

If there are a total of n electrons per m^3 in the metal, all with constant drift velocity \mathbf{v}_d , we have the following relation for the net electric current density \mathbf{j} :

$$\begin{aligned}\mathbf{j} &= -nev_d = \left(\frac{ne^2\tau}{m} \right) \mathbf{E} \\ &= \sigma \mathbf{E}\end{aligned}\quad (6.2)$$

where the positive scalar quantity

$$\sigma = \frac{ne^2\tau}{m} \quad (6.3)$$

is defined as the electrical conductivity (the reciprocal of the electrical resistivity ρ). The form of the expression (6.3) remains the same in all models including those based on quantum physics. The difference lies only in the way n , m , and τ are defined.

We are now beginning to accomplish our aim of highlighting the gains of the Drude model. The first and the foremost of these is the derivation of the Ohm's law contained in (6.2). The law was initially set up purely on an empirical basis.

It must be, however, stated that the electrical conductivity σ is not universally scalar since in some complicated situations \mathbf{j} becomes nonlinear with respect to \mathbf{E} making σ to behave as a tensor. The electrical conductivity is often expressed in terms of the drift mobility of electrons μ as

$$\sigma = ne\mu \quad (6.4)$$

where

$$\mu = \frac{|\mathbf{v}_d|}{\mathbf{E}} \quad (6.5)$$

Drude put (6.3) in a further useful form by exploiting the ideas of kinetic theory such as to define τ in terms of thermal velocity v_{th} expressed by the following relations:

$$\tau = \frac{\Lambda}{v_{\text{th}}} \quad (6.6)$$

$$\frac{3}{2} k_B T = \frac{1}{2} m v_{\text{rms}}^2 = \frac{1}{2} m v_{\text{th}}^2 \quad (6.7)$$

where Λ is the electron mean free path. Using these relations, (6.3) can be written as

$$\sigma = \frac{ne^2\Lambda}{(3mk_B T)^{1/2}} \quad (6.8)$$

The relation (6.8) is another achievement of the Drude theory since it gives the right magnitude of electrical conductivity and correctly describes that conductivity increases with decrease in temperature (see Fig. 6.1). Experimental values at 77, 273 and 373 K are given in Table 6.1 for comparison. But the dependence on $T^{-1/2}$ is not in agreement with the observed T^{-1} dependence in

Table 6.1 Experimental electrical conductivities* of some metals (in $10^8 \text{ ohm}^{-1} \text{ cm}^{-1}$)

Metal	<i>At</i> 77 K	<i>At</i> 273 K	<i>At</i> 373 K
Li	0.96	0.12	0.08
Na	1.25	0.24	melts
K	0.72	0.16	melts
Cu	5.00	0.64	0.45
Ag	3.33	0.66	0.47
Au	2.00	0.49	0.35
Al	3.33	0.41	0.28
Fe	1.52	0.11	0.07
Zn	0.91	0.18	0.13
In	0.56	0.12	0.08
Tl	0.27	0.07	0.04
Pb	0.21	0.05	0.037

*From G.W.C. Kaye and T.H. Laby, *Table of Physical and Chemical Constants* (Longmans Green, 1966).

most of the common metals. The modern theory that resolves this issue attributes this inconsistency to the unrealistic assumption of electron-ion elastic collisions and the use of classical statistics to describe the electrons in the Drude model. The Drude model is further plagued by its inability to account for the low electron heat capacity and the temperature independent behaviour of paramagnetic susceptibility of conduction electrons. As we will see later in this chapter and in Chapter 13, solutions to both of these problems are achieved by applying quantum statistics.

By feeding the measured value of conductivity in (6.3), we obtain the value of relaxation time τ . The estimated value of τ is in the range 10^{-15} to 10^{-14} s. The thermal velocities estimated from (6.7) are around 10^5 m s^{-1} . Using (6.6), we find that the typical values of mean free path Λ are of the order of a few angstroms which compare with the interatomic separations. This in a way supports the electron-ion collisions in Drude theory. But the low temperature measurements on some of the purest and least imperfect crystals show the mean free path as large as a few centimetres. This strongly contradicts the Drude picture in which the electron bumps along from ion to ion. On the other hand, the modern theory tells that electron waves cannot be scattered by a perfectly periodic potential leading to an infinite mean free path in an ideal crystal. The experimental data in no way casts aspersion on the above theoretical view, simply because no body has ever produced any perfect and ideal crystal. Since every real crystal has deviations from periodicity because of the presence of imperfections and even impurities, the value of the mean free path is limited by the scattering from these centres. The measured values of electrical conductivity of a highly pure sample of copper metal is plotted as a function of temperature in Fig. 6.1.

AC conductivity

Here, we need to calculate the current that would flow under the influence of an alternating electric field (time-dependent),

$$\mathbf{E} = \mathbf{E}(\omega)e^{-i\omega t} \quad (6.9)$$

If the time-dependent momentum per electron be denoted by \mathbf{p} , we get the following equation of motion for the electron,

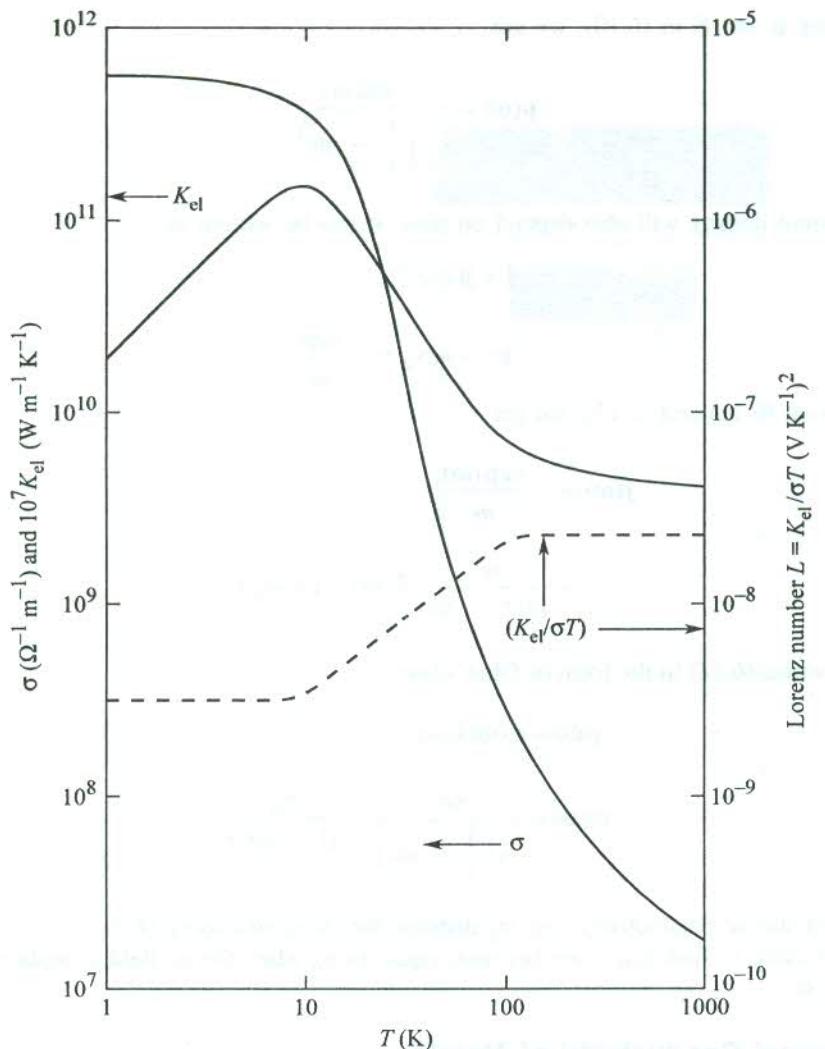


FIG. 6.1 Variations in electrical conductivity σ , electronic thermal conductivity K_{el} and Lorenz number L ($= K_{el}/\sigma T$) with change in temperature for a highly pure sample of copper metal. (Data on σ from G.T. Meaden, *Electrical Resistance of Metals* (Plenum Press, 1965); and K_{el} values from *American Institute of Physics Handbook* (McGraw-Hill, 1971).

$$\frac{dp}{dt} = - \frac{p}{\tau} - eE \quad (6.10)$$

where $\left(-\frac{p}{\tau} \right)$ represents the damping term following from collisions.

The time-dependent solution of the equation of motion is sought in the form

$$p = p(\omega)e^{-i\omega t} \quad (6.11)$$

On substituting \mathbf{p} and \mathbf{E} in (6.10), we get

$$\mathbf{p}(\omega) = - \frac{e\mathbf{E}(\omega)}{\left(\frac{1}{\tau} - i\omega\right)} \quad (6.12)$$

Since the current density will also depend on time, it can be written as

$$\mathbf{j} = \mathbf{j}(\omega)e^{-i\omega t} \quad (6.13)$$

But

$$\mathbf{j} = -nev_d = -\frac{nep}{m}$$

Therefore, using (6.11) and (6.13), we get

$$\begin{aligned} \mathbf{j}(\omega) &= -\frac{nep(\omega)}{m} \\ &= \frac{ne^2}{m\left(\frac{1}{\tau} - i\omega\right)} \mathbf{E}(\omega) \quad (\text{using 6.12}) \end{aligned} \quad (6.14)$$

One usually writes (6.14) in the form of Ohm's law:

$$\mathbf{j}(\omega) = \sigma(\omega) \mathbf{E}(\omega) \quad (6.15)$$

where

$$\sigma(\omega) = \frac{ne^2}{m\left(\frac{1}{\tau} - i\omega\right)} = \frac{\sigma_0}{(1 - i\omega\tau)} \quad (6.16)$$

$\sigma(\omega)$ is called the ac conductivity and σ_0 denotes the dc conductivity (6.3).

It is instructive to find that $\sigma(\omega)$ becomes equal to σ_0 when the ac field is replaced with a dc field (i.e. $\omega = 0$).

6.1.2 Thermal Conductivity of Metals

In the Drude model it is assumed that the major contribution to thermal conductivity of metals comes from conduction electrons. This assumption has basis in the general observation that metals are far better conductors of heat than insulators. The participation of metal ions is reflected in the phonon contribution which is neglected in Drude theory on account of its relatively much small measure.

Consider a metal bar whose two ends are maintained at a constant difference of temperature. This situation refers to the steady state when the whole of thermal energy being fed at the hot end is received at the cold end without any net absorption in the bar. Let the temperature gradient along the length of the bar be defined as $-\partial T / \partial x$, meaning thereby that the temperature decreases down the length being denoted by x . For a small temperature gradient, the thermal current \mathbf{j} , which is a measure of the thermal energy flowing per unit sectional area of the bar per unit time, is found to be proportional to the temperature gradient and thus

$$\mathbf{j} = K_{\text{el}} \left(-\frac{\partial T}{\partial \mathbf{x}} \right) \quad (6.17)$$

where the constant of proportionality, K_{el} , stands for the electronic thermal conductivity.

The relation shows that the thermal energy flows in a direction opposite to that of the gradient.

Let us confine ourselves to the one-dimensional flow of heat energy in which electrons move parallel to the length of the bar. The calculation of \mathbf{j} proceeds on lines parallel to those discussed in Section 5.3.3 in connection with the phonon current. The main difference between the two cases is that the number of heat carriers in this case (electrons) is conserved. Let us concentrate on a group of n electrons per unit volume. At a point x along the length, half of them ($n/2$) arrive from the hot end and the rest ($n/2$) from the cold end. The electron flowing from the hot end should have met its last collision at $(x - v_x \tau)$ whereas the one from the cold end would have its last collision at $(x + v_x \tau)$. If the thermal energy per electron in thermal equilibrium at T be denoted by $u[T(x)]$, the net thermal current can be written as

$$j = \frac{1}{2} n v_x [u(T(x - v_x \tau)) - u(T(x + v_x \tau))] \quad (6.18)$$

where $T(x)$ is the temperature at x and v_x is the x -component of the average electron velocity. In the approximation that the variation in temperature over a mean free path length Λ is small, (6.18) can be expanded about x giving,

$$j = n v_x^2 \tau \frac{\partial u}{\partial T} \left(-\frac{\partial T}{\partial x} \right) \quad (6.19)$$

The factor $n(\partial u / \partial T)$ may be replaced by the electron heat capacity per unit volume C_{el} and v_x^2 by $\frac{1}{3}v^2$ in (6.19) to get

$$j = \frac{1}{3} C_{\text{el}} v^2 \tau \left(-\frac{\partial T}{\partial x} \right) \quad (6.20)$$

Comparing (6.20) with (6.17), we get

$$K_{\text{el}} = \frac{1}{3} C_{\text{el}} v^2 \tau = \frac{1}{3} C_{\text{el}} v \Lambda \quad (6.21)$$

where v is the RMS speed of electrons.

The calculation of C_{el} is made by taking the thermal energy per electron at temperature T as $\frac{3}{2}k_B T$, in accordance with the law of equipartition of energy. This gives the heat capacity per electron as $\frac{3}{2}k_B$, and $C_{\text{el}} = \frac{3}{2}nk_B$. Furthermore, the value of v^2 is obtained from (6.7). Making these substitutions in (6.21), we get

$$K_{\text{el}} = \frac{3}{2} \left(\frac{nk_B^2 \tau}{m} \right) T \quad (6.22)$$

Even before the electrons were known, in 1835 Wiedemann and Franz were able to notice that all good conductors of electricity are good conductors of heat too. They established that the ratio of thermal conductivity to electrical conductivity is proportional to the temperature for a large number of metals. Later in 1881 Lorenz (different from Lorentz) observed that the proportionality constant equalling $K/\sigma T$ has almost the same value for most of the common metals. This constant is known as the Lorenz number and the empirical law became famous as the Wiedemann–Franz law. From (6.3) and (6.22), Drude obtained

$$\frac{K}{\sigma} = \frac{3}{2} \left(\frac{k_B}{e} \right)^2 T \quad (6.23)$$

The coefficient of T in (6.23) is a constant. Drude was thus able to derive the Wiedemann–Franz law which is acknowledged as the biggest success of the Drude model. The Lorenz number L as defined in the theory is given by

$$\begin{aligned} L &= \frac{K}{\sigma T} = \frac{3}{2} \left(\frac{k_B}{e} \right)^2 \\ &= 1.11 \times 10^{-8} \text{ W } \Omega \text{ K}^{-2} \\ &= 1.11 \times 10^{-8} (\text{V K}^{-1})^2 \end{aligned} \quad (6.24)$$

But this number is about half the experimental value given by Table 6.2 and Fig. 6.1. Though Drude theory correctly describes the Wiedemann–Franz law, this disagreement points to certain weak links in the theory. However, it is amusing to learn that Drude obtained the correct Lorenz number because of the mistake of getting half the value of σ given by (6.3). The impressive success as it sounded at that time turns out to be misconceived further because of the errors in the calculation of thermal conductivity. The value of the electron heat capacity $\left(\frac{3}{2} nk_B\right)$ used is exactly half the Dulong–Petit value, which is quite large compared to its negligible contribution observed at room

Table 6.2 Experimental thermal conductivities* and Lorenz number for some metals

Metal	273 K		373 K	
	Thermal conductivity (W cm ⁻¹ K ⁻¹)	Lorenz number (V K ⁻¹) ²	Thermal conductivity (W cm ⁻¹ K ⁻¹)	Lorenz number (V K ⁻¹) ²
Na	1.38	2.12×10^{-8}	melts	
Cu	3.85	2.20	3.82	2.29×10^{-8}
Ag	4.18	2.31	4.17	2.38
Au	3.10	2.32	3.1	2.36
Al	2.38	2.14	2.30	2.19
Fe	0.80	2.61	0.73	2.88
Tl	0.50	2.75	0.45	2.75
Bi	0.09	3.53	0.08	3.35

*From G.W.C. Kaye and T.H. Laby, *Table of Physical and Chemical Constants* (Longmans Green, 1966).

temperature. The actual value is about a hundred times smaller. Another error appears in the magnitude of electrons thermal velocity which, in fact, is a hundred times larger. It is simply a pleasant coincidence that the two errors almost cancel each other in (6.21), giving an overall success to the Drude model.

The experimental values of K and L for some metals at 273 K and 373 K are given in Table 6.2. The observed behaviour of these quantities for a highly pure sample of copper metal is shown in Fig. 6.1 over a wide range of temperature. The pronounced fall in the value of Lorenz number below 100 K (see Fig. 6.1) turns out to be a valid fact for most of the metals to which Drude could not offer an explanation. As will be discussed later in Sections 6.6 and 6.7, the two extreme temperature ranges are characterized by two different mechanisms of electron collisions. These different mechanisms provide different temperature dependences to the mean free path that controls both the electrical and thermal conductivities. This explains why Lorenz number decreases significantly on cooling below 100 K.

The most instructive aspect of the Drude's calculations is the elimination of τ in the derivation of the Wiedemann-Franz law. The astounding success achieved in this respect brings out the inference that the Drude theory is capable of explaining the behaviour of physical quantities that are τ independent. The inference made above was confirmed by Lorentz for Hall coefficient found independent of relaxation time. A quantitative account of the matter is given in the following section without going into the theory of Hall coefficient slated for discussion at a later stage in this chapter.

6.2 LORENTZ MODIFICATION OF THE DRUDE MODEL

Lorentz took up the task of modifying the oversimplified Drude model and constructed his theory on the basis of following points:

1. The assumption that all electrons move with the same thermal velocity is abandoned.
2. The classical Maxwell-Boltzmann velocity distribution is perturbed by the presence of an electric field or a thermal gradient. Both of these tend to displace the equilibrium velocity distribution and distort its symmetry.
3. The approach of Boltzmann transport equation is followed to describe the transport of charge and kinetic energy of electrons by a statistical distribution of mobile electrons constituting the electron gas.

To avoid duplication, the derivation of the expression for electrical conductivity will be taken up at a proper stage for the quantum model of the free electron gas, i.e. for the free electron Fermi gas. Nevertheless, we give below the relation as obtained by Lorentz:

$$\sigma_L = \left(\frac{8}{3\pi} \right)^{1/2} \frac{ne^2 \Lambda}{(3mk_B T)^{1/2}} \quad (6.25)$$

If we denote the Drude conductivity of (6.8) as σ_D , then

$$\begin{aligned} \sigma_D &= \left(\frac{3\pi}{8} \right)^{1/2} \sigma_L \\ &= 1.09 \sigma_L \end{aligned} \quad (6.26)$$

From (6.25) and (6.26) we learn that the Lorentz modifications provide neither any noteworthy change in the quantitative measure nor any change in the temperature dependence of electrical conductivity. Lorentz, however, went on to analyze the effect of a uniform magnetic field on current carrying conductors. This effect is known as the Hall effect. The net force on an electron moving with velocity \mathbf{v} under the action of a static electric field \mathbf{E} and a uniform magnetic field \mathbf{B} is

$$\mathbf{F} = -e[\mathbf{E} + \mathbf{v} \times \mathbf{B}] \quad (6.27)$$

where \mathbf{F} is called the Lorentz force.

Lorentz solved the Boltzmann transport equation under the above conditions and obtained the following expression for the Hall coefficient,

$$R_H = -\left(\frac{3\pi}{8}\right) \frac{1}{ne} \quad (6.28a)$$

whereas according to Drude theory,

$$R_H = -\frac{1}{ne} \quad (6.28b)$$

The electron density n can be estimated from (6.28) by feeding the measured value[†] of Hall coefficient in it. These are given in Table 6.3 together with the atomic concentrations in certain metals. The estimate is consistent with the contribution of one conduction electron per atom in sodium. The result is not so good for silver and is worse for other metals. This points to an otherwise established fact that the free electron approximation is best applicable to alkali metals with the lone loosely bound outermost $3s^1$ electron. The subject is effectively treated by the band theory that also accounts for the observed positive Hall coefficient in certain metals.

Table 6.3 Hall coefficient and electron density of some metals

Metal	Hall coefficient (R_H) [*] ($m^3 C^{-1}$)	Electron density (n) deduced from R_H ^{**} (m^{-3})	Atomic concentration deduced from lattice constant (m^{-3})
Na	-2.5×10^{-10}	2.9×10^{28}	2.5×10^{28}
Ag	-8.4×10^{-11}	8.8×10^{28}	5.8×10^{28}
Al	-3.0×10^{-11}	2.5×10^{29}	6.0×10^{28}
Cd	$+6.0 \times 10^{-11}$	Negative	4.6×10^{28}
Fe	$+2.5 \times 10^{-11}$	Negative	8.5×10^{28}
Ni	-6.0×10^{-10}	1.2×10^{28}	9.2×10^{28}

* From *Handbook of Physics*, Ed., E.U. Condon and H. Qdishaw (McGraw-Hill, 1958).

** Using $R_H = \frac{3\pi}{8ne}$

† Determined from $R_H = \frac{E_y}{j_x B_z}$, whose details can be found in Section 6.6.1.

Other important consequences of the Lorentz solution of Boltzmann transport equation are:

1. *The Thomson effect.* An electromotive force is developed in conductors between whose two ends a temperature gradient is maintained. This effect is treated in Section 6.8.2.
2. *The magnetoresistance effect.* The resistivity of a metal increases when placed in a magnetic field. The Drude model does not predict this effect.

Where the Thomson effect is exploited in several thermoelectric devices, the magnetoresistance effect is immensely useful in the study of electrical transport properties. The magnetoresistance is defined as the fractional increase in the zero field resistivity when the sample is placed in the region of a magnetic field. In the Lorentz model when the Boltzmann equation is solved without ignoring terms in B of higher powers, the magnetoresistance increases as the square of the magnetic field. Experiments have confirmed this square-law dependence. The study of magnetoresistance has proved of tremendous value in the characterization of semiconductors. The knowledge about its behaviour is very crucial in the study of transformation between the metallic and insulating phases of composites. Some of these have interesting properties making them attractive for applications.

6.3 THE FERMI-DIRAC DISTRIBUTION FUNCTION

As already stated, the Sommerfeld theory differs from the classical theories in the replacement of the Maxwell-Boltzmann statistics by the Fermi-Dirac statistics. The distribution of electron thermal speeds in thermal equilibrium at temperature T is described by the function,

$$f(v) = \frac{1}{\exp\left[\left(\frac{1}{2}mv^2 - \mu\right)/k_B T\right] + 1} \quad (6.29)$$

It is more commonly expressed as a function of electron energy,

$$f(\varepsilon) = \frac{1}{\exp[(\varepsilon - \mu)/k_B T] + 1} \quad (6.30)$$

The $f(\varepsilon)$ is known as the Fermi-Dirac distribution function. It denotes the probability that an orbital of energy ε be populated in an ideal electron gas in thermal equilibrium. The quantity μ is a function of temperature and is called the *chemical potential*. The highest energy level populated at absolute zero is designated as ε_F and is called the Fermi level. Its value is decided so as to get the correct number for the total number of particles (here, electrons), using the distribution function (6.30). At this temperature, none of the electrons have an energy greater than ε_F . All the electrons are confined to levels below ε_F .

The temperature dependence of $f(\varepsilon)$ may be summarized as follows:

At $T = 0$

$$\begin{aligned} f(\varepsilon) &= 0 && \text{for } \varepsilon > \varepsilon_F \\ &= 1 && \text{for } \varepsilon < \varepsilon_F \end{aligned} \quad (6.31a)$$

At $T > 0$

$$f(\varepsilon) = \frac{1}{2} \quad \text{for } \varepsilon = \varepsilon_F \quad (6.31b)$$

As $T \rightarrow 0$

$$\begin{aligned} \lim_{T \rightarrow 0} f(\varepsilon) &= 1 && \text{for } \varepsilon < \mu \\ &= 0 && \text{for } \varepsilon > \mu \end{aligned} \quad (6.31c)$$

For these conditions to be true, it is required that

$$\lim_{T \rightarrow 0} \mu = \varepsilon_F \quad (6.32)$$

But observations on metals show that to a high degree of precision the chemical potential remains equal to the Fermi energy well up to room temperature. This allows μ to be safely replaced by ε_F in (6.30) while dealing with metals at room temperature and below. This practice, however, should not relent us to ignore the difference of μ from its value at absolute zero in general, since at times it may yield grossly inaccurate results.

The variation of $f(\varepsilon)$ with the change in electron energy ε at certain temperatures is shown in Fig. 6.2. For $(\varepsilon - \mu) \gg k_B T$, (6.30) reduces to the form

$$f(\varepsilon) \approx \exp[(\mu - \varepsilon)/k_B T] \quad (6.33)$$

which is simply the classical Boltzmann distribution. This indicates that for high values of electron energy the Fermi-Dirac function approaches the classical distribution function (see Fig. 6.2).

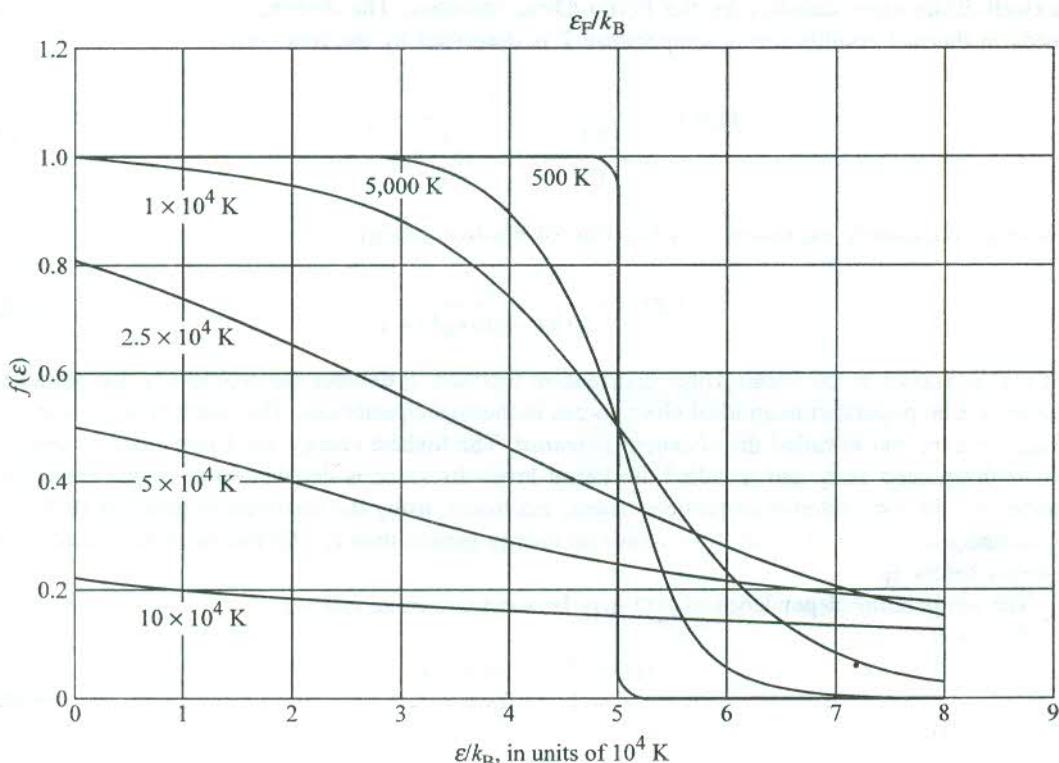


FIG. 6.2 Temperature dependence of the Fermi-Dirac distribution function for a free electron gas in three dimensions, with $T_F = \frac{\varepsilon_F}{k_B} = 50,000$ K.

Therefore, in this limit the choice between the classical and the quantum statistics is meaningless as both give similar results.

The Fermi energy ε_F is often expressed in terms of temperature by the relation,

$$\varepsilon_F = k_B T_F \quad (6.34)$$

where T_F , called the *Fermi temperature*, turns out to be of the order of tens and thousands of kelvin because the Fermi energy per electron is several electron volts.

6.4 THE SOMMERFELD MODEL

The electron-ion interaction cannot be totally ignored even in metals because the conduction electrons remain confined within the volume of the metal. Since no emission of electrons is observed at room temperature, the potential energy of an electron at rest inside the metal must be lower than that of an electron at rest outside. An electron needs to be excited to a certain level so as to leave the metal and escape to infinity. This level is called the *vacuum level*. The electrons move in the field of the periodic potential of positive ions. Therefore, the potential energy of electrons should be periodic in the crystal lattice. Sommerfeld, however, described the cube of a metal crystal as a three-dimensional potential box with an infinite barrier at the surfaces taking the potential energy as a constant within the box. But it is quite interesting that even such an oversimplified model gives a good agreement with the experiment, though the work function of metals is in the range of 5 eV.

Consider a metal cube of side L . The potential energy of electrons in the metal in the form of an infinite square-well is shown in Fig. 6.3.

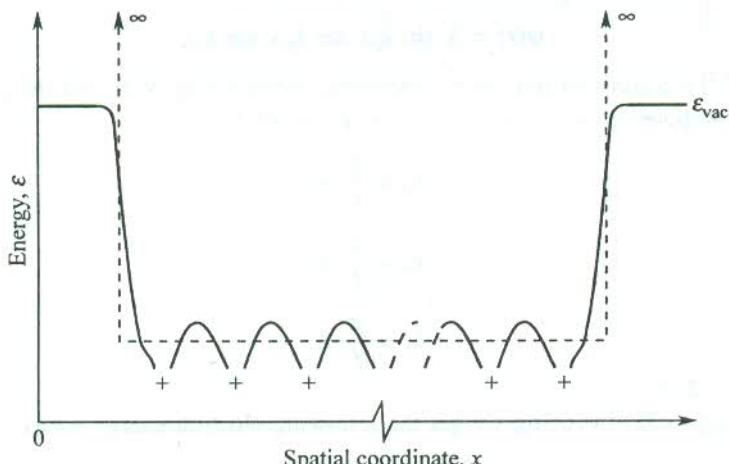


FIG. 6.3 General form of the potential energy of an electron moving in a periodic metallic crystal of positive cores. This is approximated by a square-well potential shown.

Let $V(\mathbf{r})$ denote the potential of an electron at point \mathbf{r} and ε' be its total energy. Then, we can write the time-independent Schrödinger wave equation in the one-electron approximation as

$$\left[\frac{\left(-i\hbar \frac{d}{dr} \right)^2}{2m} + V(\mathbf{r}) \right] \psi(\mathbf{r}) = \varepsilon' \psi(\mathbf{r}) \quad (6.35)$$

where $-i\hbar(d/d\mathbf{r})$ stands for the linear momentum per electron and $\psi(\mathbf{r})$ is the electron wavefunction. Sommerfeld defined $V(\mathbf{r})$ as

$$\begin{aligned} V(x, y, z) &= V_0, && \text{a constant for } 0 \leq x, y, z \leq L \\ &= \infty, && \text{otherwise} \end{aligned}$$

Let the kinetic energy per electron be denoted by ε , such that $\varepsilon = \varepsilon' - V_0$, which gives

$$-\frac{\hbar^2}{2m} \nabla^2 \psi(\mathbf{r}) = \varepsilon \psi(\mathbf{r}) \quad (6.36)$$

The infinite barrier at the bounding surfaces ($x, y, z = 0$ and L) keeps the electrons confined within the crystal and gives the following boundary conditions for the wavefunction:

$$\begin{aligned} \psi &= 0 && \text{for } x = 0 \text{ and } L; \text{ and } 0 < y, z < L \\ & && y = 0 \text{ and } L; \text{ and } 0 < z, x < L \\ & && z = 0 \text{ and } L; \text{ and } 0 < x, y < L \end{aligned} \quad (6.37)$$

These fixed boundary conditions require the electrons to remain confined within the potential box. Therefore, the free electron plane waves moving along the x -edge of the cube [$\exp(ik_x x)$] have to be reflected back just on reaching the potential barrier. The two oppositely moving plane waves may form standing waves [$\exp(ik_x x) \pm \exp(-ik_x x)$] identified as cosine and sine waves, respectively. Since only sine waves satisfy the boundary condition (6.37), they serve as the proper solutions to the Schrödinger equation (6.36). The solution is expressed as

$$\psi(r) = A \sin k_x x \sin k_y y \sin k_z z \quad (6.38)$$

where $A [= (2/L)^{3/2}]$ is a constant and can be determined normalizing $\psi(r)$ over the potential box. The three Cartesian components of the wavevector \mathbf{k} are given by

$$\begin{aligned} k_x &= \frac{\pi}{L} n_x \\ k_y &= \frac{\pi}{L} n_y \\ k_z &= \frac{\pi}{L} n_z \end{aligned} \quad (6.39)$$

with $n_x, n_y, n_z = 1, 2, 3, \dots$

On substituting (6.38) in (6.36), we get the following electron energy states:

$$\varepsilon_{\mathbf{k}} = \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2) \quad (6.40)$$

$$= \frac{\hbar^2 k^2}{2m} \quad (6.40a)$$

$$= n^2 \frac{1}{2m} \left(\frac{\hbar}{2L} \right)^2 \quad (6.40b)$$

with $n_x^2 + n_y^2 + n_z^2 = n^2$.

For an electron moving in the free space, we use (6.40a) to express the energy. The variation of energy with wave vector as plotted in Fig. 6.4 is known as the free electron dispersion curve. But when a free electron moves in a potential box, as in the present case, the electron energy is meaningfully given by (6.40b).

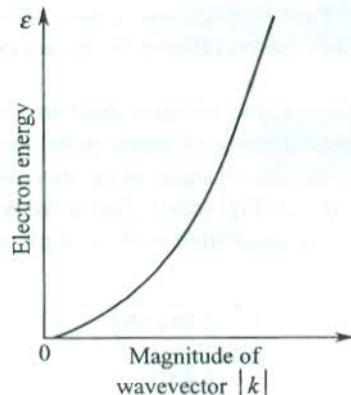


FIG. 6.4 Variation of the free electron energy as a function of the magnitude of its wavevector.

The components (k_x, k_y, k_z) or (n_x, n_y, n_z) together with the spin quantum number m_s are the quantum numbers of the problem. The wavefunctions and the energy states of a free electron in a square-well potential of length L in the x -direction are depicted in Fig. 6.5. The allowed energy values in the three-dimensional space produce constant spherical energy surfaces,

$$\epsilon_{\mathbf{k}} = \frac{\hbar^2 k^2}{2m} = \text{constant}$$

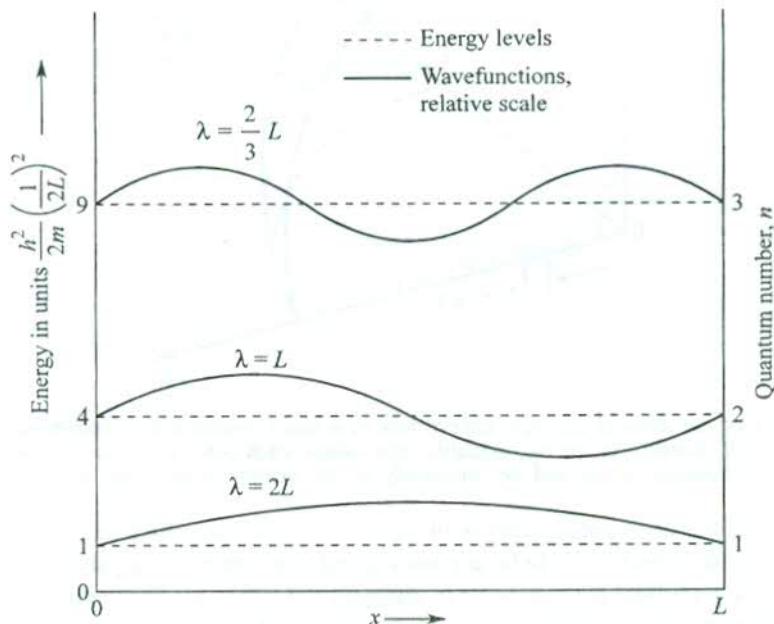


FIG. 6.5 Eigenvalues and eigenfunctions of the first three states of a free electron confined in a potential box of length L .

6.4.1 The Density of States

An insight into the possible values of k_x, k_y, k_z given by (6.39) provides the clue to the determination of density of electron states. We have excluded the zero value of n_x, n_y and n_z since the respective solutions cannot be normalized over the potential box. Further, no new linearly independent solutions are found for the negative values of these integers and as such these values are not considered. Thus for the fixed boundary condition (6.37), the useful possible values of \mathbf{k} totally lie in the positive octant of the k -space.

Equations (6.39) show that there is one electron state in volume $(\pi/L)^3$ of the k -space. The principle of calculation of the electron density of states is the same as that for phonons described in Section 5.2.3. Let us calculate the number of states in the thin shell of the octant enclosed between the energy surfaces $\epsilon(\mathbf{k})$ and $\epsilon(\mathbf{k}) + d\epsilon$ [see Fig. (6.6)]. This number, obtained by dividing the volume of this part of the shell by the volume associated with a single state, is

$$\frac{\left(\frac{1}{8}\right)\left(4\pi k^2 dk\right)}{\left(\frac{\pi}{L}\right)^3}$$

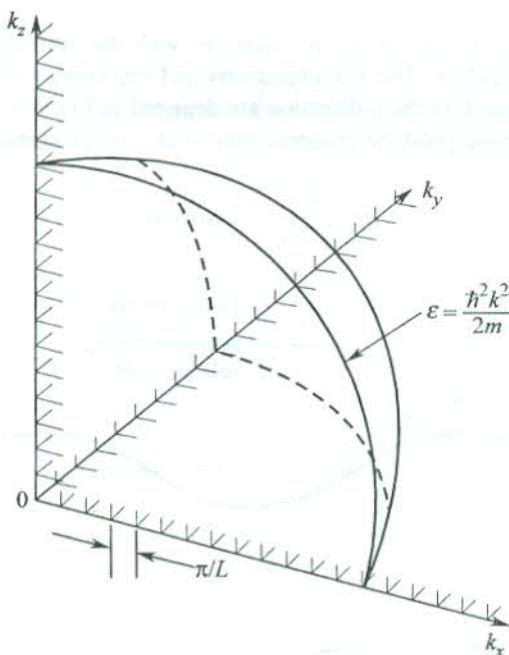


FIG. 6.6 A lattice of the allowed electron wavevectors in k -space representing electron states in an infinite square well. Because of the two possible spin values, each point gives two states. The states have a linear separation of πL and lie completely in one octant for the fixed boundary conditions.

The correct number of possible states is twice this number since the Pauli principle allows two electrons with opposite spins ($\pm 1/2$) to be accommodated in a single orbital designated by a \mathbf{k} -value. The two electrons in an orbital are said to be in different states that are degenerate (degenerate states have a common eigenvalue). Therefore, if $D(\epsilon)$ denotes the density of states,

$$\begin{aligned}
 D(\varepsilon)d\varepsilon &= 2 \frac{\left(\frac{1}{8}\right)(4\pi k^2 dk)}{\left(\frac{\pi}{L}\right)^3} \\
 &= \frac{L^3}{\pi^2} k^2 dk \\
 &= \frac{V}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} \varepsilon^{1/2} d\varepsilon \quad \left(\text{using } \varepsilon_k = \frac{\hbar^2 k^2}{2m}\right)
 \end{aligned}$$

Therefore,

$$D(\varepsilon) = \frac{V}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} \varepsilon^{1/2} \quad (6.41)$$

where V is the volume of the crystal. We may express the density of states per unit volume of the crystal as

$$\frac{D(\varepsilon)}{V} = \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} \varepsilon^{1/2} \quad (6.42)$$

The density of states as a function of energy is graphically represented in Fig. 6.7.

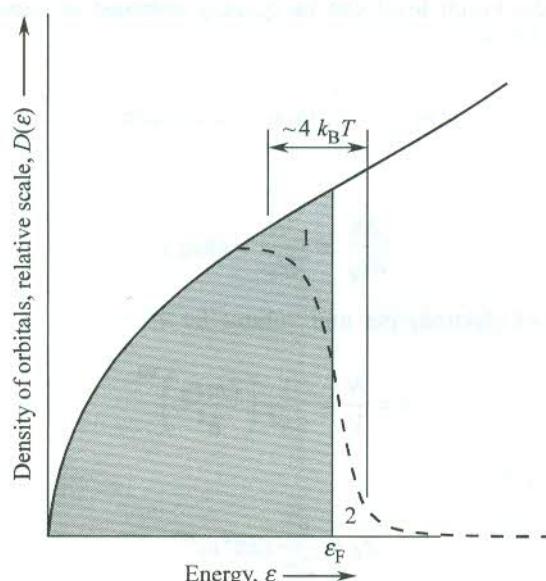


FIG. 6.7 Variation of the density of single-particle states as a function of temperature for a free electron gas in three dimensions. For $k_B T < \varepsilon_F$, the density of the filled orbitals $D(\varepsilon)f(\varepsilon, T)$ at finite temperature is given by the dashed curve. The shaded area represents the filled orbitals at 0 K. On heating to temperature T , the average energy increases and the electrons from region 1 (below ε_F) are thermally excited to region 2 (above ε_F).

It is left as an exercise for the reader to show that the periodic boundary conditions,

$$\psi(x + L, y + L, z + L) = \psi(x, y, z)$$

that yield plane wave solutions to the Schrödinger equation (6.36), lead to the same expression for the density of states as given by (6.41).

6.4.2 The Free Electron Gas at 0 K

The state of the free electron gas at the absolute zero temperature forms the basis for studying the properties of metals. We know that at this temperature all the electrons are filled in orbitals below the Fermi level, ε_F . Using this concept, we calculate the total number of electrons,

$$N = \int_0^{\varepsilon_F} D(\varepsilon) f(\varepsilon) d\varepsilon$$

$$= \frac{V}{2\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \int_0^{\varepsilon_F} \varepsilon^{1/2} d\varepsilon \quad (\text{for } \varepsilon < \varepsilon_F, f(\varepsilon) = 1)$$

or

$$N = \frac{V}{3\pi^2} \left(\frac{2m\varepsilon_F}{\hbar^2} \right)^{3/2} \quad (6.43)$$

The density of states at the Fermi level can be quickly obtained in terms of the total number of electrons if we rewrite (6.43) as

$$\ln N = \frac{3}{2} \ln \varepsilon_F + \text{a constant} \quad (6.44)$$

Differentiating it, we get

$$\frac{dN}{d\varepsilon_F} = \frac{3N}{2\varepsilon_F} = D(\varepsilon_F) \quad (6.45)$$

If we denote the number of electrons per unit volume by n ,

$$n = \frac{N}{V} = \frac{1}{3\pi^2} \left(\frac{2m\varepsilon_F}{\hbar^2} \right)^{3/2} \quad (6.46)$$

giving the Fermi energy ε_F as

$$\varepsilon_F = \frac{\hbar^2}{2m} (3\pi^2 n)^{2/3} \quad (6.47)$$

Thus the Fermi energy may be estimated using (6.47) with the knowledge of the electrons density, n . Since metals have about 10^{28} electrons per m^3 , the magnitude of ε_F is several eV per electron. From (6.34) we get a Fermi temperature, T_F in the range of tens of thousands of kelvin which

is enormously higher than the room temperature. On this scale of temperature the room temperature is far more closer to 0 K. Furthermore, experiments show that most of the properties of metals at room temperature are not much different than those at 0 K. This permits us to use the value of the density of states near the Fermi level for all calculations by replacing ϵ with ϵ_F in (6.41).

