n_{p0} at some distance from the junction.

Reverse bias affects the carrier densities as shown in Fig. 4.4b. The holes arriving at the junction from the *p*-side encounter a very high potential barrier and very few cross into the *n*-side. Also, the few minority carriers (holes) arriving at the junction from the *n*-side are swept away by the field into the *p*-side so that there is a very low concentration of holes on the *n*-side, well below the normal equilibrium concentration of p_{n0}. In a similar manner, the number of electrons at the junction on the *p*-side is well below the equilibrium concentration in the bulk material n_{p0}. تكنس



---- Hole concentration near *p-n* junction with forward bias
—·— Electron concentration near *p-n* junction with forward bias



—·— Hole concentration near junction with reverse bias
---- Electron concentration near junction with reverse bias

Fig. 4.4 Carrier concentration at a *p–n* junction with: (a) forward bias; (b) reverse bias

For holes in the *n*-type material, the concentration at the junction is given by

$$p_n = p_p \exp -\frac{e(V_d - V_b)}{kT} \tag{12}$$

If the equilibrium hole density in the *n*-side is p_{no} we have from eqn. 6.

$$p_{no} = p_p \exp -\frac{eV_d}{kT} \tag{13}$$

therefore

$$p_n = p_{no} \exp \frac{eV_b}{kT} \tag{14}$$

r.h.s = right hand side.

Using the continuity equation (eqn. 52, Chapter 3) and assuming that both generation of holes and drift currents are negligible, we have for the decay of holes with time

$$\frac{\mathrm{d}p_n}{\mathrm{d}t} = -\frac{(p_n - p_{n0})}{\tau_p} + D_p \frac{\mathrm{d}^2 p_n}{\mathrm{d}x^2} \tag{15}$$

where the first term on the r.h.s. is the rate of decay through recombination, and the second term represents the diffusion current. Hence

$$\frac{\mathrm{d}p_n}{\mathrm{d}t} = \frac{p_n}{\tau_p}\left(\exp - \frac{eV_b}{kT} - 1\right) + D_p \frac{\mathrm{d}^2 p_n}{\mathrm{d}x^2} \tag{16}$$

and in the steady state

$$D_p \frac{\mathrm{d}^2 p_n}{\mathrm{d}x^2} + \frac{p_n}{\tau_p}\left(\exp - \frac{eV_b}{kT} - 1\right) = 0 \tag{17}$$

Integrating, and using the boundary conditions,

$$p_n = p_{n0} \exp \frac{eV_b}{kT} \quad \text{at } x = 0, \qquad p_n = p_{n0} \quad \text{at } x \to \infty$$

we obtain

$$p_n = p_{n0}\left\{1 + \left(\exp \frac{eV_b}{kT} - 1\right) \exp - \frac{x}{(D_p \tau_p)^{1/2}}\right\}. \tag{18}$$

Now, from eqn. 41a, Chapter 3, we have

$J_p = -e D_p \frac{dP}{dx}$

$$I_p = -AeD_p \left(\frac{\mathrm{d}p_n}{\mathrm{d}x}\right)_{x=0}$$

where A is the area of the junction. Therefore

$$I_p = \frac{AeD_p p_{n0}}{L_p}\left(\exp \frac{eV_b}{kT} - 1\right) \tag{19}$$

A similar expression can be derived for the electron current, giving

$$I_n = \frac{AeD_n n_{p0}}{L_n}\left(\exp \frac{eV_b}{kT} - 1\right) \tag{20}$$

therefore

$$I_F = Ae\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right)\left(\exp \frac{eV_b}{kT} - 1\right) \tag{21}$$

Comparing this equation with the diode equation, eqn. 11, we have

$$I_0 = Ae\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right) \tag{22}$$

$\because L_p = (\tau_p D_p)^{\frac{1}{2}}$

and the constants C_1, C_2 are defined; or, in terms of lifetime,

$$I_0 = Ae\left(\frac{L_p p_{n0}}{\tau_p} + \frac{L_n n_{p0}}{\tau_n}\right) \tag{23}$$

Equation 23 may be written in terms of the acceptor and donor impurity concentrations assuming that they are completely ionized.

$$I_0 = en_i^2\left(\frac{D_p}{L_p N_d} + \frac{D_n}{L_n N_a}\right) \tag{24}$$

Since $p_{n0} = n_i^2/N_d$ and $n_{p0} = n_i^2/N_a$ eqns. 23 and 24 show that the saturation current depends on the minority carrier concentrations. For example, if the p-type region is heavily doped and the n-type region lightly doped, then large numbers of holes will be injected into the n-region where they are, of course, minority carriers. This type of junction design is known as *minority-carrier injection* and is used in many semiconductor devices.

4.1.5. Width of the depletion layer and depletion layer capacity

We have seen that space-charge layers form on either side of a p–n junction and that this disturbed region effectively controls the behaviour of the p–n junction, while the bulk of the crystal remains unchanged. We now calculate the width of the depletion layer for two cases assuming (a) that the transition from p-type to n-type material and the other boundaries of the depletion layer are abrupt, and (b) that the junction is linearly graded, with the impurity doping changing linearly from p-type at some point within the p-region to zero at the junction, and then to n-type at some point within the n-region. The two conditions are shown in Fig. 4.5a and b respectively. (Ref. 2 discusses non-uniformly graded junctions.)

Case 1. Abrupt junction

Let the space-charge density be greater in the n-type crystal than in the p-type crystal and let the extent of the space-charge layer be d_p on the p-side and d_n on the n-side.

Since the total charge on either side of the junction must be the same,

$$eN_d d_n = eN_a d_p \tag{25}$$

We also use Poisson's equation in one-dimensional form

$$\frac{\mathrm{d}^2 V}{\mathrm{d}x^2} = -\frac{\rho}{\varepsilon\varepsilon_0}$$

For the p-region this becomes

$$\frac{\mathrm{d}^2 V}{\mathrm{d}x^2} = \frac{eN_a}{\varepsilon\varepsilon_0} \tag{26}$$

Integrating,

$$\frac{\mathrm{d}V}{\mathrm{d}x} = \frac{eN_a x}{\varepsilon\varepsilon_0} + K_1 \tag{27}$$

at $x = d_p$ the electric field $-\mathrm{d}V/\mathrm{d}x = 0$, therefore $\therefore K_1 = -\frac{eN_a d_p}{\varepsilon\varepsilon_0}$

$$\frac{\mathrm{d}V}{\mathrm{d}x} = \frac{eN_a}{\varepsilon\varepsilon_0}(x - d_p)$$

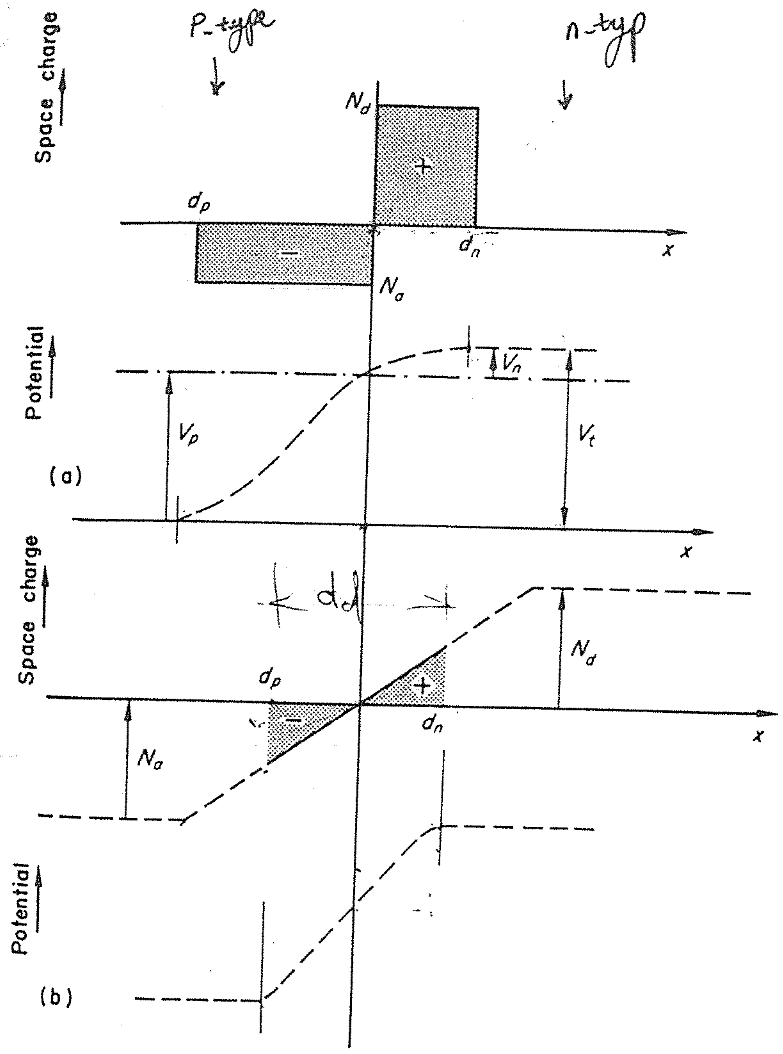


Fig. 4.5 The depletion layer: (a) abrupt junction; (b) graded junction

Integrating again $V = \frac{eN_a}{\varepsilon\varepsilon_0}\left(\frac{x^2}{2} - d_p x + K_2\right)$