The size of ϵ_F is also measured in terms of the Fermi wavevector k_F , whose magnitude equals the radius of the spherical Fermi energy surface in the three-dimensional k -space. From (6.47), we have

$$k_F = (3\pi^2 n)^{1/3} \quad (6.48)$$

The electrons move on the Fermi surface with a constant speed, v_F expressed as

$$v_F = \frac{\hbar k_F}{m} = \frac{\hbar}{m} (3\pi^2 n)^{1/3} \quad (6.49)$$

6.4.3 Energy of Electron Gas at 0 K

This estimate is invaluable to several calculations involving free electrons. The energy per unit volume is expressed as

$$U_0 = \frac{1}{V} \int_0^{\epsilon_F} \epsilon D(\epsilon) d\epsilon \quad (6.50)$$

Substituting $D(\epsilon)$ from (6.41), we obtain

$$U_0 = \frac{\epsilon_F^{5/2}}{5\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \quad (6.51)$$

It is more useful to know the energy per electron, also referred to as the average kinetic energy per free electron. We obtain this value by dividing (6.51) by (6.46). Therefore,

$$\text{Average kinetic energy per electron} = \frac{U_0}{n} = \frac{3}{5} \epsilon_F \quad (6.52)$$

6.5 THE ELECTRON HEAT CAPACITY

The most glaring thermal anomaly concerning the electron heat capacity, as encountered in the Drude model, was resolved by Sommerfeld with the use of quantum statistics. He applied Fermi-Dirac statistics to study the properties of the free electron gas. The density of states that describes the number of energy states per unit energy interval is shown in Fig. 6.7 as a function of energy. Sommerfeld pointed out that the smeared out region of the distribution function has a width of the order of $4k_B T$. According to Fig. 6.7 the electrons in the energy range $2k_B T$ below the Fermi level undergo excitation to cross it when the metal is heated from absolute zero to a temperature T . The electrons in still deeper levels are unable to scatter to a slightly different energy on account of all levels of comparable energy being already occupied with no room for additional occupancy as dictated by the Pauli principle. From Fig. 6.7 we find that only the uppermost electrons equalling a fraction of roughly $(4k_B T/\epsilon_F)$ are affected by the presence of a field gradient (thermal or electrical).

Sommerfeld thus concluded that electrons occupying states only close to the Fermi level ϵ_F contribute to most of the properties of metals. We can make a rough estimate of the electron heat capacity based on the above picture. Taking the rise in thermal energy per electron as $k_B T$ on heating the system from 0 K to T K, the change in the net thermal or internal energy per unit volume may be given by

$$\begin{aligned} U &= n \left(\frac{4k_B T}{\epsilon_F} \right) \cdot k_B T \\ &= \frac{4nk_B T^2}{T_F} \quad (\text{since } \epsilon_F = k_B T_F) \end{aligned} \quad (6.53)$$

Therefore, the electron heat capacity C_{el} is written as

$$C_{el} = \frac{\partial U}{\partial T} = 8nk_B \left(\frac{T}{T_F} \right) \quad (6.54)$$

It is definitely a very small number since T_F is around 10^5 K. The order of magnitude is consistent with the experimental observation that there is hardly a deviation from the Dulong–Petit value of heat capacity at room temperature. This confirms absurdity of the Drude's estimate of $\frac{3}{2} nk_B$ which is exactly half the Dulong–Petit value.

Now, we take up a proper calculation of the heat capacity contributed by the conduction electrons. The net increase in internal energy per unit volume as the temperature rises from 0 K to a temperature T is written as

$$U = \int_0^\infty \epsilon D(\epsilon) f(\epsilon) d\epsilon - \int_0^{\epsilon_F} \epsilon D(\epsilon) d\epsilon \quad (6.55)$$

with the total number of electrons per unit volume given by

$$n = \int_0^\infty D(\epsilon) f(\epsilon) d\epsilon \quad (6.56a)$$

So, we write

$$\epsilon_F \cdot n = \epsilon_F \int_0^\infty D(\epsilon) f(\epsilon) d\epsilon \quad (6.56b)$$

On differentiating (6.55) and (6.56b), we get

$$C_{el} = \frac{\partial U}{\partial T} = \int_0^\infty \epsilon D(\epsilon) \frac{\partial f}{\partial T} d\epsilon$$

$$0 = \varepsilon_F \frac{\partial n}{\partial T} = \int_0^\infty \varepsilon_F D(\varepsilon) \frac{\partial f}{\partial T} d\varepsilon$$

Subtracting the latter from the former of the above equations, we get

$$C_{el} = \int_0^\infty (\varepsilon - \varepsilon_F) D(\varepsilon) \frac{\partial f}{\partial T} d\varepsilon \quad (6.57)$$

An examination of Fig. 6.2 and Fig. 6.7 shows that at temperatures of interest the derivative $\partial f/\partial T$ has significant values in the region of $\pm 2k_B T$ about ε_F . The variation of $D(\varepsilon)$ in this region is not much and, therefore, we can use $D(\varepsilon_F)$ for $D(\varepsilon)$ in the first approximation getting,

$$C_{el} = D(\varepsilon_F) \int_0^\infty (\varepsilon - \varepsilon_F) \frac{\partial f}{\partial T} d\varepsilon \quad (6.58)$$

Differentiating the Fermi-Dirac function (6.30), we obtain

$$\frac{\partial f}{\partial T} = \frac{\varepsilon - \varepsilon_F}{k_B T^2} \frac{\exp [(\varepsilon - \varepsilon_F)/k_B T]}{\{ \exp [(\varepsilon - \varepsilon_F)/k_B T + 1] \}^2} \quad (6.59)$$

Putting $\frac{\varepsilon - \varepsilon_F}{k_B T} = x$, we get

$$C_{el} \approx k_B^2 T D(\varepsilon_F) \int_{-\varepsilon_F/k_B T}^\infty \frac{x^2 e^x}{(e^x + 1)^2} dx \quad (6.60)$$

The e^x factor in the integrand is negligible for $x \leq -\varepsilon_F/k_B T$, i.e. at low temperatures. Therefore, the lower limit of integration may be safely extended to $-\infty$ transforming the integral into a standard form:

$$\int_{-\infty}^\infty \frac{x^2 e^x}{(e^x + 1)^2} dx = \frac{\pi^2}{3} \quad (6.61)$$

When the integral in (6.60) is substituted with this value, obtained from tables of standard integrals, we get

$$C_{el} = \frac{\pi^2}{3} D(\varepsilon_F) k_B^2 T \quad (6.62)$$

Using (6.45), we express the density of states per unit volume at Fermi energy as

$$D(\varepsilon_F) = \frac{3n}{2\varepsilon_F} \quad (6.63)$$

Substituting $D(\varepsilon_F)$ from (6.63) in (6.62), we obtain

$$C_{el} = \frac{\pi^2}{2} \frac{n k_B^2 T}{\varepsilon_F} = \frac{\pi^2}{2} n k_B \left(\frac{T}{T_F} \right) \quad (6.64)$$

The measured values of electron heat capacity of metals at low temperature are in consistency with the magnitude and the linear temperature dependence expressed by (6.64). Also, it can be checked that with the use of (6.64) for the value of heat capacity and the Fermi velocity v_F for the electrons thermal speed, we get a Lorenz number $K/\sigma T$ which is in excellent agreement with the experimental value. It is another improvement on the Drude theory. It may, however, be further observed that with the use of Fermi–Dirac statistics the heat capacity has been degenerated or degraded from its large classical value by a factor of about $(\varepsilon_F/3k_B T)$. It is in this sense that the free electron gas is referred to as the *degenerate electron gas*. We will see that the use of *degenerate* and *non-degenerate* in respect of the semiconductors has a different meaning.

At low temperatures below the Debye temperature and the Fermi temperature, the heat capacity of a metal is expressed as

$$C_V = \gamma T + \alpha T^3 \quad (6.65)$$

where the first term represents the electron contribution and the second term is the usual Debye term contributed by phonons. The phonon or ionic contribution dominates at high temperature but drops spectacularly at low temperatures where it becomes even smaller than the electron contribution. In metals at low temperatures the linear term dominates. The parameters γ and α can be determined from the conventional graph between C_V/T and T^2 using the measured values of C_V . The graph is a straight line and drawn for copper in Fig. 6.8. The intercept with the ordinate axis is the measure of γ whereas the slope gives the value of α . The experimental values of γ together with their ratio to the theoretical values for a number of metals are given in Table 6.4.

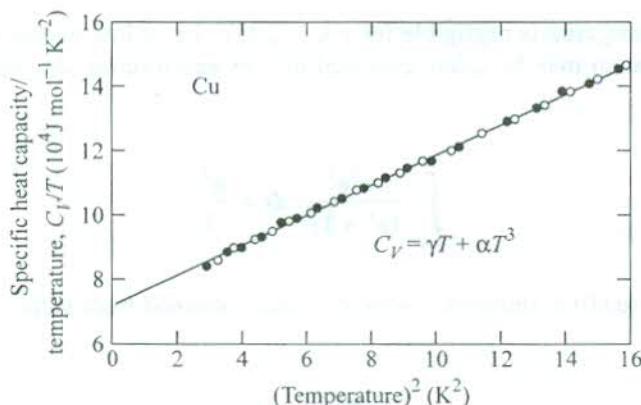


FIG. 6.8 Variation of C_V/T as a function of T^2 for copper. Two sets of experimental points refer to two separate measurements. [After C.A. Bailey, P.L. Smith, *Phys. Rev.*, 114, 1010 (1959).]

Table 6.4 Comparison of the experimental and free electron values of heat capacity constant γ for some metals

Metal	γ_{exp} (mJ mol ⁻¹ K ⁻²)	$\gamma_{\text{exp}}/\gamma_{\text{theo.}}$
Li	1.7	2.3
Na	1.7	1.5
K	2.0	1.1
Cu	0.69	1.37
Ag	0.66	1.02
Au	0.73	1.14
Al	1.35	1.6
Fe	4.98	10.0
Co	4.98	10.3
Ni	7.02	15.3
Zn	0.64	0.85
Pb	2.98	1.97

We observe a reasonably good agreement for most of the metals shown in Table 6.4 except for the transition metals, Mn, Co and Ni where the estimates from the free electron model show large deviations. This can be interpreted as a result of the strong localized nature of d-electrons occupying a partially filled shell. Since there is little overlap among electron waves, the problem cannot be solved in the free electron model. The solution is provided by the band theory which is the subject of the next two chapters. The point that can be appreciated here is the irrelevance of using the electron rest mass for calculations because the d-electrons behave as heavier ones on account of their tight binding. This fact is effectively considered in the band theory.

6.6 THE SOMMERFELD THEORY OF ELECTRIC CONDUCTION IN METALS

Sommerfeld followed Lorentz approach for calculating the electrical conductivity of metals. This method is based on the use of Fermi–Dirac distribution function for solving the Boltzmann transport equation. A complete rigorous treatment of electrical conduction following this approach is fairly lengthy and complicated. A simplified treatment is outlined in this section to explain the principles involved.

In the absence of a field or thermal gradient, the energy and momentum distribution of the free electron gas is described by the equilibrium distribution function,

$$f^0(p) = \frac{1}{\exp\left[\left(\frac{p^2}{2m} - \varepsilon_F\right)/k_B T\right] + 1} \quad (6.66)$$

As a simplification it is assumed that the metal is homogeneous, so that f^0 is independent of the spatial coordinates. The application of any field deforms this function leading to a changed function. In transport phenomena there are two opposite mechanisms that compete with each other. In electric

conduction, the electric field distorts the distribution function from its equilibrium form whereas at the same time the scattering of electrons by phonons and defects tries to restore the equilibrium form of the function. If there is no thermal gradient in the system, the net time rate of change of the distribution function can be expressed as

$$\frac{df}{dt} = \left(\frac{\partial f}{\partial t} \right)_{\text{field}} + \left(\frac{\partial f}{\partial t} \right)_{\text{scatt}} \quad (6.67)$$

where the field term may be treated in general sense so as to include the effect of both electric and magnetic fields depending on the situation. In the presence of a thermal gradient there is an additional term on the RHS of (6.67) contributed by the diffusion process and (6.67) is rewritten as

$$\frac{df}{dt} = \left(\frac{\partial f}{\partial t} \right)_{\text{field}} + \left(\frac{\partial f}{\partial t} \right)_{\text{scatt}} + \left(\frac{\partial f}{\partial t} \right)_{\text{diff.}} \quad (6.68)$$

The steady state condition requires the net rate of change of the function to vanish, i.e. $df/dt = 0$, giving

$$\left(\frac{\partial f}{\partial t} \right)_{\text{field}} + \left(\frac{\partial f}{\partial t} \right)_{\text{scatt}} + \left(\frac{\partial f}{\partial t} \right)_{\text{diff.}} = 0 \quad (6.69)$$

The above equation is the required Boltzmann transport equation. The steady state must not be confused with the equilibrium state that refers to the state when no field or thermal gradients are present.

Now, let us apply (6.69) to electronic conduction in a metal assuming that all points within its volume remain at a constant temperature. Then, (6.67) reduces to the form

$$\left(\frac{\partial f}{\partial t} \right)_{\text{field}} + \left(\frac{\partial f}{\partial t} \right)_{\text{scatt}} = 0 \quad (6.70)$$

Under the influence of the electric field \mathbf{E}_x that acts along the negative x -direction, the velocity distribution of electrons would be drifted towards the positive x -direction. The magnitude of drift velocity v_x depends on the strength of the field. The drift of the spherical Fermi distribution of velocities is shown in Fig. 6.9. As a result the equilibrium distribution f^0 approaches a deformed distribution f . Again, on the ground of simplicity we consider the consequences of a uniform field \mathbf{E}_x , so that the spatial derivative of $(f - f^0)$ is zero.

The force on an electron owing to the electric field \mathbf{E} along the x -direction is

$$-e\mathbf{E}_x = \frac{\partial \mathbf{p}_x}{\partial t} \quad (6.71)$$

where \mathbf{p}_x denotes the electron's linear momentum along the x -direction.

In view of (6.71), we have

$$\left(\frac{\partial f}{\partial t} \right)_{\text{field}} = \frac{\partial \mathbf{p}_x}{\partial t} \cdot \frac{\partial f}{\partial \mathbf{p}_x} = (-e\mathbf{E}_x) \frac{\partial f}{\partial \mathbf{p}_x} \quad (6.72)$$

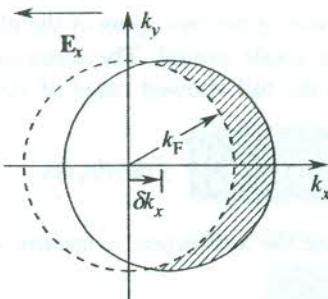


FIG. 6.9 The effect of a steady electric field E_x on the k -space distribution of quasi-free electrons. The Fermi distribution of equilibrium distribution (dashed) centred at $(0, 0, 0)$ is displaced in the stationary state by an amount $\delta k_x = -etE_x/\hbar$.

The structure of the scattering term in (6.70) is quite complex and cannot be discussed here in completeness as the related theory involves concepts at an advanced level. This term is modelled in the so-called relaxation time approximation. Lorentz made the postulate that the rate at which f returns to the equilibrium distribution f^0 because of scattering is proportional to the deviation of f from f^0 , i.e.

$$\left(\frac{\partial f}{\partial t} \right)_{\text{scatt}} \propto f(\mathbf{p}) - f^0(\mathbf{p})$$

which is reduced to

$$\left(\frac{\partial f}{\partial t} \right)_{\text{scatt}} = \frac{f(\mathbf{p}) - f^0(\mathbf{p})}{\tau} \quad (6.73)$$

It should be noted that the relaxation time τ is defined for each location in the momentum space. Assuming that f does not depend on position (i.e. $\nabla_{\mathbf{r}} f = 0$) and using (6.70), (6.72) and (6.73), we can express the stationary non-equilibrium distribution function by

$$f(\mathbf{p}) = f^0(\mathbf{p}) + eE_x \tau \left(\frac{\partial f}{\partial p_x} \right) \quad (6.74)$$

The problem is now reduced to solving this differential equation for f . For this, we make use of the Heisenberg uncertainty principle relating uncertainties in the position (x, y, z) and the momentum (p_x, p_y, p_z) giving

$$\Delta x \cdot \Delta y \cdot \Delta z \approx \frac{\hbar^3}{\Delta p_x \Delta p_y \Delta p_z} \quad (6.75)$$

where \hbar is the Planck constant.

Since there is one state in volume $\Delta x \cdot \Delta y \cdot \Delta z$, the number of states per unit volume of the crystal in the momentum range $p + dp$ and $p = \frac{\Delta p_x \cdot \Delta p_y \cdot \Delta p_z}{\hbar^3} \times 2$.

A factor of 2 is brought in, because of the two spins of the electron ($\pm 1/2$). The values of spin create two degenerate states from a single orbital. The electron density that expresses the total number of states per unit volume in the full allowed range of momenta may thus be written as

$$n = \frac{2}{h^3} \iiint f dp_x dp_y dp_z \quad (6.76)$$

The electric current density along the x -direction is usually written as

$$\begin{aligned} j_x &= -nev_x \\ &= -\frac{2e}{h^3} \iiint \left[f^0 + eE_x \tau \left(\frac{\partial f}{\partial p_x} \right) \right] v_x dp_x dp_y dp_z \quad [\text{using (6.74) and (6.76)}] \\ &= -\frac{2e^2}{h^3} E_x \iiint \tau \left(\frac{\partial f}{\partial p_x} \right) v_x dp_x dp_y dp_z \end{aligned} \quad (6.77)$$

since f^0 is spherically symmetric and does not contribute to the current.

The volume element $dp_x dp_y dp_z$ in (6.77) equals the volume of a thin spherical shell in the momentum space centred at the sphere of radius p , i.e.

$$dp_x dp_y dp_z = 4\pi p^2 dp \quad (6.78)$$

Further,

$$\frac{\partial f}{\partial p_x} = \frac{\partial f}{\partial \epsilon} \cdot \frac{\partial \epsilon}{\partial p_x} \quad (6.79)$$

with the electron energy

$$\epsilon = \frac{p^2}{2m} = \frac{p_x^2 + p_y^2 + p_z^2}{2m}$$

which gives

$$\frac{\partial f}{\partial p_x} = \left(\frac{\partial f}{\partial \epsilon} \right) v_x \quad (6.80)$$

Making use of the relations (6.78) to (6.80), j_x from (6.77) can be expressed as

$$j_x = \frac{8\pi e^2 E_x}{h^3} \int_0^\infty \tau v_x^2 \left(-\frac{\partial f}{\partial \epsilon} \right) m(2m\epsilon)^{1/2} d\epsilon \quad (6.81)$$

In order to carry out the integration we make the following assumptions to simplify the treatment:

- (i) τ depends only on the magnitude of the electron velocity and not on the direction of motion.

(ii) For small fields,

$$\frac{\partial f}{\partial \epsilon} = \frac{\partial f^0}{\partial \epsilon}$$

and

$$v_x^2 \approx \frac{1}{3} v^2$$

(iii) Since $-\partial f/\partial \epsilon$ has an appreciable value only in an energy range of a few $k_B T$ about the Fermi level ϵ_F (see Fig. 6.10), the values of τ and ϵ that matter in calculations are those at the Fermi energy (ϵ_F). Therefore, to a good approximation τ and ϵ under the integral sign may be replaced by $\tau(\epsilon_F)$ and ϵ_F in that order making them thus independent of energy.

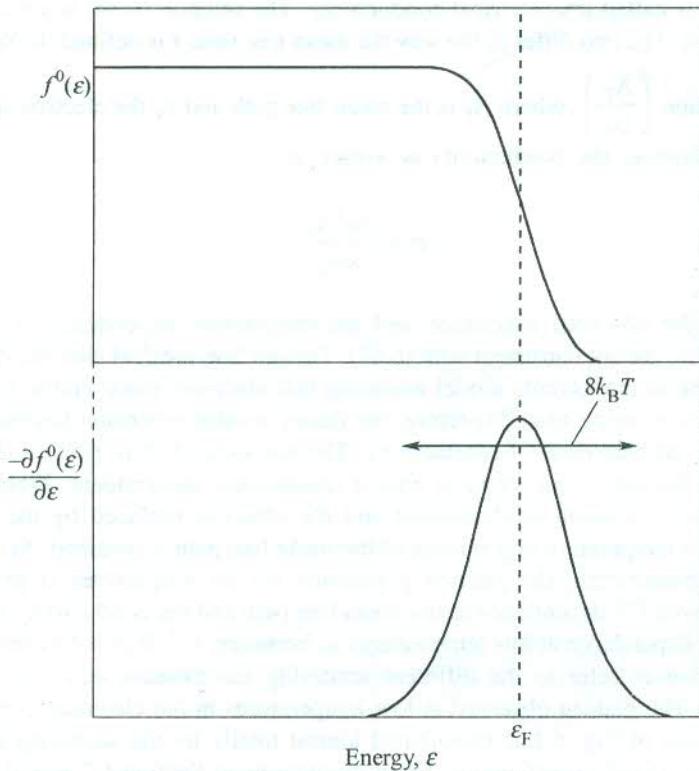


FIG. 6.10 Variation in the Fermi-Dirac distribution function (the occupancy factor) and in its derivative with change in energy.

Applying these assumptions, (6.81) is transformed to:

$$j_x = \frac{16\pi e^2 E_x \tau(\epsilon_F) \epsilon_F^{3/2}}{3h^3} (2m)^{1/2} \int_0^\infty \left(-\frac{\partial f^0}{\partial \epsilon} \right) d\epsilon \quad (6.82)$$

Using (6.47) and $\int_0^\infty \left(-\frac{\partial f^0}{\partial \epsilon} \right) d\epsilon = 1$, we have

$$j_x = \frac{ne^2 \tau(\epsilon_F)}{m} E_x$$

or

$$j_x = \sigma E_x \quad (6.83)$$

where

$$\sigma = \frac{ne^2 \tau(\epsilon_F)}{m} \quad (6.84)$$

The quantity σ is called the electrical conductivity. The relation (6.84) is similar in form to the Drude's relation (6.3). The two differ in the way the mean free time τ is defined. In Sommerfeld model

it is defined as the ratio $\left(\frac{\Lambda_F}{v_F} \right)$, where Λ_F is the mean free path and v_F the electron speed at the Fermi level. With this definition, the conductivity is written as

$$\sigma = \frac{ne^2 \Lambda_F}{mv_F} \quad (6.85)$$

Unfortunately, the observed magnitude and the temperature dependence of conductivity (see Table 6.1 and Fig. 6.1) are not consistent with (6.85). Though Sommerfeld used the quantum statistics, he continued to work in the Lorentz model assuming that electrons made elastic collisions with ion cores, which is far from being true. Therefore, the theory needed extension beyond this assumption to explain the observed behaviour of conductivity. The variation of σ in (6.85) is determined by the change in Λ_F since the magnitude of v_F is almost temperature independent. When the assumption of electron-ion elastic collision is abandoned and the effect is replaced by the electron-phonon scattering, the correct temperature dependence of the mean free path is obtained. At temperatures well above the Debye temperature, the phonon population for all frequencies is proportional to the temperature. This gives T^{-1} dependence to the mean free path and the conductivity in agreement with the experiment. The dependence at low temperatures is, however, T^{-5} (Bloch-Grüneisen law). The two temperature dependences refer to the different scattering mechanisms active in the two extreme temperature ranges. The plateau observed at low temperatures in the electrical conductivity and the Lorenz number curves of Fig. 6.1 is contributed almost totally by the scattering of electrons from impurities and static crystal imperfections. From discussions in Section 6.7 it will be clear that this scattering is temperature independent. In the high temperature range the electron-phonon collisions dominate over other processes and account alone for the observed behaviour.

Nevertheless, the Sommerfeld model has a unique status in the theory of metals since it identified for the first time the difference between a large number of 'free electrons' and 'the much smaller number of conduction electrons'.

6.6.1 The Hall Coefficient (R_H)

The calculation of the Hall coefficient by Sommerfeld theory is more of academic interest as the estimates from Lorentz theory are also consistent with the measured values. The use of quantum

statistics is of no advantage simply because the expression for Hall coefficient does not finally involve the mean free time. The phenomenon of Hall effect, to which the reader is supposed to have been introduced in an earlier course, needs to be discussed for the calculation of Hall coefficient.

Consider a metal crystal in the form of a strip as shown in Fig. 6.11. A steady current of density j_x flows owing to an electric field E_x applied along the x -direction in the strip. When a uniform static magnetic field B directed along the z -direction is switched over the region of the crystal, a small difference of potential ($\sim \mu\text{V}$) develops across the crystal's faces along the y -direction. This happens because of the deflection of electrons by the Lorentz force (6.27). These electrons create an electric field along the y -direction in the crystal. As soon as the force owing to this field on electrons becomes equal to the Lorentz force, the deflection of electrons stops resulting in the saturation of the field E_y , known as the Hall field. The effect is known as the Hall effect.

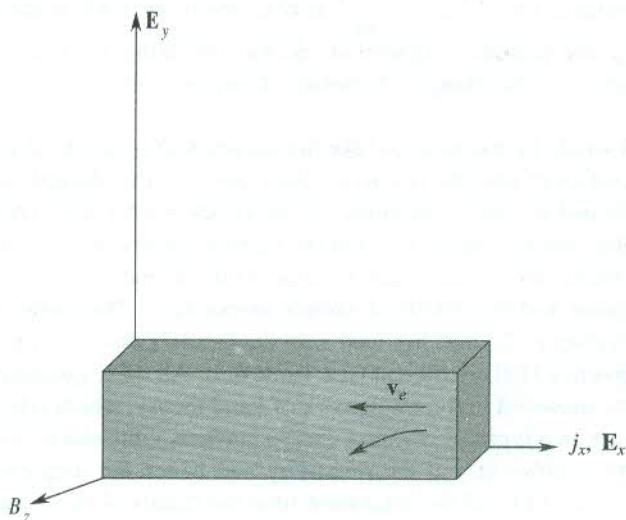


FIG. 6.11 Geometry of the Hall effect. The applied electric field E_x is along the x -direction. The E_y denotes the developed Hall voltage. The v_e represents the electron velocities.

The solution of the Boltzmann transport equation in the above problem is tedious and complex and as such is given here. We straightaway quote the solution. Under the combined influences of E_x and E_y , the general steady-state solution to the first order in B is

$$\mathbf{E} = \frac{\mathbf{j}}{\sigma} + \frac{\omega_c \tau}{\sigma |B|} \mathbf{B} \times \mathbf{j} \quad (6.86)$$

There are two components of \mathbf{E} , one parallel to \mathbf{j} , i.e. E_x and the other perpendicular to \mathbf{j} , i.e. E_y .

$$|E_x| = \frac{|j_x|}{\sigma} \quad (6.87)$$

$$|E_y| = R_H |B| |j_x| \quad (6.88)$$

with

$$R_H = \frac{\omega_c \tau}{\sigma B_z} = -\frac{1}{ne} = \frac{E_y}{j_x B_z} \quad (6.89)$$

where ω_c is the cyclotron frequency given by eB_z/m .

Relation (6.89) is the usual expression derived on the basis of much simple arguments. The experimental determination of R_H is based on this relation. The quantities E_y , j_x and B_z are all actually measured. The method of measurement follows from the geometry of Fig. 6.11, described above. Whereas B_z is obtained from the calibration curve of the electromagnet's field for the used pole spacing, E_y and j_x are derived from the measured voltage and current respectively along the y and x directions with the knowledge of sample dimensions. For a certain concentration of electrons, the calculated value of R_H ($= -\frac{1}{ne}$) as obtained from (6.89) is about 15 per cent lower than the value given by the Lorentz expression (6.28a), bringing it closer to the measured value. The flagrant discord for the transition metals, however, obscures this marginal success (see Table 6.3).

Furthermore, it would be too naive to take the accuracy of relations (6.86) and (6.89) for granted. In fact, the Hall coefficient and the resistivity both are found to depend on magnetic field and are not constant as indicated by these equations. In the Drude model one expects the Hall field (E_y) to be proportional to the magnetic field (B_z) and the current density (j_x). This keeps the Hall coefficient (R_H) constant. It is rather unbelievable that R_H should not depend on the relaxation time that strongly depends on temperature and the details of sample preparation. The carrier concentration as derived from the measured values of R_H matches well with the Drude's estimates only for alkali metals. Some metals show even positive Hall coefficient (see Table 6.3). All these questions including the magnetic field dependence are answered in the framework of band theory, which relies heavily on the fact that in real solids there are two types of constant energy surfaces—the electron and hole surfaces. In this model both the Hall coefficient and the resistivity and hence the magnetoresistance are shown to depend on the magnetic field and the relaxation time the details of which can be found in Chapter 9. The discussion at this stage, however, cannot be closed without stating that the Hall coefficient and the magnetoresistance do saturate at high magnetic fields under certain conditions to be specified in Chapter 9. An elaborate theory demonstrates that for many metals, this limiting value of Hall coefficient is precisely the same as given by Drude theory.

6.7 MATTHIESSEN'S RULE

The knowledge of different mechanisms of electron collisions is most crucial to the calculation of transport properties of solids. Electrons are mostly scattered by lattice phonons and electrons; and defects involving the impurities and the immobile crystal imperfections. The experimental behaviour of most of the metals reveals that the former processes dominate at room temperature and the latter at low temperatures. The two distinct mechanisms, as we identify, to a good approximation may be treated as independent. This allows us to write the net rate at which the momentum distribution would relax back to the equilibrium distribution after the electric field is switched off as

$$\frac{1}{\tau} = \frac{1}{\tau_l} + \frac{1}{\tau_i} \quad (6.90)$$

Using (6.84), we get

$$\frac{1}{\sigma} = \frac{1}{\sigma_l} + \frac{1}{\sigma_i}$$

or

$$\rho = \rho_l + \rho_i \quad (6.91)$$

where ρ denotes the resistivity with ρ_l and ρ_i referring to the contributions from the lattice and the impurities, respectively.

Therefore, in the presence of several scattering mechanisms the resistivity is simply the sum of the resistivities that we would have if each were present alone.

The conductivity plot of Fig. 6.1 and the relative resistance plots of Fig. 6.12 show a plateau below 10 K where the electron collisions with thermal phonons and electrons are almost frozen. The resistivity in this temperature range, that appears as temperature independent, is totally contributed by electron collisions with the impurities and the static imperfections. The value of ρ_i is estimated by extrapolating the curve back to 0 K in Fig. 6.12. The values of ρ_l at different temperatures are obtained by subtracting ρ_i from the measured values of ρ (6.91). The temperature dependence of ρ_l is in accordance with the theory of collisions of electrons with phonons and electrons. For small concentrations, ρ_l shows no dependence on the defect concentration.

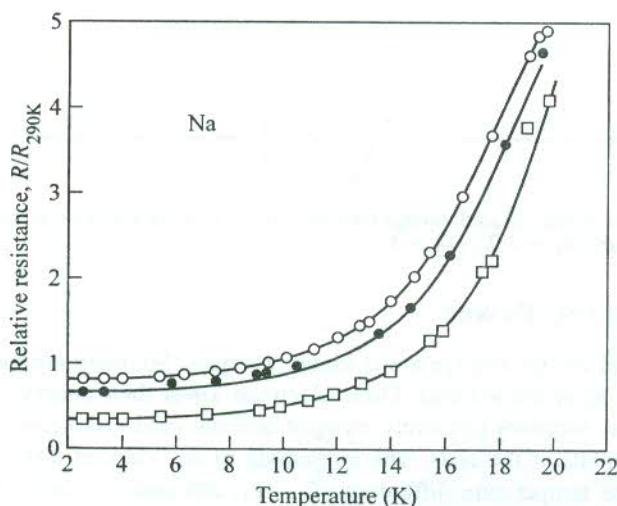


FIG. 6.12 Behaviour of electrical resistance of sodium compared to the value at 290 K with change in temperature. The experimental points marked as \circ , \bullet , \square belong to three different samples with differing defect concentrations. [After D.K.C. McDonald, K. Mendelssohn, *Proc. R. Soc., Edinburgh Sect., A* 202, 103 (1950).]

The observation that ρ_i is temperature independent is established as the Matthiessen's rule. Matthiessen observed that the resistivity of a metal increases with the increase in defect concentration. This is confirmed by the experimental curves of sodium in Fig. 6.12 where the curve with higher intercept with the resistance axis refers to the sample of higher defect content.

The assumption that mechanisms of the two collision processes under consideration are independent is by and large vindicated by observations on whose basis the Matthiessen's rule is founded. But it is not difficult to show that the Matthiessen's rule breaks down even in the relaxation

time approximation which depends on the electron wavevector \mathbf{k} . On the other hand, the realistic picture of collisions warrants the assumptions of the relaxation time approximation to be set aside, as a result of which the two collision mechanisms have a rare chance of being independent and the Matthiessen's rule is in general rendered invalid. The presence of the two competing collision mechanisms can seriously affect the configuration of electrons that control the actual collision rate. It could only be fortuitous that the distribution function in the presence of each separate mechanism be the same and therefore, the failure of the Matthiessen's rule be prevented.

6.8 THERMOELECTRIC EFFECTS

The thermoelectric behaviour of metals is described mainly in terms of three parameters: (i) the thermoelectric power, (ii) the Thomson coefficient, and (iii) the Peltier coefficient. These parameters will be defined in this section on the basis of the thermoelectric response of a conducting rod whose ends are maintained at different temperatures T_1 and T_2 ($T_1 < T_2$), as shown in Fig. 6.13.

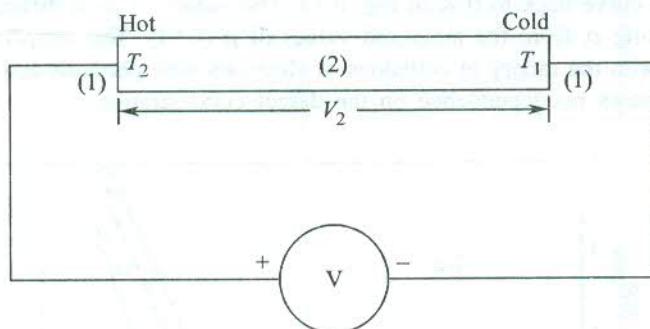


FIG. 6.13 Measurement of the induced voltage (thermo e.m.f.) in a conducting rod whose ends are maintained at temperatures T_1 and T_2 ($T_2 > T_1$).

6.8.1 Thermoelectric Power

Electrons at the hot end on the average are at higher energies. So, there are more electrons in levels above the Fermi level ϵ_F at the hot end. These electrons lower their energy by drifting to the cold end.[†] The hot end thus becomes positively charged and the cold end negatively charged, creating different electric potentials at the ends. The magnitude of the induced difference of potential (V_2) depends entirely on the temperature difference ($T_2 - T_1$) and remains unaffected by any change in the temperature distribution along the rod. A voltmeter connected to measure the induced voltage shows no deflection if the connection wires (1) are of the same material as that of the rod. If this material is different, the induced potential difference between the ends of the wires (1), in contact with the ends of the rod, is no more V_2 but some other value, say V_1 . The voltmeter reading in this case will be $(V_1 - V_2)$, where V_1 and V_2 refer to the difference of potentials in open circuit.

The difference of potential $(V_1 - V_2)$ denoted by V_{12} , rises monotonically with the temperature difference $(T_2 - T_1)$. The rate of change of the induced voltage with the temperature difference defines

[†] Even in metals a part of heat is conducted by phonons, which are the main carriers of heat in insulators [see Section 5.3.3].

the thermoelectric power S_{12} of the junction (1, 2). If ΔV_{12} be the rise in V_{12} because of the small increase ΔT in $(T_2 - T_1)$, then

$$S_{12} = \frac{dV_{12}}{dT} = \frac{dV_1}{dT} - \frac{dV_2}{dT} = S_1 - S_2 \quad (6.92)$$

where S_1 and S_2 are the bulk thermoelectric powers of materials to which the wires and the rod, respectively, belong. The V_{12} is often referred to as the *Seebeck potential*. Relation (6.92) most significantly emphasizes that the thermoelectric power of a junction is not a property of the junction. It is immaterial whether the junction is soldered, brazed, spot-welded or fused; S_{12} depends only on the bulk properties S_1 and S_2 of the materials in contact at the junction.

6.8.2 The Thomson Effect

Let an electric current be forced into the rod (Fig. 6.14) between its ends. When the conventional current flows down the rod from the hot end to the cold end, the electrons reaching the hot end raise their potential energy by absorbing heat from the hot end. Similarly, for the reverse conventional current, the electrons lower their potential energy at the cold end by emitting heat as they arrive there from the hot end. Thus there occurs an absorption or evolution of heat on the passage of current through a rod whose ends are maintained at different temperatures. This amount of heat is in addition to the power loss caused by Joule's heating and the heat flowing because of the presence of a temperature gradient. This absorption or evolution of heat is known as the *Thomson effect*. It does not depend on the nature of the junction involved.

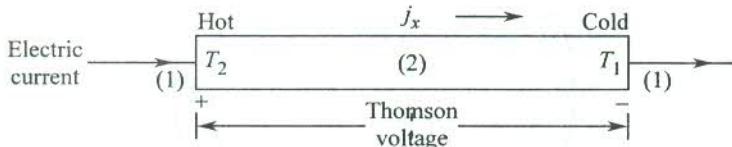


FIG. 6.14 Schematic of the Thomson effect. The Thomson voltage produced because of the temperature gradient is in addition to the Ohm's law voltage.

It is observed that the heat evolved per unit volume per unit time dQ/dt is proportional to the temperature gradient $-dT/dx$ for a constant current density j_x down the length of the rod (see Fig. 6.14). Also, under the condition of a fixed temperature gradient the dQ/dt is found to vary linearly with j_x . These variations may accordingly be expressed as

$$\frac{dQ}{dt} \propto -\frac{dT}{dx}, \text{ with current density } j_x \text{ maintained at a constant value.}$$

$$\frac{dQ}{dt} \propto j_x, \text{ with } -\frac{dT}{dx} \text{ maintained at a fixed value.}$$

Therefore, we may express dQ/dt as

$$\frac{dQ}{dt} = -\mu_T j_x \left(\frac{dT}{dx} \right) \quad (6.93)$$

where the constant of proportionality μ_T is called the *Thomson coefficient*.

The sign of Q , the heat evolved, is taken as positive here.

6.8.3 The Peltier Effect

An electric current density j_x through a conductor down its length (in the x -direction) at constant temperature always has a heat current density j_Q associated with it such that

$$j_Q \propto j_x$$

or

$$j_Q = \Pi j_x \quad (6.94)$$

where Π is called the *Peltier coefficient*.

Relation (6.94) implies that the heat current density will undergo a change at a junction of two conductors 1 and 2 when an electric current flows across the junction because Π is different for the two conductors. The change in heat flux results in absorption or evolution of heat at the junction, depending on the direction of the electric current. When an electric current is forced into a thermocouple, the heat is absorbed at one junction and evolved at the other because the direction of the electric current between conductor 1 and conductor 2 is reversed at the other junction (Fig. 6.15). This is known as the *Peltier effect*. The Peltier effect essentially involves the pumping of heat energy from one junction to the other by making an electric current flow through a thermocouple.

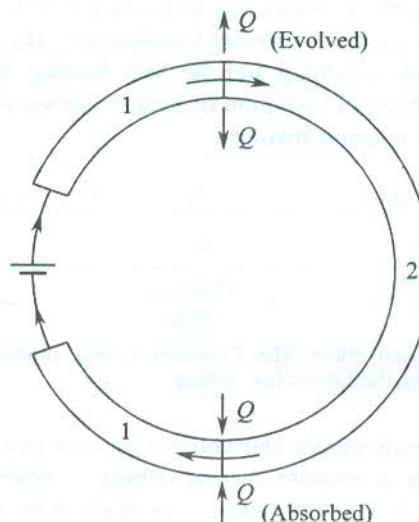


FIG. 6.15 Schematic of the Peltier effect. The heat is shown to be evolved at a thermocouple junction 1–2 when a current flows from the conductor 1 to the conductor 2. The heat is absorbed at the junction 2–1 when the current flows from the conductor 2 to the conductor 1 (the current direction is reversed).

It is appropriate to denote the Peltier coefficient at a junction by Π_{12} and define it as the reversible heat evolved or absorbed per unit time per unit electric current flowing from conductor 1 to conductor 2 at their junction. According to this definition, we have

$$\Pi_{12} = -\Pi_{21} \quad \text{with} \quad \Pi_{12} = \Pi_1 - \Pi_2$$

Like the Thomson effect, neither the Peltier effect nor the Seebeck effect depends on the nature of the junction.

6.8.4 Kelvin (Thomson) Relations

Thomson (famous as Lord Kelvin) used the principles of classical thermodynamics to show that the three thermoelectric parameters S , μ_T and Π are not independent and that only one is needed to specify the other two. He derived the following relations, commonly known as Kelvin relations:

$$\mu_T = T \frac{dS}{dT} \quad (6.95)$$

$$\Pi = TS \quad (6.96)$$

In the above relations, S stands for the absolute thermoelectric power and not the entropy. Although the laws of classical thermodynamics are not directly applicable to thermoelectricity, the Kelvin relations stand the test of experiments extremely well. If either S or μ_T is known as a function of temperature, the remaining two parameters can be obtained from (6.95) and (6.96). In the latter case, relation (6.95) is integrated to get the absolute value of S at any temperature, i.e.

$$S(T) = \int_0^T \frac{\mu_T}{T} dT \quad (6.97)$$

Here, we have ignored $S(0)$ because the third law of thermodynamics implies that all thermoelectric effects vanish at absolute zero.

A few striking gains notwithstanding, the limitations of the free electron theory are pronounced by its varying degree of success in accounting for several properties of metals. Some of these might not have even figured in this chapter. Satisfactory explanations to all such properties and those of solids in general are provided by a less naive theory which forms the subject of the next two chapters.

SUMMARY

- According to the Drude model, the d.c. electrical conductivity of metals is

$$\sigma = \frac{ne^2\Lambda}{(3m k_B T)^{1/2}}, \text{ where } \Lambda \text{ is the electron mean free path.}$$

Experimentally, it is observed that

$$\sigma \propto T^{-1}.$$

- The Drude model correctly describes the Wiedemann–Franz law, but gives a wrong magnitude of the Lorenz number L . The value of L as defined in the Drude theory is about half the measured value and is given by

$$L = \frac{K}{\sigma T} = \frac{3}{2} \left(\frac{k_B}{e} \right)^2, \text{ where } K \text{ is the thermal conductivity.}$$

3. The dc electrical conductivities in the Drude model and the Lorentz model are related as

$$\sigma_D = \left(\frac{3\pi}{8} \right)^{1/2} \sigma_L, \quad D: \text{Drude}$$

L: Lorentz

4. According to the Lorentz model, the Hall coefficient of metals is given by

$$R_H = - \left(\frac{3\pi}{8} \right) \frac{1}{ne}$$

The measured values of the Hall coefficient are close to the values given by

$$R_H = - \frac{1}{ne}$$

R_H is measured using the formula:

$R_H = \frac{E_y}{j_x B_z}$, where the Hall field E_y , the current density j_x and the magnetic field B_z are mutually perpendicular along the y , x and z directions, respectively.

5. The Fermi-Dirac distribution function is

$$f(\varepsilon) = \frac{1}{\exp [(\varepsilon - \mu)/k_B T] + 1}$$

where μ is the *chemical potential*, defined by $\lim_{T \rightarrow 0} \mu = \varepsilon_F$; ε_F is the Fermi energy.

6. In the Sommerfeld model for metals:

- (i) The density of states is

$$D(\varepsilon) = \frac{V}{2\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \varepsilon^{1/2} \left(= \frac{3N}{2\varepsilon_F} \text{ at } \varepsilon = \varepsilon_F \right)$$

with $\varepsilon_F = \frac{\hbar^2}{2m} (3\pi^2 n)^{2/3}$, where $n = \frac{N}{V}$.

- (ii) The electron heat capacity is

$$C_{el} = \frac{\pi^2}{2} n k_B \left(\frac{T}{T_F} \right)$$

- (iii) The dc electrical conductivity is

$$\sigma_0 = \frac{ne^2 \tau(\varepsilon_F)}{m} = \frac{ne^2 \Lambda_F}{mv_F}$$

and the ac conductivity is,

$$\sigma = \frac{\sigma_0}{(1 - i\omega\tau)}$$

7. The resistivity of a metal is expressed as

$$\rho = \rho_l + \rho_i$$

where l and i distinguish the lattice (or phonon) and impurity contributions.

The observation that ρ_i is temperature independent is established as the Matthiessen's rule.

8. Heat absorbed or evolved in the Thomson effect is in addition to the energy loss caused by Joule's heating (produced owing to the electric current, forced externally) and the heat flowing by virtue of the temperature gradient.
 9. The commonly known Kelvin relations are:

$$\mu_T = T \frac{dS}{dT}$$

$$\Pi = TS$$

where μ_T is the Thomson coefficient and S and Π denote the thermoelectric power and the Peltier coefficient, respectively.

PROBLEMS

6.1 Assuming Drude model:

- (a) Show that the probability for an electron, picked at random at any time, not to have made a collision during the preceding t is given by $e^{-t/\tau}$.
- (b) Show that at any moment, the time between the last and the next collision averaged over all electrons is 2τ . In view of this result, comment on the Drude's mistake of a factor of $1/2$ in the calculation of electrical conductivity.

6.2 Consider a spherical body of mass m and charge q moving through a viscous fluid with a constant velocity ($\propto qE_x$) under the action of an electric field E_x . A small magnetic field B_z is now applied. Obtain equations of motion in the xy -plane and find the radius of curvature of the trajectory.

6.3 (a) Plot $f(\epsilon)$ as a function of $(\epsilon - \epsilon_F)$ in units of $k_B T$ at room temperature in the range $-3k_B T \leq \epsilon \leq 3k_B T$.
 (b) Examine at what values of ϵ , the $f(\epsilon)$ differs appreciably from unity.

6.4 Give a comparison of the equilibrium distribution functions representing Maxwell-Boltzmann and Fermi-Dirac statistics. Explain why only the former depends on the particle density.

6.5 Calculate the density of states for a two-dimensional gas of free electrons in a so-called quantum well. The boundary conditions for the electronic wave function are: $\psi(x, y, z) = 0$ for $|x| > a$, where a is of atomic dimensions.

- 6.6** Show that the fraction of electrons within $k_B T$ of the Fermi level is equal to $3k_B T/2\epsilon_F$, if $D(\epsilon) \sim \epsilon^{1/2}$.
- 6.7** The magnitude of the electron wavevector corresponding to the top most level filled in a system of free electrons is equal to $(2\pi^2)^{1/3}/a$, where a is the interatomic separation. If each level is three-fold degenerate, calculate the number of conduction electrons per atom.
- 6.8** Derive relations for the electron density at which the Fermi surface first touches the zone boundary in Na and Cu metals.
- 6.9** Use the free electron theory to calculate the Fermi energy of Na and Al metals. Their lattice constants are 4.3 Å and 4.0 Å, respectively.
- 6.10** Calculate the Fermi energy and Fermi temperature of liquid He³ whose density near 0 K is 81 kg m⁻³. (He³ is a fermion with atomic spin 1/2.)
- 6.11** At what temperature does the electron heat capacity become larger than the phonon heat capacity? Express this temperature in terms of the Debye temperature and the electron density.
- 6.12** Calculate the electron heat capacity at 1000 K in Na, Al and Cu metals using the following data:

	Electron density $n(\times 10^{28} \text{ m}^{-3})$	Fermi energy $\epsilon_F(\text{eV})$
Na	2.5	3.1
Al	18.0	11.7
Cu	8.5	7.1

Estimate the fraction it forms of the total heat capacity, assuming that 1000 K is well above the Debye temperature in each case.

- 6.13** The electrical conductivity of a pure gold crystal is $5 \times 10^7 \text{ ohm}^{-1} \text{ m}^{-1}$ at 273 K. Gold has a close-packed FCC structure with atomic radius as 1.44 Å. Each atom contributes one conduction electron. If a gold crystal has 0.1 per cent of randomly distributed vacancies, estimate the electron mean-free path as determined by scattering from vacancies alone. Calculate the mean-free path for phonon-scattering at 273 K from the given value of conductivity. What is the revised value of conductivity according to Matthiessen's rule?

SUGGESTED FURTHER READING

- Ashcroft, N.W. and N.D. Mermin, *Solid State Physics* (Saunders College, 1988).
- Ibach, H. and H. Lüth, *Solid State Physics* (Springer-Verlag 1995).
- Jain, J.P., in *Solid State Physics*, Vol. 5, F. Seitz and D. Turnbull (Eds.) (Academic Press, 1957).
- Lorentz, H.A., *The Theory of Electrons* (Dover, 1952).
- Meaden, G.T., *Electrical Resistance of Metals* (Plenum Press, 1965).
- Resenberg, H.M., *Low Temperature Solid State Physics* (Oxford, 1963).
- Ziman, J.M., *Principles of the Theory of Solids* (Cambridge, 1972).

Electron Energy Bands

In Chapter 6, we gave an account of the degree of success achieved by the free electron theory in explaining some properties of metals. At the same time we also came across observations to which even the quantized free electron theory did not have the answers. To name a few, the positive Hall coefficient and the wrong hint at the number of conduction electrons per atom (Table 6.3) in some metals exposed weak links in the theory. In addition, there are more vital questions like why some materials are metals while others are good insulators. Their electrical conductivities may differ by a magnitude of order 10^{30} which is unusually high for any physical property. Complex transport phenomena in the presence of a magnetic field too, do not find satisfactory explanation in the frame of the free electron theory. The continuous parabolic energy curve (see Fig. 6.4) renders us helpless in finding a clue to the sharp resonance-like structures observed in the optical spectra of solids. This implies that the discreteness of the atomic energy levels must be accounted for in the theory.

Solutions to the above problems may be sought by correcting the free electron theory for certain simplifications that form its basis. Luckily, efforts made in this direction bring about a remarkable transformation of the theory in terms of success. The theory that emerges as a result of pruning the simplifications in the free electron model has come to be known as the *band theory of solids*. To begin with, the constant potential energy in the Sommerfeld theory is replaced by a non-vanishing potential energy term that is periodic in the crystal, i.e. the rectangular potential box in Section 6.4 is replaced by an infinite periodic potential. This ignores the surface effects and requires the electron to observe the periodicity of the crystal lattice.

In this chapter, we solve the one electron time-independent Schrödinger wave equation to obtain the allowed electron energy states. The states appear as bunches of closely spaced levels forming energy bands. Two successive energy bands are separated by a region of forbidden energy, known as the *forbidden energy gap* or simply the *band gap*. One of the most special features of the band theory is that it distinguishes between electrons in the outermost shell and those in the inner shells by treating the latter as heavier. This proves enormously profitable in dealing with properties that are predominantly controlled by electrons in a certain single shell.

7.1 CONSEQUENCES OF PERIODICITY

In light of the above arguments, we consider a periodic potential energy function as

$$V(\mathbf{r}) = V(\mathbf{r} + \mathbf{t}_n) \quad (7.1)$$

where \mathbf{t}_n is an arbitrary translation vector in the direct crystal lattice usually defined by (1.2) as

$$\mathbf{t}_n = n_1 \mathbf{a} + n_2 \mathbf{b} + n_3 \mathbf{c}$$

Being periodic in the crystal lattice, the potential $V(\mathbf{r})$ may be expanded in a Fourier series and expressed as

$$V(\mathbf{r}) = \sum_{\mathbf{g}} V_{\mathbf{g}} \exp(i\mathbf{g} \cdot \mathbf{r}) \quad (7.2)$$

where \mathbf{g} denotes a reciprocal lattice vector.

Our first aim is to look for an appropriate solution to the following one electron time-independent Schrödinger equation:

$$H\Psi(\mathbf{r}) = \left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \right] \Psi(\mathbf{r}) = \varepsilon \Psi(\mathbf{r}) \quad (7.3)$$

Let us try a general plane-wave expansion,

$$\Psi(\mathbf{r}) = \sum_{\mathbf{k}} C_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (7.4)$$

as the solution to (7.3). Substituting (7.2) and (7.4) in (7.3), we get

$$\sum_{\mathbf{k}} \left(\frac{\hbar^2 k^2}{2m} \right) C_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) + \sum_{\mathbf{k}} \sum_{\mathbf{g}} C_{\mathbf{k}} V_{\mathbf{g}} \exp[i(\mathbf{k} + \mathbf{g}) \cdot \mathbf{r}] = \varepsilon \sum_{\mathbf{k}} C_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (7.5)$$

Equating the coefficients of $\exp[i\mathbf{k} \cdot \mathbf{r}]$ on both sides, we obtain

$$\left(\frac{\hbar^2 k^2}{2m} - \varepsilon \right) C_{\mathbf{k}} + \sum_{\mathbf{g}} V_{\mathbf{g}} C_{\mathbf{k}-\mathbf{g}} = 0 \quad (7.6)$$

In analogy with phonons (Section 4.2.1), the total number of allowed k -values for the electrons waves, when subjected to Born-von Karman periodic boundary condition, will be equal to the total number of unit cells, say N . Relation (7.6) represents one of the N algebraic equations which as a set have replaced the wave equation (7.3). Each of the equations (7.6) that corresponds to a certain value of k couples only those expansion coefficients $C_{\mathbf{k}}$ in (7.4) whose k -values differ from each other by a reciprocal vector \mathbf{g} , i.e. $C_{\mathbf{k}}$ is coupled to $C_{\mathbf{k}-\mathbf{g}}$, $C_{\mathbf{k}-\mathbf{g}'}$, $C_{\mathbf{k}-\mathbf{g}''}$. The solution to each of the N equations (7.6) may be interpreted as a superposition of plane waves of wavevectors $\mathbf{k} - \mathbf{g}$, $\mathbf{k} - \mathbf{g}'$, $\mathbf{k} - \mathbf{g}''$, etc. The wavefunction with eigenvalue $\varepsilon_{\mathbf{k}}$ may thus be written as

$$\Psi_k(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{k}-\mathbf{g}} \exp[i(\mathbf{k} - \mathbf{g}) \cdot \mathbf{r}] \quad (7.7)$$

or

$$\begin{aligned} \Psi_k(\mathbf{r}) &= \sum_{\mathbf{g}} C_{\mathbf{k}-\mathbf{g}} \exp(-i\mathbf{g} \cdot \mathbf{r}) \cdot \exp(i\mathbf{k} \cdot \mathbf{r}) \\ &= u_k(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \end{aligned} \quad (7.8a)$$

with

$$u_k(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{k}-\mathbf{g}} \exp(-i\mathbf{g} \cdot \mathbf{r}) \quad (7.8b)$$

From (7.8a) we infer that $u_k(\mathbf{r})$ as defined by (7.8b) modulates the plane waves $\exp(i\mathbf{k} \cdot \mathbf{r})$ to a

form that serves as a solution to the wave equation (7.3) for a single electron state k . It is essentially the statement of the Bloch theorem defining the Bloch function $\Psi_k(r)$ which is written as

$$\Psi_k(r) = u_k(r) \exp(ik \cdot r) \quad (7.8c)$$

Relation (7.8c) represents the proper wavefunction that we were looking for. We notice that $u_k(r)$ is periodic in the direct crystal lattice. That is,

$$u_k(r + t_n) = u_k(r) \quad (\text{since } \exp(-ig \cdot t_n) = 1). \quad (7.9)$$

7.1.1 Proof of the Bloch Theorem

The k -value associated with the Bloch function $\Psi_k(r)$ belongs to the set of values (4.9) allowed by the periodic boundary condition. The validity of the Bloch theorem would be proved if this set of k -values is reproduced by subjecting the Bloch function (7.8c) to the periodic boundary condition. For the sake of simplicity we provide the proof for a non-degenerate wavefunction in a monatomic one-dimensional crystal. Consider a crystal of length L along the x -direction having N atoms with interatomic spacing a . The eigenstate of an electron with wavevector k near an atom positioned at x is given by the Bloch function,

$$\Psi_k(x) = u_k(x) \exp(ik \cdot x) \quad (7.10)$$

When the electron moves the distance a along the x -direction, the wavefunction changes to

$$\begin{aligned} \Psi_k(x + a) &= u_k(x + a) \exp[ik \cdot (x + a)] \\ &= Cu_k(x) \exp(ik \cdot x) \quad (\text{from 7.9}) \\ &= C\Psi_k(x) \end{aligned} \quad (7.11)$$

where

$$C = \exp(ik \cdot a) \quad (7.12a)$$

If we imagine the crystal to be in the form of a ring, N steps of the above displacement or translation will bring the electron back to the same atom at x from where it started. This means that

$$\Psi_k(x + Na) = \Psi_k(x + L) = \Psi_k(x) \quad (7.12b)$$

But, the net phase-shift in the wavefunction on the basis of (7.11) requires,

$$\Psi_k(x + L) = C^N \Psi_k(x) \quad (7.13)$$

For (7.12b) to be satisfied,

$$C^N = 1 = \exp(i2\pi n)$$

or

$$C = \exp\left(\frac{i2\pi n}{N}\right) \quad (7.14)$$

Using (7.12a), we have

$$k = \frac{2\pi n}{Na} = n \frac{2\pi}{L} \quad (7.15)$$

with

$$n = 0, \pm 1, \pm 2, \dots, \pm \left(\frac{N}{2} - 1 \right), \frac{N}{2} \quad (7.16)$$

$$\text{for } n = \pm \frac{N}{2}, \quad k = \pm \frac{\pi}{a}$$

These two values of k differ by a reciprocal lattice vector $2\pi/a$ (shortest for a one-dimensional lattice). We now show immediately below that the Bloch function is periodic in the reciprocal lattice ($\Psi_{k+g} = \Psi_k$). Because of this reason the two values of \mathbf{k} correspond to the same wavefunction and, therefore, only one sign is included for this value of n in (7.16). The set of k -values (7.15) takes the following form for a three-dimensional crystal.

$$k_x = n_x \frac{2\pi}{L}; \quad k_y = n_y \frac{2\pi}{L} \quad \text{and} \quad k_z = n_z \frac{2\pi}{L} \quad (7.17)$$

with n_x, n_y, n_z defined by (7.16).

The set (7.17) is identical with (4.9) derived for phonons. This confirms that the Bloch function (7.8c) is an appropriate wavefunction to describe electrons moving in the periodic crystal potential.

7.1.2 The Periodicity of the Bloch Functions and Their Eigenvalues

It is straightforward to show that the Bloch functions are periodic in the reciprocal lattice, the periodicity being given by the reciprocal lattice vectors. When the electron wavevector changes from \mathbf{k} to $\mathbf{k} + \mathbf{g}$, the new wavefunction is

$$\begin{aligned} \Psi_{\mathbf{k}+\mathbf{g}}(\mathbf{r}) &= \sum_{\mathbf{g}'} C_{\mathbf{k}+\mathbf{g}-\mathbf{g}'} \exp(-i\mathbf{g}' \cdot \mathbf{r}) \exp(i(\mathbf{k} + \mathbf{g})) \cdot \mathbf{r} \\ &= \left(\sum_{\mathbf{g}''} C_{\mathbf{k}-\mathbf{g}''} \exp(-i\mathbf{g}'' \cdot \mathbf{r}) \right) \cdot \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (\text{with } \mathbf{g}' - \mathbf{g} = \mathbf{g}'') \end{aligned}$$

or

$$\Psi_{\mathbf{k}+\mathbf{g}}(\mathbf{r}) = \Psi_{\mathbf{k}}(\mathbf{r}) \quad (\text{using 7.8a}) \quad (7.18)$$

Relation (7.18) states that the wavefunction repeats for a change in \mathbf{k} by a reciprocal lattice vector.

Now, the Schrödinger equation (7.3) for the eigenstate $\Psi_{\mathbf{k}+\mathbf{g}}(\mathbf{r})$ is

$$H\Psi_{\mathbf{k}+\mathbf{g}}(\mathbf{r}) = \varepsilon_{\mathbf{k}+\mathbf{g}} \Psi_{\mathbf{k}+\mathbf{g}}(\mathbf{r}) \quad (7.19)$$

But in view of (7.18), we have

$$H\Psi_{\mathbf{k}}(\mathbf{r}) = \varepsilon_{\mathbf{k}} \Psi_{\mathbf{k}}(\mathbf{r}) \quad (7.20)$$

Comparing (7.20) with (7.3), we get

$$\varepsilon_{\mathbf{k}+\mathbf{g}} = \varepsilon_{\mathbf{k}} \quad (7.21)$$

Relations (7.18) and (7.21) express that the Bloch functions and their eigenvalues are periodic in the reciprocal lattice space. Energy value $\varepsilon_{\mathbf{k}}$ is obtained in terms of \mathbf{k} which may be

interpreted as a quantum number. In view of the periodic character of $\Psi_{\mathbf{k}}(\mathbf{r})$ and $\varepsilon_{\mathbf{k}}$, we need to know the allowed values of the quantum number k only within the first Brillouin zone for a complete description of these functions throughout the k -space. The periodic energy surfaces ε_k give rise to the electron energy bands. We discuss the formation of energy bands and band gaps in the next two sections.

7.2 WAVE MECHANICAL INTERPRETATION OF ENERGY BANDS

A crystal may be regarded to have been formed by bringing a large number of isolated atoms into a close assembly. An isolated atom has a set of discrete energy levels. Let ϕ_A represent the wavefunction of an electron in an atom A and ϕ_B in another atom B of a crystal with monatomic basis. When atoms A and B come closer, the electron shares its presence between the regions of A and B. The electron oscillates with a well-defined frequency between the two atoms. In the language of wave mechanics, we say that the electron waves ϕ_A and ϕ_B overlap. The two states between which the shared electron resonates may be approximately expressed as

$$\Psi(\pm) = \frac{1}{\sqrt{2}} (\phi_A \pm \phi_B) \quad (7.22)$$

If the electron energy in an isolated atom be ε_0 , the eigenvalues of the above eigenfunctions are written as

$$\varepsilon(\pm) = \varepsilon_0 \pm \varepsilon_1 \quad (7.23)$$

The two values differ by $2\varepsilon_1$, such that

$$2\varepsilon_1 = \hbar\omega \quad (7.24)$$

where ω is the frequency with which the electron shuttles back and forth between the two states $\psi(\pm)$.

The ε_1 is a measure of the electron's interaction with the core of the two atoms. We must be conscious of the fact that we continue to work in the independent electron approximation. For atoms A and B at large distances (isolated), ε_0 is a doubly degenerate level associated with eigenfunctions ϕ_A and ϕ_B . When a large number of atoms N are assembled to make the crystal, an electron is supposed to be shared among N atoms. This splits the N -fold degenerate ε_0 level into N closely-spaced levels. A similar splitting occurs for each of the other discrete levels of an isolated atom. The group of closely-spaced N levels is called an *energy band* (Fig. 7.1). For a silicon crystal of 1 cm^3 , the spacing between the successive levels in a band is of the order of 10^{-23} eV , a negligibly small magnitude. Therefore, the variation of energy within a band may be taken as continuous in the first approximation. The splitting increases with the decrease in the interatomic spacing (see Fig. 7.1).

7.3 THE KRONIG-PENNEY MODEL

In view of the discrete energy level scheme of isolated atoms, it is an unlikely possibility that the energy bands would be infinitely continuous. This is really the case, in general, as there is observed a region of forbidden energies between the two successive bands. However, the bands overlap in some cases. An energy band is almost centred around its parent level, the N -fold degenerate level in an isolated atom of a crystal composed of N atoms. Kronig and Penney (1930) demonstrated, after

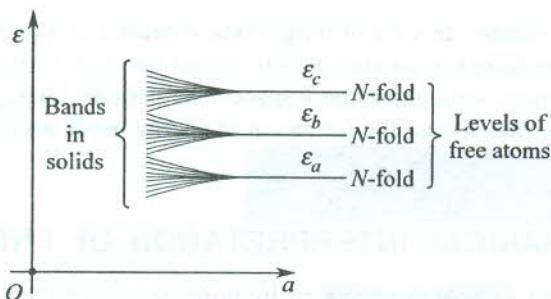


FIG. 7.1 Formation of energy bands in solids. When N atoms are brought together to form a solid, each atomic level is N -fold degenerate for large interatomic separations. The atomic levels ϵ_a , ϵ_b , ϵ_c , etc. spread into bands at a small interatomic separation.

the Bloch theorem (1928) came into existence, that regions of forbidden energies intervene the regions of allowed energies. They accomplished the task using a one-dimensional square-well crystal potential as depicted in Fig. 7.2.

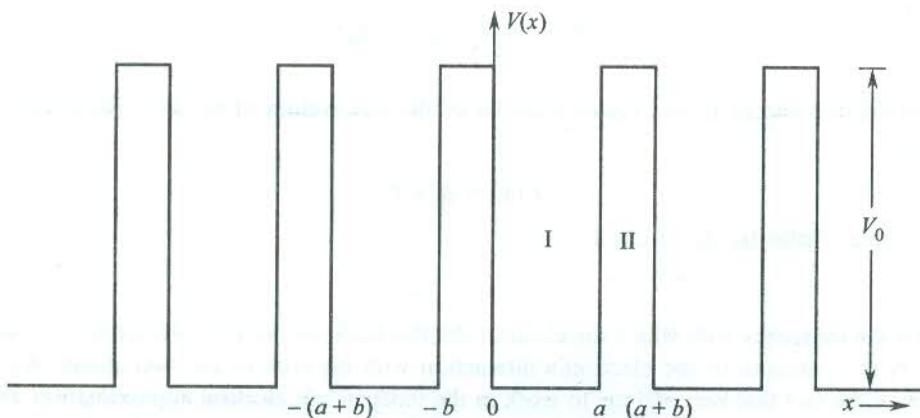


FIG. 7.2 One-dimensional periodic electron potential energy used in the Kronig–Penney model.

There may arise two extreme situations: One in which the barrier is thin, such that bV_0 is vanishingly small and the other when the barrier is thick. In the first case, the electron can move freely through the potential at any electron energy given by the free electron dispersion relation (6.40a). In the second case, the electron with energy less than V_0 has little chance of tunnelling from one atom space to the next and (6.40b) serves as the valid dispersion relation with ‘ a ’ in place of L . This extreme case refers to the problem of an electron in the potential box (Section 6.4).

The Kronig–Penney model describes an intermediate situation where V_0 may be large but ‘ b ’ is so small that an electron with energy less than V_0 can tunnel the potential barrier. The solution of the time-independent Schrödinger wave equation involving this potential is a simple exercise in quantum mechanics. It is based on the following points:

- (i) The wavefunctions in regions I and II match at the interface and so do their derivatives.
- (ii) As V_0 increases, ‘ b ’ decreases so as to keep the product bV_0 at a constant value. Even in the limiting case of large V_0 , the product remains a finite quantity.

The attendant calculations are, however, tedious and involve complex algebra that is of secondary interest to us. We directly quote the energy–wavevector relation as:

$$P \left(\frac{\sin \alpha a}{\alpha a} \right) + \cos \alpha a = \cos ka \quad (7.25)$$

where α is related to energy by

$$\alpha = \left(\frac{2m\varepsilon_k}{\hbar^2} \right)^{1/2} \quad (7.26)$$

and $\frac{(bV_0) ma}{\hbar^2}$ has been replaced by P as we pass to the limit, $b = 0$ and $V_0 = \infty$.

The LHS of (7.25) is plotted as a function of α in Fig. 7.3 for $P = 2$. The absolute magnitude of the ordinate crosses unity when α equals any multiple of $\pm\pi/a$. This is found to be consistent and independent of the value of P . Since the ordinate is equal to $\cos ka$, its magnitude cannot cross the

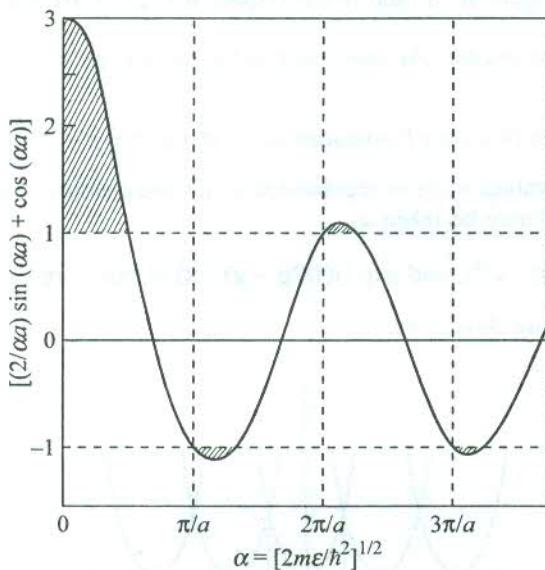


FIG. 7.3 A plot of the left-hand side of (7.25) for $P = 2$ as a function of α which depends on the electron energy in the Kronig–Penney model. The hatched regions correspond to the ranges of forbidden energies.

limits ± 1 . The solution to (7.25) yields alternating ranges of real and imaginary k -values. The ranges of imaginary k correspond to the ranges of α in which the absolute magnitude of the ordinate exceeds the permitted limit of unity. Thus the energies that correspond to these α values are forbidden. This results in the alternation of regions of allowed and forbidden energies. The regions of allowed energies refer to the energy bands whereas those of forbidden energies form the band gaps. In the extreme limits $P \rightarrow 0$ and $P \rightarrow \infty$, we approach the two extreme cases of dispersion relations discussed in the beginning of this section. The elegance with which the Kronig–Penney model predicts the occurrence of band gaps enhances the significance of the model well beyond its historic value.

7.4 THE NEARLY FREE ELECTRON MODEL

The energy bands of solids have generally complex structures. Therefore, it would be appropriate to study the influence of a crystal potential on the electron energy in a certain limiting case and then extend the ideas to any general situation. For example, we may assume that the crystal potential grows slowly from zero value (corresponding to an empty lattice). In the limit of a vanishingly small potential, the Fourier coefficients V_g may be equated to zero in the first approximation. But it would be wrong to ignore totally the effect of periodicity in view of its vital consequences already described. The description of free electron states as given by the energy parabola (Fig. 7.4) is bound to alter in view of (7.21). The allowed energy states are no more confined to a single parabola in the k -space. The states are represented by other parabolae as well, that are displaced by any reciprocal lattice vector \mathbf{g} :

$$\varepsilon_k = \varepsilon_{k+g} = \frac{\hbar^2}{2m} |k + g|^2 \quad (7.27)$$

On the ground of simplicity, the parabolae for a one-dimensional crystal are drawn in Fig. 7.4. The periodicity in the real space is ' a ' and in the k -space it is given by the reciprocal lattice vector $\mathbf{g} \left(= n \cdot \frac{2\pi}{a}\right)$; n being an integer. At zone boundaries the energy values are degenerate as two parabolae intersect here. The first zone boundaries occur at $k = \pm \frac{1}{2} g = \pm \frac{\pi}{a}$. Therefore, the electron wavefunction with these k -values must be represented by a superposition of at least two plane waves which for a small potential may be taken as

$$\exp(i\mathbf{g} \cdot \mathbf{x}/2); \text{ and } \exp[i(1/2\mathbf{g} - \mathbf{g}) \cdot \mathbf{x}] = \exp(-i\mathbf{g} \cdot \mathbf{x}/2) \quad (7.28)$$

The waves move in opposite directions.

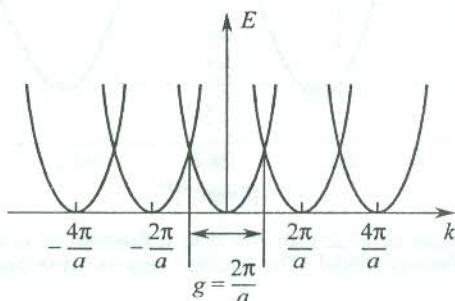


FIG. 7.4 Periodic occurrence of the parabolic energy curves of a free electron in one-dimensional reciprocal space. The periodicity in the real space is ' a '. The electron is supposed to be placed in a periodic lattice with a vanishing potential.

But equation (7.6) requires that reciprocal lattice vectors larger than $2\pi/a$ should also be considered. The value of C_k as determined from (7.6) is appreciably large when ε_k and ε_{k-g} both approach the value $\left(\frac{\hbar^2 k^2}{2m}\right)$. In this the absolute magnitude of C_{k-g} is approximately equal to that

of C_k . The two plane waves (7.28) at the first zone boundaries ideally correspond to this condition. Hence, other reciprocal vectors can be ignored in the approximation for the construction of wavefunctions at the zone boundary. The wavefunctions may be expressed as

$$\Psi(+)\sim e^{ig \cdot \mathbf{x}/2} + e^{-ig \cdot \mathbf{x}/2} \sim \cos \frac{\pi x}{a} \quad (7.29)$$

$$\Psi(-)\sim e^{ig \cdot \mathbf{x}/2} - e^{-ig \cdot \mathbf{x}/2} \sim \sin \frac{\pi x}{a} \quad (7.30)$$

These standing waves appear as a result of the Bragg reflection occurring at $k = \pm g/2 = \pm \pi/a$, as described in Chapter 3. The electron plane waves when Bragg reflected superpose the waves moving towards the same zone edge where the former suffered Bragg reflection. The probability densities of the two sets of standing waves are

$$\Psi^*(+)\Psi(+)=|\Psi(+)|^2 \sim \cos^2 \frac{\pi x}{a} \quad (7.31)$$

$$\Psi^*(-)\Psi(-)=|\Psi(-)|^2 \sim \sin^2 \frac{\pi x}{a} \quad (7.32)$$

The electron potential energy in a one-dimensional crystal is drawn in Fig. 7.5. The potential field belongs to the positive ion cores whose valence electrons move in this field. The figure also

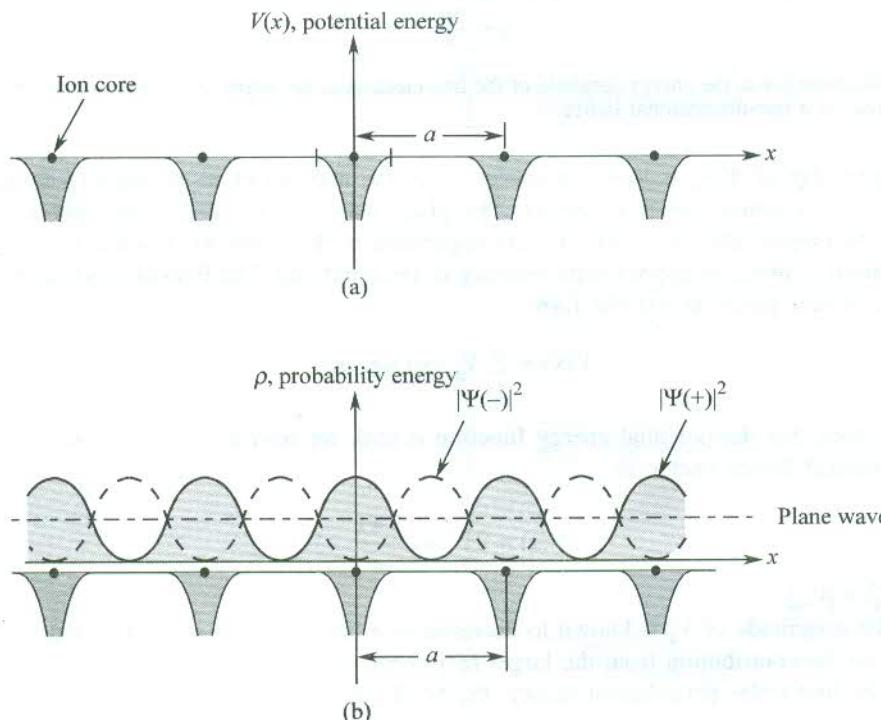


FIG. 7.5 (a) Qualitative form of the electron potential energy $V(x)$ in a one-dimensional crystal lattice. Dark circles denote the positions of the positive ion cores. (b) Distribution of probability density for the standing waves $\Psi(\pm)$ and the plane wave inside the crystal lattice.

depicts the probability distributions of the standing waves and the simple plane waves. The plane waves $\exp(ik \cdot x)$ have the same probability density at all points, since $\exp(-ik \cdot x) \cdot \exp(ik \cdot x) = 1$. The distribution for $\Psi(+)$ favours the piling of electronic charge exactly above the ion cores. On the other hand, the $\Psi(-)$ waves push the electronic charge away from the ion cores. The eigenvalues of $\Psi(+)$ and $\Psi(-)$ differ, though both correspond to the same k -value (π/a or $-\pi/a$). The energy dispersion curve which is continuous throughout the zone shows a gap with two unequal roots at the zone boundaries (see Fig. 7.6). This explains the origin of the band gaps observed in the energy band structure.

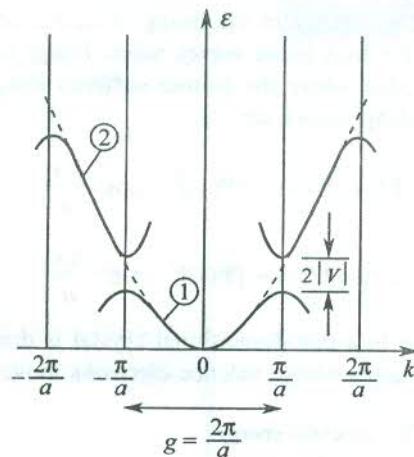


FIG. 7.6 Discontinuity in the energy parabola of the free electron at the edges ($k = \pm \pi/a$) of the first Brillouin zone in a one-dimensional lattice.

The eigenvalue of $\Psi(+)$ is lower in energy since the maxima of its probability density occur at the points of minimum potential energy. The plane wave energy at the zone edges is centred between the two eigenvalues $E(+)$ and $E(-)$ corresponding to $\Psi(+)$ and $\Psi(-)$ in that order. Now, we are in a position to make an approximate measure of the band gap. The Fourier expansion of a one-dimensional crystal potential has the form

$$V(x) = \sum_g V_g \exp(ig \cdot x) \quad (7.33)$$

Appreciating that the potential energy function is real, we rewrite (7.33) considering only the shortest reciprocal lattice vector as

$$V(x) = 2V \cos \frac{2\pi x}{a} \quad (7.34)$$

with $V = |V_g| = |V_{-g}|$.

Since the magnitude of V_g is known to decrease as g increases, for an approximate calculation we can ignore the contribution from the larger reciprocal vectors.

Using the first order perturbation theory, the band gap is written as

$$\Delta E = E(+) - E(-) = 2V \int \cos \frac{2\pi x}{a} [\Psi^*(+) \Psi(+) - \Psi^*(-) \Psi(-)] dx \quad (7.35)$$

Normalizing the wavefunctions at $k = \pm \pi/a$ over the crystal's length L , we have

$$\begin{aligned}\Psi(+)&\sim\left(\frac{2}{L}\right)^{1/2} \cos \frac{\pi x}{a} \\ \Psi(-)&\sim\left(\frac{2}{L}\right)^{1/2} \sin \frac{\pi x}{a}\end{aligned}\quad(7.36)$$

Using (7.36), we have from (7.35)

$$\Delta E = \frac{2V}{L} \int_0^L \left(1 + \cos \frac{4\pi}{a} x\right) dx$$

or

$$\Delta E = 2V \quad (7.37)$$

Thus the band gap is equal to twice the magnitude of the leading Fourier coefficient of the crystal potential. The range of allowed energy states covered by the dispersion curve (see Fig. 7.6) in the first Brillouin zone constitutes the first energy band. Similarly, the higher bands are identified with the respective dispersion curves in other zones.

7.5 ZONE SCHEMES FOR ENERGY BANDS

There are three zone schemes in which the energy bands are drawn:

- (i) The extended zone scheme
- (ii) The reduced zone scheme
- (iii) The periodic zone scheme

The extended zone scheme

In this scheme different energy bands are drawn in different zones in the k -space. Figure 7.6 gives the plot of dispersion curves in the extended zone scheme. The first band is shown in the first zone ($-\pi/a \leq k \leq \pi/a$) and the next higher in the second zone ($\pi/a \leq k \leq 2\pi/a$ and $-2\pi/a \leq k \leq -\pi/a$), and so on.

The reduced zone scheme

All energy bands are shown in the first Brillouin zone in this scheme. As an example, the free electron parabola (see Fig. 6.4) is shown in this scheme by Fig. 7.7. The curves in the two segments of the second zone are translated to the first zone by reciprocal vectors $2\pi/a$ and $-2\pi/a$, separately. Similarly, the energy band pictures in other zones are translated to the first zone by appropriate reciprocal lattice vectors.

The periodic zone scheme

Every band is drawn in every zone in this scheme.

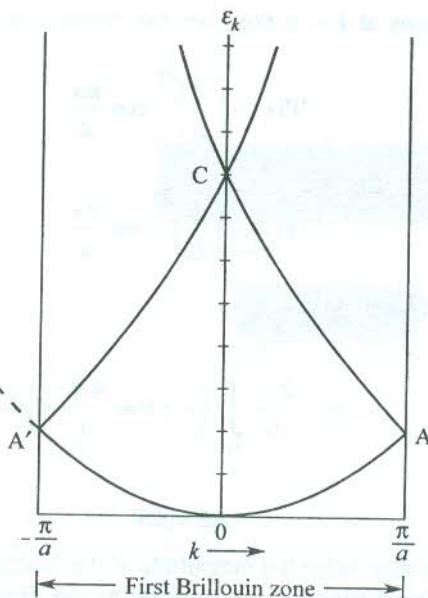


FIG. 7.7 The free electron energy parabola plotted in the reduced zone scheme for a one-dimensional lattice. The curve AC, when translated by $-2\pi/a$, reproduces the usual free electron curve for the negative k -values represented by the dashed curve. Similarly, the translation of A'C by $2\pi/a$ will give the curve for the positive k -values. This often gives a useful description of the band structure of a crystal.

The first three energy bands of a linear crystal as drawn in the three schemes are shown in Fig. 7.8 for the purpose of comparison.

7.6 ENERGY BANDS IN A GENERAL PERIODIC POTENTIAL

Having solved the one-electron Schrödinger wave equation (7.3) in the vanishing limit of a periodic crystal potential (NFE model), we have gained insight into the problem. The wave equation (7.3) in principle can be replaced by (7.6) without any loss of generality. Hence, the process of treating a general periodic potential proceeds in continuation with (7.6), which we rewrite as

$$(\lambda_{\mathbf{k}} - \varepsilon) C_{\mathbf{k}} + \sum_{\mathbf{g}} V_{\mathbf{g}} C_{\mathbf{k}-\mathbf{g}} = 0 \quad (7.38)$$

with

$$\lambda_{\mathbf{k}} = \frac{\hbar^2 k^2}{2m} \quad (7.39)$$

Equation (7.38) is called the *central equation*. It represents a set of simultaneous linear equations that couple the coefficients $C_{\mathbf{k}-\mathbf{g}}$ for all reciprocal lattice vectors. The number of these equations equals the number of reciprocal lattice vectors. The equations are consistent if the determinant of the coefficients of $C_{\mathbf{s}}$ vanishes. Considering only one Fourier component $V_{\mathbf{g}}$, as in the NFE model, we write three consecutive equations which yield the following determinant.

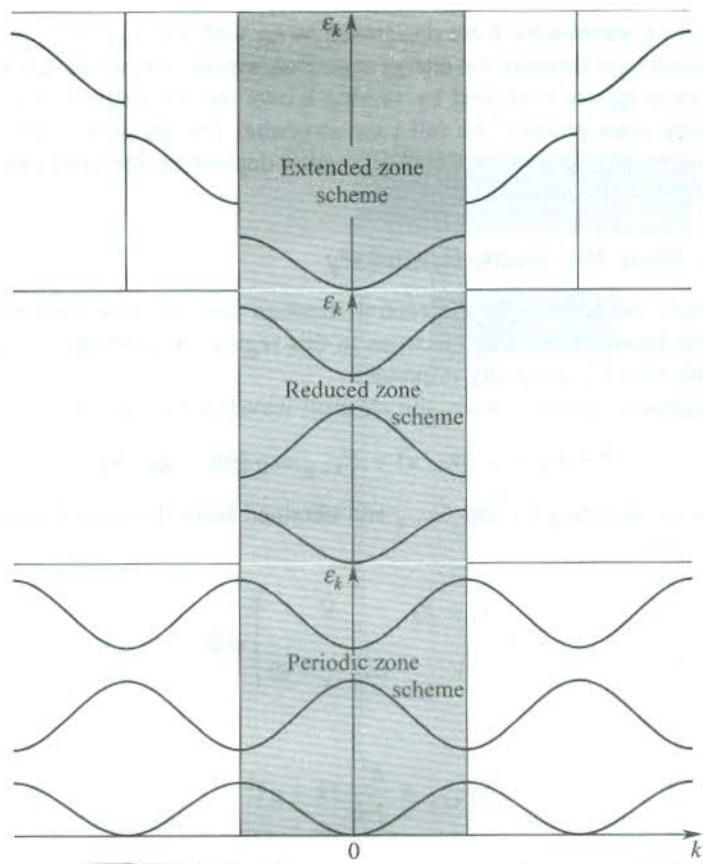


FIG. 7.8 First three energy bands of a one-dimensional crystal in the extended, reduced and periodic zone schemes.

$$\begin{vmatrix} (\lambda_{k-g} - \varepsilon) & V & 0 \\ V & (\lambda_k - \varepsilon) & V \\ 0 & V & (\lambda_{k+g} - \varepsilon) \end{vmatrix} \quad (7.40)$$

with

$$|V_g| = |V_{-g}| = V.$$

It is a small portion of a huge determinant that is evolved from the set (7.38). By equating (7.40) to zero and solving for ε we get three roots that fall in three different energy bands at a certain value of k . The size of the determinant is chosen according to the extent of the energy spectrum that is required. But the size factor is no more a consideration now, with the availability of fast computers. On choosing k that differs from a value in the first Brillouin zone by a reciprocal lattice vector, there occurs no change in the energy spectrum as the same set of equations in a different order appear, giving the same roots of energy. Therefore, very often the k -values within the first zone alone are considered.