Using an arbitrary reference $V = 0$ at $x = d_p$

$$V = \frac{eN_a}{\varepsilon\varepsilon_0}\left(\frac{x^2}{2} - d_p x + \frac{d_p^2}{2}\right) \tag{28}$$

Therefore, the height V_p from our reference level to the junction is given by putting $x = 0$ in eqn. 28 giving

$$V_p = \frac{eN_a}{2\varepsilon\varepsilon_0} \cdot d_p^2 \tag{29}$$

Similarly, using Poisson's equation for the n-side, where x is positive, we have the part of the barrier voltage given by V_n

$$V_n = \frac{eN_d}{2\varepsilon\varepsilon_0} \cdot d_n^2 \tag{30}$$

The total barrier voltage is

$$V_t = V_p + V_n = \frac{e}{2\varepsilon\varepsilon_0}(N_a d_p^2 + N_d . d_n^2) \tag{31}$$

Using eqns. 25 and 31 and solving for d_n and d_p, we have

$$d_n = \left(\frac{2\varepsilon\varepsilon_0 V_t}{e}\right)^{1/2}\left(\frac{N_a}{N_a N_d + N_d^2}\right)^{1/2} \tag{32}$$

$$d_p = \left(\frac{2\varepsilon\varepsilon_0 V_t}{e}\right)^{1/2}\left(\frac{N_d}{N_a N_d + N_a^2}\right)^{1/2} \tag{33}$$

and d_d, the total depletion layer width, is

$$d_d = d_n + d_p = \left(\frac{2\varepsilon\varepsilon_0 V_t}{e}\right)^{1/2}\left[\left(\frac{N_a}{N_a N_d + N_d^2}\right)^{1/2} + \left(\frac{N_d}{N_a N_d + N_a^2}\right)^{1/2}\right] \tag{34}$$

In practice it is often found that one side of the junction is much more heavily doped than the other. Considering the case where the p-type side is again heavily doped, we have $N_a \gg N_d$ and therefore

$$d_d \doteqdot \left(\frac{2\varepsilon\varepsilon_0 V_t}{eN_d}\right)^{1/2} \tag{35}$$

From eqns. 32 and 33 it is apparent that under these conditions d_n is much greater than d_p and most of the depletion layer lies in the lightly doped region.

The width of the depletion layer depends on the impurity doping and gets narrower as the doping is increased. It also depends on $V^{1/2} = (V_a + V_b)^{1/2}$ and therefore increases when the junction bias is in the reverse direction and decreases when the bias is in the forward direction.

As an example, we calculate the width of the depletion layer in the case of a Ge junction with $N_a \gg N_d$, from eqn. 35.

If all the donors are ionized

$$\sigma_n = \mu_n e N_d \tag{36}$$

Thus

$$d_d = \left(\frac{2\varepsilon\varepsilon_0 \mu_n V_t}{\sigma_n}\right)^{1/2} \tag{37}$$

Permitivity of Free space (ε_0) = 8.854×10^{-14} farad/cm

$\left(\frac{10^9}{4\pi c^2}\right)$

Velocity of light in free space = 2.998×10^{10} cm/sec

For Ge we have the following data:

relative permittivity = 16
electron mobility = 3,900 cm^2/volt sec
conductivity assumed = 0·02 mho/cm
bias + V_d = 0·5 volt

From these figures substituted in eqn. 37 we find

$$d_d = 5{\cdot}24\ \mu$$

THE JUNCTION CAPACITANCE

The depletion layer acts as a parallel-plate capacitor, and its capacitance may be written as

$$C_d = \frac{\mathrm{d}Q}{\mathrm{d}V_t} = \frac{A\varepsilon\varepsilon_0}{d_d} \qquad (38)$$

therefore $$C_d = A\left(\frac{\varepsilon\varepsilon_0 N_a N_d}{2}\right)^{1/2}(N_d + N_a)^{-1/2}V_t^{-1/2} \qquad (39)$$

Numerically for the case we have just considered

$$C_d = 30\ \mu\mathrm{F/m^2}$$

Case 2. Linearly graded junction

The variation of charge-density across the junction is illustrated in Fig. 4.5b. This type of variation is found in many practical junctions at low bias voltages and is therefore of interest.

We have, in the p-region,

$$\frac{\mathrm{d}^2V}{\mathrm{d}x^2} = \frac{ex\theta}{\varepsilon\varepsilon_0} \qquad (40)$$

where θ is the impurity-density gradient $\theta = (N_a + N_d)/l$.

Integrating

$$\frac{\mathrm{d}V}{\mathrm{d}x} = \frac{e\theta x^2}{2\varepsilon\varepsilon_0} + K_1 \qquad (41)$$

Now, $$\frac{\mathrm{d}V}{\mathrm{d}x} = 0 \quad \text{at } x = d_p$$

hence $$\frac{\mathrm{d}V}{\mathrm{d}x} = \frac{e\theta}{2\varepsilon\varepsilon_0}(x^2 - d_p^2) \qquad (42)$$

Integrating again, $$V = \frac{e\theta}{2\varepsilon\varepsilon_0}\left(\frac{x^3}{3} - d_p^2x\right) + K_2$$

But, $$V = 0 \quad \text{at } x = 0, \qquad \therefore\ K_2 = 0$$

and $$V = \frac{e\theta}{2\varepsilon\varepsilon_0}\left(\frac{x^3}{3} - d_p^2x\right)$$

Therefore $$V_p = \frac{e\theta}{2\varepsilon\varepsilon_0}\left(\frac{2d_p^3}{3}\right)$$ $V_t =$

From symmetry we have $$d_p = d_d/2$$

therefore $$V_p = \frac{e\theta}{2\varepsilon\varepsilon_0}\cdot\frac{d_d^3}{12} \qquad (43)$$

$V_n = \frac{e\theta}{2\varepsilon\varepsilon_0}\left(\frac{2d_n}{3}\right)^3$

and $$V_t = \frac{e\theta}{\varepsilon\varepsilon_0}\cdot\frac{d_d^3}{12} \qquad (44)$$

so that $$d_d = \left(\frac{12V_t\varepsilon\varepsilon_0}{e\theta}\right)^{1/3} \qquad (45)$$

From the symmetry of the arrangement,

$$\theta = (N_a + N_d)/l$$

THE JUNCTION CAPACITY

The junction capacity is again $\varepsilon\varepsilon_0A/d_d$, therefore

$$C_d = A(\varepsilon\varepsilon_0)^{2/3}\left(\frac{e\theta}{12V_t}\right)^{1/3} \qquad (46)$$

4.1.6. Diode resistance and diffusion capacity

We derive in this section linear variational models for the junction diode which hold, first, at low frequencies where capacity effects may be neglected and, then, at higher frequencies where diode capacities must be taken into account. The diode current equation is

$$I = I_0\left(\exp\frac{eV_b}{kT} - 1\right)$$

$\therefore \frac{(I + I_0)}{I_0} = \exp\frac{eV_b}{kT}$

Now $$\frac{\mathrm{d}I}{\mathrm{d}V_b} = \frac{1}{r_d} = g_d$$

gives the conductance of the diode, hence

$$g_d = \frac{eI_0}{kT}\exp\frac{eV_b}{kT} = \frac{e(I + I_0)}{kT}$$

or $$r_d = \frac{kT}{e(I + I_0)} \qquad (47)$$

This equation shows that when the diode is reverse biased, $I = -I_0$ and ideally the resistance is infinite. In practice, this is not so because there is always some current flow in the reverse direction due to thermal carrier generation in the depletion region, and there are also surface leakage currents.

In the forward direction $I \gg I_0$ in normal operation and we may write

$$r_d \doteqdot kT/eI \doteqdot 0{\cdot}026/I \text{ ohms} \tag{48}$$

At higher frequencies we must take capacity effects into account. We already have an expression for depletion layer capacity, but there is another effect in the p–n junction which behaves as a capacitative effect and is known as the *diffusion capacity*.

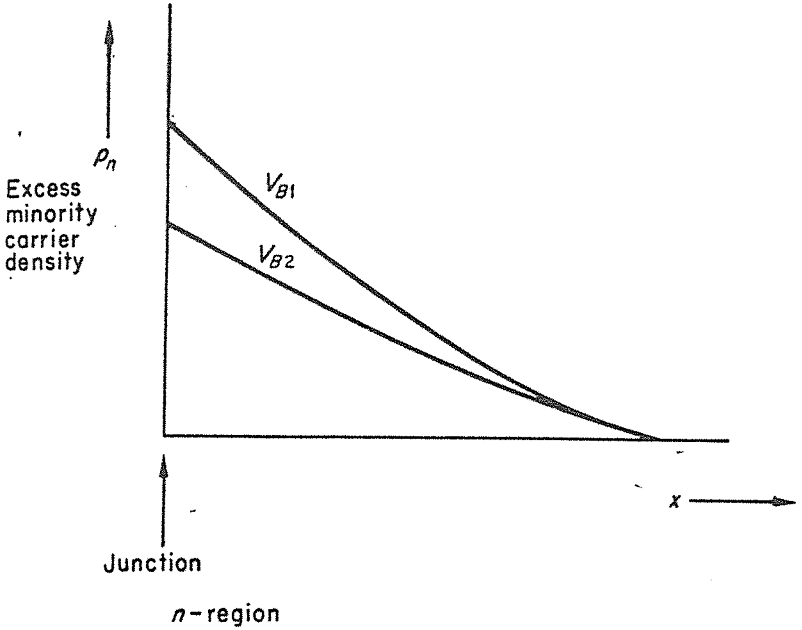


Fig. 4.6 Variation of carrier concentration with forward bias

Consider a forward biased p–n junction. The excess minority carrier (hole) concentration in the n-type side is as shown in Fig. 4.6, for two values of forward bias. The total charge stored is a function of V_b and hence acts as a capacitance. The *excess* holes as a function of distance from the junction may be written as

$$p_n = p_{n0}\left(\exp\frac{eV_b}{kT} - 1\right)\exp\frac{-x}{L_p} \tag{49}$$

as in eqn. 18.