Roots that refer to a wavevector k are designated as ε_{nk} with $n = 1, 2, 3, \dots$ for the first, second, third, ..., bands, respectively, forming the energy spectrum whose different levels belong to different energy bands. The roots ε_{nk} are evaluated by varying k over the allowed set of values. These when arranged on an energy scale produce the full band structure. The pictures in the three-dimensional space are fairly complex. On account of k being direction dependent, the band structures, in general, look different in different directions.

7.6.1 Solution Near the Zone Boundary

It is important to study the solution of the central equation near the zone boundary because of the large deviation in the behaviour of free electrons in this region. According to (7.38), if C_{k-g} is an important coefficient, then C_k is equally important.

In the two-component approximation, the electron wavefunction is taken as

$$\Psi = C_k \exp(i\mathbf{k} \cdot \mathbf{x}) + C_{k-g} \exp[i(\mathbf{k} - \mathbf{g}) \cdot \mathbf{x}] \quad (7.41)$$

The two equations, coupling C_k and C_{k-g} and obtained from the central equation (7.38), have a solution if

$$\begin{vmatrix} (\lambda_k - \varepsilon) & V \\ V & (\lambda_{k-g} - \varepsilon) \end{vmatrix} = 0$$

with

$$\lambda_{k-g} = \frac{\hbar^2}{2m} |k - g|^2$$

Solving it for ε , we obtain

$$\varepsilon_k(\pm) = \frac{1}{2} (\lambda_{k-g} + \lambda_k) \pm \left[\frac{1}{4} (\lambda_{k-g} - \lambda_k)^2 + V^2 \right]^{1/2} \quad (7.42)$$

The roots $\varepsilon_k(\pm)$ when plotted as a function of wavevector give dispersion curves for the first two energy bands as shown in Fig. 7.9.

The corresponding roots at the zone boundary ($k = \frac{1}{2}g$) may be written as

$$\varepsilon(\pm) = \lambda \pm V \quad (7.43)$$

where

$$\lambda = \frac{\hbar^2}{2m} \left(\frac{1}{2} g \right)^2 \quad (7.44)$$

The λ is the free electron energy at the zone boundary. We find that the band gap obtained from (7.43) is identical with that given by (7.37).

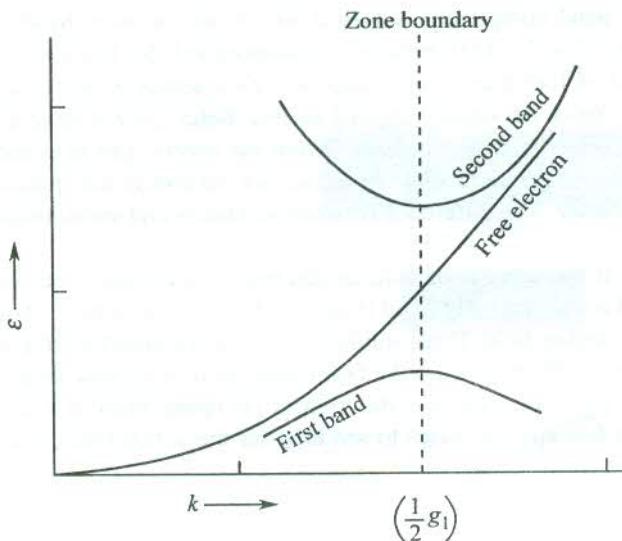


FIG. 7.9 Qualitative form of solutions of (7.42) in the periodic zone scheme near the boundary of the first Brillouin zone. The free electron curve is drawn for comparison.

It is instructive, however, to express the roots near the zone boundary in terms of the wavevector \tilde{k} as measured from the zone boundary. The k and \tilde{k} are related by

$$\tilde{k} = k - \frac{1}{2} g \quad (7.45)$$

Using (7.45), (7.42) can be expressed as

$$\varepsilon_{\tilde{k}}(\pm) = \varepsilon(\pm) + \frac{\hbar^2 \tilde{k}^2}{2m} \left(1 \pm \frac{2\lambda}{V} \right) \quad (7.46)$$

Thus in this definition of the wavevector, the dependence of energy on wavevector is similar to that for the free electrons. When V is negative, $\varepsilon_{\tilde{k}}(-)$ corresponds to the upper band.

7.7 INSULATORS, SEMICONDUCTORS AND METALS

The electronic transport in solids is found to be closely controlled by their band structure. With the knowledge of the band structure, it is possible to predict whether a solid is a good or bad conductor of electricity. We discuss below how insulators, semiconductors and metals may be distinguished on the basis of qualitative differences in their band structures.

We know that the number of energy states in a band is equal to the number of primitive cells (say N) in the crystal. Since each state can accommodate two electrons of opposite spins, a number of $(2N)$ electrons would be required to fill the band completely. Therefore, if a primitive cell contributed an even number of electrons to a band, the band would be completely filled. For example, in a monatomic crystal in which each atom has one valence electron, the band will be fully occupied if there are two atoms in the primitive cell. With two valence electrons per atom, only one atom

in the cell will fill the band completely. Solids in which one or more bands are completely filled and all others are empty [Fig. 7.10(a)], behave as insulators at 0 K. The lowest allowed empty band is preceded by a region of forbidden energy gap (> 5 eV), across which electrons need be excited for electric conduction. Since the usually applied electric fields are not large enough to provide this excitation, the material behaves as an insulator. When the energy gap is in the range of an eV, the material shows a conductivity intermediate to values for insulators and metals and the material is classified as a semiconductor. The difference between insulators and semiconductors is one of degree and not of type.

On the other hand if the number of valence electrons per primitive cell is an odd number, the top-most occupied band is only partially filled [Fig. 7.10(b)]. Such a material shows the flow of current on the application of an electric field. These solids are thus good conductors of electricity and are called metals. Alkali metals and noble metals are the best examples of this class where one valence electron is contributed by each primitive cell and the highest occupied band is exactly half filled.

Now, we take up a few specific cases to see how far the actual band structures conform to the above principle.

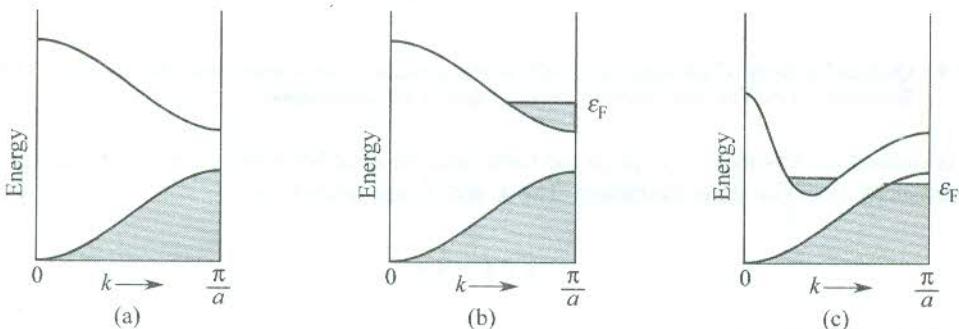


FIG. 7.10 Qualitative band schemes for insulators and metals: (a) insulators—the lower band and all below it are completely filled; all higher bands are empty, (b) metals—a partially occupied band, and (c) metals of relatively low conductivity (semi-metals)—overlap between, a filled band and an empty band.

Sodium. It belongs to the group of alkali metals all having the BCC structure with a rhombohedral primitive cell that contains effectively one atom. The electronic configuration of a sodium atom is $[1s^2 2s^2 2p^6] 3s^1$. Since there is a correspondence between the discrete states of an isolated atom and the Brillouin zones in a solid, we expect that the 10 electrons in the closed inner shells form narrow bands in the sodium metal occupying the first five zones (1 for each s-shell and 3 for the p-shell) in the extended zone scheme. The single outermost ($3s^1$) electron produces the half-filled band in the next zone. Accordingly, the solid sodium is metallic and so are other alkali metals.

Magnesium. It is a member of the alkaline earth metals all of which have two valence electrons per primitive cell, irrespective of the symmetry. The magnesium atom has electronic configuration $[1s^2 2s^2 2p^6] 3s^2$ which would apparently give the insulating behaviour to the solid magnesium, contrary to the observed metallic character. The metallic character arises because of the overlapping of the empty 3p-band with the filled 3s-band. Thus the 3s-electrons can be almost continuously excited to states in the 3p-band. The overlap only saves the alkaline earth metals from being branded insulator. They are not the same good conductors of electricity as the alkali metals. They are rather

classified as semi-metals. A qualitative demonstration of the band overlap is shown in Fig. 7.10(c). The Fermi energy ϵ_F in magnesium occurs at an energy which fills the 3s band about 90 per cent, with just a small per cent occupancy for the overlapping sp band.

Diamond. The isolated carbon atoms have the electronic configuration $-1s^2\ 2s^2\ 2p^2$. The mixing of 2s and 2p wavefunctions in the tetrahedrally bonded diamond crystal is well known to result in the sp^3 hybridization. The sp^3 hybrid band further splits into two because of the modification of the s- and p-level in the crystal. Each of the two hybrid sub-bands can accommodate four electrons. The four outer electrons belonging to the 2s- and 2p-states fill the lower sub-band, leaving the upper one empty. There exists a forbidden gap E_g of about 5 eV at 0 K between the two sub-bands. These features that account for the insulating property of diamond are shown in Fig. 7.11. Band structures of the semiconductors Ge and Si are characterized by similar features. A proper discussion on semiconductors is deferred to Chapter 9.

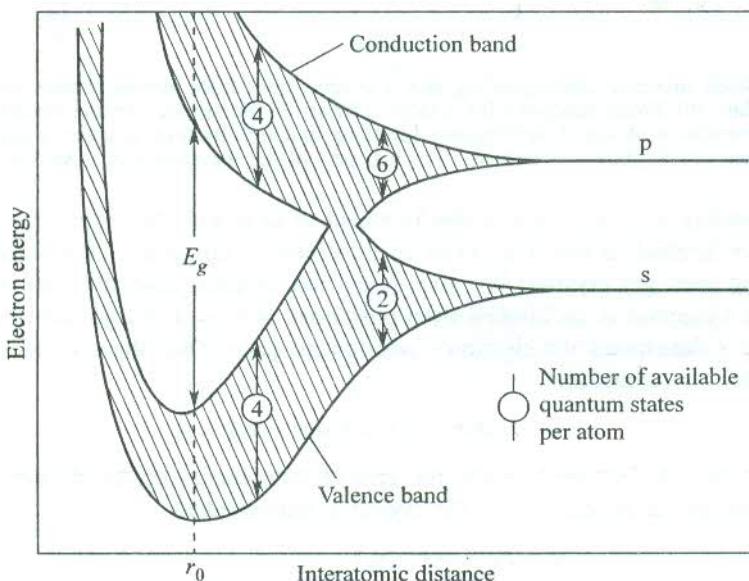


FIG. 7.11 Behaviour of energy bands as a function of interatomic separation for diamond (C), Si and Ge. The band gap E_g is defined at the equilibrium separation r_0 . The figure shows that the band gap is not tied to the periodicity of the lattice. Amorphous solids can also show a band gap. [After W. Shockley, *Electrons and Holes in Semiconductors* (van Nostrand, 1950).]

7.8 THE TIGHT-BINDING APPROXIMATION

In the NFE model the plane wave part, $\exp(i\mathbf{k} \cdot \mathbf{r})$, of the Bloch function, $[u_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r})]$, is emphasized and the atomic part is overlooked. But electrons in the low-lying inner core levels of a free atom are strongly localized in space. The property is largely retained by these electrons when atoms are assembled to form the solid. It points to the inadequacy in describing every band structure in terms of quasi-free electrons. In order to deal with the localized electron, an alternative approach is followed in which the atomic part of the Bloch function is stressed. This approach is known as the *tight-binding approximation*. The single electron wavefunction in the crystal is expressed as a

linear combination of the atomic orbitals (LCAO) that the electron occupies in a free atom. The forms of Bloch functions in a linear crystal for $k = 0$ and $k \neq 0$ are drawn in Fig. 7.12.

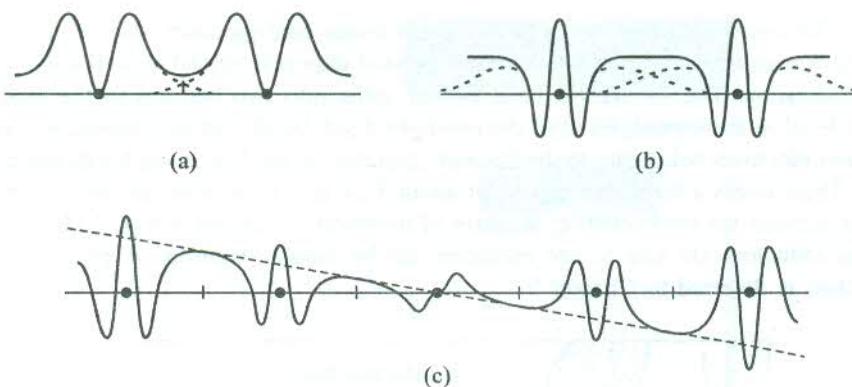


FIG. 7.12 (a) Bloch functions corresponding to $k = 0$ state, built from atomic orbitals (dashed) with small overlap. (b) Bloch functions for a large overlap of the atomic orbitals (dashed). (c) Schematic representation of wavefunctions corresponding to $k \neq 0$. Its form is approximately the product of (b) and $\cos k$ (shown as dashed). The real part of the wavefunction is shown.

The tight-binding approximation is ideally suited to deal with the inner core electrons. It has successfully been applied to the d-electrons in transition metals and to the valence electrons in diamond like and inert gas crystals. We give below the simplest case of s-state electrons.

Consider an s-electron in an isolated atom positioned at \mathbf{r}_n , with the ground state wavefunction $\phi(\mathbf{r} - \mathbf{r}_n)$, where \mathbf{r} determines the electron's position in space. Then the one-electron Schrödinger wave equation for free atoms is

$$H_0\phi(\mathbf{r} - \mathbf{r}_n) = \varepsilon_0\phi(\mathbf{r} - \mathbf{r}_n) \quad (7.47)$$

where H_0 and ε_0 are the Hamiltonian and the ground state energy of the electron in a free atom.

The Hamiltonian for an electron in the crystal is expressed as

$$\begin{aligned} H &= H_0 + v(\mathbf{r} - \mathbf{r}_n) \\ &= \left[-\frac{\hbar^2}{2m} \nabla^2 + v_0(\mathbf{r} - \mathbf{r}_n) + v(\mathbf{r} - \mathbf{r}_n) \right] \end{aligned} \quad (7.48)$$

with

$$v(\mathbf{r} - \mathbf{r}_n) = \sum_{m \neq n} v_0(\mathbf{r} - \mathbf{r}_m) \quad (7.49)$$

In the above equations, $v_0(\mathbf{r} - \mathbf{r}_n)$ denotes the potential energy of an electron when localized at the isolated atom positioned at \mathbf{r}_n . The influence of atoms in the vicinity of \mathbf{r}_n , where the electron in question is strongly localized on relative terms, is treated as a perturbation on H_0 and represented by $v(\mathbf{r} - \mathbf{r}_n)$. Now, our aim is to look for solutions of the following Schrödinger equation:

$$H\Psi_k(\mathbf{r}) = \varepsilon_k\Psi_k(\mathbf{r}) \quad (7.50)$$

where ε_k is the electron energy in the crystal and $\Psi_k(\mathbf{r})$ the Bloch wavefunction.

Let us try the solution,

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \sum_n \exp(i\mathbf{k} \cdot \mathbf{r}_n) \phi(\mathbf{r} - \mathbf{r}_n) \quad (7.51)$$

The above function satisfies the properties of a Bloch function. For example, it is quite simple to check that the function is periodic in the k -space.

According to the perturbation theory, the first order energy is given by

$$\varepsilon_k = \frac{\int \Psi_{\mathbf{k}}^*(\mathbf{r}) H \Psi_{\mathbf{k}}(\mathbf{r}) dV}{\int \Psi_{\mathbf{k}}^*(\mathbf{r}) \Psi_{\mathbf{k}}(\mathbf{r}) dV} \quad (7.52)$$

Using (7.51), we get

$$\int \Psi_{\mathbf{k}}^*(\mathbf{r}) \Psi_{\mathbf{k}}(\mathbf{r}) dV = \sum_{n,m} \exp[i\mathbf{k} \cdot (\mathbf{r}_n - \mathbf{r}_m)] \int \phi^*(\mathbf{r} - \mathbf{r}_m) \phi(\mathbf{r} - \mathbf{r}_n) dV \quad (7.53)$$

For a strongly localized electron, $\phi(r - r_m)$ is significant only in the proximity of r_m . Therefore, we evaluate (7.53) by putting $m = n$ in the first approximation. If there are N atoms in the crystal, we have

$$\int \Psi_{\mathbf{k}}^*(\mathbf{r}) \Psi_{\mathbf{k}}(\mathbf{r}) dV = \sum_n 1 = N \quad (7.54)$$

Making use of (7.47), (7.48) and (7.54), electron energy ε_k is written as

$$\varepsilon_k \approx \frac{1}{N} \sum_{n,m} \exp[i\mathbf{k} \cdot (\mathbf{r}_n - \mathbf{r}_m)] \int \phi^*(\mathbf{r} - \mathbf{r}_m) [\varepsilon_0 + v(\mathbf{r} - \mathbf{r}_n)] \phi(\mathbf{r} - \mathbf{r}_n) dV \quad (7.55)$$

For the term containing ε_0 , we again neglect the overlap between the nearest neighbours, putting $m = n$. Therefore,

$$\frac{\varepsilon_0}{N} \sum_{n,m} \exp[i\mathbf{k} \cdot (\mathbf{r}_n - \mathbf{r}_m)] \int \phi^*(\mathbf{r} - \mathbf{r}_m) \phi(\mathbf{r} - \mathbf{r}_n) dV = \varepsilon_0 \quad (7.56)$$

Including the overlap up to the nearest neighbours for the perturbation term, we write

$$\int \phi^*(\mathbf{r} - \mathbf{r}_n) v(\mathbf{r} - \mathbf{r}_n) \phi(\mathbf{r} - \mathbf{r}_n) dV = -\alpha \quad (7.57)$$

(on the same atom)

$$\int \phi^*(\mathbf{r} - \mathbf{r}_m) v(\mathbf{r} - \mathbf{r}_n) \phi(\mathbf{r} - \mathbf{r}_n) dV = -\gamma \quad (7.58)$$

(between the nearest neighbours)

All terms in the summation over n , each of which is evaluated over all m (either the same

atom or the nearest neighbours), are equal in magnitude on the demand of periodicity. Since the summation over n runs over all atoms in the crystal, the sum is simply N times the value of a single term. This factor of N cancels with the factor of N in the denominator. In view of this and relations (7.56), (7.57), (7.58), the electron energy ε_k assumes the form

$$\varepsilon_k \sim \varepsilon_0 - \alpha - \gamma \sum_m e^{ik \cdot (r_n - r_m)} \quad (7.59)$$

The sum in (7.59) is carried only over the nearest neighbours. In a simple cubic crystal with lattice constant ' a ', the nearest-neighbour atoms are at

$$r_n - r_m = (\pm a, 0, 0); \quad (0, \pm a, 0); \quad (0, 0, \pm a)$$

This gives the s-state energy in the crystal as

$$\varepsilon_k \approx \varepsilon_0 - \alpha - 2\gamma(\cos k_x a + \cos k_y a + \cos k_z a) \quad (7.60)$$

When the atoms are brought together to form a crystal, the single atomic energy level ε_0 broadens to constitute an energy band whose component levels are defined by (7.60). We can determine the bandwidth as follows.

Energy of the band at its bottom:

The bottom lies at $k = 0$, giving

$$\varepsilon_{\text{bott}} \approx \varepsilon_0 - \alpha - 6\gamma \quad (7.61)$$

Energy of the band at its top:

The top occurs at $k = \pm \pi/a$, giving

$$\varepsilon_{\text{top}} \approx \varepsilon_0 - \alpha + 6\gamma \quad (7.62)$$

Therefore, from (7.61) and (7.62),

$$\text{the bandwidth} = 12\gamma \quad (7.63)$$

Thus the bandwidth is proportional to γ which represents the overlap of atomic orbitals (7.58). A qualitative illustration of the results of tight-binding calculation for a simple cubic crystal is made in Fig. 7.13. Whereas γ determines the bandwidth, α is interpreted as the lowering of the centre of gravity of the free atomic level on forming the solid. As one proceeds from inner to outer shells in an atom, the width of the respective energy bands goes on increasing because of more and more overlap. This is consistent with the result of the Kronig-Penney model and also confirmed by experiments. It is amply clear from even Fig. 7.13 where the second band appears as much wider.

Furthermore, the effect of tight binding on the energy surfaces of a simple cubic crystal can be examined with the help of (7.60). It is instructive to do this for the limiting k -values in the reduced zone scheme, i.e. near the centre and the boundary of the first zone. Near the centre $ka \leq 1$ and we expand the cosine function to obtain

$$\varepsilon_k \approx \varepsilon_0 - \alpha - 6\gamma + \gamma k^2 a^2 \quad (7.64)$$

These values refer to the bottom region of the band and conform to the constant spherical energy surfaces. But as the wavevector increases, the shape gets distorted and deviates sufficiently from the spherical nature at large values of k [see Fig. 7.14(a)].

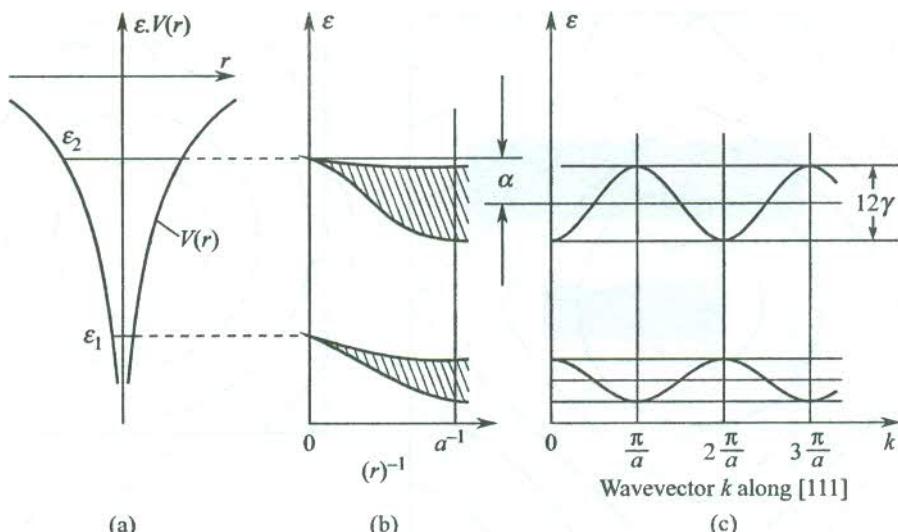


FIG. 7.13 Illustration of the result of a tight-binding approximation for a primitive cubic lattice of lattice constant ' a '. (a) Position of energy levels ϵ_1 and ϵ_2 in the potential $V(r)$ of the free atom. (b) Variation in broadening of the levels ϵ_1 and ϵ_2 as a function of the reciprocal atomic separation r^{-1} . (c) Variation in one-electron energy ϵ as a function of the wavevector $k(1, 1, 1)$ in the direction of the body diagonal [111].

In order to investigate the region near the zone boundary, we express k in terms of its value k' as measured from the zone boundary:

$$k = \frac{\pi}{a} - k' \quad (7.65)$$

Substituting (7.65) in (7.60) and appreciating that $k'a \leq 1$ near the zone boundary, we obtain

$$\epsilon_k \approx \epsilon_0 - \alpha + 6\gamma - \gamma k'^2 a^2 \quad (7.66)$$

The above result is similar to that of (7.64) with the difference that the spherical surfaces are centred at the corners of the zone [Fig. 7.14(a)]. We may compare the energy surfaces in the TBA with those in the NFE model, shown in Fig. 7.14(b). The main difference lies in the fact that the spherical shape is maintained to much larger values of k in the NFE model. This only shows that the results in the two models are almost similar for small wavevectors. In view of this fact, the k -dependent part of the energy dispersion (7.64) near the zone's centre is comparable with $\hbar^2 k^2 / 2m$ (for the free electrons), and we have

$$\gamma k'^2 a^2 = \frac{\hbar^2 k^2}{2 m^*} \quad (7.67)$$

giving

$$m^* = \frac{\hbar^2}{2\gamma a^2} \quad (7.68)$$

where m^* is identified as the effective mass of the electron, implying that the electron mass should be treated as variable. Relation (7.68) predicts that m^* is larger for electrons in the inner shells whose

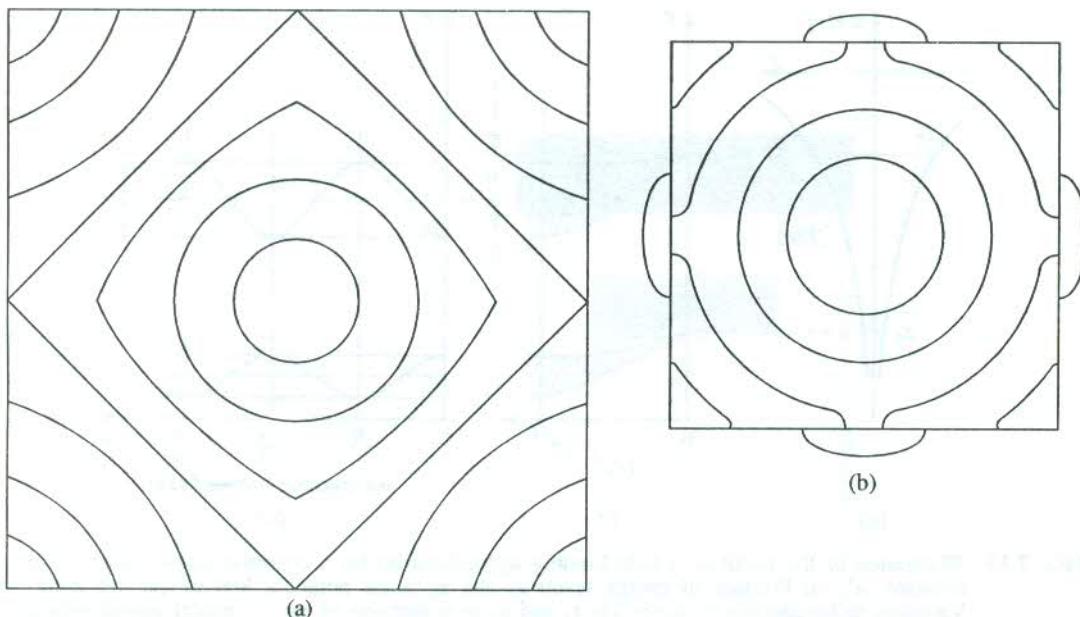


FIG. 7.14 (a) Constant energy curves for the tight-binding approximation for a simple cube crystal in the plane $k_z = 0$. Energy surfaces are spherical around the zone centre for only small k -values. The surfaces are again spherical with centres at the corners of the zone for small k -values, measured from the corners. (b) Constant energy curves in the NFE model. Energy surfaces are spherical up to fairly large values of k . Notice the changes caused by the zone boundary.

overlap is far less. We consider this as an invaluable result of the theory of tight-binding approximation, especially because the concept of effective mass has remarkably improved the understanding of several physical properties ranging from electronic conduction to the complex optical phenomena.

7.9 THE WIGNER–SEITZ CELLULAR METHOD

The significance of the Wigner–Seitz model is substantiated by the impressive success it achieved in accounting for the band structure and the cohesive energy of alkali metals. The first systematic calculation of energy bands appeared in the form of this model. The technique of calculation, referred to as a *cellular method*, is based mainly on the symmetry properties of a certain primitive cell designed by Wigner and Seitz themselves. The method of construction of this cell has already been described in Section 1.2. The alkali metals have the BCC structure for which the Wigner–Seitz cell is a polyhedron.

The entire crystal volume is imagined to have been filled up with identical polyhedra, assuming that there is only one electron in a given polyhedron at a time, together with the positively charged metal ion at the centre. The polyhedra being neutral, the interactions among themselves are initially neglected and only the interaction within each polyhedron is considered. The electron is supposed to move in the spherically symmetric potential field of the ion. The field is assumed not to extend past the boundaries of the polyhedron.

We consider the extreme case of the $\mathbf{k} = 0$ state. Then the Bloch wavefunction $\Psi_{\mathbf{k}}(\mathbf{r})$ has the form

$$\Psi(\mathbf{r}) = u_0(\mathbf{r}) \quad (7.69)$$

This wavefunction itself is periodic in the crystal, i.e. the wavefunction remains unchanged when translated from one face to the opposite face of the cell. It requires that $\frac{\partial \Psi}{\partial \mathbf{n}} = 0$ at the boundary of the polyhedron, with \mathbf{n} as a direction normal to any of the faces of the polyhedron. The above condition in the crystal replaces the free atom boundary condition, $\Psi(r) \rightarrow 0$ as $r \rightarrow \infty$, by

$$\left(\frac{\partial \Psi}{\partial r} \right)_{r=r_s} = 0 \quad (7.70)$$

where r_s is the radius of a sphere to which the polyhedron may be approximated. Accordingly, the volume of the polyhedron may be given by $4\pi r_s^3/3$.

The solution of the one-electron Schrödinger wave equation is much easier for $k = 0$ than for any general k -value because $u_0(\mathbf{r})$ is non-degenerate and observes the full symmetry of the crystal. Wigner and Seitz gave an accurate estimate of $u_0(\mathbf{r})$. In view of the boundary condition (7.70), the exercise simply reduces to solving the radial Schrödinger wave equation:

$$\left[\frac{1}{\mathbf{r}^2} \frac{\partial}{\partial \mathbf{r}} \left(\mathbf{r}^2 \frac{\partial}{\partial \mathbf{r}} \right) + \frac{2m}{\hbar^2} (\varepsilon_0 - V_0(\mathbf{r})) \right] \Psi(\mathbf{r}) = 0 \quad (7.71)$$

where $V_0(r)$ is the potential energy of an s-electron and ε_0 is the energy eigenvalue in the field of a crystalline ion within one of the polyhedra.

However, for a general wavevector we have to solve,

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V_0(\mathbf{r}) \right] u_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) = \varepsilon_{\mathbf{k}} u_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (7.72)$$

But,

$$\begin{aligned} \nabla^2 [u_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r})] &= \nabla [i\mathbf{k} u_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) + \exp(i\mathbf{k} \cdot \mathbf{r}) \nabla u_{\mathbf{k}}(\mathbf{r})] \\ &= \exp(i\mathbf{k} \cdot \mathbf{r}) [\nabla^2 u_{\mathbf{k}}(\mathbf{r}) - \mathbf{k}^2 u_{\mathbf{k}}(\mathbf{r}) + 2i\mathbf{k} \cdot \nabla u_{\mathbf{k}}(\mathbf{r})] \end{aligned} \quad (7.73)$$

Putting (7.73) in (7.72), we get

$$\left[-\frac{\hbar^2}{2m} (\nabla^2 + 2i\mathbf{k} \cdot \nabla) + V_0(\mathbf{r}) \right] u_{\mathbf{k}}(\mathbf{r}) = \left(\varepsilon_{\mathbf{k}} - \frac{\hbar^2 k^2}{2m} \right) u_{\mathbf{k}}(\mathbf{r}) \quad (7.74)$$

It must be observed that $u_0(\mathbf{r})$ is not an exact solution to (7.74) which is in fact satisfied by $u_{\mathbf{k}}(\mathbf{r})$, the periodic part of the general Bloch function.

We treat the $\mathbf{k} \cdot \nabla$ terms as a perturbation and insist that $u_{\mathbf{k}}(\mathbf{r})$ obey the boundary condition (7.70). This gives electron energies $\varepsilon_{\mathbf{k}}$ in the form

$$\varepsilon_{\mathbf{k}} = \varepsilon_0 + \frac{\hbar^2 k^2}{2m} \quad (7.75)$$

which gives energies in the shape of a band as measured from the level of ε_0 .

We must appreciate that these calculations depend only on the atomic volume and are independent of the crystal structure. Therefore, for a solid metal and a liquid of equal density we expect the same results in this model.

The 3s radial wavefunction in sodium metal as estimated in the Wigner–Seitz model at $k = 0$ (the Brillouin zone centre) and at the Brillouin zone boundary is plotted in Fig. 7.15. The 3s atomic wavefunction is also drawn for comparison. When the wavefunction is subject to the boundary condition (7.70), midway between the neighbouring atoms ($r = r_s$), the solution of the radial Schrödinger equation (7.71) for $k = 0$ yields the eigenvalue (ϵ_0) as -8.2 eV. This value* is considerably lower than the ground state energy of the free atom (-5.15 eV), obtained by applying the boundary condition $\Psi(r) \rightarrow 0$ as $r \rightarrow \infty$. The calculated energy of 3s orbitals at the zone boundary is found to be $+2.7$ eV. It should be realized that these orbitals are empty because the 3s energy band in sodium is only half-filled (refer to Section 7.7) and the corresponding states are located near the top of the band. The negligible amplitude of the Wigner–Seitz wavefunction (often referred to as the cellular wavefunction) at the zone boundary is consistent with this fact since it gives little probability for the states in this region to be occupied. It is significant to notice that all the three wavefunctions are identical in the core of the metal ion.

The shape of the cellular wavefunction at the zone centre ($k = 0$) carries the most vital information about the behaviour of the 3s electron in sodium metal. The plot in Fig. 7.15 shows the variation of the wavefunction as a function of the distance of the electron from the centre of the atomic polyhedron. The wavefunction is flat over about 90 per cent of the atomic volume. The total charge distribution in the flat region corresponds to the charge on an electron. This takes us to the conclusion that $u_k(r)$ remains constant (u_0) over most of the atomic volume and the plane wave part of the wavefunction alone determines the electron motion in this region. Thus the valence electrons of sodium behave mostly as free electrons. This is found to be true for other alkali metals too. But

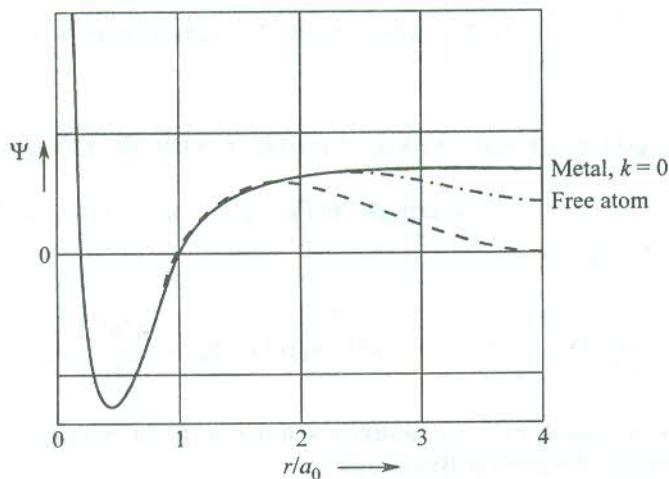


FIG. 7.15 Variation of the radial wavefunction for the 3s electron in sodium. The continuous curve describes the Wigner–Seitz wavefunction (cellular wavefunction) at the Brillouin zone centre ($k = 0$). The lower of the dashed curve represents the cellular wavefunction at the zone boundary. The dashed curve in between the cellular plots denotes the wavefunction in the free atom. The distance r of the electron from the centre of the atomic polyhedron is measured in units of the Bohr radius a_0 .

* E.P. Wigner and F. Seitz, *Phys. Rev.*, **43**, 804 (1933); **46**, 509 (1934).

the results for the noble metals, also monovalent, are on the other extreme. The ratio of the ionic to the atomic radius is close to unity, making them act like hard spheres. Because of this reason, the noble metals in contrast to the alkali metals cannot be treated in the framework of the free electron approximation.

7.9.1 Estimation of Cohesive Energy

Concepts of cohesive energy were discussed in Chapter 2. We know that binding of atoms results in the lowering of their ground orbital energy. Thus the ground state energy of an electron in the crystal is lower than that in the free atom. This lowering in energy, taken as a measure of the cohesive energy, is a consequence of replacing the Schrödinger boundary condition in the free atom by the periodic boundary condition (7.70). On the demand of energy conservation, an increase in the binding energy is offset by the Fermi energy contribution to the kinetic energy of valence electrons. The spherical approximation of the Wigner–Seitz theory works satisfactorily in the BCC and FCC crystals. We take sodium metal, a BCC structure, for our discussion.

The electron energy in the crystal is given by (7.75). The value of ε_0 is estimated from the Wigner–Seitz model. The second term in (7.75), denoting the average kinetic energy per electron, is obtained from the free electron theory using (6.52). The ground state energy of the valence electron in the crystal is then,

$$\varepsilon_k = \varepsilon_0 + \frac{\hbar^2 k^2}{2m} = \varepsilon_0 + \frac{3}{5} \varepsilon_F \quad (7.76)$$

For sodium, $\varepsilon_0 = -8.2$ eV; $\varepsilon_F = 3.1$ eV (from tables). These values give, $\varepsilon_k = -6.34$ eV.

The above value of ε_k when subtracted from the corresponding value in the free atom (-5.15 eV) gives the cohesive energy as equal to 1.19 eV. The close experimental value (1.13 eV) demonstrates the success of the Wigner–Seitz approximation in alkali metals.

The model provides additional information from the ε_k versus r_s plot. The minimum value of ε_k (7.76) defines the theoretical lattice parameter. For this particular value of r_s , the cohesive energy and compressibility of alkali metals have been calculated and found in good agreement with the experiments.*†

7.10 METHODS OF BAND STRUCTURE CALCULATION IN USE: A QUALITATIVE VIEW

In the last two sections we discussed two extreme cases of band structure calculation in the form of the tight-binding and Wigner–Seitz approximations. While one overemphasizes the atomic aspect, the other overemphasizes the plane wave aspect of the Bloch function. The tight-binding approximation is useful for interpolation and the Wigner–Seitz approximation gives a good account of several properties of alkali metals. But the methods that actually work for a variety of solids are somewhat much different. Though these methods are mathematically tedious, the problem has eased considerably with access to modern computers. No clear-cut prescription can be handed out before beginning the

* F. Seitz, op. cit., p. 365.

† J. Bardeen, *J. Chem. Phys.*, **6**, 367, 372 (1938).

exercise for a certain material. We grow wiser with experience in suitably executing the computer programme which is generally complex.

The scope of this book, however, does not permit us to give a full account of the recent techniques of calculation. Of these the orthogonalized plane wave (OPW) and the augmented plane wave (APW) methods are most prominent and advanced. The pseudopotential method is also often used on account of its ability to predict the energy–wavevector relationship with acceptable accuracy.

We are aware that a single plane wave describes the wavefunction outside the core regions quite well. But a complete wavefunction must be represented by an appropriate combination of a large number of plane waves because of its rapid variation in the vicinity of the core. The knowledge of atomic orbitals at each site readily enables us to distinguish between a plane wave and the complete wavefunction. Based on this idea, Herring formulated the OPW method. The orthogonalized plane waves are in fact linear combinations of plane waves and mixtures of atomic wavefunctions of the occupied states of the cores. This takes care of the electron behaviour both within and outside the core regions. The method has been applied to several metals and non-metals with reasonable success in getting the band shapes.

With a large value of the parameter P in (7.25), the Kronig–Penney model changes over to a one-dimensional form of the APW method. It is another way to improve upon the NFE model by approximating the periodic potential suitably in the regions within and outside the cores. The potential within a sphere around each ion core is taken as the usual atomic potential and assumed constant outside the core regions. The Schrödinger wave equation, when solved in the two regions, yields two separate solutions that are matched on the spherical boundaries between the regions. The wavefunction within the core is expanded in spherical harmonics. But outside the core region it is represented by a combination of plane waves, known as an augmented plane wave. The matching of plane waves onto the atomic functions is the most difficult aspect of the exercise. But with the availability of fast computers these days, the technique is made to work for a large number of metals and semiconductors.

The pseudopotential approach has its roots essentially in the effectiveness of the NFE model (Section 7.4) in many solids. The admixture of core states shows, surprisingly, little effect on the energy of higher states in many metallic materials. In this approach, the periodic potential energy function is replaced by a modified potential energy function with a few Fourier coefficients V_g that refer to only short reciprocal lattice vectors. The modified potential is called the *pseudopotential*. In the well-known Empirical Pseudopotential Method (EPM), the Fourier coefficients are deduced from theoretical fits to the optical reflectance and absorption data of the crystal of interest. The potential is fairly smooth and free from the deep wells of the free potential. Success in the interpretation of band structures notwithstanding, difficulties in the description of some other properties of electron come to surface. It is likely to happen since the method may yield an incorrect wavefunction. Such a problem is resolved by correcting the wavefunction for the due representation of the atomic component.

The band structures as derived from calculations are too complicated to be discussed at this stage. We will refer to these for some simple materials in the next two chapters while gaining more familiarity with the concept of energy surfaces.

SUMMARY

1. The statement of the Bloch theorem is

$$\Psi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r})$$

where $\Psi_{\mathbf{k}}(\mathbf{r})$ denotes an electron wave describing an electron with wavevector \mathbf{k} at \mathbf{r} in the crystal, and $u_{\mathbf{k}}(\mathbf{r})$ is the periodic in the crystal lattice. That is, $u_{\mathbf{k}}(\mathbf{r} + \mathbf{t}_n) = u_{\mathbf{k}}(\mathbf{r})$, where \mathbf{t}_n is an arbitrary translation vector. The function $\Psi_{\mathbf{k}}(\mathbf{r})$ is called the Bloch function.

2. Bloch functions $\Psi_{\mathbf{k}}(\mathbf{r})$ and their eigenvalues $\varepsilon_{\mathbf{k}}$ are periodic in the reciprocal lattice. That is,

$$\Psi_{\mathbf{k} + \mathbf{g}}(\mathbf{r}) = \Psi_{\mathbf{k}}(\mathbf{r})$$

$$\varepsilon_{\mathbf{k} + \mathbf{g}} = \varepsilon_{\mathbf{k}}$$

3. In a crystal of N primitive cells, there are $2N$ independent orbitals in an energy band.
4. Energy bands are separated by regions in which no solutions to electron wave equation exist. These regions are called *band gaps*.
5. The bandwidth is proportional to the overlap of atomic orbitals.
6. The electron effective mass (m^*) is inversely proportional to overlap, i.e. m^* is large for electrons in the inner shells.
7. The cohesive energy of simple metals is estimated by calculating the lowering of the $\mathbf{k} = 0$ orbital in the conduction band. In the calculations, the boundary condition on the wavefunction is changed from Schrödinger ($\Psi \rightarrow 0$) as ($r \rightarrow \infty$) to Wigner–Seitz [$(d\Psi/dr)_{r=r_s} = 0$] condition.

PROBLEMS

- 7.1 Solve the Schrödinger equation for the potential

$$V(x, y, z) = \begin{cases} \infty, & x \leq 0 \\ 0, & x \geq 0 \end{cases}$$

and calculate the charge density

$$\rho(x, y, z) = \frac{e}{V} \sum_{k_x} \sum_{k_y} \sum_{k_z} |\Psi_{\mathbf{k}}|^2$$

- 7.2 For $P \ll 1$ in the Kronig–Penney model, show that the energy of the lowest band at $k = 0$ is given by

$$\varepsilon = \frac{\hbar^2 P}{ma^2}$$

- 7.3 Make an approximate estimate of the band gap at the corner point of the Brillouin zone for a square lattice in two dimensions, assuming the crystal potential to be given by

$$V(x, y) = -4V_0 \cos\left(\frac{2\pi x}{a}\right) \cos\left(\frac{2\pi y}{a}\right)$$

- 7.4 Find the energy–wavevector relationship for a one-dimensional monatomic crystal of lattice constant ‘ a ’ using the tight-binding approximation. Use this relation to obtain an expression for the effective mass. For which value of k is the electron velocity maximum?
- 7.5 Using the tight-binding approximation for the overlap up to the nearest neighbours, derive an expression for the energy of an s-band in (i) a BCC crystal and (ii) in an FCC crystal.
- 7.6 Prove that the density of states in the tight-binding approximation of a simple cubic crystal near the bottom of an s-band is given by

$$D(\varepsilon) = \frac{1}{2\pi^2 a^3 \gamma^{3/2}} (\varepsilon - \varepsilon(0))^{1/2}$$

where $\varepsilon(0)$ is the energy at $k = 0$.

- 7.7 For a square lattice in two dimensions:

- (a) Show that the kinetic energy of a free electron in a corner of the first Brillouin zone is a factor of two larger than that of an electron in the middle of the Brillouin zone edge.
- (b) Determine the corresponding factor for a primitive lattice in three dimensions.
- (c) Find out if the bands can overlap in a two-dimensional lattice by plotting the following dispersion relations for free electrons in a weak periodic potential.

$$\varepsilon(k_y) \text{ for } k_x = 0$$

$$\varepsilon(k_y) \text{ for } k_x = \pi/a$$

$$\varepsilon(k) \text{ for } k_x = k_y$$

- 7.8 Explain the variation of forbidden energy gap E_g with temperature in diamond.

SUGGESTED FURTHER READING

Ashcroft, N.W. and N.D. Mermin (Saunders College, 1988).

Brillouin, L., *Wave Propagation in Periodic Structures* (Academic Press, 1960).

Harrison, W.A., *Electronic Structure and the Properties of Solids* (Freeman, 1980).

Madelung, O., *Introduction to Solid State Theory*, Vol. 2 (Springer, 1978).

Ziman, J.M., *Principles of the Theory of Solids* (Cambridge, 1972).

Mobile Electrons and Fermi Surfaces

Electronic transport properties of metals and semiconductors depend basically on the characteristics of mobile electrons. However, it must be recalled that there are some phenomena which can be explained by the free electron theory only if the existence of positive charge carriers is accepted. The observed positive Hall coefficient for certain metals (see Table 6.3) is an evidence to this effect. The band theory provides a satisfactory explanation to such phenomena by assigning credibility to the concept of positive charge carriers, named as holes. Hence, to be able to account for the transport of charge carriers we need to be familiar with characteristics of holes. In this chapter, we will learn that the motion of charge carriers is strongly controlled by the shapes of the constant energy surfaces. Since the mobile electrons are close to the Fermi surface, a precise knowledge of its shape is a prerequisite to the description of the motion of charge carriers. We begin our task with the discussion on the concept of holes.

8.1 CONCEPT OF HOLES

Let us elaborate on how the concept of a hole evolves. An account of the electronic behaviour, that is linked to the concept, together with the characteristics of holes is given in a series of points as follows.

1. Consider a crystal of unit volume under the action of a steady electric field. Since there can be two electrons in each state, the number of electrons per unit volume of the \mathbf{k} -space will be equal to $2/8\pi^3$. If $v(\mathbf{k})$ denotes the velocity of an electron with wavevector \mathbf{k} , the current density owing to electrons in a band occupied up to a certain level can be written as

$$j = -e \int_{\text{occupied}} \frac{v(\mathbf{k}) d^3\mathbf{k}}{4\pi^3} \quad (8.1)$$

where $d^3\mathbf{k}/4\pi^3$ represents the number of electrons in the small volume element $d^3\mathbf{k}$ around \mathbf{k} .

Had the unoccupied states in the band been occupied by electrons, their contribution to the current density would be

$$-e \int_{\text{unoccupied}} \frac{v(\mathbf{k}) d^3\mathbf{k}}{4\pi^3} \quad (8.2)$$

As there is no current in a completely filled band, we have

$$-e \int_{\text{occupied}} \frac{v(\mathbf{k}) d^3 \mathbf{k}}{4\pi^3} - e \int_{\text{unoccupied}} \frac{v(\mathbf{k}) d^3 \mathbf{k}}{4\pi^3} = 0$$

or

$$j = -e \int_{\text{occupied}} \frac{v(\mathbf{k}) d^3 \mathbf{k}}{4\pi^3} = e \int_{\text{unoccupied}} \frac{v(\mathbf{k}) d^3 \mathbf{k}}{4\pi^3} \quad (8.3)$$

Relation (8.3) suggests that the current owing to electrons in the occupied levels is equal to the current that would flow if only the presently unoccupied states in the band were populated, all with positive charges e . These charges are called *holes*. Holes represent essentially the empty states in a band. Holes are thus generally expected to occur in the region close to the top of a band.

2. We refer back to the picture of electron energy bands depicted in Fig. 7.9 and use it to grasp the characteristics of holes. Conventionally, the lower band in the figure is known as the valence band and the upper as the conduction band. Figure 8.1 shows the excitation in which an electron with wavevector \mathbf{k}_e at state E in the valence band has been excited to state Q in the conduction band, on the absorption of a photon of negligible wavevector. This creates a hole at E with a wavevector usually denoted by \mathbf{k}_h in the graphical representation. The relationship between \mathbf{k}_e and \mathbf{k}_h is easily derived with the following arguments. The momentum of a completely filled band is zero. That is,

$$\hbar \sum \mathbf{k} = 0 \quad (8.4)$$

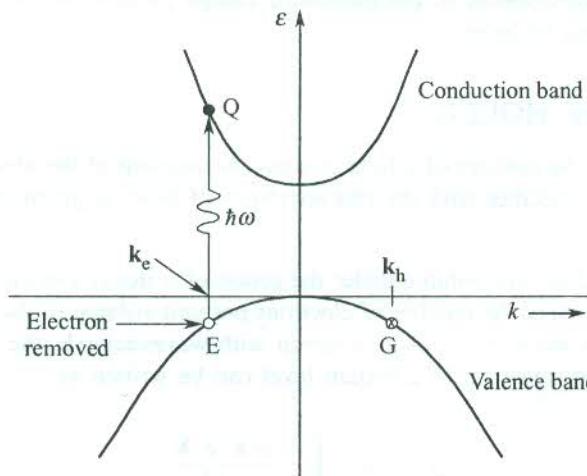


FIG. 8.1 Excitation of an electron with wavevector \mathbf{k}_e from state E in the valence band to state Q in the conduction band by the absorption of a photon of negligible wavevector. The excitation creates a hole with wavevector \mathbf{k}_h ($= -\mathbf{k}_e$) in the state G of the valence band.

The symmetry of the band in Fig. 8.1 suggests that in a filled band for an electron with wavevector \mathbf{k}_e on the left side of the energy axis there is an electron with wavevector $-\mathbf{k}_e$ on the other side to ensure that (8.4) holds good. When an electron with wavevector \mathbf{k}_e is excited to the conduction band on absorption of an optical photon, the momentum of the valence band becomes

$$\hbar [\sum \mathbf{k} - \mathbf{k}_e] = -\hbar \mathbf{k}_e \quad (8.5)$$

which is rightly equal to the momentum of the electron whose counterpart on the other side has been excited to the conduction band. In consistency with our arguments on unoccupied states, the momentum of the band as expressed by (8.5) can be totally associated with the positive charge (hole) occupying the lone empty state of the band. Therefore, if \mathbf{k}_h denotes the hole wavevector, we have

$$\hbar\mathbf{k}_h = -\hbar\mathbf{k}_e$$

or

$$\mathbf{k}_h = -\mathbf{k}_e \quad (8.6)$$

where \mathbf{k}_e is the wavevector of the missing electron from the valence band.

According to (8.6), the wavevector of the hole is equal to that of the unpaired electron at point G in the valence band whose counterpart (the missing electron) has been transferred to the conduction band.

3. The inversion operator in a symmetric band takes \mathbf{k}_e to $-\mathbf{k}_e$ which denotes the hole wavevector (8.6). This way we can locate the tips of the wavevectors of all the holes that would be created by exciting electrons from the valence band. We thus effectively get the dispersion curve of holes. Taking valence band as the electron band, we associate this curve with the hole band. It is simply the electron band turned upside down as shown in Fig. 8.2. This figure is helpful in simulating the motion of a hole. For a symmetric band,

$$\epsilon_e(\mathbf{k}_e) = \epsilon_e(-\mathbf{k}_e) = -\epsilon_h(-\mathbf{k}_e) = -\epsilon_h(\mathbf{k}_h)$$

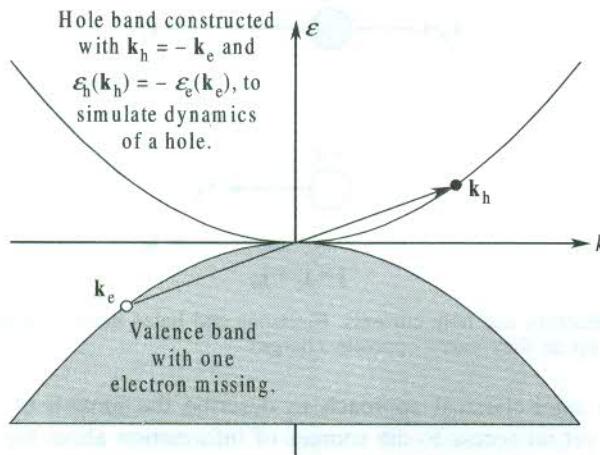


FIG. 8.2 Simulation of hole dynamics by performing the inversion operation in a symmetric band. With the valence band as the electron band, the hole band is constructed using $\mathbf{k}_h = -\mathbf{k}_e$ and $\epsilon_h(\mathbf{k}_h) = -\epsilon_e(\mathbf{k}_e)$.

4. In the presence of an electric field the current contribution by the unpaired electron at G is

$$\begin{aligned} -ev(-\mathbf{k}_e) &= -[-ev(\mathbf{k}_e)] \\ &= -(current of the missing electron) \\ &= current contributed by the hole with charge +e \end{aligned}$$

Thus the velocity of the hole is equal in magnitude and direction to that of the missing electron, i.e.

$$\mathbf{v}(\mathbf{k}_h) = \mathbf{v}(\mathbf{k}_e)$$

5. It will be shown in the next section that the effective mass is inversely proportional to the curvature, $\frac{d^2\varepsilon}{dk^2}$. Since the curvature of the hole band is opposite to that of the electron band (Fig. 8.2), we get

$$m_h^* = -m_e^* \quad (8.7)$$

where m_h^* and m_e^* denote, respectively, the effective mass of the hole and that of the electron missing from the valence band. For example, we will see that m_e^* is negative near the top of a band. Relation (8.7) implies that m_h^* is positive near the top of the band.

Lastly, we comment on the role of holes in the flow of current. It is sheer for convenience, especially in semiconductors, that we consider the current in the valence band to be carried entirely by holes—the fictitious particles of positive charge—even though the current is actually driven by electrons. In this model, contributions to the total current come from the flow of electrons in the conduction band and from the flow of holes in the valence band. The holes, being centres of the positive charge, move along the direction of the electric field whereas the electrons move opposite to the field direction (Fig. 8.3). Since the direction of the current owing to the two types of carriers, becomes the same, the two components of the current add up to give the net current.

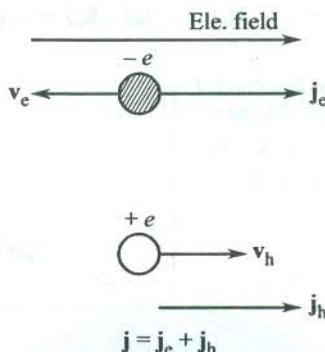


FIG. 8.3 The state of electron and hole currents. Electrons and holes move in opposite directions, but their currents add up as they carry opposite charges.

We will follow a semi-classical approach to describe the motion of electrons and holes in crystals. But we have yet no access to the sources of information about the mass, the velocity and the acceleration for applying the Newton's laws to the motion of these particles. Due caution is demanded in view of the concept of effective mass introduced in Section 7.8. It will be shown below that both the effective mass and the velocity of concern (the group velocity) directly depend on the shape of the Fermi surface.

8.2 EFFECTIVE MASS

The actual motion of an electron in crystals does not conform to its phase velocity, defined as

$$v_p = \frac{\omega_k}{k} \quad (8.8)$$

where ω_k and k denote the angular frequency and the wavevector of the electron, respectively.

The requirement of the value of total wavevector in (8.8) makes it unsuitable for use in the reduced zone scheme which we have adopted for most of the descriptions. On the other hand, the motion of the centre of charge of the wavepacket comprising an electron traces out the actual path of the electron. The wavepacket of energy $\epsilon_{\mathbf{k}}$ moves around \mathbf{k} in the real space through the crystal at the group velocity, given by

$$\mathbf{v}_g = \frac{\partial \omega}{\partial \mathbf{k}} = \frac{1}{\hbar} \nabla_{\mathbf{k}} \epsilon \quad (8.9)$$

This definition requires the value of the wavevector as measured from the centre or the boundary of the zone which exactly fits into our scheme. Relation (8.9) shows that the direction of the group velocity is determined by the energy gradient, directed along the surface normal at the point of interest to the constant energy surface. Figure 8.4 shows that the directions of the group velocity \mathbf{v}_g and phase velocity \mathbf{v}_p do not coincide on a non-spherical energy surface. The influence of the periodic crystal potential on electrons causes deviation from the spherical shape of the free electron surfaces.

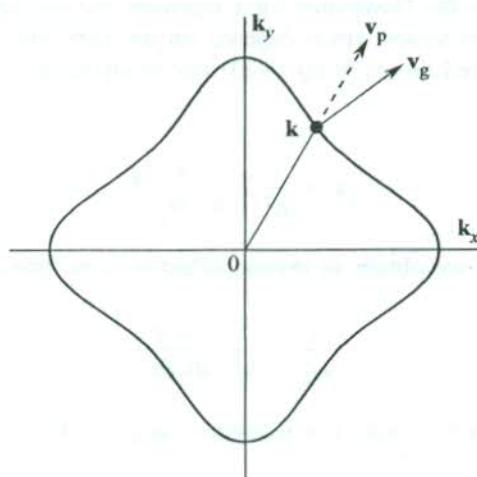


FIG. 8.4 Directions of the phase and group velocities of an electron on a non-spherical constant energy surface.

Under the action of a steady electric field, an electron experiences a force \mathbf{F} , equalling $-e\mathbf{E}$. The energy of the electron changes at the rate,

$$\frac{d\epsilon}{dt} = \mathbf{v}_g \cdot \mathbf{F} \quad (8.10)$$

But

$$d\epsilon = \left(\frac{d\epsilon}{dk} \right) \cdot dk = \nabla_{\mathbf{k}} \epsilon \cdot dk$$

or

$$\frac{d\epsilon}{dt} = \frac{1}{\hbar} \nabla_{\mathbf{k}} \epsilon \cdot \hbar \frac{dk}{dt} \quad (8.11)$$

From (8.10) and (8.11), we get

$$\mathbf{F} = \hbar \frac{d\mathbf{k}}{dt} \quad (8.12)$$

Relation (8.12) is an important relation which in accordance with the Newton's second law establishes that the electron momentum is given by $\hbar\mathbf{k}$.

Further, from (8.9) the acceleration of the electron \mathbf{a} can be written as

$$\begin{aligned} \mathbf{a} &= \frac{d\mathbf{v}_g}{dt} = \frac{1}{\hbar^2} \nabla_{\mathbf{k}} \nabla_{\mathbf{k}} \epsilon \cdot \hbar \frac{d\mathbf{k}}{dt} \\ &= \frac{1}{\hbar^2} \nabla_{\mathbf{k}} [\nabla_{\mathbf{k}} \epsilon \cdot \mathbf{F}] \end{aligned} \quad (8.13)$$

The above equation tells us that \mathbf{F} can produce a change in \mathbf{v}_g in directions other than the direction of \mathbf{F} . Comparing (8.13) with the Newtonian force equation, we see that the closest thing to a mass of the electron is an inverse tensor which depends on the curvature of the ϵ versus \mathbf{k} dispersion curve. This can be determined by resolving (8.13) into components along the three arbitrary axes. Then,

$$\mathbf{a}_i = \frac{1}{\hbar^2} \sum_j \frac{\partial^2 \epsilon}{\partial k_i \partial k_j} \mathbf{F}_j \quad (8.14)$$

with $i, j = x, y, z$; and thus we obtain an inverse effective mass tensor with components

$$\frac{1}{m_{ij}^*} = \frac{1}{\hbar^2} \frac{\partial^2 \epsilon}{\partial k_i \partial k_j} \quad (8.15)$$

where m_{ij}^* is called the *effective mass*. For isotropic energy surfaces, the effective mass is simply expressed by

$$\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{d^2 \epsilon}{dk^2} \quad (8.16)$$

In view of (8.15) or (8.16), the effective mass of an electron in an energy band is positive near the bottom of the band where the curvature is upward or $d^2\epsilon/dk^2$ is positive. Similarly, it is negative near the top of the band (see Fig. 7.6). The change in sign of the effective mass occurs at the point of inflexion on the dispersion curve. In the two-component approximation of the wavefunction (7.41), the component $\exp[i(k - g)x]$ represents the reflected wave component whose increase owing to an increase in k is a measure of the momentum transferred to the crystal lattice. At the Brillouin zone boundary (the top of band) it exactly cancels the forward moving component $\exp(ikx)$, resulting in Bragg reflection. Near the top of the band, the approach to Bragg reflection causes an overall decrease in the forward momentum of the electron even though the applied electric field is kept on increasing. It means that the electron transfers more momentum to the lattice than what is imparted to it by the field on increasing \mathbf{k} to $\mathbf{k} + \Delta\mathbf{k}$. This situation demands that the effective mass be treated as negative to satisfy the Newton's second law.

Since the curvature of the hole band is opposite to that of the electron band (see Fig. 8.2), the effective mass of a hole is positive near the top and negative near the bottom of a band. The condition (8.7) is based on the observations made in this section.

8.3 CONSTRUCTION OF THE FERMI SURFACES

As stated in the beginning, the knowledge of the shapes of the Fermi surfaces is helpful in investigating the motion of electrons in crystals. The exercise of deriving the form of the Fermi surface involves the following steps:

- (i) Construction of the Brillouin zones
- (ii) Construction of the Fermi surface in the extended zone scheme
- (iii) Reproducing pictures in the reduced and periodic zone schemes.

The condition of Bragg reflection, $2\mathbf{k} \cdot \mathbf{g} + \mathbf{g}^2 = 0$, determines the boundaries of different Brillouin zones. This condition requires that the tip of the vector $\mathbf{g}/2$ should touch the plane perpendicular to it at the mid-point of the vector \mathbf{g} . Considering a square lattice of lattice constant 'a' we define the sets of the three shortest wavevectors as

$$\mathbf{g}_1 = \frac{2\pi}{a} \hat{\mathbf{k}}_x$$

and the three equivalent vectors as

$$-\frac{2\pi}{a} \hat{\mathbf{k}}_x; \quad \pm \frac{2\pi}{a} \hat{\mathbf{k}}_y \quad (8.17)$$

$$\mathbf{g}_2 = \frac{2\pi}{a} (\hat{\mathbf{k}}_x + \hat{\mathbf{k}}_y)$$

and the three equivalent vectors as

$$-\frac{2\pi}{a} (\hat{\mathbf{k}}_x + \hat{\mathbf{k}}_y); \quad \frac{2\pi}{a} (-\hat{\mathbf{k}}_x + \hat{\mathbf{k}}_y); \quad \frac{2\pi}{a} (\hat{\mathbf{k}}_x - \hat{\mathbf{k}}_y) \quad (8.18)$$

$$\mathbf{g}_3 = \frac{4\pi}{a} \hat{\mathbf{k}}_x$$

and the three equivalent vectors as

$$-\frac{4\pi}{a} \hat{\mathbf{k}}_x; \quad \pm \frac{4\pi}{a} \hat{\mathbf{k}}_y \quad (8.19)$$

where $\hat{\mathbf{k}}_x$ and $\hat{\mathbf{k}}_y$ denote unit vectors along the \mathbf{k}_x and \mathbf{k}_y axes, respectively.

The perpendicular bisectors of \mathbf{g}_1 and its three equivalent vectors by symmetry enclose the area of the first Brillouin zone. These bisectors join the similar bisectors of \mathbf{g}_2 and its allies to bound four separate parts of the second zone. Continuing the exercise for (8.19), we find that the third zone appears as eight separate pieces (Fig. 8.5).

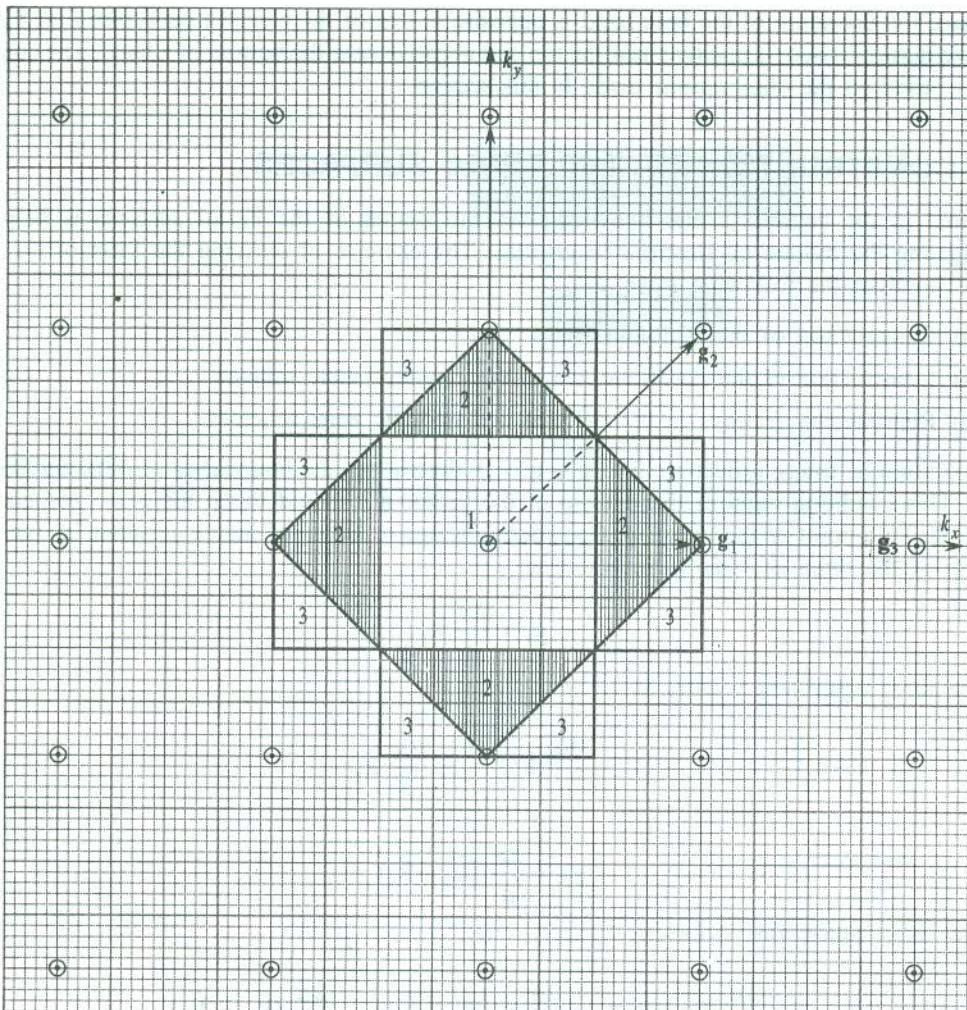


FIG. 8.5 Boundaries and locations of the first three Brillouin zones of a square lattice. \mathbf{g}_1 , \mathbf{g}_2 and \mathbf{g}_3 are the first three shortest reciprocal lattice vectors whose perpendicular bisectors form the boundaries of zones. The second and third zones appear in separated pieces.

The Fermi surfaces are constructed either by calculations or by experimental methods to be described later in this chapter. We consider at this stage the case of free electrons whose Fermi surfaces are the simplest to construct by using the Harrison's method. After determining the reciprocal lattice points of the crystal (ideally a metal) of interest, spheres around each lattice point (as centre) are drawn, taking the radius as a proportionate measure of the electron density. An occupied state in the first zone is represented by a point in the k -space that lies within at least one sphere. Similarly, the points lying within at least two spheres correspond to the occupied states in the second zone and the interpretation continues for points in three or more spheres.

The free electron Fermi surface for a square lattice is shown in Fig. 8.6. It extends into the third Brillouin zone. It is straightforward to show that parts of the second and third zones can be translated to the first zone by reciprocal lattice vectors, each with the same magnitude ($2\pi/a$). For example, the

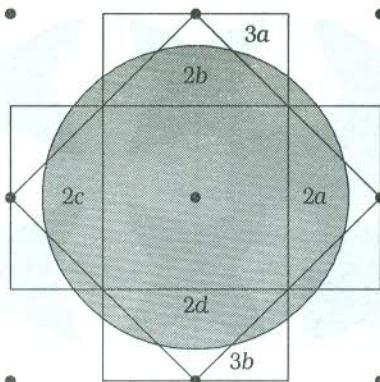


FIG. 8.6 The constant energy surface (the Fermi surface) of free electrons for a certain electron concentration of a square lattice with its first three Brillouin zones. The Fermi surface encloses only the first zone completely. The symbols a , b , c , d refer to the different parts of a zone.

part $2a$ is translated by $-2\pi\hat{\mathbf{k}}_x/a$. The first three zones in the reduced zone scheme are drawn in Fig. 8.7. The Fermi surfaces as viewed in the reduced zone scheme are pictured in Fig. 8.8. The parts of the Fermi surface in the third zone are still not connected. When drawn in the periodic zone scheme (Section 7.5), these parts join to form a lattice of rosettes that symbolizes the Fermi surface in this zone (Fig. 8.9).

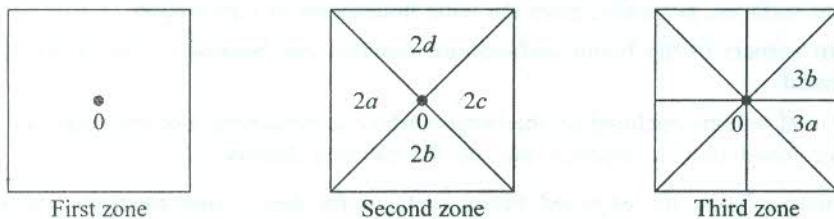


FIG. 8.7 The first, second and third Brillouin zones as pictured in the reduced zone scheme. Parts of the second and third zones are put together in a square on translating them by reciprocal lattice vectors of appropriate size.

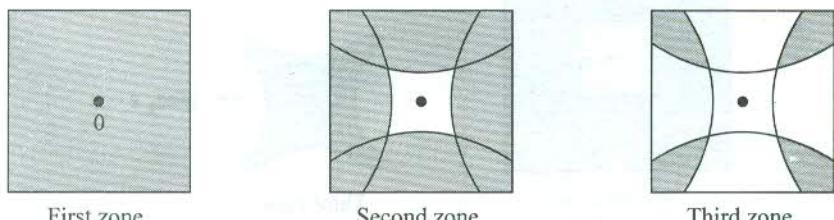


FIG. 8.8 The free electron Fermi surface of Fig. 8.6 in the reduced zone scheme. The occupied electron states are represented by shaded areas. Note that only the first zone is completely occupied.

Taking a realistic step forward, we explore how the Fermi surfaces of nearly free electrons can be derived. The proposal closely corresponds to the case of alkali metals which respond to the NFE model most positively. The Fermi surfaces of nearly free electrons are derived from those of the free electrons by taking the following into consideration:

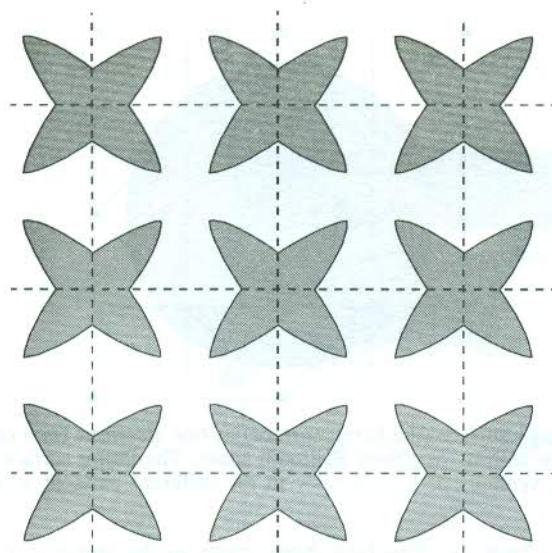


FIG. 8.9 The third zone picture of the Fermi surface in the periodic zone scheme.

- (i) Energy gaps are created at the zone boundaries because of the interaction of the electron with the periodic crystal potential.
- (ii) Fermi surfaces, generally, meet the zone boundaries at right angles.
- (iii) Sharp corners in the Fermi surfaces are rounded out because of the effect of the crystal potential.
- (iv) The total volume enclosed by the Fermi surface is independent of the details of the electron lattice interaction. It depends only on the electron density.

On qualitative basis, the expected Fermi surfaces for nearly free electrons are as shown in Fig. 8.10. These are only marginally modified from those of the free electrons (see Figs. 8.8 and 8.9).

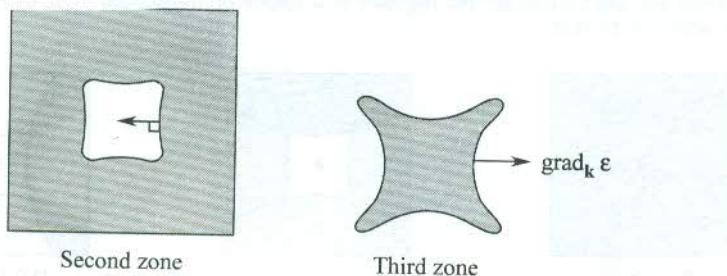


FIG. 8.10 Fermi surfaces in second and third zones for nearly free electrons. The direction of $\text{grad}_k \epsilon$ shows that in the second zone the energy increases towards the interior of the figure and in the third zone the energy increases towards the exterior of the figure. The Fermi surface in the second zone is holelike and in the third zone electronlike.

The Fermi surfaces of alkali metals as determined in the NFE approximation are nearly spherical, in agreement with the experiments. To quote, the surface is closely spherical for sodium; but in cesium a deviation of 10 per cent from the spherical shape is observed. In noble metals too,

the shape is spherical except that the surface approaches the surface of the zone most closely along [111] directions (see Fig. 8.22). The divalent metals beryllium and magnesium have weak lattice interactions and their Fermi surfaces are nearly spherical. For the trivalent metal aluminium, the figures in the first three zones are drawn in Fig. 8.11. The section of the Fermi surface in the third zone is funny and known as 'monster', typified by eight tentacles as shown in Fig. 8.11(c). Generally, the predictions of the Harrison free electron model are in good agreement with the features of the Fermi surfaces as revealed by experiments. The monster, however, needs be modified to the form shown in Fig. 8.11(d).

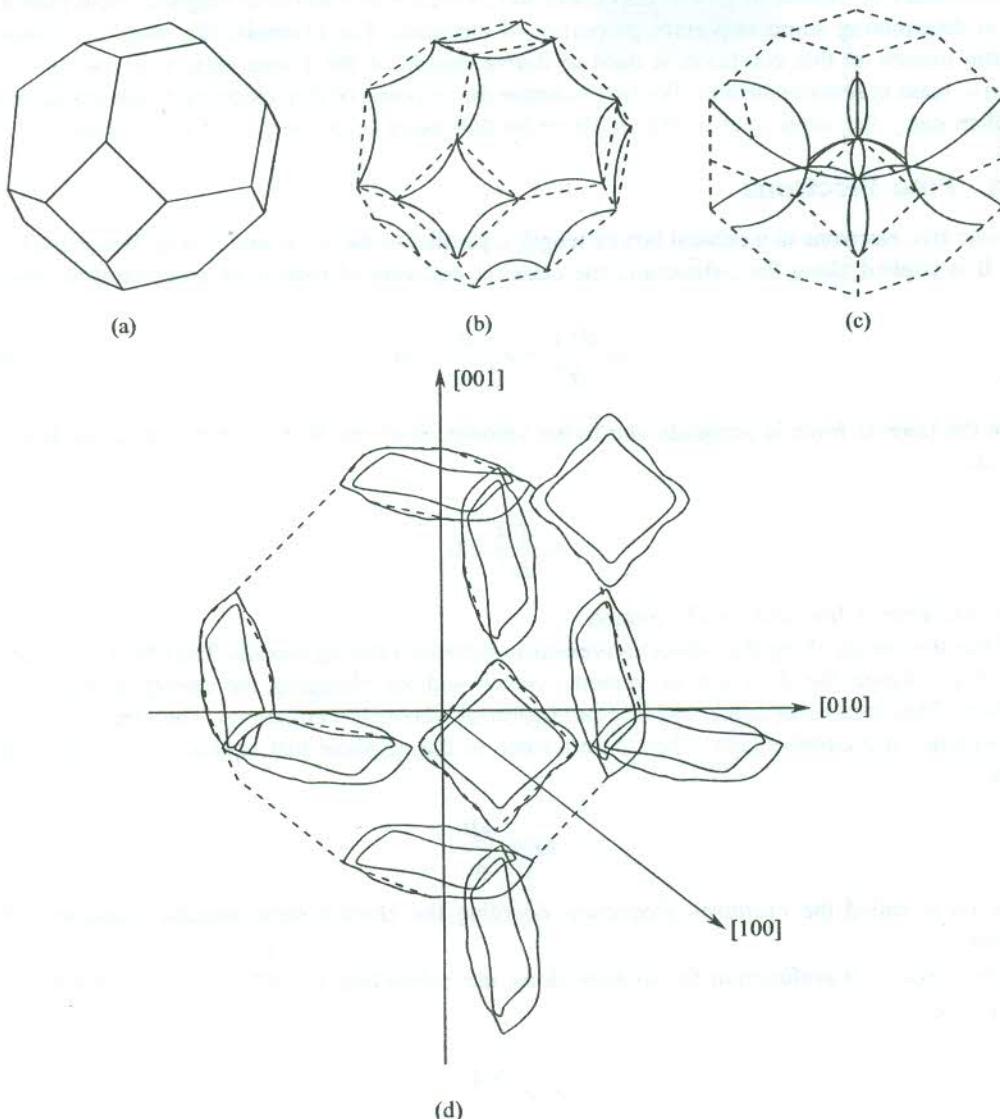


FIG. 8.11 The Fermi surface of aluminium in the (a) first zone, (b) second zone and (c) third zone showing the monster with eight tentacles. Figure 8.11(d) shows the modified form of the monster as derived with the use of experimental data. [After E.P. Volskii, *J.E.T.P.*, **46**, 123 (1963).]

Now, we need to be familiar with the motion of electrons in the presence of magnetic fields as it is crucial to experiments that are conducted to supplement all discussions carried out so far. This will largely enable us to have a picture of the dynamics of electron motion in crystals. Features of the motion are crucial to the determination of the effective mass and the Fermi surface, slated for discussion at a later stage in this chapter.

8.4 ELECTRONS IN A UNIFORM MAGNETIC FIELD

Characteristics of motion of an electron under the influence of a uniform magnetic field play a key role in determining some important properties of electrons. For example, the study of changes in electron motion in this condition is used to derive details of the Fermi surface in metals and the effective mass in semiconductors. We first examine the response of free electrons to the magnetic field and then carry the ideas over to Bloch electrons that describe properties of real solids.

8.4.1 Free Electrons

Consider free electrons in a cubical box of length L parallel to the x -, y - and z -axes. When a magnetic field \mathbf{B} is applied along the z -direction, the classical equation of motion of an electron is given by

$$m \frac{d^2 \mathbf{r}}{dt^2} = e \left(\frac{d\mathbf{r}}{dt} \times \mathbf{B} \right) \quad (8.20)$$

where the Lorentz force is perpendicular to the velocity $d\mathbf{r}/dt$ and \mathbf{B} . Its component along \mathbf{B} is zero. That is,

$$m \frac{d^2 \mathbf{z}}{dt^2} = 0 \quad (8.21)$$

which describes a free and steady motion.

Thus the energy along the z -direction remains unchanged during motion. The effect of a magnetic field is to change the direction of velocity vector without changing the energy along the field direction. This is in accordance with the prediction of Maxwell's equations. The electron moves in the xy -plane on a circular path. The Lorentz force in the xy -plane just equals the centripetal force giving,

$$\omega_c = \frac{eB}{m} \quad (8.22)$$

where ω_c is called the *cyclotron frequency*, denoting the characteristic angular frequency of the electron.

The electron wavefunction for motion along the z -direction is $\frac{1}{\sqrt{L}} \exp(i k_z z)$ which has the eigenvalue,

$$\epsilon_z = \frac{\hbar^2 k_z^2}{2m} \quad (8.23)$$

with k_z taking the same values as in the absence of a magnetic field, i.e.

$$k_z = \frac{2\pi n_z}{L}; \quad n_z \text{ being any integer} \quad (8.24)$$

But the solution of the one-electron Schrödinger wave equation in the xy -plane yields an energy that is quantized in steps of $\hbar\omega_c$ and is not simply $\frac{\hbar^2}{2m}(k_x^2 + k_y^2)$ as expected in the absence of the field. The electron energy in the xy -plane is expressed as

$$\varepsilon_{\perp} = \left(n + \frac{1}{2}\right) \hbar\omega_c \quad (8.25)$$

where n is a positive integer. The total electron energy is

$$\begin{aligned} \varepsilon_n(k_z) &= \varepsilon_z + \varepsilon_{\perp} \\ &= \frac{\hbar^2 k_z^2}{2m} + \left(n + \frac{1}{2}\right) \hbar\omega_c \end{aligned} \quad (8.26)$$

The relation (8.26) expresses the orbit quantization. We will see later that an energy level with a given n and arbitrary k_z is highly degenerate. Each group of degenerate energy levels that corresponds to a magnetic level as defined by a given pair of quantum numbers n and k_z is called a *Landau level*.

The area of the electron orbit is another quantity which is involved in certain remarkable physical phenomena to be discussed in this chapter. With this purpose, we rewrite (8.20) as

$$\hbar \frac{d\mathbf{k}}{dt} = e \left(\frac{d\mathbf{r}}{dt} \times \mathbf{B} \right) \quad (8.27)$$

The above equation expresses force in terms of the rate of change of the wavevector under the assumption that the electron moves without scattering. We must appreciate that the wavevector changes only in the xy -plane (\perp_r to \mathbf{B}) with k_z remaining constant throughout the motion. The tip of the wavevector (\mathbf{k}_{\perp}) in the xy -plane traces out a circle whose area in the \mathbf{k} -space is πk_{\perp}^2 . According to quantum mechanics the electron energy in the xy -plane (ε_{\perp}) can be expressed as $\frac{\hbar^2 k_{\perp}^2}{2m}$. On equating it to the quantized energy as given by (8.25), we get

$$S = \pi k_{\perp}^2 = \left(n + \frac{1}{2}\right) \frac{2\pi eB}{\hbar} \quad (8.28)$$

where S denotes the area of the orbit in the \mathbf{k} -space.

The difference between the areas of the consecutive orbits is

$$S_{n+1} - S_n = \left(\frac{2\pi eB}{\hbar} \right) \quad (8.29)$$

Hence the areas enclosed by the electron path in a magnetic field are quantized in units of $2\pi eB/\hbar$. The quantized areas (8.28) are cylindrical in shape. They have come to be known as *Landau tubes*. The magnetic field constantly changes the direction of the velocity vector in such a way that the tips of the allowed \mathbf{k} vectors touch a set of Landau tubes as shown in Fig. 8.12.

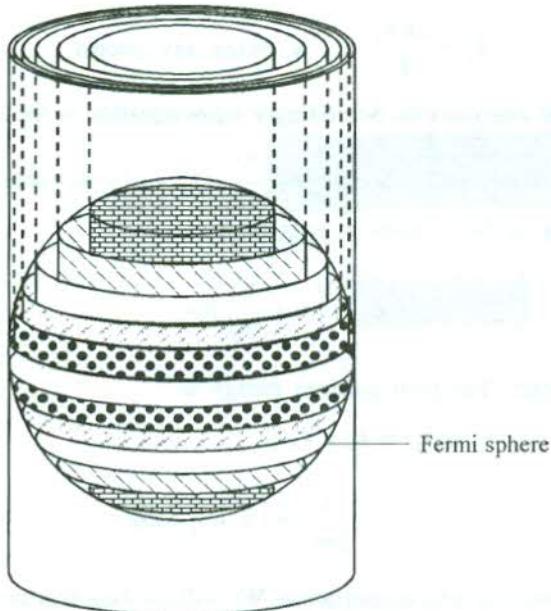


FIG. 8.12 Landau tubes (occupied regions of the \mathbf{k} -space) for a free electron gas in a magnetic field. The allowed states lie on Landau cylinders, cut off by the Fermi sphere.