To obtain the total charge enclosed we can integrate e times eqn. 49 from $x = 0$ to $x = \infty$, giving

$$Q_p = ep_{n0}L_p\left(\exp\frac{eV_b}{kT} - 1\right) \tag{50}$$

The current at $x = 0$ is found from eqn. 41a, Chapter 3

$$J_p = -eD_p\frac{dp_n}{dx}$$

$$= eD_p\frac{p_{n0}}{L_p}\left(\exp\frac{eV_b}{kT} - 1\right) \tag{51}$$

From eqns. 50 and 51

$$Q_p = \frac{L_p^2 J_p}{D_p} = \tau_p J_p \tag{52}$$

and

$$C_{p\,\text{diff}} = \frac{dQ_p}{dV_b} = \tau_p\frac{dJ_p}{dV_b} = \frac{eI}{kT}\tau_p = \frac{\tau_p}{r_p} \tag{53a}$$

Similarly for electrons, we have

$$C_{n\,\text{diff}} = \frac{\tau_n}{r_n} \tag{53b}$$

Therefore the total diffusion capacity is

$$C_{\text{diff}} = \frac{\tau_p}{r_p} + \frac{\tau_n}{r_n} \tag{54}$$

If, however, the p-type region is heavily doped, then we can write the diffusion capacity as

$$C_{\text{diff}} \doteqdot \frac{\tau_p}{r_p}$$

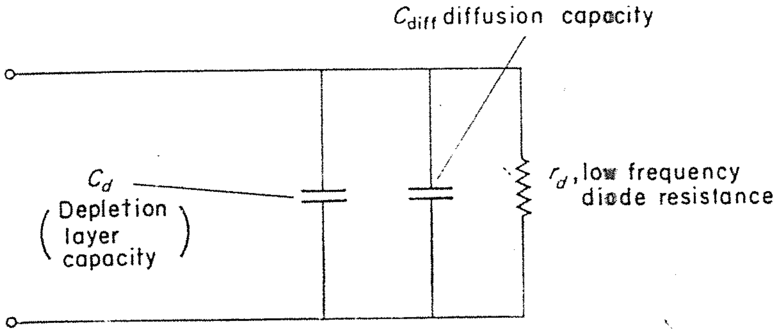


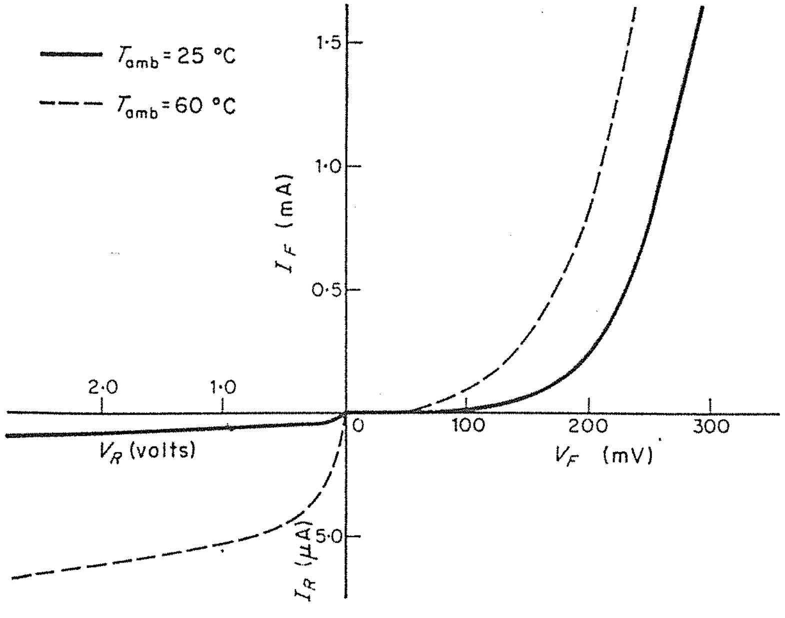
Fig. 4.7 High frequency model for the junction diode with forward bias

Using the results for the depletion-layer capacity, the low-frequency diode resistance and the diffusion capacity, we are in a position to derive a model for the diode which holds at high frequencies as shown in Fig. 4.7. It must be emphasized that the two capacitative components in the model depend on the voltage across the diode and therefore the exact frequency response is very complicated. Ref. 2 gives a more detailed account of p–n junction theory.

4.1.7. *p–n* Junction rectifiers

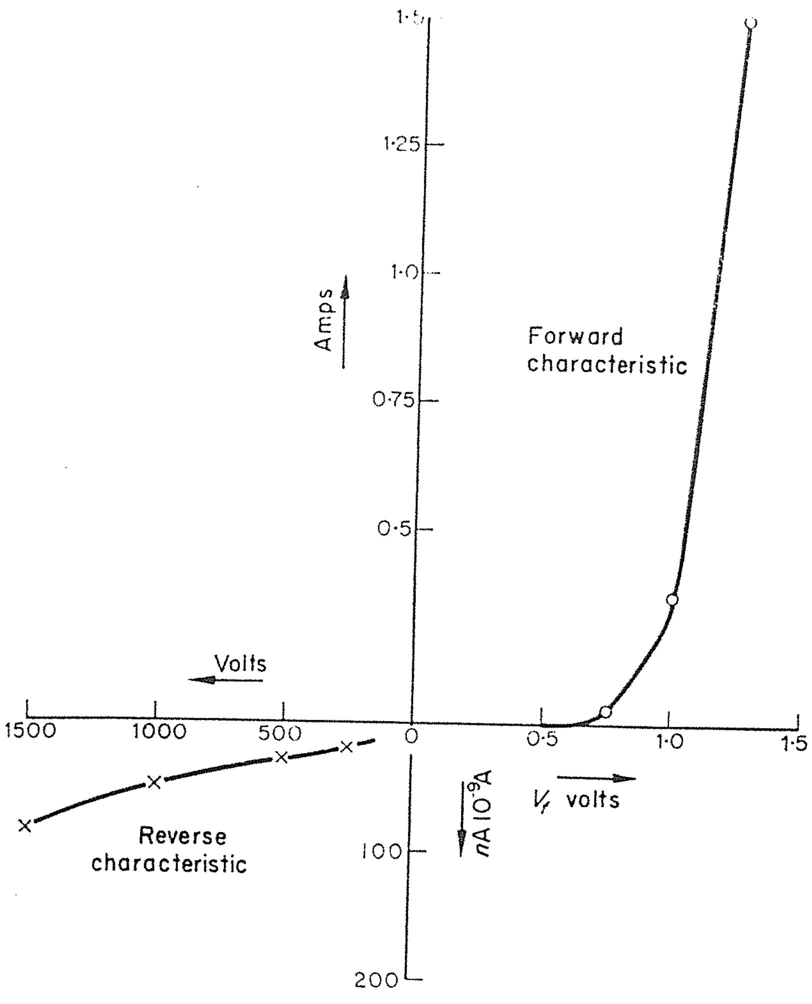
The equations we have derived for the diode current show clearly that the *p–n* junction may be employed as a rectifier with a low forward resistance and a high reverse resistance. Ge and Si diodes are extensively used as rectifiers in power supply circuits. The characteristics of a typical low-power Ge diode and a medium-power Si diode are shown in Fig. 4.8a and b. Ge diodes are employed at low operating power levels where the device temperature does not much exceed 50 °C. Silicon diodes, however, may be operated at temperatures in excess of 100 °C and therefore, with adequate cooling arrangements, at very high power levels. Si diodes capable of delivering several hundred amperes have been constructed.

The forward current characteristic of a *p–n* junction diode fits the curve predicted by the theory we have outlined in this chapter, except at very high forward currents where ohmic drops in the semiconductor material begin to have an appreciable effect. The reverse current characteristic, however, does not agree so well. The theory predicts a reverse current, I_0, which should remain constant for values of reverse voltage more negative than, say, 0·1 V. In practice, the reverse current, especially for Ge diodes, is found to be greater than predicted theoretically and there is an increase with reverse voltage. The increase is thought to be caused by surface leakage, but is not clearly understood. Modern devices have been much improved in this respect.