We are using the Onsager's generalization of Landau free electron results which is valid for magnetic levels with only high quantum numbers. Thus only for large n is the difference $(S_{n+1} - S_n)$ a constant and independent of k_z . Electrons occupy the surface of Landau tubes. All the allowed states in the \mathbf{k} -space coalesce onto these tubes as soon as even a small magnetic field is switched on. The distance between the tubes is very small for normal magnetic fields. The total number of states remains unchanged as the states for each given k_z that lie between the tubes prior to coalescence migrate onto the tube after coalescence. But it must be borne in mind that in a magnetic field the occupied states lie on only those parts of the Landau tubes that are within the Fermi sphere.

8.4.2 Bloch Electrons

It is simple to extend the above results to electrons in an isotropic band. The motion in the \mathbf{k} -space is generalized by replacing m by the effective mass m^* . The velocity in (8.27) is taken as $\hbar\mathbf{k}/m^*$. All calculations proceed on the same lines using,

$$\omega_c = \frac{eB}{m^*} \quad (8.30)$$

In order to deal with motion in general bands, we make use of (8.27) which remains valid. The electron orbits may not be circular but remain closed, lying on the surfaces of constant energy in the \mathbf{k} -space. We write the period of the orbit as

$$T_c = \oint dt \quad (8.31)$$

The integral is evaluated over one complete orbit in the plane perpendicular to \mathbf{B} . From (8.27), we have

$$\begin{aligned}\delta t &= \frac{\hbar \delta k}{ev_1 B} \\ &= \frac{\hbar^2 \delta k}{eB} \cdot \frac{\Delta k_\perp}{\Delta \epsilon} \quad (\text{using 8.9})\end{aligned}\quad (8.32)$$

With $\oint k_\perp dk = S$ (the area of the orbit), we have

$$T_c = \frac{\hbar^2}{eB} \cdot \frac{\partial S}{\partial \epsilon} \quad (8.33)$$

$$\omega_c = \frac{2\pi eB}{\hbar^2} \cdot \frac{1}{\partial S / \partial \epsilon} \quad (8.34)$$

Comparing (8.33) with (8.30), we define a *cyclotron mass* as

$$m_c^* = \frac{\hbar^2}{2\pi} \frac{\partial S}{\partial \epsilon} \quad (8.35)$$

It is emphasized that this effective mass is not necessarily the same as defined in other contexts, e.g. specific heat effective mass.*

Now, we make use of the Bohr's correspondence principle according to which the difference in energy of the two adjacent levels is equal to the product of Planck's constant and the frequency of classical motion at the energy of the levels. As k_z is a constant of the semi-classical motion, we apply this condition to levels with a given k_z .

$$\epsilon_{n+1} - \epsilon_n = \frac{\hbar}{T_c} \quad (8.36)$$

Substituting the value of T_c from (8.33), we get

$$(\epsilon_{n+1} - \epsilon_n) \frac{\partial S}{\partial \epsilon} = \frac{2\pi eB}{\hbar} \quad (8.37)$$

Based on the free electron results, the energy difference between the neighbouring Landau levels is expected to be of order $\hbar\omega_c$. Since the energies of levels themselves are at least 10^4 times greater than this, it is an excellent approximation to replace the differential quotient $\partial S / \partial \epsilon$ with the difference quotient $(S_{n+1} - S_n) / (\epsilon_{n+1} - \epsilon_n)$. Doing so in (8.37), we get

$$S_{n+1} - S_n = \frac{2\pi eB}{\hbar} \quad (8.38)$$

*N.W. Ashcroft and N.D. Mermin, *Solid State Physics*, Ch. 2 (Saunders College, 1988).

which is the same result as for free electrons (8.29). By rewriting (8.38) as

$$S_{n+1} = S_n + \frac{2\pi eB}{\hbar}$$

we conclude that for large n the area of the semi-classical orbit may be expressed as

$$S_n = (n + \gamma) \frac{2\pi eB}{\hbar} \quad (8.39)$$

where γ denotes the phase correction that is independent of n . It is equal to 1/2 for free electrons. The orbit areas are quantized in units of $2\pi eB/\hbar$. Relation (8.39) is the famous Onsager's quantization condition.

The features of electron dynamics in electric and magnetic fields, as treated so far, have been exploited to plan experiments for evaluating the effective mass, the shape and the size of Fermi surfaces. The method of cyclotron resonance is generally used to determine the effective mass of charge carriers. But wide ranging experimental methods are used for the study of Fermi surfaces. Some of the principles that form the basis of experiments being conducted are: anomalous skin effect, cyclotron resonance, de Haas-van Alphen effect, Shubnikov-de Haas effect, magneto-resistance. On the basis of relative experimental and theoretical significance we have selected the methods of anomalous skin effect, cyclotron resonance and de Haas-van Alphen effect for our discussion.

8.5 ANOMALOUS SKIN EFFECT

In one of the earliest experiments on the determination of the Fermi surfaces, Pippard (1957) studied the reflection and absorption of microwaves in copper. It is observed that the electric field of microwave radiation makes only a limited penetration into the metal. Tendency of a metal to prevent the penetration of an alternating electric field into it is called the *skin effect*. If the microwave frequency is not too high, the depth to which the field penetrates into the metal can be calculated using the classical theory. The penetration depth is called the *classical skin depth* and is given by

$$\delta_0 = \frac{c}{\sqrt{2\pi\sigma\omega}} \quad (8.40)$$

where c is the speed and ω the angular frequency of the electromagnetic radiation; and σ denotes the electrical conductivity of the metal. δ_0 is also defined as the distance inside the metal over which the amplitude of the electric field of radiation is damped to 1/e of its initial value.

The derivation of (8.40) is valid when the skin depth δ_0 is far more greater than the electron mean free path Λ (i.e. $\delta_0 \gg \Lambda$). This condition is met in metals at room temperature and above. The resulting skin effect in these conditions is called the *normal skin effect*. But, at low temperatures, especially when the metal sample is very pure the electron mean free path becomes greater than the skin depth. The skin depth may be about 100 times less than the value predicated by (8.40). Under these conditions the classical theory breaks down and the decrease in the amplitude of radiation field will no longer be exponential with distance. This effect is known as the *anomalous skin effect*. In conditions of the anomalous effect only a small number of electrons that move almost tangentially to the metal surface, interact with the radiation field and complete one free path totally within the skin depth.

The most interesting property of anomalous skin effect is, however, its dependence on the geometry of the Fermi surface. Some features of the Fermi surface are known to depend only on the orientation of the Fermi surface relative to the actual surface of the metal sample. Further, the data on reflectivity and absorption of microwaves show correlation with these features making it possible to derive some information about the Fermi surface. This technique is, however, rarely used on account of the limited information provided by it.

8.6 CYCLOTRON RESONANCE

We saw above that the band electrons circulate with the cyclotron frequency ω_c under the action of a magnetic field. The electrons circulate around the field in a plane perpendicular to it. The electron energy in the plane is quantized in units of $\hbar\omega_c$. Therefore, a resonance in the absorption of energy takes place when the crystal is exposed to a suitably phased electromagnetic radiation of frequency ω_c . This is called *cyclotron resonance* on whose principle a cyclotron operates. The resonant magnetic field is searched by varying the field. The useful range of frequencies of the electro-magnetic radiation falls in the microwave region.

For a sharp resonance, the electron scattering has to be minimum. Our assumption that there are no centres of scattering in the crystal may not be justified in absolute sense. Even the most perfect and the purest available crystals are known to have some imperfections and impurities to scatter electrons. Hence, for the measurement of ω_c , experiments must be performed at low temperatures and high magnetic fields. This suppresses the scattering and thermal fluctuations. The sharpness of resonance, however, depends primarily on the broadening of energy levels. From the uncertainty principle, the broadening $\Delta\varepsilon$ is given by

$$\Delta\varepsilon \approx \frac{\hbar}{\tau} \quad (8.41)$$

Therefore, if $\frac{\hbar}{\tau} > \hbar\omega_c$, no resonance features will be observed as the electron would be scattered before it could complete one revolution. Thus the basic condition for the cyclotron resonance is

$$\omega_c\tau \gg 1 \quad (8.42)$$

Another important factor of concern is the skin depth available to the microwaves used as the electromagnetic radiation. The larger the conductivity of the crystal, the smaller the skin depth. The skin depth should be larger than both the electron mean free path and the size of the orbit in the resonant magnetic field. This ensures that the electron feels the electromagnetic field once per revolution. In view of this fact, the experimental requirements for semiconductors and metals are different needing separate discussions.

8.6.1 Semiconductors

The cyclotron resonance in semiconductors is conveniently observed because of their low electrical conductivity which facilitates the penetration of the radio frequency field deeper into the sample. The arrangement of the fields is shown in Fig. 8.13. Using (8.30) we make a rough estimate of ω_c . Taking m^* as 10^{-31} kg for a magnetic field of 1 tesla,

$$\omega_c \approx 10^{12} \text{ s}^{-1} \quad \text{or} \quad v_c \approx 2 \times 10^{11} \text{ Hz} = 200 \text{ GHz}$$

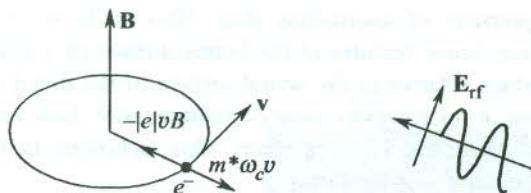


FIG. 8.13 Description of cyclotron resonance for electrons in a semiconductor. The radio frequency field E_{rf} lies in the plane of the orbit around the static magnetic field B . The fields B and E_{rf} are mutually perpendicular. Holes circulate in the clockwise direction in this set-up.

This is an inconveniently high frequency that corresponds to a wavelength of about 1 mm. Generally, the experiment is performed at liquid helium temperature with pure samples to ensure a large value of τ . Then, we can use microwaves of about 1 cm wavelength in a field of 0.1 tesla and satisfy the requirement $\omega_c \tau \gg 1$.

The theory of cyclotron resonance is fairly complicated for both semiconductors and metals. Only a simple account of the same is given here. For a parabolic energy surface, the orbit area is proportional to k^2 . Notwithstanding its dependence on the direction of \mathbf{k} , the energy ϵ_k is thus a linear function of S . The gradient $\partial\epsilon/\partial S$ in the expression for ω_c (8.34) then remains constant in the whole region of the band. As an example, we take the conduction band of Ge that is a spheroid. When the magnetic field lies in the xz -plane and is directed at angle θ with the z -direction, the form of the ellipse perpendicular to the field direction is given by

$$\epsilon = \frac{\hbar^2}{2} \left[\frac{k_t^2}{m_t^*} + k_l^2 \left(\frac{\cos^2 \theta}{m_t^*} + \frac{\sin^2 \theta}{m_l^*} \right) \right] \quad (8.43)$$

and

$$\frac{\epsilon}{S} = \frac{\hbar^2}{2\pi} \left[\frac{\cos^2 \theta}{m_t^{*2}} + \frac{\sin^2 \theta}{m_t^* m_l^*} \right]^{1/2} \quad (8.44)$$

The effective masses m_t^* and m_l^* are determined with the knowledge of ω_c and θ , available from the experiment. Here t stands for transverse and l for longitudinal.

In practice, the cyclotron resonance is essentially used to determine the effective mass of charge carriers in semiconductors. The resonance absorption spectra give distinct evidence for the presence of electrons and holes. The direction of the hole orbit is, however, opposite to that of the electron orbit. Hence the distinction between electrons and holes can be made by using a circularly polarized high frequency microwave field incident along the axis of \mathbf{B} field. Figure 8.14 gives the cyclotron resonance spectrum of a germanium crystal where the magnetic field is oriented in the (110) plane at an angle of 60° with the [100] direction. Each peak corresponds to a certain effective mass. The explanation to the observation of light and heavy holes in the spectrum may be sought in the band structure at the valence band maximum where the hole states exist. But, a simple treatment of the band structure, which is rather complicated in this region, fails to meet the objective. The exercise is deferred to the next chapter for a complete answer.

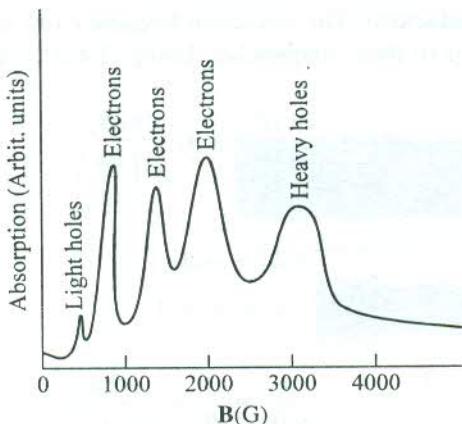


FIG. 8.14 Cyclotron resonance in Ge with magnetic field oriented in a (110) plane at an angle of 60° with the [100] direction. The two valence bands, degenerate at $\mathbf{k} = 0$ with different curvatures, give rise to two peaks marked as ‘light’ and ‘heavy’ holes. [After G. Dresselhaus, A.F. Kip, C. Kittel, *Phys. Rev.*, **98**, 368 (1955).]

8.6.2 Metals

The presence of a large number of mobile electrons endows metals with a high electrical conductivity. As a result the skin depth available to the incident microwaves is smaller than both the electron mean free path and the dimensions of the electron's real space orbit in the magnetic field corresponding to ω_c . Due to this reason in this case a different geometry is used, as proposed by Azbel and Kaner and shown in Fig. 8.15. The special feature of the geometry is that the steady magnetic field \mathbf{B} is parallel to the specimen's surface. It makes the electrons available in the skin depth region while performing their spiral orbits. These electrons absorb energy from microwaves if they see the electric field in right direction once during each revolution. On account of small skin depth, the resonance is observed only at harmonics of the cyclotron frequency:

$$\omega = n\omega_c$$

$$= n \frac{2\pi eB}{\hbar^2 \left(\frac{\partial S}{\partial \epsilon} \right)} \quad (\text{from 8.34}) \quad (8.45a)$$

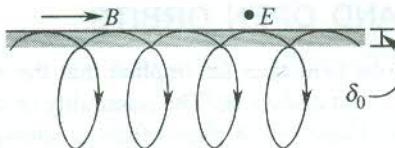


FIG. 8.15 Parallel-field Azbel-Kaner geometry for cyclotron resonance study in metals.

In a normal experimental set-up, a microwave source of fixed frequency is used and the magnetic field is varied to locate the resonance. Suppose two successive resonances are observed for the magnetic fields B_1 and B_2 . It implies that B_1 and B_2 must correspond to fields that could have shown resonance for two separate fundamental frequencies of microwaves in a material with

large skin depth (e.g. semiconductors). The resonance frequency (ω) in the present case represents two harmonics (say n_1 and n_2) of these frequencies. Using (8.45a), we can write

$$\omega = n_1 \frac{2\pi e B_1}{\hbar^2 \left(\frac{\partial S}{\partial \epsilon} \right)} = n_2 \frac{2\pi e B_2}{\hbar^2 \left(\frac{\partial S}{\partial \epsilon} \right)} \quad (8.45b)$$

or

$$n_1 B_1 = n_2 B_2$$

If

$$n_2 = n_1 + 1$$

$$\frac{1}{B_2} = \frac{1}{B_1} + \frac{1}{n_1 B_1} = \frac{1}{B_1} + \frac{2\pi e}{\hbar^2 \left(\frac{\partial S}{\partial \epsilon} \right) \omega}$$

or

$$\delta \left(\frac{1}{B} \right) = \frac{1}{B_2} - \frac{1}{B_1} = \frac{2\pi e}{\hbar^2 \left(\frac{\partial S}{\partial \epsilon} \right) \omega}$$

or

$$= \frac{e}{m_c^* \omega} \quad (\text{from 8.35}) \quad (8.46)$$

Thus we see that the resonant absorption is periodic in $(1/B)$ with a period of $\left(\frac{1}{B_2} - \frac{1}{B_1} \right)$. With

the knowledge of B_1 and B_2 in the relation (8.46), the estimation of the cyclotron effective mass (m_c^*) becomes straightforward.

The study of surface impedance Z in this set-up provides another example of the above periodicity. Considering the microwave source as the voltage source, the surface impedance can be measured at different values of the magnetic field. It is observed that the field derivative of the impedance $\partial Z / \partial B$ oscillates with the period $\delta(1/B)$. Figure 8.16 shows this periodic effect in copper.

8.7 CLOSED ORBITS AND OPEN ORBITS

Although not stated, the discussions held thus far implied that the energy surface of electrons, or holes or of both are closed. But it is not always so. The possibility of open orbits must be appreciated before the interpretation of the de Hass–van Alphen effect is attempted.

Consider a Fermi surface shown in Fig 8.17(a). The Fermi surface touches the zone boundary. An electron on this plane section starts moving from point D' towards point A. Its arrival at point A is the same as reaching point A' which is equivalent to point A in the k -space on account of being translationally connected by a reciprocal lattice vector. Since the same is true for B and B', C and C', and D and D', the electron goes round a closed path—from A' to B (or B'), from B' to C (or C') and finally from C' to D (or D').

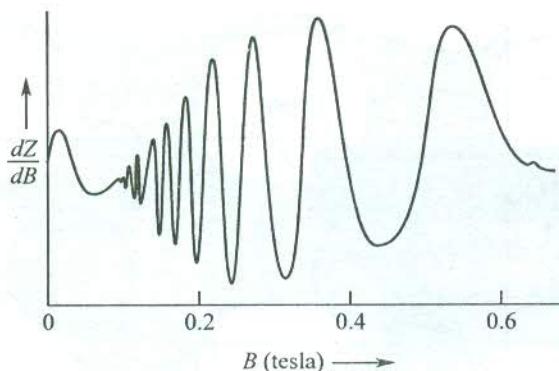


FIG. 8.16 Oscillation of the field derivative of the surface impedance dZ/dB in copper as a function of the magnetic field at 24 GHz. [After A.F. Kip, D.N. Langenberg, T.W. Moore, *Phys. Rev.*, **124**, 359 1961.]

The orbit in the reduced zone scheme [Fig. 8.17(a)] shows discontinuities at the zone boundaries, which is only illusory. The electron has the same wavefunction at points A and A', since they are equivalent in the \mathbf{k} -space. The orbit is actually continuous or closed as is clear from the picture in the periodic zone scheme [Fig. 8.17(b)]. Without having to shift the representative point through a reciprocal lattice vector each time the electron crosses a zone boundary, the continuity in path is maintained by providing another cell of the reciprocal lattice on the other side. Since the orbit encloses the empty region of the \mathbf{k} -space, it is befitting to name it as a *hole orbit*. Its characteristic cyclotron frequency is defined accordingly. A three-dimensional view of a multiply-connected Fermi surface as depicted in Fig. 8.18 shows that on sectional planes both electron orbits and hole orbits can be found.

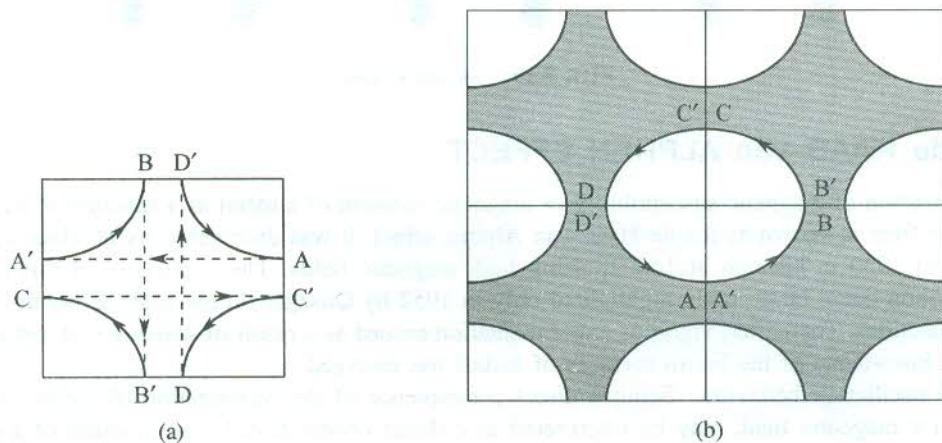


FIG. 8.17 A closed orbit in the (a) reduced zone scheme and (b) periodic zone scheme.

The possibility of an open orbit is exhibited in Fig. 8.19. It is clear that orbits in a plane, whose section with the Fermi surface is not a closed curve, are open. The representative point would not be brought back to the starting point by a magnetic field. In this case the electron continues to move with wavevector increasing towards infinitely large values in the periodic zone scheme.

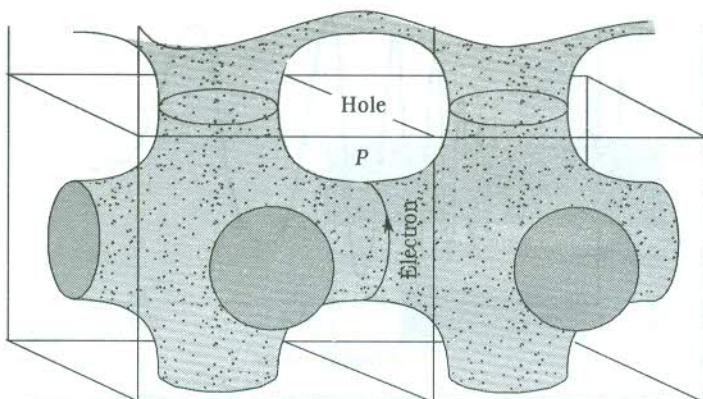


FIG. 8.18 Electron and hole orbits on a multiply-connected Fermi surface.

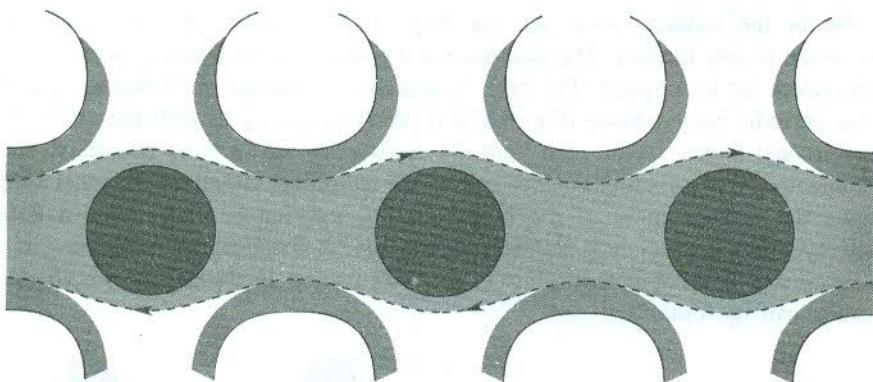


FIG. 8.19 An open orbit.

8.8 de HAAS–van ALPHEN EFFECT

The oscillation of magnetic susceptibility or magnetic moment of a metal as a function of the static magnetic field is known as the de Haas–van Alphen effect. It was discovered by de Haas and van Alphen in 1930 in bismuth at 14.2 K using high magnetic fields. The usefulness of this curious phenomenon could be properly highlighted only in 1952 by Onsager on the basis of his theory for Bloch functions. Thereafter, vigorous experimentation ensued as a result of which a vast and precise body of knowledge of the Fermi surfaces of metals has emerged.

The oscillatory behaviour, being a direct consequence of the quantization of closed electron orbits in a magnetic field, may be interpreted as a direct observational manifestation of a purely quantum phenomenon.

In order to get well-defined oscillations, the experiments are performed with pure crystals at low temperatures where the Fermi surface is sharp and $k_B T \ll \hbar\omega_c$, keeping the electron scattering at minimum. We make use of the Onsager's theory, developed in Section 8.4.2, to explain the dHvA effect. Our arguments, that are strictly valid at 0 K, are tenable as the experiments are conducted actually at very low temperatures.

Let us rewrite the quantization condition (8.38) as

$$\Delta S = \frac{2\pi eB}{\hbar} \quad (8.47)$$

It tells us that as the magnetic field is increased the Landau tubes (see Fig. 8.12) expand. When a Landau tube touches the Fermi surface, the average energy of electrons is maximum because the electrons are drawn up to the Fermi level, the highest occupied level at 0 K. On increasing the field further, the Landau tube moves beyond the Fermi surface and gets depopulated resulting in the decrease of the average energy. The average energy is minimum when the contact contour of the Fermi surface is exactly halfway in the space between two successive Landau tubes, since the highest occupied states fall below the Fermi level. The total average energy of mobile electrons (the electron gas) in this situation is lower than that in the absence of a magnetic field by an amount of the order

of $\frac{1}{2}\hbar\omega_c$ per electron at the Fermi level. Thus a slow variation of the magnetic field causes oscillations in the average energy. A quantitative treatment of this oscillatory behaviour has been provided by Ziman.[†] It involves unusually large and complicated integrals whose evaluation requires the use of certain tricky methods of pure mathematics. The limited scope of the book, obviously, does not allow us to go into the details of this treatment.

It is rather easy to specify the period of oscillation if we appreciate that orbits belonging to different quantum numbers can be manipulated to have a common value of the area by adjusting the magnetic field. The areas of the two successive orbits are:

$$S_n = (n + \gamma) \frac{2\pi eB_1}{\hbar}$$

$$S_{n+1} = (n + 1 + \gamma) \frac{2\pi eB_2}{\hbar} \quad (8.48)$$

If B_1 and B_2 are such that

$$S_n = S_{n+1} = S \quad (8.49)$$

then

$$\left(\frac{1}{B_2} - \frac{1}{B_1} \right) = \frac{2\pi e}{\hbar S}$$

or

$$\delta \left(\frac{1}{B} \right) = \frac{2\pi e}{\hbar S} \quad (8.50)$$

Since $n \propto \frac{1}{B}$ [from (8.48)], $B_2 < B_1$. Thus by decreasing B slowly, we can proceed to orbits of higher quantum numbers. The orbit area changes with variations in B in such a way that for every change of $2\pi e/\hbar S$ in the value of $1/B$, the orbit area is repeated. This is the same as the oscillation

[†] J.M. Ziman, *Principles of the Theory of Solids*, 2nd ed., Ch. 9 (Cambridge, 1972).

of the orbit area with a period $2\pi e/\hbar S$. These periodic quantum oscillations in principle are expected to be observed in all solid state properties. Similar oscillations of electrical and thermal conductivities in a strong magnetic field are some examples to quote. The phenomenon involving the electrical conductivity is known as de Haas–Shubnikov effect.

The oscillation of magnetic moment in dHvA effect follows from the oscillatory behaviour of the free energy as discussed above. Being a magnetic field derivative of the free energy, the magnetic moment too oscillates as a function of the static magnetic field. The orbit area S in (8.50) is essentially a measure of the cross-sectional area of the Fermi surface. Periods of the sections of a Fermi surface of general shape at different values of the wavevector along \mathbf{B} are different. The contributions from all sections add up and account for the net response. But, the dominant contribution to $\delta(1/B)$ in (8.50) comes from orbits whose time periods remain unaffected by small changes in the magnitude of \mathbf{k} along \mathbf{B} . These orbits are known as *extremal orbits*. The extremal orbits on a general Fermi surface for two orientations of \mathbf{B} are shown in Fig. 8.20. If the magnetic field be applied along the z -direction, the extremal areas in (8.50) would be represented by those cross-sectional areas at whose contours on the Fermi surface the condition,

$$\frac{dS}{dk_z} = 0 \quad (8.51)$$

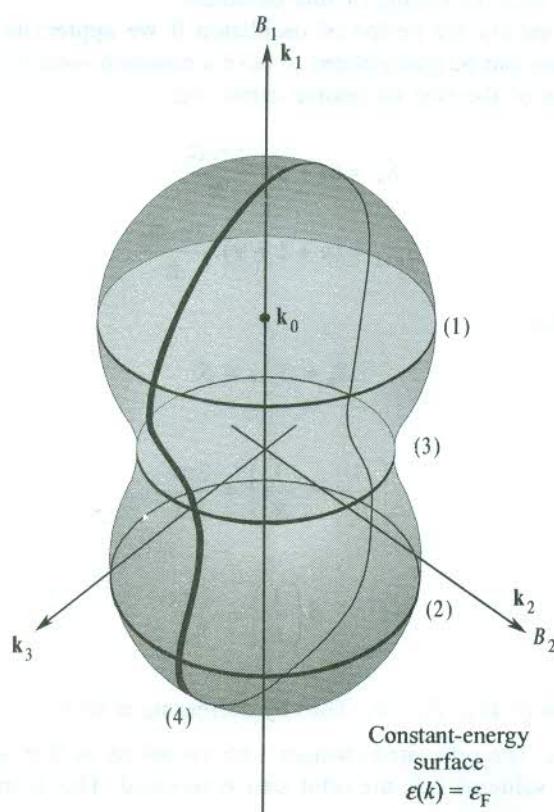


FIG. 8.20 Illustration of extremal orbits. When B is along the k_1 -axis, there are two maximum extremal orbits (1) and (2), and one minimum extremal orbit (3). For B along the k_2 -axis, only one extremal orbit (4) is observed.

is satisfied. The condition may be met for several values of k_z (at different heights of the Fermi surface along B). It should be abundantly clear (see Fig. 8.20) that maximum and minimum cross-sections both refer to the extremal orbits.

The oscillations occur even when only the magnetic field orientation is varied and the strength kept constant. It is so because for some directions the Landau tubes touch the Fermi surface along the contact contour while for other directions they do not. The oscillations disappear for field directions, giving open orbits. The complete theoretical structure of dHvA effect is based on the existence of closed orbits with definite periods. Thus by changing the magnetic field direction all the extremal cross-sections of the Fermi surface can be mapped out, making the reconstruction of the Fermi surface feasible.

Experimental set-up

In a typical set-up, a few mm long single crystal wire is mounted in a small pick-up coil. There are two basic methods of measurement. In one method the magnetic field is produced by discharging a large capacitor through a solenoid cooled to liquid nitrogen temperature. Within a few milliseconds the field increases to about 10 T and drops to zero. This method is known as the impulsive method. With changing magnetic field, an emf is induced which is proportional to

$$\frac{dM}{dt} = \frac{dM}{dB} \cdot \frac{dB}{dt} \quad (8.52)$$

where M is the magnetic moment.

Since metals are only weakly diamagnetic, the difference between the applied field and the average field experienced by electrons in a metal may be ignored. This point can be properly appreciated after having lessons on magnetism (Chapters 13 and 14).

In accordance with (8.52) the induced emf produces an oscillatory signal which is proportional to the differential susceptibility dM/dB .

The second basic method employs superconducting coils[†] that provide highly stable steady fields of the order of 5 T. Once the solenoids get energized, a persistent current is set up with no external source of emf (because of zero resistance). The steady field is modulated at a convenient frequency using auxiliary coils. With the modulation being done at a fixed frequency, we have the advantage of using phase sensitive detection techniques for improving the signal to noise ratio.

Example

The original spectrum^{††} showing the oscillation of magnetic susceptibility of gold is reproduced in Fig. 8.21. The picture of the Fermi surface, as constructed by making use of the information derived from this measurement, is shown in Fig. 8.22. The d-states in gold are occupied and the Fermi level is placed in the region of the sp-band. So, the Fermi surface appears approximately spherical. The picture shows that the surface in one Brillouin zone is connected to that in the next by a neck along the <111> direction showing distortions along this direction. Therefore, with a magnetic field along

[†] The superconducting property is observed at low temperatures. This implies that coils used here are dipped in a low temperature bath. This property will be discussed in Chapter 15. The d.c. resistance of a superconductor is zero.

^{††} B. Lengeler, *Springer Tracts Mod. Phys.*, 82, 1, p. 1 (Springer, 1978).

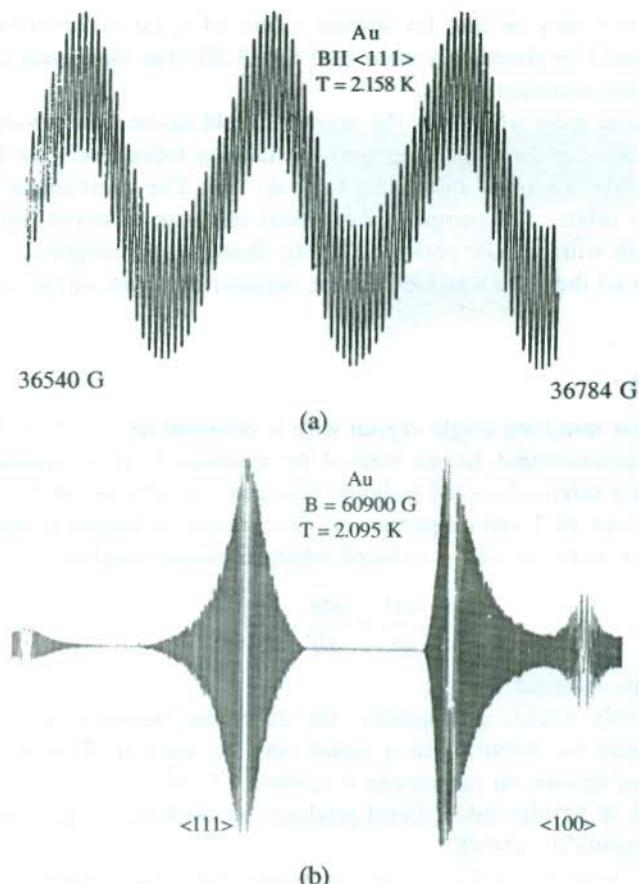


FIG. 8.21 The oscillation of the magnetic susceptibility of gold: (a) as a function of the magnetic field, (b) with variation in the direction of the magnetic field. These quantum oscillations can be observed only for very sharp Fermi distribution ($k_B T < \hbar\omega_c$).

the $\langle 111 \rangle$ direction there are two types of closed orbits along which the Landau tubes peel off from the Fermi surface on increasing the magnetic field slowly. These are identified as neck and belly orbits. By comparing the periods (1:29) in the oscillogram [Fig. 8.21(a)], we directly obtain the ratio of the cross-sectional areas of these orbits. The absolute magnitudes are calculated from the data by making use of (8.50). These calculations yield the values of areas as $1.5 \times 10^{19} \text{ m}^{-2}$ and $4.3 \times 10^{20} \text{ m}^{-2}$ for the neck and belly orbits, respectively.

It is useful to make an estimate of the Fermi surface of gold by approximating its surface to the free electron spherical Fermi surface. The value of the Fermi wavevector k_F is taken from tables. Using $k_F = 1.2 \times 10^{10} \text{ m}^{-1}$, we get an extremal area of $4.5 \times 10^{20} \text{ m}^{-2}$. Interestingly, this is in excellent agreement with the experimental value for the extremal belly orbit that represents the average shape and size of the Fermi surface of gold.

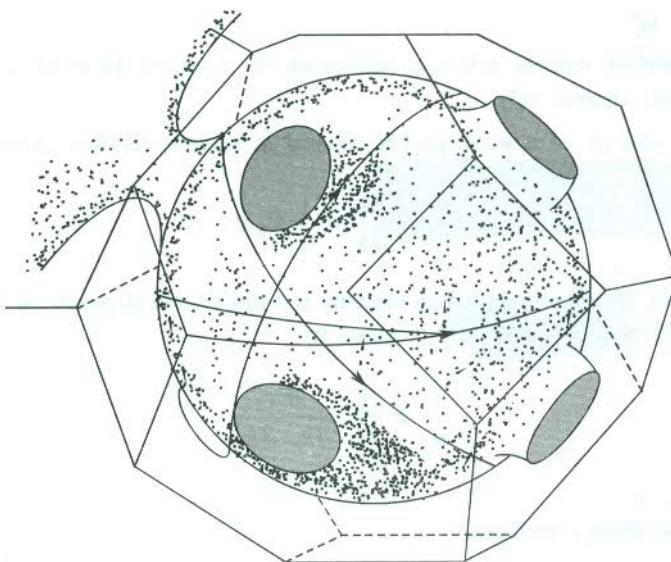


FIG. 8.22 Model of the Fermi surface of noble metals Cu, Ag and Au. The surface makes contact with the Brillouin zone boundary along the [111] directions. There are two extremal orbits perpendicular to the [111] directions.

SUMMARY

1. An empty electron state in an otherwise filled band of a crystal is identified with a hole. If there is only one hole in a band, the properties of the hole are those of $(N - 1)$ electrons, where N is the number of electrons in the completely filled band.
2. Properties of holes:
 - (i) Current density contributed by holes is

$$j = -e \int_{\text{occupied}} \frac{v(\mathbf{k}) d^3 \mathbf{k}}{4\pi^3} = e \int_{\text{unoccupied}} \frac{v(\mathbf{k}) d^3 \mathbf{k}}{4\pi^3}$$

- (ii) If an electron with wavevector \mathbf{k}_e is missing from a band, the hole thus created will have a wavevector \mathbf{k}_h such that

$$\mathbf{k}_h = -\mathbf{k}_e$$

- (iii) Taking the energy of a filled band as zero of the energy scale, the energy of the hole is positive:

$$\epsilon_h(\mathbf{k}_h) = -\epsilon_e(\mathbf{k}_e)$$

- (iv) The velocity of the hole created in the valence band owing to the excitation of an electron from the valence band to the conduction band is equal in magnitude and direction to the velocity of the electron now missing from the valence band.

$$v_h(\mathbf{k}_h) = v_e(\mathbf{k}_e)$$

(v) $m_h^* = -m_e^*$

where m_h^* and m_e^* are the effective masses of the hole and the electron (missing from the valence band), respectively.

3. The group velocity of a wave packet centred at wavevector \mathbf{k} is given by

$$\mathbf{v}_g = \frac{\partial \omega}{\partial \mathbf{k}} = \frac{1}{\hbar} \nabla_{\mathbf{k}} \epsilon$$

4. Generally, the effective mass of an electron is a tensor that depends on the curvature of the ϵ versus k dispersion curve and is given by

$$\frac{1}{m_{ij}^*} = \frac{1}{\hbar^2} \frac{\partial^2 \epsilon}{\partial k_i \partial k_j}$$

with $i, j = x, y, z$.

For isotropic energy surfaces,

$$\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{\partial^2 \epsilon}{\partial k^2}$$

5. The constant energy surface in the \mathbf{k} -space that separates the filled states from the empty states at absolute zero is called the Fermi surface. The total volume enclosed by the Fermi surface depends only on the electron concentration and is independent of the details of the lattice interaction.
6. The total electron energy in the presence of a magnetic field applied along the z -direction is written as

$$\epsilon_n(k_z) = \frac{\hbar^2 k_z^2}{2m} + \left(n + \frac{1}{2}\right) \hbar \omega_c$$

where ω_c is the cyclotron frequency (eB/m^*).

Each group of degenerate energy levels that corresponds to a magnetic level as defined by a given pair of quantum numbers n and k_z is called a *Landau level*.

7. The basic condition for the cyclotron resonance is

$$\omega_c \tau \gg 1.$$

8. The de Haas–van Alphen effect shows periodicity of $\delta\left(\frac{1}{B}\right)$ given by

$$\delta\left(\frac{1}{B}\right) = \frac{2\pi e}{\hbar S}$$

where S is the extremal cross-sectional area of the Fermi surface in the \mathbf{k} -space.

The study of the dHvA effect thus provides S that is considered as perpendicular to \mathbf{B} .

PROBLEMS

- 8.1** (a) Show that the group velocity is twice the phase velocity for any state of an ideal free electron gas.
 (b) Assuming that electron velocities in a system are represented by spherical distribution and ϵ - k relationship as, $\epsilon = ak^2 - bk^4$, find the group and phase velocities in the system. What is the value of energy in terms of a and b at which the two types of velocities are equal?
- 8.2** Construct the first two Brillouin zones of a primitive rectangular two-dimensional lattice with axes a , b ($= 2a$).
 If the lattice belongs to a metal in which there is one atom of valency one in a primitive cell, calculate the radius of the free electron Fermi sphere in cm^{-1} . Take $a = 2 \text{ \AA}$, $b = 4 \text{ \AA}$. Draw the Fermi sphere to scale on a drawing of the first Brillouin zone.
- 8.3** In potassium at 68 GHz three consecutive cyclotron resonances are observed at magnetic fields of 0.74 T, 0.59 T and 0.49 T. What is the cyclotron resonance mass of electrons in potassium?
- 8.4** In a cyclotron resonance set-up, a klystron radiation of $2.4 \times 10^{10} \text{ Hz}$ is used. For a sample, the resonance absorption occurs at a magnetic field of $8.6 \times 10^{-2} \text{ T}$. Calculate the effective mass of the respective charge carriers. Determine the range of relaxation time over which a resonance is observable.
- 8.5** (a) Apply the Onsager's quantization condition (8.39) to the orbits of free electron levels and show that it leads directly to the free electron levels (8.26) if $\gamma = 1/2$.
 (b) Derive an expression for the degeneracy of Landau levels in the case of a two-dimensional free electron gas. Show that the degeneracy is equal to the number of zero-field free electron levels with a given k_z , and with k_x and k_y within a planar region of area ΔS given by (8.29).
- 8.6** An electron in sodium at the Fermi level moves initially in the xy -plane. Calculate the radius of its cyclotron orbit when a magnetic field of 1 T is applied. What is the ratio of the orbit area in real space to that in \mathbf{k} -space?
- 8.7** The expression for the Fermi energy in a non-isotropic solid is

$$\epsilon = (\alpha k_x^2 + \beta k_y^2)$$

Describe the rate at which \mathbf{k} changes during an extremal cyclotron orbit at the magnetic field \mathbf{B}_z . Assume that α and β are unequal. How is the orbit in real space?

- 8.8** Make an estimate of the expected value of period $\delta(1/B)$ in a dHvA effect study of potassium on the free electron model. What is the order of the maximum temperature below which the experiment should be done to obtain good results in a magnetic field of strength 1 T?

SUGGESTED FURTHER READING

- Falikov, L.M., *Fermi Surface Studies: Electrons in Crystalline Solids* (IAEA, Vienna, 1973).
 Harrison, W.A., *Electronic Structure and the Properties of Solids* (Freeman, 1980).
 Shoenberg, D., *Magnetic Oscillations in Metals* (Cambridge, 1984).
 Ziman, J.M., *Principles of the Theory of Solids*, 2nd ed. (Cambridge, 1972).

Semiconductors

As mentioned in the beginning of Chapter 7, the values of electrical conductivity of solids are spread over almost the widest range for any common physical property. Hence these form a basis for classifying solids. Solids characterized by extremely high and extremely low values of electrical conductivity are identified as metals and insulators, respectively. A pure metal at 1 K may have a conductivity of the order of $10^8 \text{ ohm}^{-1} \text{ m}^{-1}$ against a low of $10^{-20} \text{ ohm}^{-1} \text{ m}^{-1}$ for an extreme insulator. Materials with conductivity values intermediate to these extreme orders of magnitude are called *semiconductors*. Typical conductivity values for semiconductors lie in the range from 10^{-7} to $1 \text{ ohm}^{-1} \text{ m}^{-1}$. The most useful feature of semiconductors is that their electrical conductivity generally decreases with increasing purification in contrast to metals where the conductivity always increases with increasing purification.

Notwithstanding the enormous technological importance of semiconductors, their study is even more crucial to the understanding of electronic properties of solids. We will see that it is possible to apply the Maxwell–Boltzmann statistics to deal with charge carriers in semiconductors. This gives exact analytical solutions of many problems that can be solved only by approximate or numerical methods in metals where the Fermi–Dirac statistics has to be used. Furthermore, with remarkable progress having been made in the technology of growing semiconductor crystals, the degree of purity and perfection achieved in growing single crystals of semiconductors is much higher than that in metals and insulators. It is a matter of absolute importance to the study of electronic properties, some aspects of which are obscured by effects arising because of the presence of impurities and crystalline imperfections. It simply amounts to say that the quantitative studies of some phenomena that are generally difficult or cannot be made accurately in metals, are easily carried out in semiconductors with the required precision. Thus the study of semiconductors is helpful in interpreting the electronic properties of solids in general. In this chapter we concentrate on shaping the basic theoretical ideas in the framework of the band theory of solids. Nevertheless, a brief account of the structure and the principle of working of some simple and popular devices is included in Section 9.10.

9.1 CLASSIFICATION OF SEMICONDUCTORS

The classification of solids made on the basis of their band structures gives a lead to exploiting the band theory for the interpretation of many solid state properties. We learnt in Chapter 7 that completely filled and completely empty bands do not contribute to the flow of current. The highest filled band (the valence band) in metals is only partially filled and the flow of current occurs on account of almost continuous excitation of valence band electrons to empty states of the band under the influence of an electric field. Therefore, a material that has only completely filled and completely empty bands behaves as a perfect insulator at absolute zero when there can hardly be found any electrons in the lowest empty band (the conduction band) as a result of the thermal excitation of

electrons in the valence band. But if the difference between the upper edge of the valence band and the lower edge of the conduction band is not large and less than 2 eV, there is a finite probability for a small fraction of electrons occupying the uppermost states of the valence band to be thermally excited to the conduction band at moderate and high temperatures. At these temperatures the width of the energy range over which the Fermi distribution function rapidly changes is relatively substantial. This enables the consequences of the change in the distribution function easily observable. A representative band scheme for metals, semiconductors and insulators is drawn in Fig. 9.1. The figure portrays a qualitative difference in the band structures of these solids.

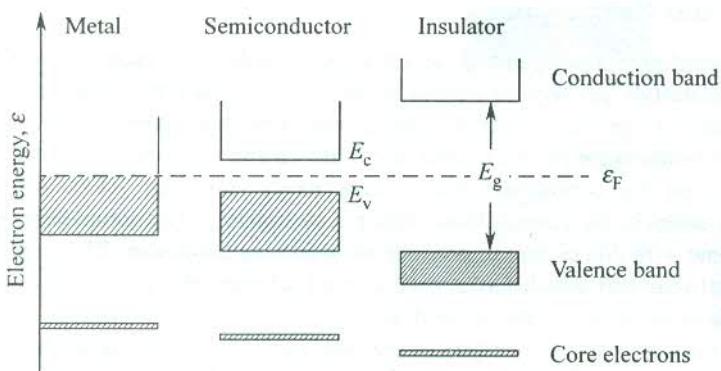


FIG. 9.1 Comparative energy band schemes for metals, semiconductors and insulators at $T = 0$ K. The valence band is completely filled in semiconductors and insulators but only partially filled (shaded) in metals. A relatively much smaller band gap E_g in semiconductors distinguishes them from insulators.

There is a long list of semiconductor materials. Only a few of them are elements. Si, Ge, grey Sn and grey Se are some examples. The first three are in Group IV and Se is in Group VI of the periodic table. Rest of them are mostly binary compounds, mainly of two types. In one type (e.g. GaAs, InSb, GaP) one element from Group III (e.g. B, Al, Ga, In) is combined with an element from Group V (e.g. N, P, As, Sb). The other type of binary compounds are formed with one element from Group II (e.g. Zn, Cd, Pb) and the other from Group VI (e.g. S, Se, Te). Some important examples of this type are ZnS, CdSe, PbTe. SiC and SiGe are the only known binary semiconductors whose both components are from Group IV. Some oxides also show semiconducting behaviour— TiO_2 , Cu_2O and ZnO are prominent examples of this class. A list of technologically important semiconductor materials is given in Table 9.1 which also contains the relevant data on these materials.

Semiconductors are primarily of two types—*intrinsic* and *extrinsic*. The intrinsic semiconductors are usually pure monatomic or diatomic solids. An intrinsic material is converted into the extrinsic type (or the impurity type) by adding traces (~ 1 part in one million) of a suitable impurity with the aim to enhance the level of the charge carrier density. Values of electrical conductivity of intrinsic semiconductors lie far below the range that is useful for the purpose of applications. But semiconductors are gifted with the unique quality that their electrical conductivity can be increased by several orders of magnitude by mixing with them suitable impurities in small concentration. This has enormously increased their technological importance. Thus it is mostly the extrinsic semiconductors that form the basis for semiconductor devices. The details of formation and model of extrinsic semiconductors are given in Section 9.4.

9.2 EXAMPLES OF BAND STRUCTURE

We know from Chapter 7 that the band structure of a solid is closely related with its crystal structure. Not many crystalline structures are favourable to the semiconducting behaviour. Most of the thoroughly investigated semiconductors have a diamond type lattice. We take examples of Si, Ge and GaAs which are distinguished for their large number of applications. The Si and Ge crystals have diamond structure. The crystal lattice of GaAs is zinc blende type which is only a modified diamond lattice, as discussed in Chapter 1. We describe below the band structures of these crystals.

9.2.1 Silicon and Germanium

The outer electron configurations of Si and Ge are $3s^23p^2$ and $4s^24p^2$, respectively. The origin of band structure of these materials has been discussed in Section 7.7. The tetrahedral bonding orbitals (sp^3) are formed because of the mixing of s- and p-wavefunctions. Near the bonding distance, at equilibrium, these orbitals were shown to split into bonding and antibonding orbitals which constitute the valence band and the conduction band, respectively (see Fig. 7.11). All the four s- and p-electrons occupy states in the valence band, filling it completely. The completely empty conduction band should combine with this picture to produce an insulating behaviour. This is really the case with the diamond crystal (carbon) which has a similar band scheme. But on account of small band gap (E_g), Si and Ge show semiconducting properties.

Figure 7.11 reveals an important feature of the band gap regarding its dependence on temperature. The observation that the size of the energy gap (or the splitting) between the valence and conduction bands decreases with increase in the interatomic separation, indicates that the gap is smaller at higher temperatures where the interatomic separation becomes larger because of thermal expansion. This fact is confirmed by the measured values of the band gap E_g at different temperatures (see Table 9.1).

Band structures are calculated by fitting the measured physical quantities, such as the band gap, the positions of points of high symmetry in the Brillouin zone (critical points), and the curvature of energy surfaces (the effective mass). The calculated band structures of Si and Ge are shown in Fig. 9.2. The features of the two band schemes appear quite different in contrast to the qualitative similarity as expected on the basis of Fig. 7.11. This is obviously the effect of the difference in electron wavefunctions associated with $3s^23p^2$ and $4s^24p^2$ configurations.

The symbols, Γ , X and L stand for positions of certain points of high symmetry in the Brillouin zone. They refer to the points at the zone centre (000), $\frac{2\pi}{a}(100)$ and $\frac{2\pi}{a}\left(\frac{1}{2}\frac{1}{2}\frac{1}{2}\right)$, respectively, where ' a ' is the lattice constant. The valence band maximum occurs at the zone centre, i.e. $k = 0$ for both Si and Ge. But the conduction band minimum occurs for k along the [100] direction in Si and for k along the [111] direction in Ge. This amounts to saying that electrons of the lowest energy in the conduction band have their wavevectors oriented along the [100] direction in Si and along the [111] direction in Ge. In both materials the valence band maximum and the conduction band minimum thus occur at different values of k . Semiconductors having this type of band structure are called the *indirect gap* semiconductors and those for which the maximum and minimum in question fall at the same value of k are referred to as the *direct gap* semiconductors. The discussion on consequences of the indirect gap nature of band structure will be taken up in Section 9.2.3.

The first Brillouin zone of Si and Ge crystals is a truncated octahedron appropriate to the FCC symmetry of their unit cell. In parabolic approximation (i.e. retaining terms up to the order k^2 in the expression for ϵ_k), the surfaces of constant energy are ellipsoids as confirmed by cyclotron resonance studies (see Section 8.6.1).

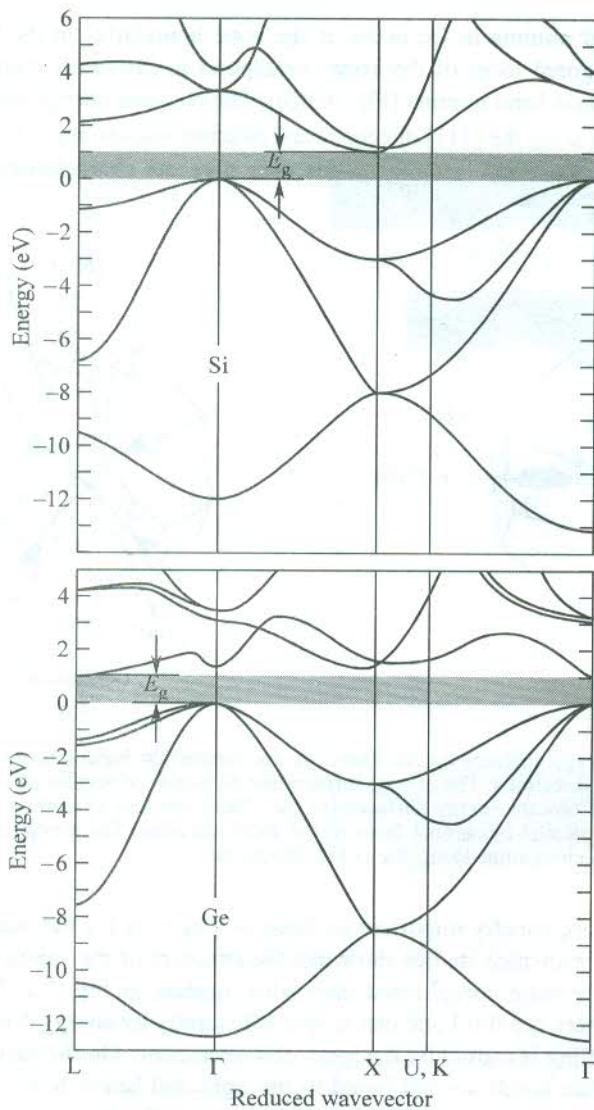


FIG. 9.2 Calculated band structures of Si and Ge [After J.R. Chelikowsky, M.L. Cohen, *Phys. Rev.* **B14**, 556 (1976)]. The structures confirm the indirect gap nature for both Si and Ge. For Ge, the spin-orbit splitting is also considered.

In Si there are six symmetry-related minima of the conduction band at points in the [100] directions. Each of the six ellipsoids is an ellipsoid of revolution about a cubic axis by symmetry. They appear as cigars elongated along the cube axes as shown in Fig. 9.3(a). The electron has two effective masses—the longitudinal m_l^* (along the axis) and the transverse m_t^* (perpendicular to the axis). Their values in terms of the free electron mass m are given as— $m_l^* = 0.98m$ and $m_t^* = 0.19m$. The valence band shows two degenerate maxima both located at $k = 0$ with spherical symmetry within validity of the ellipsoidal expansion (Fig. 9.2). The two effective masses are $0.49m$ and $0.16m$.

The conduction band minima in Ge occur at the zone boundaries in the [111] directions. The minima on parallel hexagonal faces of the zone correspond to the same energy levels giving four symmetry-related conduction band minima [Fig. 9.3(b)]. The constant energy surfaces are ellipsoids of revolution with elongation along the [111] directions and effective masses, $m_l^* = 1.57m$ and $m_t^* = 0.082m$. The two degenerate valence band maxima in this case give the characteristic effective masses of $0.28m$ and $0.44m$.

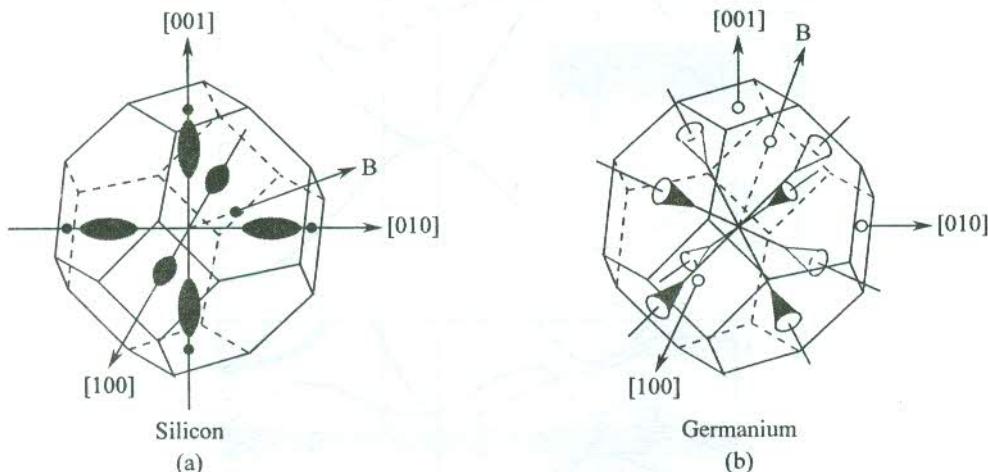


FIG. 9.3 (a) Constant energy surfaces for Si: There are six conduction band minima of cigar shape at points along the [100] directions. The energy surfaces are ellipsoids of revolution with elongation along the cube axes. (b) Constant energy surfaces for Ge: There are four symmetry related conduction band minima on the parallel hexagonal faces of the Brillouin zone. The energy surfaces are ellipsoids of revolution with elongation along the [111] directions.

Properties of holes are equally important as those of electrons for the study of semiconductors. Details of the cyclotron resonance studies show that the structure of the valence band maximum near $\Gamma(k = 0)$ in Si and Ge is more complicated than what appears in Fig. 9.2. In addition to the two degenerate bands at Γ , there is a third one that is split off slightly by energy Δ towards a lower energy band (Fig. 9.4). The splitting is caused by the spin-orbit interaction. On the basis of the effective mass values, the two degenerate bands are attributed to the light and heavy holes (Fig. 8.14). The holes associated with the identity of the third band are named as ‘split-off holes’. The split-off energy Δ has been estimated at 0.044 eV in Si and 0.29 eV in Ge.

9.2.2 Gallium Arsenide

GaAs crystal enjoys a special status, firstly on account of its direct-gap and secondly because of its band-gap energy being closely below the energy range of visible radiation. These properties render it most suitable for the fabrication of efficient optical devices.

The crystal has zinc blende (ZnS) structure. It is an example of the mixed ionic and covalent bonding. The chemical bonding is interpreted as the superposition of these two extreme cases of bonding. In ionic bonding, an electron is transferred from Ga to As to give the ionic structure $\text{Ga}^+ \text{As}^-$. On the other hand in the second extreme case, with the displacement of an electron from As to Ga, the number of electrons in the outer shell of both Ga and As atoms becomes four which results in the

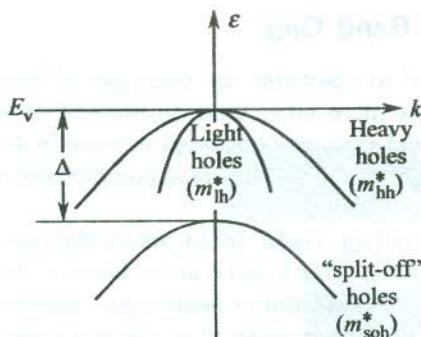


FIG. 9.4 Structure of the valence band (qualitative) in Si or Ge near the top showing the hole bands with the spin-orbit interaction taken into account. The split-off hole band is separated from the top at $k = 0$ by Δ , contributed by the spin-orbit interaction.

sp^3 hybridization as in the case of Si and Ge. The observed tetrahedrally coordinated ZnS structure of GaAs serves as a certain proof to the effect that the effects of covalent bonding dominate.

The band structure of GaAs is shown in Fig. 9.5. All the valence band maxima and conduction band minima occur at $\Gamma(k = 0)$ showing its direct gap nature with a gap of 1.43 eV at 300 K. The constant energy surfaces are accordingly spherical. The conduction band effective mass m_e^* is $0.07m$. There are three distinct valence bands similar in form to those for Si and Ge at Γ . The three respective effective masses are given as $m_{lh}^* = 0.12m$, $m_{hh}^* = 0.68m$ and $m_{soh}^* = 0.2m$ with $\Delta = 0.34$ eV.

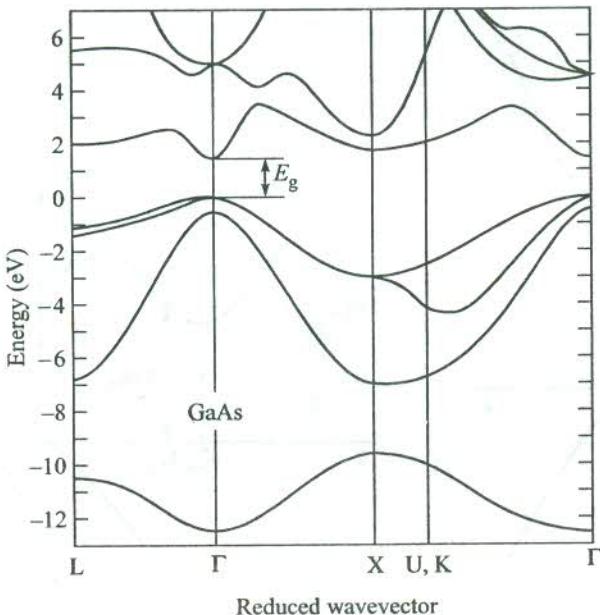


FIG. 9.5 Calculated band structure of GaAs, a representative of III-V semiconductors. The conduction band minima and the valence band maxima occur at $k = 0$, showing the direct gap behaviour. [After J.R. Chelikowsky, M.L. Cohen, *Phys. Rev.*, **B14**, 556 (1976).]

9.2.3 Determination of Band Gap

A number of methods are used to determine the band gap of semiconductors. The technique of continuous optical absorption is more often used on account of its accuracy and the important information gained about the band structure. An abrupt increase in the absorption of optical radiation takes place as soon as the energy of the incident radiation $\hbar\omega$ becomes large enough to exceed the energy gap.

In direct-gap semiconductors (e.g. GaAs, InSb), where the conduction band minimum and the valence band maximum occur at the same \mathbf{k} -value in the \mathbf{k} -space, the optical threshold at $\omega = E_g/\hbar$ directly gives the band gap [Fig. 9.6(a)]. But in indirect-gap materials (e.g. Si, Ge, GaP), the direct photon absorption accompanied by the transfer of an electron from the top of the valence band to the bottom of the conduction band would not conserve the crystal momentum because the initial and final points of the transition in the \mathbf{k} -space do not have the same k -value. Hence such a direct transition is not allowed. The transition process will have to be indirect in which the absorption of an optical photon must be accompanied by some other process with whose involvement the condition of momentum conservation may be satisfied. The intensity of a continuous absorption spectrum gets sufficient contribution from phonons. With the involvement of a phonon in the indirect transition under discussion, the sum of wavevectors before the transition becomes equal to their sum after the transition, showing the momentum conservation. There can be two possibilities—one in which a phonon is emitted (created) after the transition and the other in which a phonon is absorbed (destroyed) along with the optical photon to materialize the transition.

Let us denote the wavevector of the optical photon by \mathbf{K}_{op} and that of the phonon by \mathbf{K}_{ph} . If

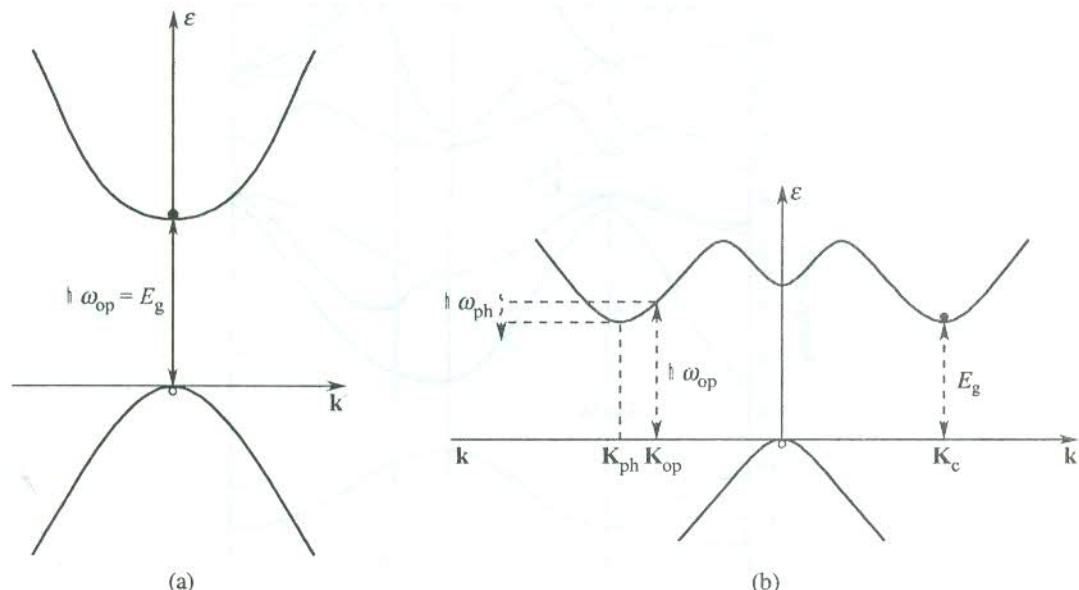


FIG. 9.6 (a) Absorption of a photon in a direct-gap semiconductor, where an electron from the valence band is lifted into the conduction band and the crystal momentum is conserved without the involvement of any other process. (b) The photon absorption in an indirect-gap material where it is shown that a phonon is emitted on the photon absorption to satisfy the momentum conservation. The \mathbf{K}_{op} and \mathbf{K}_{ph} , respectively, denote the wavevectors of the absorbed optical photon and the emitted phonon.

the wavevector at the conduction band minimum as measured from the valence band maximum be represented by \mathbf{K}_c , then in the case of phonon emission [Fig. 9.6(b)], we have

$$\begin{array}{l} \mathbf{K}_{\text{op}} = \mathbf{K}_c + \mathbf{K}_{\text{ph}} \\ (\text{before the transition}) \quad (\text{after the transition}) \end{array} \quad (9.1)$$

But the photon wavevectors \mathbf{K}_{op} are negligibly small in the range of energies concerned. Therefore, from (9.1),

$$\mathbf{K}_c \approx -\mathbf{K}_{\text{ph}} \quad (9.2)$$

Relation (9.2) states that the increase of crystal momentum by $\hbar\mathbf{K}_c$ during the transition is offset by an equal amount owing to the emission of a phonon with wavevector $(-\mathbf{K}_c)$. Since the phonon takes away a part of the energy of the incident photon (though very small), the optical threshold energy is greater than E_g [see Fig. 9.6(b)]. In this case,

$$\hbar\omega_{\text{op}} = E_g + \hbar\omega_{\text{ph}} \quad (9.3)$$

At high temperatures a good number of phonons are present in the crystal. With the absorption of a phonon during the transition, the momentum conservation is given by

$$\begin{array}{l} \mathbf{K}_{\text{op}} + \mathbf{K}_{\text{ph}} = \mathbf{K}_c \\ (\text{before the transition}) \quad (\text{after the transition}) \end{array} \quad (9.4)$$

This gives the optical threshold as

$$\hbar\omega_{\text{op}} = E_g - \hbar\omega_{\text{ph}} \quad (9.5)$$

So, the optical threshold is lower than E_g . The change in optical threshold in these processes is generally of little consequence because the phonon energies are characteristically small (\sim a few hundredths of an eV) as compared to the size of the band gap. This change, however, is of considerable value in semiconductors having a small band gap.

Since the indirect transitions generate heat when phonons are created, the direct-gap materials are preferred in order to have efficient devices. A mixture of the two types of materials is a desired manipulation in some devices such as injection lasers where it is practised on account of the demand for higher power.

Another method commonly used for determining the band gap is based on the temperature-dependent study of the electrical conductivity σ . We will see later in this chapter that electrical conductivity of intrinsic semiconductors depends on temperature according to the following proportionality:

$$\sigma \propto \exp(-E_g/2k_B T) \quad (9.6)$$

An approximate value of E_g is given by the $\ln(\sigma)$ versus $1/T$ graph. The gap is also determined from the intrinsic carrier concentrations derived from the experimental values of Hall coefficient. The values of band gap and nature of gap for a number of important semiconductors are given in Table 9.1.

9.3 INTRINSIC CARRIER DENSITIES

The difference between the electrical conductivity of semiconductors and metals in the form of

Table 9.1 Data on the band gap of some important semiconductors (D = direct gap, I = indirect gap)

Crystal	Type of gap	E_g (eV)	
		0 K	300 K
Si	I	1.17	1.11
Ge	I	0.75	0.67
grey Sn	D	0.00	0.00
GaAs	D	1.52	1.43
GaSb	D	0.81	0.68
GaP	I	2.32	2.25
InSb	D	0.23	0.17
InAs	D	0.43	0.36
InP	D	1.42	1.27
AlSb	I	1.65	1.6
PbS	D	0.29	0.34–0.37
PbSe	I	0.17	0.27
PbTe	I	0.19	0.29
CdS	D	2.58	2.42
CdSe	D	1.84	1.74
CdTe	D	1.61	1.44
ZnO	—	3.44	3.2
ZnS	—	3.91	3.6
SnSe	D	0.3	0.18
TiO ₂	—	3.03	—
Cu ₂ O	D	2.17	—

former's strong dependence on temperature indicates that thermal excitations control the conductivity of semiconductors in a big way. The thermal excitation of an electron from the valence band across the forbidden energy gap E_g to the conduction band creates a hole in the valence band. The number of these carriers that contribute to the flow of electric current increases more and more with a continued thermal excitation. In this section we calculate the density of these carriers in the state of thermal equilibrium in an intrinsic semiconductor considered as highly pure such that the contribution to the carrier concentrations from impurities may be neglected. Calculations based on appropriate statistics show that the electron and hole concentrations equal in the present case, are strongly temperature dependent which thus accounts for the conductivity behaviour. On the basis of what we learnt in Chapter 6, the conductivity of a semiconductor may be expressed as

$$\sigma = |e|(n\mu_n + p\mu_p) \quad (9.7)$$

where n and p are electron and hole densities (per unit volume), respectively, and μ_n and μ_p represent the corresponding mobilities. The contributions from electrons and holes in (9.7) simply add up because of the opposite sign of their charge and opposite directions of their drift velocities.

It is easy to appreciate that in excitations of our interest the carriers near the band edges or the parabolic part of the valence and conduction bands are involved. In this region the effective mass can be treated as constant in the first approximation. This takes care of the neglect of energy dependence of mobility in (9.7). For calculations in semiconductors the chemical potential μ appearing

in the Fermi distribution function (6.30) is replaced by the Fermi energy ε_F . But the level of chemical potential always lies in the region of the forbidden energy gap where no single-electron energy levels exist. The very definition of the Fermi level which is considered as the highest occupied level at absolute zero, thus becomes redundant since no single-electron energy level is available to coincide with the Fermi level. Therefore, in the fitness of things the Fermi level in semiconductors should be interpreted as a synonym to the chemical potential.

The occupancy of energy levels in semiconductors must be described by the Fermi-Dirac distribution function $f(\varepsilon, T)$ as in other solids:

$$f(\varepsilon, T) = \frac{1}{1 + \exp[(\varepsilon - \varepsilon_F)/k_B T]} \quad (9.8)$$

If $D_c(\varepsilon)$ and $D_v(\varepsilon)$ denote the density of states in the conduction and valence band, respectively, the charge carrier densities are usually written as

$$n = \int_{E_c}^{\infty} D_c(\varepsilon) f_e(\varepsilon, T) d\varepsilon \quad (9.9)$$

with f_e as the electron occupancy, given by (9.8) and

$$p = \int_{-\infty}^{E_v} D_v(\varepsilon) f_h(\varepsilon, T) d\varepsilon = \int_{-\infty}^{E_v} D_v(\varepsilon) [1 - f_e(\varepsilon, T)] d\varepsilon \quad (9.10)$$

with f_h as the hole occupancy.

Here E_c and E_v refer to the energy values at the edges of the conduction and valence band, respectively (Fig. 9.7). The electron energy as measured from the bottom (or edge) of the conduction band is $(\varepsilon - E_c)$, where ε is the absolute value of electron energy. Similarly, the hole energy when measured from the top (or edge) of the valence band is equal to $(E_v - \varepsilon)$. Applying the parabolic approximation, i.e. assuming the effective mass to remain constant, the two density of states are written using (6.41) as

$$D_c(\varepsilon) = \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2} \right)^{3/2} (\varepsilon - E_c)^{1/2} \quad (9.11)$$

and

$$D_v(\varepsilon) = \frac{1}{2\pi^2} \left(\frac{2m_h^*}{\hbar^2} \right)^{3/2} (E_v - \varepsilon)^{1/2} \quad (9.12)$$

When electron and hole energies are such that $|(\varepsilon - \varepsilon_F)| \gg k_B T$, the respective distribution functions are effectively represented by Boltzmann distribution functions as expressed below:

$$f_e \approx \exp \left[\frac{-(\varepsilon - \varepsilon_F)}{k_B T} \right] \quad (9.13)$$

(in the conduction band)

and

$$f_h = 1 - f_e \approx \exp \left[\frac{-(\epsilon_F - \epsilon)}{k_B T} \right] \quad (9.14)$$

(in the valence band)

In this condition the carrier density is not large and the semiconductor is called a non-degenerate semiconductor.

Substituting (9.13) in (9.9), the electron density in the state of thermal equilibrium at temperature T has the form,

$$n = \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2} \right)^{3/2} \exp \left(\frac{\epsilon_F}{k_B T} \right) \int (\epsilon - E_c)^{1/2} \exp \left(-\frac{\epsilon}{k_B T} \right) d\epsilon \quad (9.15)$$

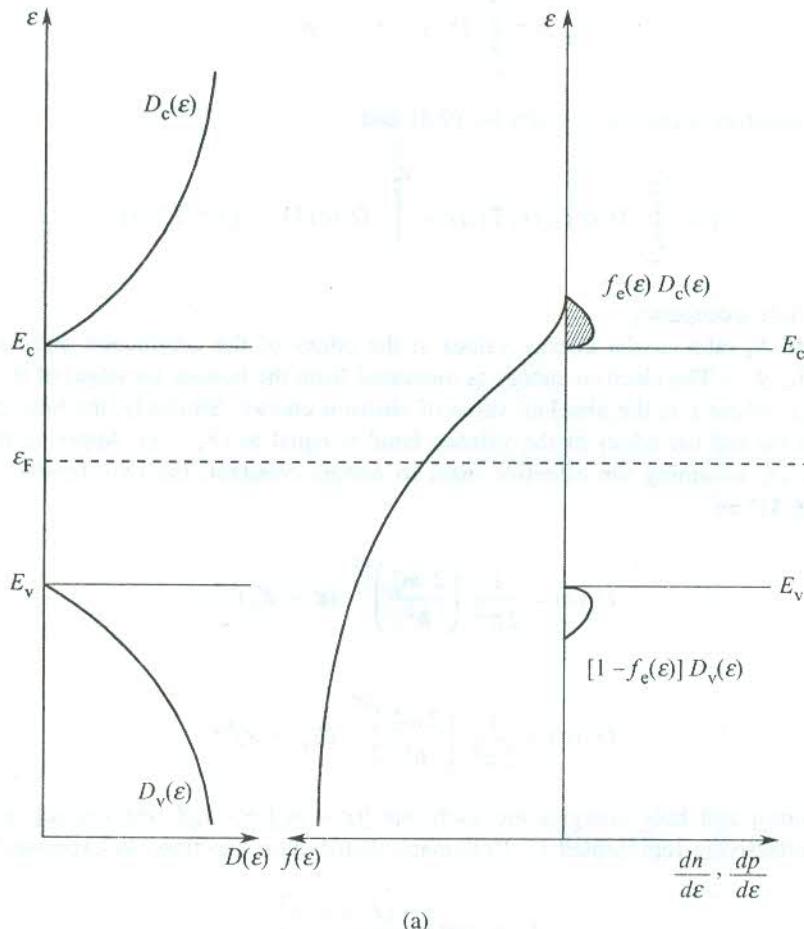


FIG. 9.7 Contd.

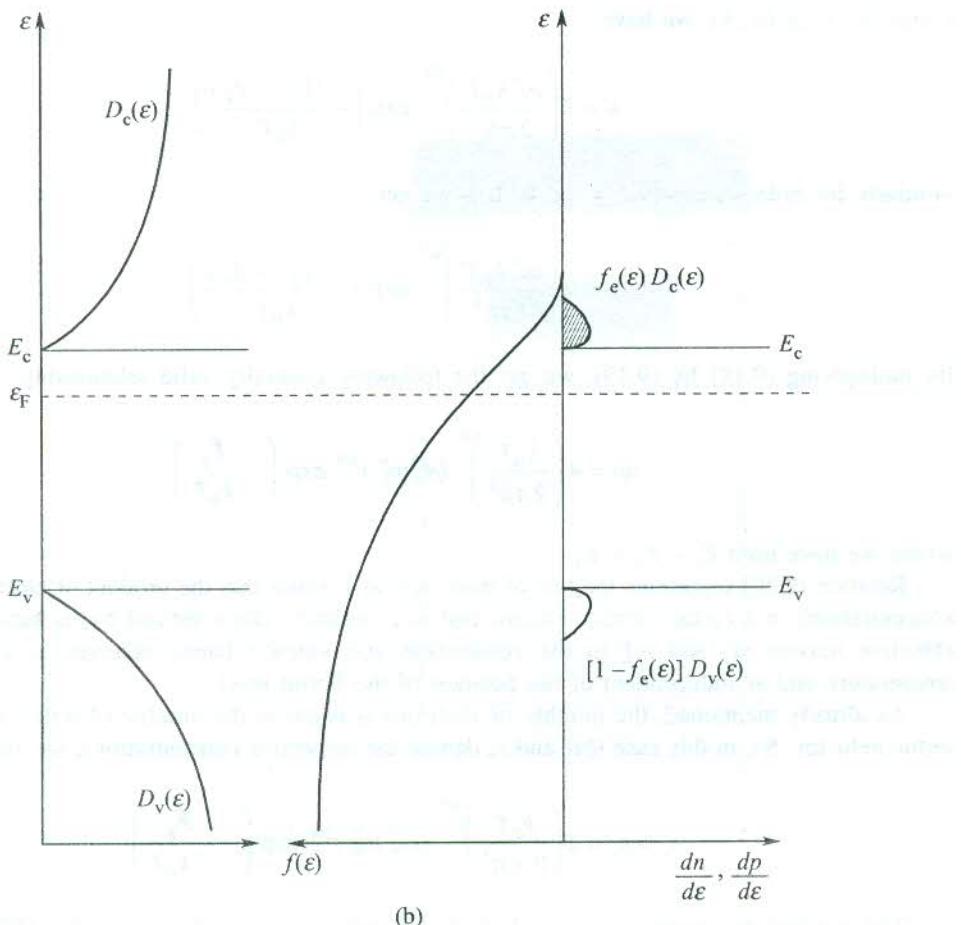


FIG. 9.7 A schematic behaviour of the Fermi function $f(\epsilon)$, the density of states $D(\epsilon)$, and electron (n) and hole (p) densities in the conduction and valence bands when: (a) the densities of states in the conduction and valence bands are equal and (b) the densities of states in the conduction and valence bands are not equal.

Using the substitution $\left(\frac{\epsilon - E_c}{k_B T}\right) = x^2$, we get

$$n = \frac{1}{\pi^2} \left(\frac{2 m_e^*}{\hbar^2}\right)^{3/2} (k_B T)^{3/2} \exp\left(\frac{-(E_c - \epsilon_F)}{k_B T}\right) \int_0^{\infty} x^2 \exp(-x^2) dx \quad (9.16)$$

But,

$$\int_0^{\infty} x^2 \exp(-x^2) dx = \frac{\sqrt{\pi}}{4} \quad (9.17)$$

(from standard tables)

Using (9.17) in (9.16), we have

$$n = 2 \left(\frac{m_e^* k_B T}{2\pi\hbar^2} \right)^{3/2} \exp \left(\frac{-(E_c - \varepsilon_F)}{k_B T} \right) \quad (9.18)$$

Similarly for holes, using (9.14) and (9.10), we get

$$p = 2 \left(\frac{m_h^* k_B T}{2\pi\hbar^2} \right)^{3/2} \exp \left(\frac{-(\varepsilon_F - E_v)}{k_B T} \right) \quad (9.19)$$

By multiplying (9.18) by (9.19), we get the following generally valid relationship:

$$np = 4 \left(\frac{k_B T}{2\pi\hbar^2} \right)^3 (m_e^* m_h^*)^{3/2} \exp \left(- \frac{E_g}{k_B T} \right) \quad (9.20)$$

where we have used $E_c - E_v = E_g$.

Relation (9.20) represents the law of mass action. It states that the product of electron and hole concentrations in a certain semiconductor that is completely characterized by its band gap E_g and effective masses m_e^* and m_h^* in the conduction and valence bands behaves as a function of temperature and as independent of the position of the Fermi level.

As already mentioned, the number of electrons is equal to the number of holes in an intrinsic semiconductor. So, in this case if n_i and p_i denote the respective concentrations, we get from (9.20)

$$n_i = p_i = 2 \left(\frac{k_B T}{2\pi\hbar^2} \right)^{3/2} (m_e^* m_h^*)^{3/4} \exp \left(- \frac{E_g}{2k_B T} \right) \quad (9.21)$$

The numbers as determined at 300 K in Si, Ge and GaAs are 2.4×10^{13} , 1.5×10^{10} and 5×10^7 cm $^{-3}$, respectively. In intrinsic materials the Fermi level at a certain temperature adjusts its position that is required to maintain the charge neutrality condition consistent with (9.18) and (9.19). This demands that

$$n = p$$

giving

$$\exp \left(\frac{2\varepsilon_F}{k_B T} \right) = \left(\frac{m_h^*}{m_e^*} \right)^{3/2} \exp \left(\frac{E_v + E_c}{k_B T} \right) \quad (9.22)$$

or

$$\varepsilon_F = E_v + \frac{1}{2} E_g + \frac{3}{4} k_B T \ln \left(\frac{m_h^*}{m_e^*} \right) \quad (9.23)$$

Thus the Fermi level shows a weak temperature dependence. But if the electron and hole effective masses are equal, $\varepsilon_F = E_v + \frac{1}{2} E_g$ and the Fermi level lies exactly in the middle of the gap. This is true at all temperatures.

But in general when the density of states of conduction and valence bands and the two effective masses are unequal, the Fermi level is asymmetrically placed with respect to the positions of E_c and E_v . It is demonstrated in Fig. 9.7(b) where the two density of states are unequal.

9.4 GENERAL FEATURES OF EXTRINSIC SEMICONDUCTORS

As mentioned in Section 9.1, the extrinsic semiconductors were developed because the electrical conductivity of intrinsic semiconductor is, generally, not large enough to meet the requirement of devices. Besides, it is hard to imagine of an absolute intrinsic material. Even the purest single crystals have impurity contents to a certain degree. But these impurity contents normally do not increase the carrier concentration to a useful level. GaAs is, however, an exception. The purest available single crystals of GaAs show a carrier density of about 10^{16} cm^{-3} which is enormously high compared to the intrinsic value ($5 \times 10^7 \text{ cm}^{-3}$). In general, the standard method of increasing the conductivity of an intrinsic material is to add to it a suitable impurity or electrically active element in small concentration. The method is known as *doping*. Impurities that enhance the carrier density by contributing additional electrons to the conduction band are called *donors* and those which create additional holes in the valence band are known as *acceptors*.

For example, let us consider the electrically active elements suitable for doping in Si and Ge crystals. When pure crystals of Si and Ge are doped with any element from Group V such as P or As, we get *n*-type semiconductors. On doping with an element from Group III (e.g. B, Al, In), *p*-type semiconductors are formed.

9.4.1 The *n*-type Semiconductors

Suppose a pure Ge crystal is doped with As, an immediate neighbour to the right of Ge in the periodic table. The arsenic atom may enter the germanium crystal lattice either by replacing a germanium atom (i.e. substitutionally) or by occupying a position where no germanium atom is supposed to be located in a pure and perfect crystal (i.e. interstitially). Data on lattice constant measurements show that the arsenic atom enters the crystal substitutionally. The electronic configurations in the outermost shells of germanium and arsenic atoms are $3s^2 3p^2$ and $4s^2 4p^3$, respectively. The germanium crystal has a diamond structure in which each atom forms tetrahedral bonds with its four neighbours. When an arsenic atom that substitutes a germanium atom finds itself surrounded by four germanium atoms, its four electrons get engaged in tetrahedral bonds with all four neighbouring atoms as depicted in Fig. 9.8. The arsenic atom is left with an extra loosely bound valence electron which may be easily freed and made available in the conduction band for the purpose of conduction. On account of having these additional electrons for conduction, the doped crystal is called an *n*-type semiconductor where *n* stands for electrons.

According to a model proposed to interpret the replacement of a germanium atom by an arsenic atom, the arsenic atom is pictured as an occasional germanium atom with an additional positive charge of e fixed at its core to which an additional electron is bound. Thus the donor impurity can be described as a hydrogen-like centre in which the Coulomb attraction between the core and the valence electron is screened by the neighbouring germanium electrons. The centre being a bound system is characterized by a set of quantized energy levels whose scheme is qualitatively similar to that of the hydrogen atom. The ionization energies of the hydrogen and the donor atoms are expressed as

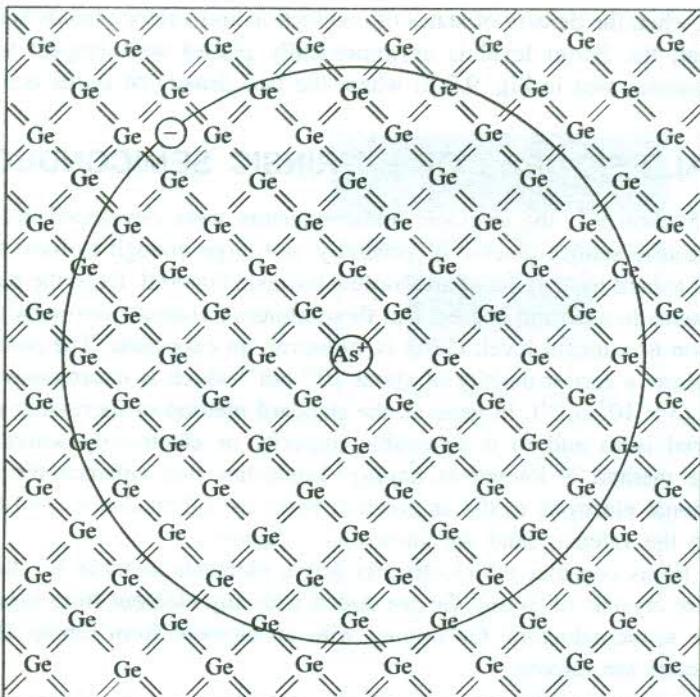


FIG. 9.8 A donor impurity atom As is shown to substitute a Ge atom in the germanium crystal. The fifth electron of As, that is not engaged in a tetrahedral bond (also interpreted as the bound valence electron of donor impurity centre), is smeared over thousands of Ge sites in the crystal. The lattice constant and the Bohr radius of the impurity centre are not drawn to scale.

$$E_H = \frac{me^4}{2(4\pi\epsilon_0\hbar)^2} \quad (9.24)$$

for the hydrogen atom

and

$$E_d = \frac{m_e^* e^4}{2(4\pi\epsilon_m\hbar)^2} = \frac{m_e^* E_H}{m_s^2} \quad (9.25)$$

for the donor atom

Here ϵ_0 is the permittivity of vacuum space and ϵ_s is the relative permittivity or the static dielectric constant of the medium of the germanium crystal. The constant $\epsilon_m (= \epsilon_0\epsilon_s)$ represents the permittivity of the medium of the crystal. The values of the static dielectric constant of some important semiconductors are given in Table 9.2.