(a)

Fig. 4.8 (a) Characteristics of Ge *p–n* junction diode at two values of ambient temperature



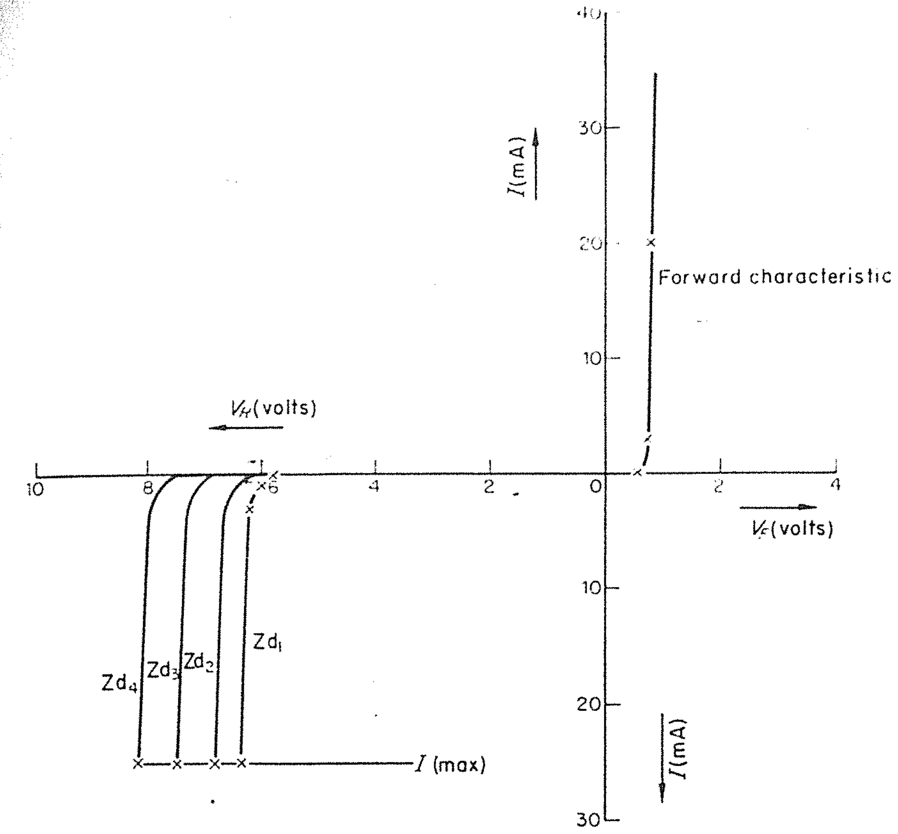
(b)

Fig. 4.8 (b) Characteristic of a Si rectifier

(*a* and *b* from *Mullard Technical Handbook*, vol. 4 (1967) Mullard, London)

The reverse current in Si devices is much lower than in Ge. This is to be expected because of the greater energy gap. There is again a slight increase in reverse current with increasing reverse voltage. In this case the increase is believed to occur because of increased carrier generation in the depletion layer which widens as the reverse bias is raised.

Temperature also has a considerable effect on the reverse current, as can be seen from the characteristics shown in Fig. 4.8a, and care must be taken to ensure that the manufacturer's limiting parameters are not exceeded.



(c)

Fig. 4.8 (c) Breakdown characteristics of four Zener diodes

Finally, there is a critical reverse voltage at which the diode breaks down and conducts heavily in the reverse direction, and obviously this voltage must not be approached in normal rectifier operation. It is, however, a feature of the *p–n* junction which may be usefully employed in other applications. The mechanism of this breakdown and its application are described in the next section.

4.2. Reverse bias breakdown and Zener diodes

When the reverse voltage on any diode approaches a particular value, the current through it begins to increase rapidly with very little further increase in voltage. The junction is said to 'break down' at this point and the current through it has to be limited by the insertion of a series resistance in the circuit. The breakdown characteristic for a diode is shown in Fig. 4.8c. Provided that the current is not allowed to increase to such a value that the junction is destroyed by overheating, the phenomenon is reproducible and is utilized in making semiconductor voltage regulators, called Zener diodes. The breakdown is caused by two different mechanisms known as *avalanche breakdown* and *Zener breakdown*.

4.2.1. Avalanche breakdown

At reverse voltages below the breakdown voltage there are only a small number of charge carriers moving through the depletion layer of a *p–n* junction. The charge carriers collide with some of the unionized atoms but the probability of making ionizing collisions is very small at low voltages. As the reverse voltage is increased, the field across the depletion layer increases and the charge carriers are accelerated and gain in kinetic energy until ionizing collisions become possible. Such collisions release a large number of electron-hole pairs in the depletion region which move apart in the high field and, in turn, are accelerated to make further collisions to produce still more free electrons and holes. At this stage an avalanche process is said to take place and the original small reverse current is multiplied many fold. On an energy level diagram, the effect is represented as a large number of electrons being excited from valence band levels into the conduction band. This type of breakdown can occur at fields of about 250 kV/cm. The voltage at which the breakdown occurs depends on the field strength and on the width of the depletion layer. If one side of the junction is heavily doped, most of the depletion layer is formed in the lightly doped side and the breakdown voltage can be controlled over a wide range by varying the impurity concentration in the highly doped side and thus changing its width.

4.2.2. Zener breakdown

This type of breakdown occurs in diodes which are heavily doped and consequently have a very narrow depletion layer. In this condition very high fields may be created and if this field strength exceeds 1 MV/cm, the diode breaks down and conducts a large current. The mechanism in this case is that covalent interatomic bonds are broken for some of the atoms in the depletion layer and electrons from the valence band jump across the energy gap and enter the conduction band. Zener breakdown depends only on the field strength and not on the path length of free carriers in the depletion layer. Provided that the depletion layer is narrow enough for the probability of ionizing collisions to be negligible, Zener breakdown is the dominant conduction mechanism.

Zener breakdown was thought to be the major factor causing breakdown in *p–n* junctions before the avalanche mechanism was recognized. In fact, since avalanche breakdown occurs at lower field strengths than Zener breakdown, it is often the prime cause. The conditions which determine the type of breakdown are the doping level and the reverse bias and, in some cases, both mechanisms play a significant part in making up the reverse current.

It is believed that Si diodes which break down at voltages below about 5V do so mainly under the influence of Zener effects. Those breaking down at voltages of 10 V or more do so almost entirely by avalanche effects. Diodes which break down at intermediate voltages are probably influenced by both effects.

The breakdown phenomenon is used to make voltage reference diodes known, rather misleadingly, as Zener diodes though most of them break down due to the avalanche process. Zener diodes are available for a large number of reference voltages ranging from a few volts to a few hundred volts.

(1)

4.3. *The junction transistor* (Ref. 3)

The schematic arrangement of an unbiased *p-n-p* junction transistor is shown in Fig. 4.9a and the corresponding potential diagram is shown in Fig. 4.9b. Viewing the transistor from the left, we have first a *p*-type region which is called the emitter, then we have an *n*-type region called the base and, finally, another *p*-type region called the collector. We *assume* that both *p*-type regions are heavily doped and have very high conductivities, and also that the *n*-type region is both comparatively lightly doped and that its width is less than the minority carrier diffusion length.

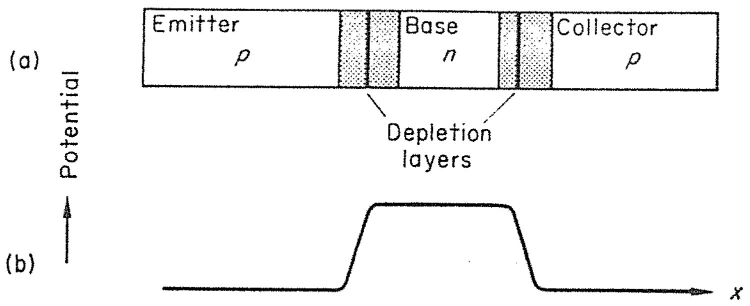


Fig. 4.9 Potential distribution on an unbiased *p–n–p* transistor

With zero applied bias, the Fermi level is continuous through the three regions. There are two depletion layers, one at each of the junctions, between emitter and base, and between base and collector. Equal numbers of holes and electrons cross these junctions in either direction and there is no net current flow.

The situation alters when bias is applied to the transistor, as shown in Fig. 4.10, and transistor action takes place in the following manner.

4.3.1. Emitter-base junction

At the emitter-base junction we have a forward-biased *p–n* junction with the *p*-type material heavily doped so that there is a high injection of holes into the *n*-type base. At the same time, there is a small number of electrons moving from the lightly doped base into the collector region and the total current is the sum of these two separate current components. The hole current, which is almost equal to the total current, plays an important part in the operation of the device, while the electron current does not have a useful role. For the emitter-base junction, an *emitter efficiency*, γ, is defined as the ratio

$$\frac{\text{hole current injected by emitter into base}}{\text{total current flow from emitter to base}}$$

This efficiency is very close to unity because the heavy doping of the *p*-type region ensures that there are many more holes being injected into the base than there are electrons capable of moving from base to emitter.

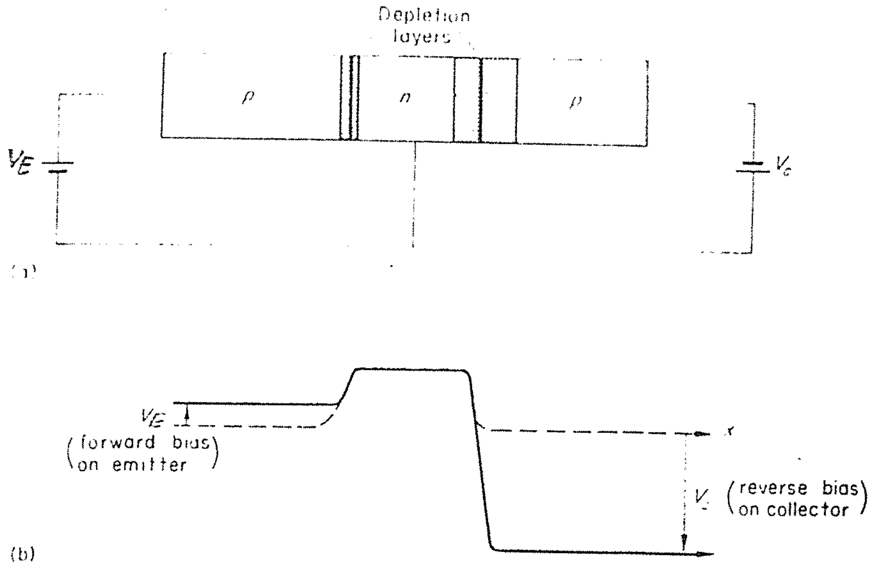


Fig. 4.10 Potential distribution on a normally biased *p–n–p* transistor

4.3.2. Base region

The holes injected into the base region have a high concentration at the injection plane near the emitter. Hence they diffuse into the base in order to try to reach a uniform concentration. (We assume there are no fields in the base region.) Holes which manage to reach the base-collector junction encounter a field which sweeps them out of the base into the collector. Those unable to reach the collector recombine with electrons in the base region, consequently there is a flow of electrons into the base region continuously replenishing the electrons lost by recombination. To express the fact that not all the injected holes reach the collector, we define a *transport factor*, β, as the ratio

$$\frac{\text{holes reaching collector}}{\text{holes injected into base}}$$

This factor, again, is only slightly less than unity in a well-designed transistor. It depends on the lifetime of holes in the *n*-type region and on the geometry of the transistor. The influence of geometry may be understood by considering that the holes diffuse in all directions once they enter the base region. Some are lost by recombination within the bulk base material and there is also substantial recombination at the surface of the base where the recombination probability is much higher. To obtain a good transport factor, the collector must be designed to have a much greater area than the emitter. Figure 4.11 illustrates the advantage of having a large collector.

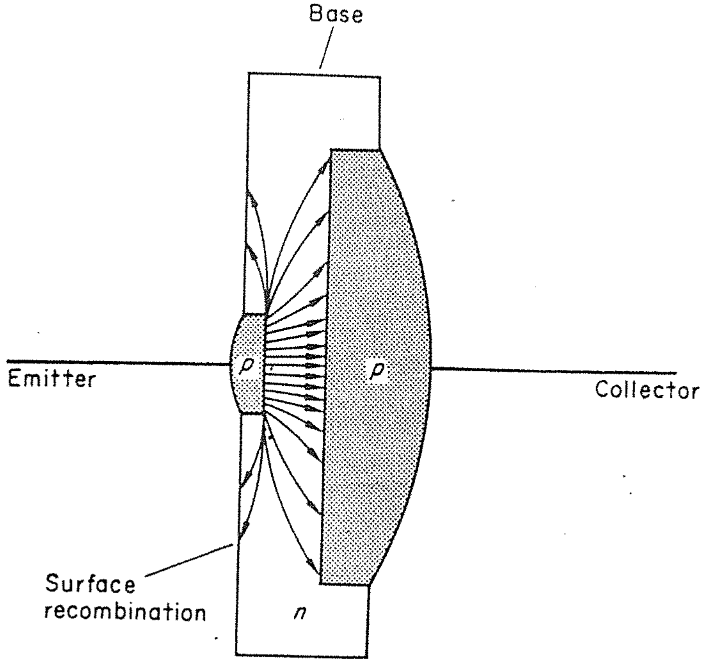


Fig. 4.11 Schematic diagram illustrating the large collector area to ensure a high transport factor in the forward direction

4.3.3. Base-collector junction

We have stated earlier that the direction of the field in the depletion layer at the base-collector junction is such that it sweeps the holes arriving at the junction into the collector. If this field is sufficiently high, there can be an enhancement in the number of holes arriving at the collector, because the avalanche type of carrier multiplication described in section 4.2.1. can occur. Hence we define a collector efficiency, δ, as the ratio

$$\frac{\text{total collector current}}{\text{incident hole current}}$$

تيار التجمع الكلي
تيار الفجوات الداخل

This factor may have a value greater than one if the collector to base voltage is very high, but it is usually close to unity at normal collector voltages.

4.3.3.1. *Collector cut-off current*

Collector cut-off current is the current which flows from the collector to base when the emitter is open-circuit, and no current is injected from it. We have at the collector, under these conditions a reverse biased p–n junction. The barrier height of this junction is $V_d + V_c$ and the depletion layer extends mainly into the n-region which is lightly doped. Thermally-generated holes move from the base to the collector and give a small current flowing from collector to base. This current is known as the collector cut-off current I_{co}. The other charge-carrier movements at the junction make negligible contributions to this current; e.g. the electrons in the n-type base are unable to surmount the high barrier and the electrons in the collector are so few in number that their effect may be discounted.

4.3.4. The current flow equations and the current amplification factor

The current amplification factor for transistors is defined as the ratio of the collector current to the emitter current. More exactly, α_0 must be written as

$$\alpha_0 = \left(\frac{\partial I_c}{\partial I_e}\right)_{V_c\,\text{const.}} \tag{55}$$

since I_c is not a linear function of I_e when V_c is a parameter. Except at very high voltages, the value of α_0 is slightly less than unity.

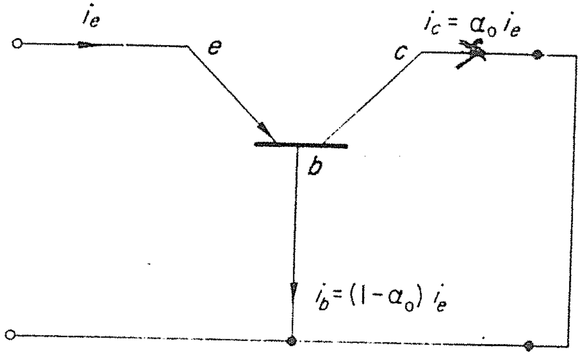


Fig. 4.12 Current directions in a transistor for a linear variational model

α_0 may also be written in terms of the parameters we have just defined as

$$\alpha_0 = \gamma\beta\delta = f(I_e, V_c) \tag{56}$$

If linear relationships are assumed between the transistor currents, we have the following equations

$$I_c = \alpha_0 I_e + I_{co} \tag{57}$$

and

$$I_b = (1 - \alpha_0)I_e - I_{co} \tag{58}$$

The current directions used in these equations are as shown in Fig. 4.12.

Another current amplification factor, β_0, is defined for a common-emitter circuit shown in Fig. 4.13

$$\beta_0 = \left(\frac{\partial I_c}{\partial I_b}\right)_{V_c \text{ const.}} = \frac{\alpha_0}{1-\alpha_0} \tag{59}$$

This shows that α_0 must be made very nearly equal to one if high current gain is to be obtained in the common-emitter circuit.

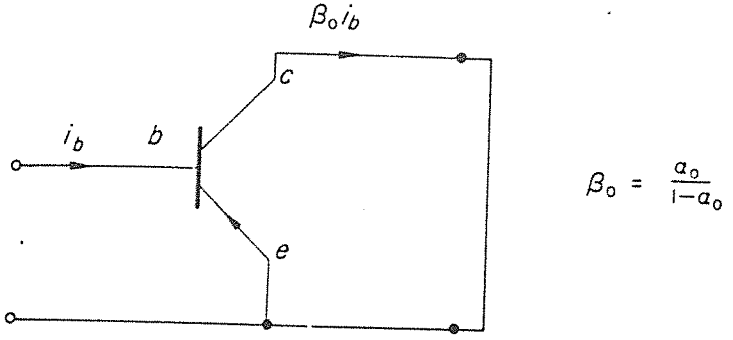


Fig. 4.13 Common-emitter configuration of a *p–n–p* transistor

4.3.5. Analysis of the junction transistor at low frequencies

An expression can be derived for the current gain of a transistor in terms of measurable physical parameters of the semiconducting materials (Ref. 3). From this expression it is possible to derive a model for the transistor in which passive circuit elements and linear voltage and current generators are used to represent the physical action. The model applies as long as the current and voltage variations are restricted to linear incremental variations at low frequencies. In carrying out the analysis we assume that the following conditions apply:

(a) there are electric fields in the depletion region only. All other regions are field-free and have very high conductivities;
(b) carrier flow through the base region is by diffusion only;
(c) there are abrupt junctions between emitter and base, and between base and collector;
(d) the temperatures are low enough for the numbers of thermally generated minority carriers to be very much smaller than the numbers of injected carriers. This condition avoids the need to use excess carrier concentrations in deriving the equations.

4.3.6. The diffusion equation for the base region

The diffusion equation for the base region may be written in the usual form

$$D_p \frac{d^2p}{dx^2} - \frac{dp}{dt} - \frac{p}{\tau_p} = 0 \tag{60}$$

At low frequencies we may assume dp/dt is negligible and write

$$\frac{d^2p}{dx^2} - \frac{p}{L_p^2} = 0 \tag{61}$$

The general solution is

$$p = A \exp -\frac{x}{L_p} + B \exp \frac{x}{L_p} \tag{62}$$

where A and B are arbitrary constants to be determined by applying boundary conditions, illustrated in Fig. 4.14. These are $p = 0$ at $x = W$, since holes reaching the collector are swept away by the field in the collector depletion layer, and

$$p = \hat{p}_{nb} = p_{nb}\left(\exp\left(eV_e/kT\right) - 1\right) \quad \text{at } x = 0$$

P_{n} → P_{n} $p = p_{nb}\left[\exp\left(\frac{eV}{kT}\right) - 1\right]$

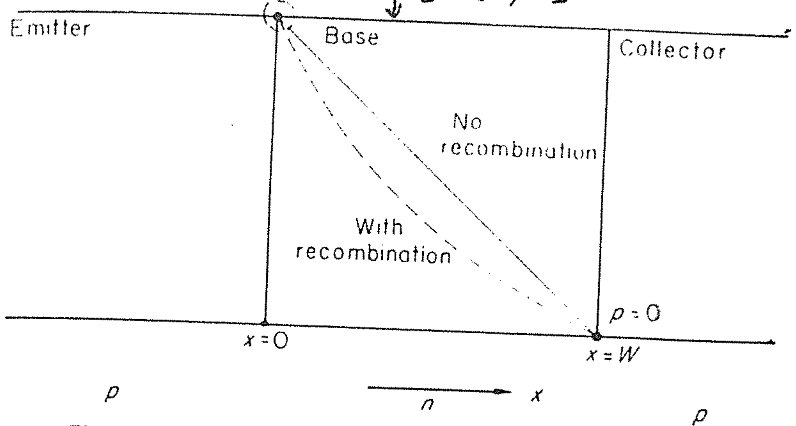


Fig. 4.14 Variation of hole density in the base of a transistor

Applying the conditions, we have

$$A = -B \exp 2W/L_p$$

and

$$B = \hat{p}_{nb}/(1 - \exp 2W/L_p)$$

hence

$$p = \frac{\hat{p}_{nb}}{1 - \exp\frac{2W}{L_p}}\left(-\exp\frac{2W}{L_p}\exp -\frac{x}{L_p} + \exp\frac{x}{L_p}\right) \tag{63}$$

or

$$p = \frac{\hat{p}_{nb}}{\sinh\frac{W}{L_p}} \sinh\left(\frac{W-x}{L_p}\right) \tag{64}$$

n_{pe} is the minority [illegible]

$n_{pe} \rightarrow$ emitt

From eqn. 41a, Chapter 3, we have $I_p = -eD_p\,\mathrm{d}p/\mathrm{d}x$, therefore

$$I_p = \frac{ep_{nb}D_p}{L_p}\,\frac{\cosh\left(\dfrac{W-x}{L_p}\right)}{\sinh\left(\dfrac{W}{L_p}\right)} \tag{65}$$

This equation gives the hole current in the base as a function of distance.

4.3.7. Electron current from emitter to base at $x = 0$

Electron current from emitter to base means, of course, a flow of electrons from the base to the emitter. The expression for this current was derived in the case of the p–n junction and is given by eqn. 20. Assuming a junction area $A = 1$

$$I_{n_{x=0}} = \frac{eD_n n_{pe}}{L_n}\left(\exp\frac{eV_e}{kT} - 1\right) \tag{66}$$

This is a small quantity because n_{pe} is usually small when the p-type emitter is heavily doped.

حقن - ادخال

4.3.8. Emitter injection efficiency

From the discussion and definition of emitter efficiency given in section 4.3.1, we have

$$\gamma = \left(\frac{I_p}{I_p + I_n}\right)_{x=0} \tag{67}$$

$$= \left(\frac{1}{1 + \dfrac{I_n}{I_p}}\right)_{x=0} \tag{68}$$

Therefore, using eqns. 65 and 66

$$\gamma = \frac{1}{1 + \dfrac{D_n L_p n_{pe}}{D_p L_n p_{nb}}\tanh\left(\dfrac{W}{L_p}\right)} \tag{69}$$

Normally, in a good transistor $W \ll L_p$ and we can write

$$\gamma \doteqdot \frac{1}{1 + \dfrac{D_n W n_{pe}}{D_p L_n p_{nb}}} \tag{70}$$

This expression for the injection efficiency is always made as close as possible to unity. This is achieved, first, by making the base width W as small as possible compared with L_n and, secondly, by making the ratio n_{pe}/p_{nb}, i.e. the ratio of equilibrium electron density in the p-type emitter to the equilibrium hole density in the n-type base, as small as possible by doping the emitter more heavily than the base.

$\beta = \dfrac{\text{عدد الفجوات التي تصل المجمع}}{\text{عدد الفجوات الخارجة من الباعث}}$

4.3.9. The transport factor

Using eqn. 65 we may find the hole current reaching the collector by putting $x = W$

$$I_{p_{x=W}} = \frac{eD_p}{L_p}\,\frac{p_{nb}}{\sinh\left(\dfrac{W}{L_p}\right)} \tag{71}$$

The transport factor β is then given by

$$\beta = \frac{I_{p_{x=W}}}{I_{p_{x=0}}} = \frac{1}{\cosh\dfrac{W}{L_p}} \tag{72}$$

$\cosh\frac{W}{L_p} = 1 + 2\sinh^2$

Usually, $W \ll L_p$ and we may expand $\cosh W/L_p$ to give

$$\beta \doteqdot \frac{1}{1 + 2\dfrac{W^2}{L_p^2}} \tag{73}$$

The transport factor is made very near unity by making the base as narrow as possible and it is always greater than 0·9 in acceptable transistors.

4.3.10 The collector current and collector efficiency

The collector current is the sum of the hole current flowing into the collector from the base and the small current due to electrons moving from the base into the collector across the reverse-biased junction. This current is given by

$$I_n = \frac{eD_n n_{pc}}{L_n}\left(\exp -\frac{eV_c}{kT} - 1\right) \tag{74}$$

or

$$I_n \doteqdot \frac{eD_n n_{pc}}{L_n} \tag{75}$$

The charge concentration and current flow in the three regions of the transistor may be illustrated as in Fig. 4.15. The difference between the total emitter and collector currents represents the current lost by recombination of carriers.

We stated earlier that the avalanche type of carrier multiplication can cause an increase in collector current when V_c is increased. It has been shown empirically that the factor which gives this multiplication ratio is given by

$$\delta = \frac{1}{1 - \left(\dfrac{V_c}{V_{\text{breakdown}}}\right)^m} \tag{76}$$

where $m \doteqdot 3$ for alloyed p–n–p transistors. The breakdown voltage is usually much greater than the operating collector voltage (for reasons that become obvious in the next section) and therefore δ is very close to unity.

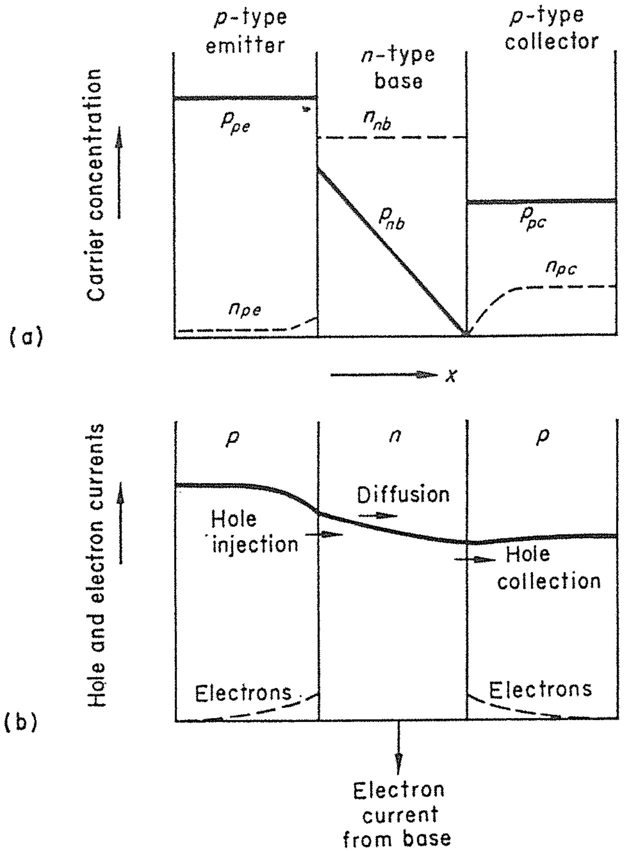


Fig. 4.15 (a) Carrier concentrations in the emitter, base and collector of a diffusion transistor. (b) Currents in a diffusion transistor

4.3.11. Base-width modulation (the Early effect)

In the analysis so far, the base width has been assumed to be constant. In fact, base width is a function of the collector voltage, V_c, and the assumption is only valid for a fixed V_c. This phenomenon was first described by J. Early (Ref. 4) and is sometimes referred to as the Early effect.

This effect may be understood by realizing that the base width used in the analysis is not the actual physical dimension of the base material, but is defined as the extent of the field-free region between the depletion layers formed at each junction of the transistor.

It was shown in section 4.1.5 that the width of the reverse-biased junction formed at the collector depends on the voltage applied between collector and base and that the width increases with increasing bias voltage. The depletion

layer then extends further into the base region, and, as a result, the field-free region, W, decreases. Conversely, if the bias decreases, the base width increases. The effect is illustrated in Fig. 4.16. For low values of bias voltage, the depletion layer is narrow and the base width is W′, and for a higher value of bias, the depletion layer extends further into the lightly-doped base region and the base width decreases to W″.

The expressions we obtained for both the injection efficiency and the transport factor were functions of the base width and therefore α_0 is also a function of W, and hence of the collector bias voltage.

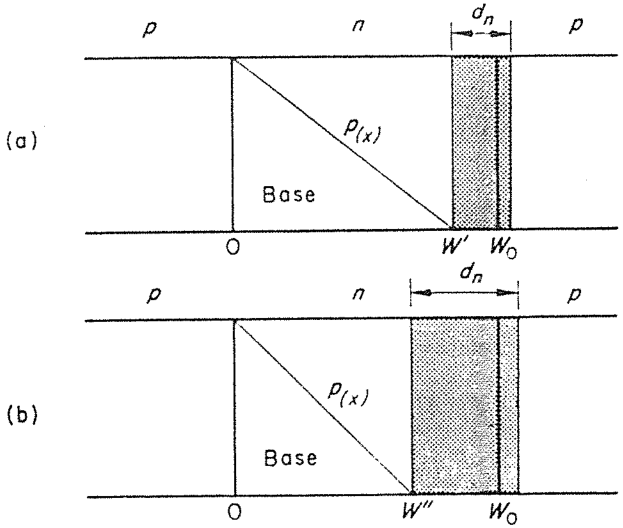


Fig. 4.16 Base-width modulation: (a) low bias voltage; (b) high bias voltage

A limiting condition can be reached for the base width when the collecto voltage is high enough for the depletion layer to extend right into the bas until it is almost at the inner boundary of the emitter-base depletion layer. I these circumstances, 'punch through' is said to occur. A large current flow from emitter to base and the conventional form of transistor action ceases The voltage at which 'punch through' may take place can be determined b considering eqn. 34, which gives the depletion layer width at a p-n junctio If the p-type region is heavily doped, we have

$$d = \left(\frac{2\varepsilon\varepsilon_0 V}{eN_d}\right)^{1/2}$$

When V_c reaches the critical punch-through value, $d = W'$ and V_{pt} is given b

$$V_{pt} = \frac{eN_d W'^2}{2\varepsilon\varepsilon_0} \qquad (7$$

The effect of base-width modulation on α_0 may be understood by considering the influence on β and γ. There is an improvement in γ with increasing V_c because the decrease in W causes an increase in $\mathrm{d}p/\mathrm{d}x$, whereas $\mathrm{d}n/\mathrm{d}x$ is almost unchanged. The transport factor β improves also because the shorter W means a small recombination loss in the base region. α_0 rises initially with increasing V_c, then remains almost constant over a range of collector voltages and then begins to rise slowly as the Early effect comes into play, and finally rises very rapidly as avalanche effects begin to dominate the operation of the transistor. Further increase of V_c causes the transistor to break down.

4.3.12. A linear variational model for the transistor applicable at low frequencies

The physical behaviour of the transistor described so far leads us to a model which holds for low-frequency linear variation in the device currents and voltages. Before we present this model, some further parameters which describe the physical nature of the transistor must be considered.

4.3.12.1. The emitter resistance

The concept of emitter resistance follows from reasoning similar to that used for deriving the low-frequency diode resistance of section 4.1.6. The relationship between current and voltage is non-linear and depends on conditions at the collector end of the device, hence we define the emitter resistance in terms of partial derivatives as

$$r_e = \frac{1}{g_e} = \left(\frac{\partial V_e}{\partial I_e}\right)_{V_c} \tag{78}$$

Now

$$I_e = I_0\left(\exp\frac{eV_e}{kT} - 1\right)$$

hence

$$g_e = \frac{eI_0}{kT}\exp\frac{eV_e}{kT} \tag{79}$$

normally with the emitter junction forward biased $I_e \gg I_0$, therefore

$$g_e \doteqdot \frac{eI_e}{kT} \tag{80}$$

or

$$r_e = \frac{kT}{eI_e} \doteqdot \frac{26}{I_e}\,\Omega, \quad \text{for } I_e \text{ in m}A \tag{81}$$

4.3.12.2. The base resistance

The flow of holes across the n-type base region is always accompanied by some recombination with free electrons. This was mentioned in section 4.3.2 where it was also pointed out that the loss of holes may be minimized by careful design of the emitter and collector geometry. The electrons lost through recombination are replenished by a flow of electrons into the base through the external base lead contact. In addition, the difference between the electron currents at the emitter and collector must also appear as an electron current in the base (see Fig. 4.15b). This total electron current flows from points within the base to the base lead contact and since the base material has very small dimensions and an appreciable resistivity, there is a noticeable voltage drop in the base regions.

The result is that the base region is not entirely field-free and the equations we derived in section 4.3.6 are not exact. However, for low injected current densities, voltage drops in the base may be neglected and a field-free base may be assumed.

The other effect which must be taken into consideration is that the voltage drop across the emitter and collector junctions cannot be assumed to be equal to the applied external battery voltages. The voltage drop across the base resistance must be taken into account and V_e, V_c, the voltage across the two junctions, cannot actually be measured, but are hypothetical voltages between the emitter and some point within the base, and between the collector and some point within the base.

The actual magnitude of the base resistance depends on the geometry of the transistor and an exact analytical calculation is difficult, but fairly good estimates are possible with simple geometries. The grown junction transistor is fairly straightforward to deal with and a good estimate is also possible for the more common alloyed-junction type using a simplified geometry as a model. In low-power transistors the value of the base resistance lies between 20 and 100 ohms and efforts are made in the design to keep the base resistance to a minimum.

4.3.12.3. Circuit representation of base-width modulation

The Early effect of base widening described in section 4.3.11 has two consequences which must appear in any model we attempt to construct for the transistor: (a) the emitter current changes as a function of the collector voltage when the emitter voltage is constant, and (b) the collector no longer appears to be a current source of infinite impedance, but has a finite resistance.

4.3.12.4. Emitter feedback

The emitter current variation is usually represented in any model for the transistor by a feedback current generator across the emitter junction, generating a current proportional to the collector voltage. The effect can be understood with reference to Fig. 4.16. The value of the hole density remains constant with V_c but $\mathrm{d}p/\mathrm{d}x$ changes and hence the emitter current varies. The emitter feedback conductance is defined as

$$g_{ef} = \left(\frac{\partial I_e}{\partial V_c}\right)_{V_e} \tag{82}$$

which may be written as

$$g_{ef} = \frac{\partial I_e}{\partial W}\cdot\frac{\partial W}{\partial V_c} \tag{83}$$

If recombination in the base is neglected, we can write

$$-\frac{\mathrm{d}p}{\mathrm{d}x} = \frac{p_{x=0}}{W}$$

assuming a linear relationship between hole density distribution and base width. Hence we may write

$$I_e = AeD_p \frac{p_{x=0}}{W} \tag{84}$$

and

$$\frac{\partial I_e}{\partial W} = AeD_p p_{x=0} \left(-\frac{1}{W^2}\right) \tag{85}$$

$$= -\frac{I_e}{W} \tag{86}$$

now $W = W' - d_n$ (see Fig. 4.16) and therefore

$$\frac{\partial W}{\partial V_c} = -\frac{\partial (d_n)}{\partial V_c} \tag{87}$$

Assuming the p-type side to be more heavily doped, we have

$$-\frac{\partial (d_n)}{\partial V_c} = -\left(\frac{\varepsilon\varepsilon_0}{2eN_d}\right)^{1/2} \frac{1}{V_c^{1/2}} \tag{88}$$

Combining eqns. 86 and 87 we get

$$g_{ef} = \frac{I_e}{W}\left(\frac{\varepsilon\varepsilon_0}{2eN_d V_c}\right)^{1/2} \tag{89}$$

from which W may be eliminated by use of eqn. 35 to give

$$g_{ef} = \frac{I_e \left(\dfrac{\varepsilon\varepsilon_0}{2eN_d V_c}\right)^{1/2}}{W' - \left(\dfrac{2\varepsilon\varepsilon_0 V_c}{eN_d}\right)^{1/2}} \tag{90}$$

4.3.12.5. *Collector resistance* .

The collector resistance is defined by the partial derivative

$$r_c = \left(\frac{\partial V_c}{\partial I_c}\right)_{I_e} = \frac{1}{g_c} \tag{91}$$

or

$$g_c = \frac{\partial (\alpha I_e)}{\partial V_c} = I_e \frac{\partial \alpha}{\partial V_c} \tag{92}$$

therefore

$$g_c = I_e \frac{\partial \alpha}{\partial W} \cdot \frac{\partial W}{\partial V_c} \tag{93}$$

The restriction in the definition that the emitter current must remain constant despite changes in V_c means that the slope of the $p(x)$ versus x curve must remain constant at $x = 0$. This condition modifies the hole distribution in the manner shown in Fig. 4.17.

We also assumed in our definition of α_0 at this stage that the emitter injection efficiency and the collector multiplication remain constant and equal unity. This restricts us to low values of collector voltage, well below avalanche breakdown values, and to low levels of current injection at the emitter. Under these conditions we have

$$g_c \doteqdot I_e \frac{\partial \beta}{\partial W} \cdot \frac{\partial W}{\partial V_c} \tag{94}$$

Now eqn. 73 gives

$$\beta = 1 - \frac{1}{2}\left(\frac{W}{L_p}\right)^2$$

hence

$$g_c \doteqdot -I_e \frac{W}{L_p^2} \cdot \frac{\partial W}{\partial V_c} \tag{95}$$

Using eqn. 88 we get

$$g_c \doteqdot I_e \frac{W}{L_p^2}\left(\frac{\varepsilon\varepsilon_0}{2eN_d V_c}\right)^{1/2} \tag{96}$$

and, as before, W may be expressed in terms of d_n and W'.

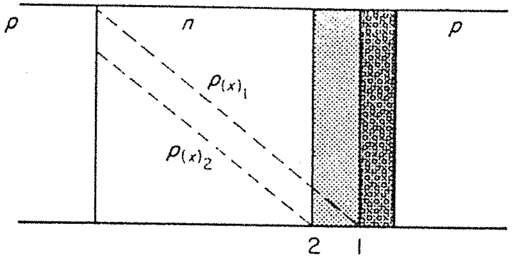


Fig. 4.17 Concentration of holes in the base with emitter current constant and V_{cb} varying

We are now in a position to present a model for the transistor using the circuit components we have obtained from a consideration of the physics of the device. It must be emphasized again that this is only one model for the transistor which holds under the restricting conditions of low-frequency linear increments in device currents and voltages.

We summarize first the parameters we must include in our model: The emitter resistance, r_e; the emitter feedback current generator, $g_{ef} V_c$; the base resistance, r_b; the current amplification factor, α; the collector conductance, g_c; using these parameters we obtain the model shown in Fig. 4.18.

This circuit model is *not* the most convenient form for use by circuit designers, but has the feature that it includes all the main physical features of the device discussed in the preceding sections of this chapter.

A more convenient form of the circuit model is shown in Fig. 4.19. This model is known as the equivalent Tee circuit for the transistor, and is much used in the design and analysis of transistor circuits. The values of the circuit elements used in this model can be calculated by writing down the mesh equafeedback current generator

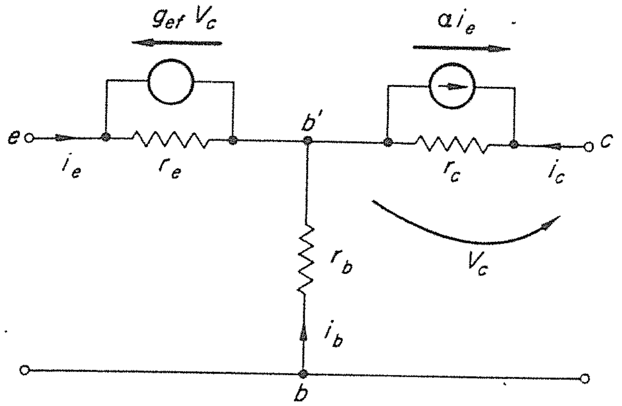


Fig. 4.18 Low frequency linear variational model for the transistor

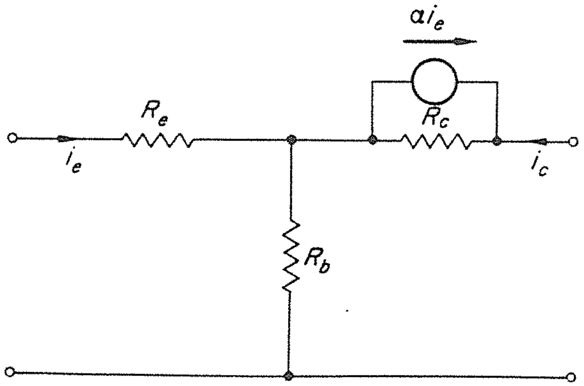


Fig. 4.19 Tee circuit model for the transistor

tions for both the circuits of Figs. 4.18 and 4.19 and comparing terms. The result gives

$$R_b = r_b + g_{ef} r_e r_c \tag{97}$$

$$R_e = r_e - g_{ef} r_e r_c (1 - \alpha) \tag{98}$$

$$R_c = r_c (1 - g_{ef} r_e) \tag{99}$$

$$a = \alpha \tag{100}$$

These expressions may be simplified by the use of suitable approximations

$$g_{ef} \cdot r_c = \left(\frac{L_p}{W}\right)^2 \tag{101}$$

$$\alpha \doteqdot \beta = 1 - \frac{1}{2g_{ef} \cdot r_c} \tag{102}$$

hence

$$R_b \doteqdot r_b + g_{ef} r_e^2 \tag{103}$$

$$R_c \doteqdot r_c \tag{104}$$

$$R_e \doteqdot r_e/2 \tag{105}$$

$$\alpha \doteqdot a \tag{106}$$

4.3.13. High-frequency model

The model of the transistor described in the last section does not prove adequate at high frequencies where a number of effects cause the performance of the transistor to deteriorate. The major effects are:

(a) the charge storage at the emitter–base junction which is similar to the effect already described in section 4.1.6 for the diode. This charge storage results in the addition of a diffusion or storage capacitor at the emitter in our model of Fig. 4.18;
(b) the current amplification factor becomes a function of frequency and its value decreases with increasing frequency.

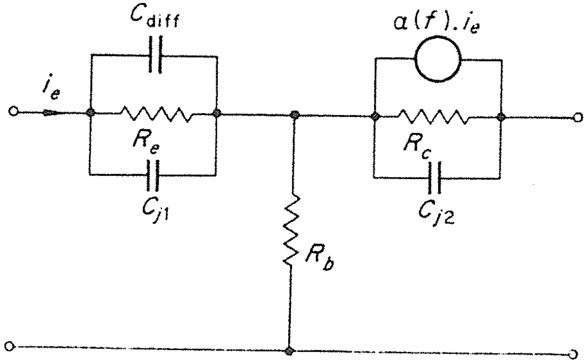
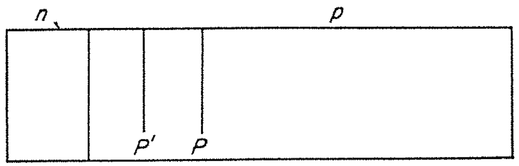


Fig. 4.20 High frequency linear variational model

The latter effect arises because of the slow transit time of minority charge carriers through the base of the transistor. As described in section 4.3.6, the transit is by the relatively slow process of diffusion and at high frequencies the gain falls. An approximate circuit model incorporating the high-frequency effects is shown in Fig. 4.20. It includes, in addition to the components described in the low-frequency model, a diffusion capacity across the emitter, a depletion-layer capacity at the emitter and, at the collector, an α which is shown as a function of frequency. However, the feedback term, $g_{ef}V_c$ which appears in the low-frequency model, is omitted because the feedback effects of C_c and r_b completely overshadow its effect. Ref. 5 discusses transistor models in detail.

Questions

1.



A small rectangular rod of germanium contains a plane p–n junction near one end, as shown in the figure. The junction is biased in the reverse direction and the resulting current is found to be 1 μA. A line of light, parallel to the junction, is focused onto the surface of the rod as at P in the figure and this causes the current to increase to 54 μA. When the line of light is moved a distance of 6×10^{-4} m nearer the junction, the current increases still further to 107 μA. The temperature throughout was 300 °K.

If the mobility of electrons in germanium is $0{\cdot}38$ m²/Vs, what is the mean lifetime of electrons in the p-portion of the rod?

2. In a plane p–n junction the p-region has 3×10^{21} acceptors/m³ and the n-region has 10^{21} donors/m³. A reverse bias of 50 V is applied to the junction.

Estimate the maximum electric field strength in the device. The relative permittivity of germanium is 16 and it may be assumed that the transition from the p-region to the n-region is abrupt.

$$\varepsilon_0 = 8{\cdot}85 \times 10^{-12} \text{ farad/m}$$

3. Show that the current density J flowing across an infinitely extended, infinitely thin, plane p–n junction is given by

$$J = J_s[\exp(eV/kT) - 1]$$

where V = bias voltage
T = absolute temperature
J_s = a constant

Calculate the incremental resistance of the junction for $V = 0{\cdot}1$ V, $T = 290$ °K and $J_s = 1$ A/m² and comment on the practical significance of this component of the total resistance. *(Grad. I.E.E.)*

4. A p–n junction in germanium is formed by a p-material having 10^{15} acceptor atoms/cm³ and an n-material with 10^{14} donor atoms/cm³. Evaluate the difference in electrostatic potential ψ_0 across this junction in the equilibrium condition at 300 °K, assuming that all the impurity atoms are fully ionized. The distribution function for the electron energies can be taken as $f(E) = \exp[-(E - E_F)/kT]$, the available energy levels for electrons in the conduction band can be represented by $N_c = 2{\cdot}5 \times 10^{19}$ levels/cm³ situated at the bottom of the conduction band, $kT = 0{\cdot}026$ eV at 300 °K, E_F is the Fermi level and the energy gap for germanium is $0{\cdot}72$ eV. *(L.U., B.Sc. Eng. Elec.)*

5. Explain in detail the operation of a p–n–p transistor, indicating the mechanism by which the current traverses the base region, and stating the factors that control the emitter, base and collector currents. Draw diagrams showing the minority-carrier concentrations in the emitter, base and collector regions as a function of distance.

If width and area of the base of a p–n–p transistor are W and A, respectively, and the hole lifetime in the base region is τ_p, derive an expression for the component of base current due to hole recombination. It may be assumed that the hole concentration in the base region near the emitter–base junction is given by

$$p_b(0) = p_{b0} \exp(eV/kT)$$

where p_{b0} = equilibrium hole concentration in the base
V = emitter–base-voltage,

the other symbols having their usual meaning.

What other components of base current are there? *(Grad. I.E.E.)*

Bibliography

1. BARDEEN, J. and BRATTAIN, W. H. (1948) *Phys. Rev.*, **74**, 230.
2. MOLL, J. L. (1964) *Physics of Semiconductors.* McGraw-Hill, New York and Maidenhead.
3. SHOCKLEY, W. (1949) *B.S.T.J.*, **28**, 435.
4. EARLY, J. M. (1952) *Proc. Inst. Radio Engrs*, **40**, 1401.
5. DEWITT, D. and ROSSOFF, A. L. (1957) *Transistor Electronics.* McGraw-Hill, New York and Maidenhead.
6. MIDDLEBROOK, R. D. (1957) *An Introduction to Junction Transistor Theory.* Wiley, New York and London.
7. HUNTER, L. P. (1966) *Introduction to Semiconductor Phenomena and Devices.* Addison-Wesley, Reading, Mass. and London.
8. GRAY, P. E. *et al.* (1964) *Physical Electronics and Circuit Models of Transistors.* Wiley, New York and London.