For calculating E_d we take the effective mass of the conduction electron in germanium as $m_e^* = 0.12m$ where m is the free electron mass. Using this substitution in (9.25), we get

$$E_d = \frac{(0.12) E_H}{\epsilon_s^2} \quad (9.26)$$

Table 9.2 Values of static dielectric constant ϵ_s for some semiconductors

Crystal	ϵ_s
Si	11.7
Ge	15.8
AlAs	10.1
AlSb	10.3
GaAs	13.13
GaSb	15.69
InSb	17.88
InAs	14.55
InP	12.37
SiC	10.2

With $E_H = 13.6$ eV and $\epsilon_s = 15.8$ for Ge, E_d is found out to be about 6.5 meV. For Si, $m_e^* = 0.3m$ and $\epsilon_s = 11.7$ and, therefore, E_d comes to about 30 meV. The position of the donor's ground energy level E_D with respect to the conduction and valence band edges is shown in Fig. 9.9(a). Its energy as measured from the conduction band edge is E_d . The energy continuum of the donor energy level scheme begins at the conduction band edge. Therefore, the electron from a donor atom can be transferred to the conduction band by simply ionizing the donor atom at the expense of only a few meV (~ 6.5 meV in Ge and ~ 30 meV in Si) and not of a few eV of energy as is the case with the intrinsic material. This makes abundantly clear why the conductivity of an *n*-type semiconductor is several orders bigger than that of the corresponding intrinsic material.

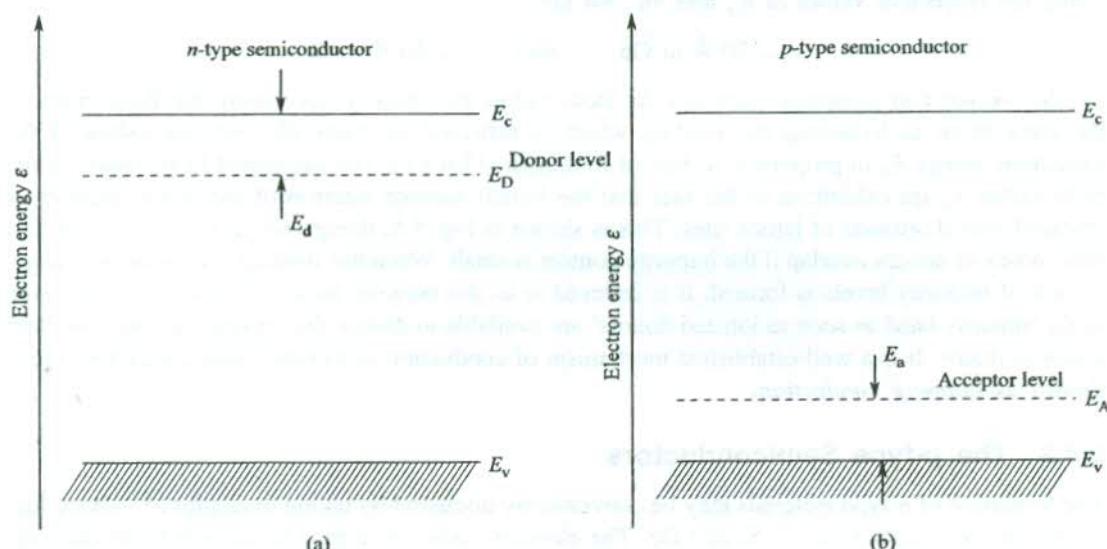


FIG. 9.9 (a) A qualitative picture showing the position of the ground donor level E_D relative to the conduction and valence band edges E_c and E_v . The E_d denotes the ionization energy of the donor atom. (b) A qualitative picture showing the position of the ground acceptor level E_A relative to the conduction and valence band edges E_c and E_v . The E_a denotes the ionization energy of the acceptor atom.

An odd feature of the model described above is that it gives the same value of E_d for all donor impurities such as P, As and Sb in a semiconductor host. The experimental values, however, show a slight variation (see Table 9.3). The description of screening in terms of the dielectric constant is rather crude and it restricts the operational domain of the model. The limitations are reflected in the model's failure to account for some of the subtle consequences of the atomic effects. The effect of screening on the Bohr radius is further revealing. The Bohr radius of the donor impurity atom is written as

$$r_d = \frac{4\pi\epsilon_0\epsilon_s\hbar^2}{m_e^*e^2} = \left(\frac{m\epsilon_s}{m_e^*}\right)r_0 \quad (9.27)$$

where r_0 is the hydrogen Bohr radius (0.53 Å).

Table 9.3 Ionization energies for a few donors and acceptors in silicon and germanium (E_d : ionization energy of donors and E_a : ionization energy of acceptors)

Impurity	E_d (meV)		Impurity	E_a (meV)	
	Si	Ge		Si	Ge
P	45	12	B	45	10.4
As	49	12.7	Al	57	10.2
Sb	39	9.6	Ga	65	10.8
			In	16	11.2

Using the respective values of ϵ_s and m_e^* , we get

$$r_d \approx 70 \text{ \AA} \text{ in Ge} \quad \text{and} \quad \approx 20 \text{ \AA} \text{ in Si}$$

So we see that screening increases the Bohr radius enormously. Increasing the Bohr radius is the same thing as loosening the binding which is reflected in drastically reduced values of the ionization energy E_d in proportion to that of hydrogen (13.6 eV). The calculated large values of the Bohr radius r_d are exhibition of the fact that the bound valence electron of the donor impurity is smeared over thousands of lattice sites. This is shown in Fig. 9.8, though not to scale. Thus the first Bohr orbits of donors overlap if the impurity content is small. When this overlap is considerably large, a band of impurity levels is formed. It is referred to as the *impurity band*. Conduction occurs even in the impurity band as soon as ionized donors[†] are available to initiate the hopping of electrons from donor to donor. It is a well-established mechanism of conduction in extrinsic semiconductors and is known as *hopping conduction*.

9.4.2 The *p*-type Semiconductors

The formation of *p*-type materials may be conveniently discussed by taking examples of valence four elemental semiconductors, i.e. Si and Ge. The elements used for doping to convert these materials into *p*-type are from Group III (e.g. B, Al, Ga, In) with three electrons in the outermost shell. The

[†] Some ionized donors always exist, firstly because of the presence of some acceptors whose affinity for an electron may ionize donors and secondly because the average thermal energy per electron ($\sim k_B T$) around room temperature is ~ 26 meV which is enough to ionize some donors.

substitutional impurity in this case completes only three of the four characteristic tetrahedral bonds of the host crystal. The doped impurity atom lacks one electron to complete the bonding with its all the neighbours. A valence electron of the host material may meet this requirement by ionizing the impurity atom negatively and creating a hole in the vicinity. Since the impurity atom is willing to accept an electron, the impurity atom is called an *acceptor*. The acceptor model is similar to the donor model. For example, a boron atom substituted in the germanium crystal is pictured as a germanium atom with a charge of $(-e)$ fixed at its core and a hole of charge (e) bound to it. The impurity acts as the centre of a bound system whose energy level scheme is similar to that of the hydrogen atom. The ground acceptor level E_A lies close to the valence band edge as shown in Fig. 9.9(b). The ionization energy E_a as measured from the valence band edge is again very small as compared to the size of E_g . The values of E_a for some acceptors in Si and Ge are given in Table 9.3. Arguments concerning large values of conductivity and hopping conduction follow the same lines as those for *n*-type materials. With the latest methods of doping, the lowest impurity concentrations that can be obtained in semiconductors are of the order of 10^{12} cm^{-3} . Hence Si that has an intrinsic concentration of $1.5 \times 10^{10} \text{ cm}^{-3}$ at 300 K does not show intrinsic conductivity though Ge does, on account of its higher intrinsic concentration (2.4×10^{13} at 300 K).

9.5 POPULATION OF DONOR AND ACCEPTOR LEVELS IN THE STATE OF THERMAL EQUILIBRIUM

Let us first define the symbols that will be used in our future descriptions:

- N_D/N_A : density of all available donors/acceptors
- N_D^0/N_A^0 : density of the neutral donors/acceptors
- N_D^+/N_A^- : density of ionized donors/acceptors
- n_D/p_A : density of electrons bound to donors/density of holes bound to acceptors
- n : density of electrons in the conduction band
- p : density of holes in the valence band.

For small impurity concentrations (i.e. for non-degenerate semiconductors) the occupancy of the conduction and valence bands is as usual described by the Boltzmann distribution function. Therefore, at these concentrations the law of mass action (9.20) which was derived on the basis of the above conditions must apply even to extrinsic semiconductors. But in the extrinsic case the value of the Fermi energy, not figuring in the law of mass action, depends on a more complicated charge neutrality condition. For homogeneous doping, the neutrality condition can be expressed as

$$n + N_A^- = p + N_D^+ \quad (9.28)$$

with

$$N_D = N_D^0 + N_D^+ \quad (9.29a)$$

$$N_A = N_A^0 + N_A^- \quad (9.29b)$$

The above terminology may be appreciated better with the help of Fig. 9.10. Let us first consider the calculation of population of the donor levels. At low impurity concentrations (up to 10^{17} cm^{-3}), the interaction between electrons bound to separate donors may be neglected. We further assume for

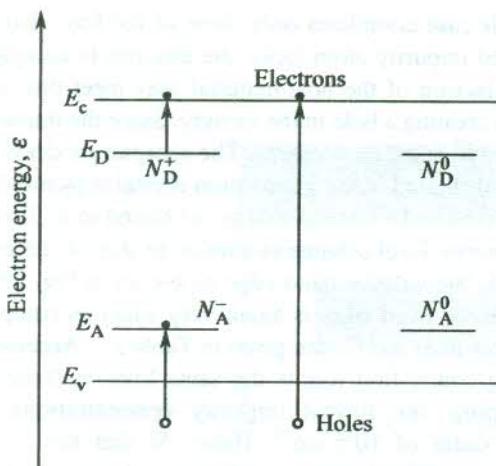


FIG. 9.10 Electrons in the conduction band and holes in the valence band made available either by inter-band excitation or by impurity ionization.

simplicity of calculations that the impurity introduces only a single one-electron level. Under these approximations the level could either be empty, or contain one electron of either spin, or contain two electrons with opposite spins. The double occupation is not favoured on account of its high energy arising from the Coulomb repulsion between two localized electrons. In the state of thermal equilibrium, the mean number of electrons in a system is expressed as

$$\langle n \rangle = \frac{\sum_j N_j \exp [-(\epsilon_j - \epsilon_F N_j)/k_B T]}{\sum_j \exp [-(\epsilon_j - \epsilon_F N_j)/k_B T]} \quad (9.30)$$

where the sum is over all states of the system; the ϵ_j and N_j denote the energy and the number of electrons in state j . With a single impurity we have just three states: one with no electrons, thereby making no contribution to the energy, and two distinguishable spin states with a single electron of energy E_D . Hence from (9.30), we have

$$\begin{aligned} \langle n \rangle &= \frac{2 \exp [-(E_D - \epsilon_F)/k_B T]}{1 + 2 \exp [-(E_D - \epsilon_F)/k_B T]} \\ &= \frac{1}{1 + \frac{1}{2} [(E_D - \epsilon_F)/k_B T]} \end{aligned} \quad (9.31)$$

The factor of $1/2$ serves to modify the Fermi-Dirac distribution in this case. However, this factor is generally ignored which amounts to treating two single electron spin states of equal energy as one state. The population density of the donor level in the state of thermal equilibrium is given by

$$n_D = N_D \langle n \rangle$$

or

$$n_D = \frac{N_D}{1 + \exp [(E_D - \epsilon_F)/k_B T]} \quad (\text{ignoring the factor of } 1/2) \quad (9.32)$$

Also note that,

$$n_D = N_D^0 \quad (9.33)$$

Similarly, it can be shown that the population density of the acceptor level in thermal equilibrium is given by

$$p_A = \frac{N_A}{1 + \exp [(\epsilon_F - E_A)/k_B T]} \quad (9.34)$$

and also,

$$p_A = N_A^0 \quad (9.35)$$

9.6 EXTRINSIC CARRIER DENSITIES

The case of an extrinsic semiconductor in which both donors and acceptors are present is difficult to handle. Because of this reason we consider a pure *n*-type semiconductor and calculate its carrier concentration that is contributed by donors alone. Generally, every *n*-type material has a few acceptors and every *p*-type material has a few donors because of practical limitations on growing 100 per cent pure crystals. But the concentration of naturally present impurities in a pure crystal is expected to be negligibly small compared to the concentration of the impurity doped. Therefore, the relation (9.28) in the present case is rewritten as

$$n = N_D^+ + p_i \quad (9.36)$$

The net electron density n has contributions of electrons from donors and from the valence band (see Fig. 9.10). The density of the latter type (n_i , being intrinsic) is equal to the density of holes (p_i). Normally $N_D^+ \gg n_i$ (or p_i). For example, in Si this condition is satisfied even at low levels of the donor concentration since the intrinsic concentration is mere $1.5 \times 10^{10} \text{ cm}^{-3}$ around room temperature. Therefore for our *n*-type material, p_i in (9.36) may be dropped giving,

$$n \approx N_D^+ = N_D - N_D^0 \quad [\text{using (9.29a)}] \quad (9.37)$$

or

$$\begin{aligned} n &= N_D - n_D \quad (\text{since } n_D = N_D^0) \\ &= N_D \left(1 - \frac{1}{1 + \exp [(E_D - \epsilon_F)/k_B T]} \right) \quad [\text{using (9.32)}] \\ &= \left(\frac{N_D}{1 + \exp [-(E_D - \epsilon_F)/k_B T]} \right) \end{aligned} \quad (9.38)$$

As already mentioned in the beginning of Section 9.5, the relations for n , p and np (the law of the mass action) as derived in Section 9.3 are applicable to extrinsic materials with low doping levels. We rewrite the relation (9.18) as

$$n = N(c) \exp\left[\frac{-(E_c - \varepsilon_F)}{k_B T}\right] \quad (9.39a)$$

with

$$N(c) = 2 \left(\frac{m_e^* k_B T}{2 \pi \hbar^2} \right)^{3/2} \quad (9.39b)$$

and (9.19) as

$$p = N(v) \exp\left(\frac{-(\varepsilon_F - E_v)}{k_B T}\right) \quad (9.40a)$$

with

$$N(v) = 2 \left(\frac{m_h^* k_B T}{2 \pi \hbar^2} \right)^{3/2} \quad (9.40b)$$

From (9.39a), we have

$$\exp\left(\frac{\varepsilon_F}{k_B T}\right) = \frac{n \exp(E_c/k_B T)}{N(c)} \quad (9.41)$$

Eliminating ε_F in (9.38) with the help of (9.41), we get

$$n = \frac{N_D}{1 + \frac{n \exp(E_d/k_B T)}{N(c)}} \quad (\text{with } E_c - E_D = E_d)$$

or

$$n^2 \left(\frac{\exp(E_d/k_B T)}{N(c)} \right) + n - N_D = 0 \quad (9.42)$$

Using the substitution,

$$\frac{\exp(E_d/k_B T)}{N(c)} = X \quad (9.43)$$

we get

$$Xn^2 + n - N_D = 0 \quad (9.44)$$

The physically meaningful solution to the above equation is

$$n = \frac{-1 + \sqrt{1 + 4 N_D X}}{2 X} \quad (9.45)$$

Rationalizing (9.45), we get

$$n = 2 N_D \left[1 + \sqrt{1 + 4 N_D X} \right]^{-1}$$

Using (9.43), we have

$$n = 2 N_D \left[1 + \sqrt{1 + 4 N_D \frac{\exp(E_d/k_B T)}{N(c)}} \right]^{-1} \quad (9.46)$$

There are three limiting cases of the above expression as discussed below:

Case I. At low temperatures,

$$4 \left(\frac{N_D}{N(c)} \right) \exp \left(\frac{E_d}{k_B T} \right) \gg 1 \quad (9.47)$$

This transforms (9.46) to the following form,

$$n \simeq \sqrt{N_D N(c)} \exp \left(\frac{-E_d}{2 k_B T} \right) \quad (9.48)$$

At these low temperatures a large number of donors stay in the unionized state. The range of temperatures over which this condition exists is known as the *freeze-out range*. We can notice the similarities between (9.48) and (9.21) that expresses the intrinsic carrier density. Both depend exponentially on temperature since the exponential dependence dominates the other dependence entering through $N(c)$. In the donors' case a much smaller quantity E_d appears in the exponential as against E_g in the intrinsic case. This accounts for a larger carrier concentration in the *n*-type material.

Case II. For temperatures at which

$$4 \left(\frac{N_D}{N(c)} \right) \exp \left(\frac{E_d}{k_B T} \right) \ll 1 \quad (9.49)$$

the relation (9.46) reduces to

$$n \simeq N_D \quad (9.50)$$

i.e. a constant.

This means that at these temperatures all donors are ionized and the electron density reaches its maximum with the excitation of electrons from the valence band considered negligible in the first approximation. This is referred to as the *saturation range* or *exhaustion range* of carriers.

Case III. When temperatures are still higher, the concentration of conduction electrons contributed by the valence band becomes appreciable. Since the concentration of donor electrons in the conduction band no more increases on account of saturation, the intrinsic electron density overtakes the density of donor electrons at some stage. In this condition an *n*-type semiconductor is characterized by intrinsic behaviour and we speak of the *intrinsic range* of carriers. The variations of electron density and energy as functions of temperature are sketched in Fig. 9.11, identifying the three ranges.

The calculation of electron density in the conduction band as carried out above is based on the assumption that there are no acceptor impurities in the *n*-type semiconductor under consideration. Experiments, however, confirm that traces of acceptors are always present. Therefore, quantitative estimates made on the basis of the above theory may differ slightly from the experimental values.

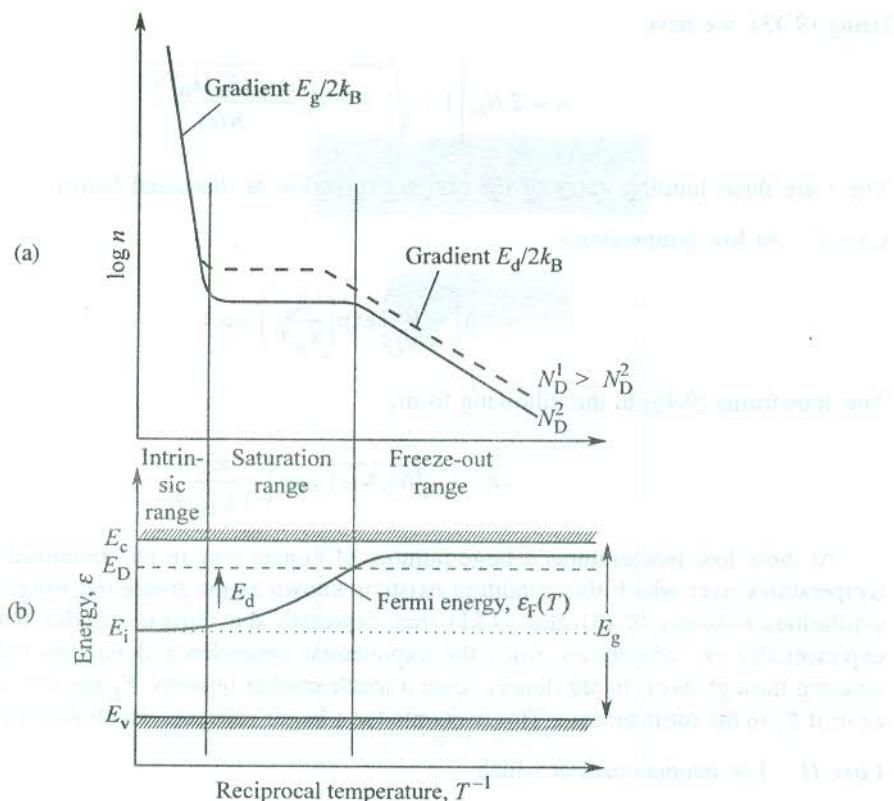


FIG. 9.11 (a) A qualitative illustration of the variation of electron concentration (n) in the conduction band with change in temperature in two samples of different donor concentrations N_D^1 and N_D^2 . (b) A qualitative illustration of the temperature dependence of the Fermi energy ϵ_F in the same semiconductor. The E_i represents the Fermi energy in the intrinsic material. All other symbols have their usual meaning.

There are several applications in which there is a need to monitor the electron and hole concentrations. Since at low doping levels the law of mass action (9.20) is valid, it is exploited to manipulate the carrier concentration in non-degenerate[†] semiconductors. When the electron concentration n is increased by adding the trace of a suitable donor impurity, the hole concentration p goes down so that the product (np) remains constant at a temperature in accordance with the law of mass action. This way it is possible to reduce the sum $(n + p)$ considerably. This method of reduction is in wide practice and known as *compensation*. The sign of majority carriers is quickly determined by detecting the sign of Hall voltage.

[†] In a degenerate semiconductor the number of impurity atoms is considerably large, due to which the number of energy levels in an impurity band is large. At these large densities the interatomic separation between impurity atoms becomes smaller resulting in the large overlap of the atomic orbitals. Because of this enhanced overlap, the width of the impurity band is increased. Under these circumstances some of the impurity levels extend into the region of the conduction band or valence band depending on whether the levels belong to the donor or acceptor impurity band. The energy value (eigenvalue) of any such impurity level is thus common to two eigen states, one belonging to the host atom and the other to the impurity atom. Hence the level in question is degenerate in the literal sense. The semiconductor with this band scheme is called a *degenerate semiconductor*.

9.7 TEMPERATURE DEPENDENCE OF ELECTRICAL CONDUCTIVITY

The electrical conductivity does not show the same type of temperature dependence over the extended and the practically useful range of temperatures. Above a certain temperature, the conductivity is not modified by the impurities in the crystal. The range extending above this temperature is called *the intrinsic temperature range* and the conductivity is referred to as the intrinsic conductivity (σ_i). Substituting the values of carrier densities (9.21) in (9.7), the intrinsic conductivity can be written as

$$\sigma_i = 2|e| \left(\frac{k_B T}{2\pi\hbar^2} \right)^{3/2} (m_e^* m_h^*)^{3/4} e^{-E_g/2k_B T} (\mu_n + \mu_p) \quad (9.51)$$

As we will see below, the mobilities depend on temperature only as a simple power law over an appropriate region. In view of this fact, the temperature dependence of conductivity will be dominated by the exponential dependence of carrier concentrations, giving

$$\sigma_i \propto e^{-E_g/2k_B T} \quad (9.52)$$

or $\rho_i = A e^{E_g/2k_B T} \quad (9.53)$

which gives

$$\ln \rho_i = \ln A + \frac{E_g}{2k_B T}$$

The above relation shows that the intrinsic resistivity is a linear function of the reciprocal of temperature which is confirmed by experiments.

However, for a proper analysis of the temperature dependence of electrical conductivity, the temperature variation of the carrier densities and that of the mobilities must be weighed against each other.

The conduction in semiconductors differs from the metallic conduction in the sense that electrons near the Fermi level contribute to the metallic conductivity whereas in semiconductors electrons near the bottom of the conduction band and holes near the top of the valence band take part in the process. There are no charge carriers near the Fermi level in semiconductors. In the approximation where the Fermi statistics can be replaced by Boltzmann statistics for non-degenerate semiconductors, the Boltzmann equation is solved to get to the expression for mobility of charge carriers. On demand of simplicity we present here only a qualitative description of the scattering processes in which electrons and holes by and large behave similarly. Several simplifications of the exact expression for mobility lead to the result,

$$\mu \propto \tau \quad (9.54)$$

where τ is the relaxation time defined in Section 6.1.1.

Since τ is proportional to the average time between successive collisions,

$$\frac{1}{\tau} \propto \langle v \rangle \sum \quad (9.55)$$

where Σ denotes the scattering cross-section of electrons and holes at a scattering centre. Relation (9.55) gives a measure of the scattering probability with $\langle v \rangle$ treated as thermal average over all electron or hole velocities in the lower conduction band or upper valence band. The use of Boltzmann statistics in semiconductors gives

$$\langle v \rangle \propto \sqrt{T} \quad (9.56)$$

Phonons happen to be the prominent source of carrier scattering in crystals. It is simpler to calculate the phonon scattering cross-section Σ_L for acoustic phonons, ($T \gg \theta_D$), where θ_D is the Debye temperature. This gives the dependence,

$$\Sigma_L \propto T \quad (L \text{ stands for lattice or phonons}) \quad (9.57)$$

Making use of (9.55), (9.56) and (9.57), we have from (9.54)

$$\mu_L \propto T^{-3/2} \quad (9.58)$$

In semiconductors, centres of ionized donors and acceptors serve as another important source of scattering. As an electron or a hole approaches such a centre, it experiences a Coulomb force and suffers scattering similar to Rutherford scattering. A rigorous treatment shows that

$$\Sigma_I \propto \langle v \rangle^{-4} \quad (9.59)$$

where I stands for ionized impurity. Therefore, with the use of (9.55), (9.56) and (9.59) in (9.54), we get

$$\mu_I \propto T^{3/2} \quad (9.60)$$

For one type of carriers, say electrons, we can write contributions of phonons and ionized impurities to resistivity as

$$\rho_L = \frac{1}{\sigma_L} = \frac{1}{ne\mu_L} \quad \text{and} \quad \rho_I = \frac{1}{\sigma_I} = \frac{1}{ne\mu_I} \quad (9.61)$$

Applying the Matthiessen's rule, the total resistivity is written as

$$\rho = \frac{1}{ne\mu} = \frac{1}{ne} \left[\frac{1}{\mu_L} + \frac{1}{\mu_I} \right] \quad (9.62)$$

or

$$\frac{1}{\mu} = \frac{1}{\mu_L} + \frac{1}{\mu_I} \quad (9.63)$$

A qualitative display of the temperature dependence of mobility in an extrinsic semiconductor is made in Fig. 9.12.

The variation of conductivity σ as a function of temperature is plotted for six samples of n -type germanium in Fig. 9.13. The conductivity shows a maximum in the exhaustion range where the electron concentration n becomes almost constant. The behaviour of σ in this range is governed by the characteristic temperature dependence of mobility as shown in Fig. 9.12. In the intrinsic range (at high temperatures) and in the freeze-out (at low temperatures), the variation in σ is effectively controlled

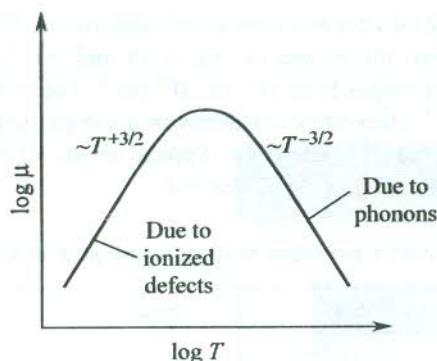


FIG. 9.12 Qualitative temperature dependence of the mobility contributions from charged impurities and phonons.

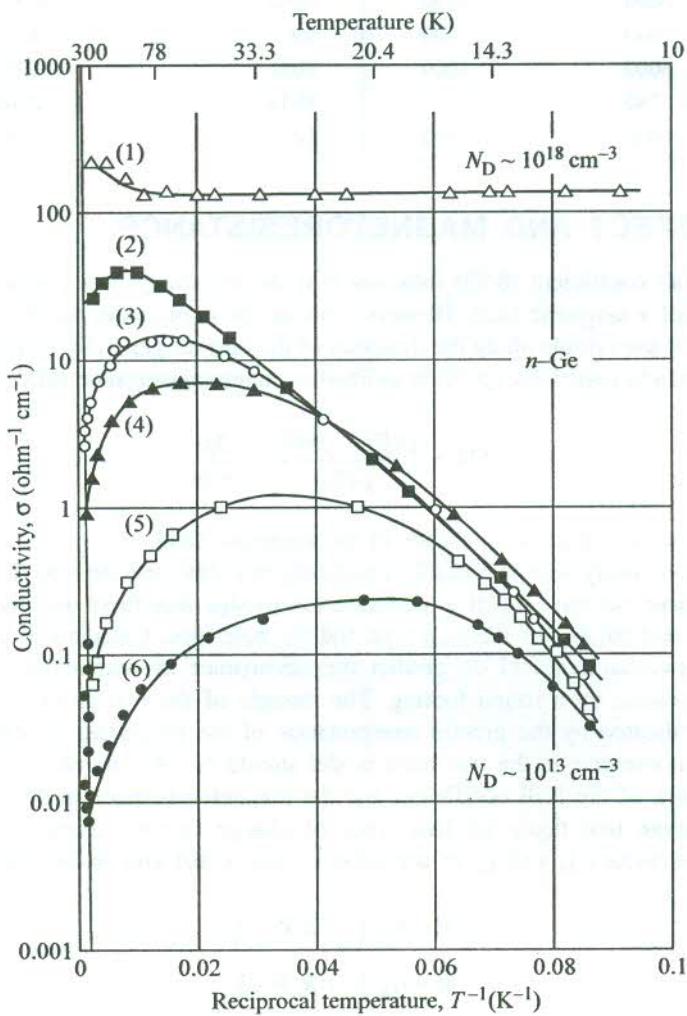


FIG. 9.13 Observed conductivity behaviour as a function of temperature for six samples of n -type germanium. The donor concentration N_D ranges from 10^{13} to 10^{18} cm^{-3} . [After E.M. Conwell, *Proc. IRE*, **40**, 1327 (1952).]

by the exponential dependence of concentration n on temperature. The temperature dependence of mobility has been derived from the curves of Fig. 9.13 and the Hall measurements. The donor concentration N_D in six samples ranges from 10^{13} to 10^{18} cm^{-3} . The mobility in the purest crystal with $N_D \approx 10^{13} \text{ cm}^{-3}$ shows ideal $T^{-3/2}$ dependence whereas in a sample with increasing values of N_D the mobility approaches the expected $T^{3/2}$ behaviour. Typical orders of electron and hole mobilities in some important semiconductors are given in Table 9.4.

Table 9.4 Typical order of carrier mobilities at room temperature in some important semiconductors

Crystal	Mobility ($\text{cm}^2/\text{V} \cdot \text{s}$)		Crystal	Mobility ($\text{cm}^2/\text{V} \cdot \text{s}$)	
	Electrons	Holes		Electrons	Holes
Si	1350	480	InSb	800	450
Ge	3600	1800	InAs	30,000	450
GaAs	8000	300	InP	4500	100
GaSb	5000	1000	PbSe	1020	930
AlAs	280	—	PbTe	2500	1000
AlSb	900	400	SiC	100	10–20

9.8 HALL EFFECT AND MAGNETORESISTANCE

The relation for Hall coefficient (6.89) indicates that the resistivity of a conductor is not affected by the application of a magnetic field. However, during the experiment on Hall effect it is noticed that the resistance of the sample along the direction of the current generally increases. The fractional change in the zero-field resistivity ($\rho(0)$) is defined as magnetoresistance (MR):

$$MR = \frac{\rho(B) - \rho(0)}{\rho(0)} = \frac{\Delta\rho}{\rho(0)} \quad (9.64)$$

where $\rho(B)$ is the resistivity in the presence of the magnetic field.

It is important to clarify here that (6.89), predicting zero MR , was derived for an isotropic single band model of charge carriers. But it is common knowledge that there are two types of constant energy surfaces in real solids: the electron type and the hole type. Calculations of galvanomagnetic properties in the two-band model do predict the occurrence of magnetoresistance, putting the experimental observation on a sound footing. The thought of the two groups of charge carriers is unambiguously vindicated by the precise interpretation of the properties of semiconductors on its basis. Therefore, an exercise in the two-band model should be very much in order. We undertake below the calculation of the Hall coefficient and the magnetoresistance in this approach.

When we assume that there are two types of charge carriers (electrons and holes), their respective current densities \mathbf{j}_e and \mathbf{j}_h , in accordance with (6.86) and (6.89) are given by

$$\begin{aligned} \mathbf{E} &= \sigma_e^{-1} \mathbf{j}_e + R_e \mathbf{B} \times \mathbf{j}_e \\ \mathbf{E} &= \sigma_h^{-1} \mathbf{j}_h + R_h \mathbf{B} \times \mathbf{j}_h \end{aligned} \quad (9.65)$$

where σ and R denote the conductivity and the Hall coefficient respectively.

The solution of the Boltzmann transport equation for single band model in the presence of magnetic field gives

$$\mathbf{j} = a_1 \mathbf{E} + a_2 \mathbf{E} \times \mathbf{B} + a_3 (\mathbf{B} \cdot \mathbf{E}) \mathbf{B} \quad (9.66)$$

where $a_1 = \frac{\sigma_0}{1 + (\omega_c \tau)^2}$, $a_2 = \frac{\sigma_0 \left(\frac{\omega_c \tau}{B} \right)}{1 + (\omega_c \tau)^2}$ and $a_3 = \frac{\sigma_0 \left(\frac{\omega_c \tau}{B} \right)^2}{1 + (\omega_c \tau)^2}$ with σ_0 as the conductivity in zero-field.

In the experiment of Hall effect, \mathbf{E} is perpendicular to \mathbf{B} . Therefore, the total current density in the two-band model according to (9.66) is,

$$\begin{aligned} \mathbf{j} &= \mathbf{j}_e + \mathbf{j}_h \\ &= \left(\frac{\sigma_0(e)}{1 + (\omega_c \tau_e)^2} + \frac{\sigma_0(h)}{1 + (\omega_c \tau_h)^2} \right) \mathbf{E} + \left(\frac{R_e \sigma_0^2(e)}{1 + (\omega_c \tau_e)^2} + \frac{R_h \sigma_0^2(h)}{1 + (\omega_c \tau_h)^2} \right) \mathbf{E} \times \mathbf{B} \end{aligned} \quad (9.67)$$

To simplify calculations, we assume that electrons (with density, n) and holes (with density, p) occupy two separate isotropic energy surfaces, have the same effective mass, m^* and same relaxation time, τ but have equal and opposite cyclotron frequencies $\mp \omega_c$. With these considerations, the current density has the form:

$$\mathbf{j} = \frac{\sigma_0}{1 + (\omega_c \tau)^2} \mathbf{E} + \left(\frac{p - n}{p + n} \right) \frac{\sigma_0 \left(\frac{\omega_c \tau}{B} \right)}{1 + (\omega_c \tau)^2} \mathbf{E} \times \mathbf{B} \quad (9.68)$$

with $\sigma_0 = \frac{(n + p) e^2 \tau}{m^*}$, the conductivity in absence of the magnetic field.

Calculation of Hall coefficient

The Hall coefficient (R_H) can be obtained by inverting Equation (9.68) and expressing \mathbf{E} in form (6.86). Comparing the equation thus obtained with (6.86) and using (6.89), one finds that the coefficient of $(\mathbf{B} \times \mathbf{j})$ in the equation is equal to the Hall coefficient having the form:

$$R_H = \frac{(p - n)(1 + (\omega_c \tau)^2)}{e[(p + n)^2 + (p - n)^2(\omega_c \tau)^2]} \quad (9.69)$$

It gives $R_H = 0$ when $n = p$. In low magnetic fields for closed orbits $\omega_c \tau \ll 1$ and in this limit, the value of the Hall coefficient is given by,

$$R_H = \frac{(p - n)}{(p + n)^2 e} = \frac{R_e \sigma_0^2(e) + R_h \sigma_0^2(h)}{(\sigma_0(e) + \sigma_0(h))^2} = \frac{p \mu_p^2 - n \mu_n^2}{e(p \mu_p + n \mu_n)^2} \quad (9.70)$$

The sign of R_H is determined by the competing opposite signs for electrons and holes. The sign determined experimentally is consistent with it. The occurrence of Hall effect in samples of n -type and p -type semiconductors is illustrated in Fig. 9.14. The scene in a real crystal is the one of the superimpositions of the two pictures shown in the figure.

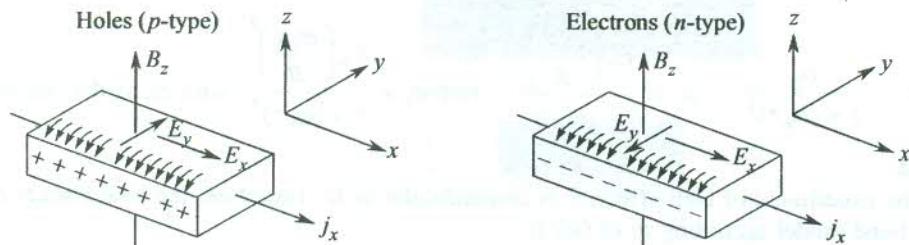


FIG. 9.14 Directions of the applied electric and magnetic fields (\mathbf{E} and \mathbf{B} respectively) relative to the direction of current flow in two samples of a semiconductor. The current density j is along the x -direction. Note that the polarity of the Hall voltage is opposite in p - and n -type samples.

Calculation of magnetoresistance

The magnetoresistance behaves as a tensor quantity. Only two components of MR are of special interest: the longitudinal (\mathbf{j} parallel to \mathbf{B}) and the transverse (\mathbf{j} perpendicular to \mathbf{B}). For isotropic two-band energy surfaces, the longitudinal component vanishes. In order to calculate the transverse MR, we are required to determine the component of \mathbf{E} along \mathbf{j} . Then, the resistivity in the presence of a magnetic field is expressible as

$$\rho(B) = \frac{\mathbf{j} \cdot \mathbf{E}}{|\mathbf{j}|^2} = \frac{1 + (\omega_c \tau)^2}{\left[1 + \left(\frac{p - n}{p + n} \right)^2 (\omega_c \tau)^2 \right]} \quad (9.71)$$

This gives

$$MR = \frac{\rho(B) - \rho(0)}{\rho(0)} = \frac{4(\omega_c \tau)^2 np}{(n + p)^2 + (n - p)^2 (\omega_c \tau)^2} \quad (9.72a)$$

which for an intrinsic material ($n = p$) will have the simple form

$$MR = \omega_c^2 \tau^2 \quad (9.72b)$$

The above pair of relations demonstrate that the magnetoresistance is a positive number. We also observe that it is proportional to B^2 (since $\omega_c \propto B$) in intrinsic materials. On increasing the magnetic field (increasing ω_c), the MR for closed orbits first increases gradually, then more slowly and finally saturates. When $n \neq p$, the approximate value at saturation is given by

$$MR = \frac{4np}{(n - p)^2} \quad (9.73)$$

One of the mysterious aspects of *MR* is the evidence of a healthy longitudinal *MR* in many semiconductors. Ideally, it should be zero in an isotropic two-band model. Its origin is attributed to the anisotropy in effective mass tensor or simply to the non-spherical shape of Fermi surfaces. As stated in Chapter 8, the measurement of *MR* is useful in deriving information about the shape of Fermi surfaces. If there are several types of charge carriers, the technique, however, fails to provide any worthwhile information.

9.9 THE *p-n* JUNCTION

The comparative ease with which the electrical characteristics of semiconductors can be monitored by controlling the traces of impurities being doped is simply striking. Because of this property, semiconductors have emerged as the indispensable material base for the development of solid state electronics. The dramatic and extensive technological consequences of this property have given a tremendous boost to the commercial interest in semiconductors.

It is of further significance to note that semiconductor devices generally exploit the characteristics of inhomogeneous semiconductors in which the donor and acceptor concentrations are not uniform. A common form in which they are used contains two separate *n* and *p* type regions with an abrupt partition boundary (a junction) within a single crystalline sample. This junction is known as the *p-n* junction. There can be more than one junction in a device. Characteristics of the *p-n* junction are crucial to the fabrication and working of most of the semiconductor devices. It is an essential component of the present electronics circuitry ranging from a simple rectifier circuit to transistor circuits and integrated circuits used in sophisticated appliances like modern computer. On the other hand, optical applications such as photo cells, LEDs and injection lasers (laser diodes) are equally important today in view of the rapidly upcoming field of optoelectronics. A proper description of all of these is beyond the scope of this book. Nevertheless, salient features of the working of a selected few devices will be presented in Section 9.10. This section is devoted to the description of the theoretical basis for the conduction behaviour of a *p-n* junction.

Consider a *p-n* junction, for example of a Si crystal whose left half is *p*-type and the right side is *n*-type. A *p-n* junction is never made by simply joining the two types of a material. It is assumed that the reader has been exposed to the methods of forming *p-n* junctions through a course in electronics. Nonetheless, the supposition that a piece of *n*-type and a piece of *p*-type material are brought into intimate contact helps in determining the electrostatic conditions present at a junction. When the two regions are treated as isolated, the levels of Fermi energy in the two regions are at different positions on the common energy scale [Fig. 9.15(a)]. Since the two regions are parts of the same crystal, the Fermi energy (representing the electrochemical potential in semiconductors) must have the same value in both halves in the state of thermal equilibrium. Near $T \approx 0$ K the position of the Fermi level is near the acceptor level in the *p*-region and near the donor level in the *n*-region. Because the impurity levels are at the extremes of the gap, the Fermi level would not have maintained a constant level unless the bending of conduction and valence band edges occurred in the transition zone. [Fig. 9.15(b)]. Immediately after the junction comes into existence, electrons from the *n*-region begin to diffuse into the *p*-region and combine with the holes present in the vicinity of acceptors. This leaves behind the positively charged donors in the *n*-region and produces negatively charged acceptors in the *p*-region. The immobile ionized impurities are considered to form a charged double layer of space charges across the jump in the doping profile [Fig. 9.15(c)]. The charged donors and acceptors create an electric field across the transition region whose direction is such as to oppose the diffusion of free charge carriers. Therefore, in the absence of any external field the diffusion

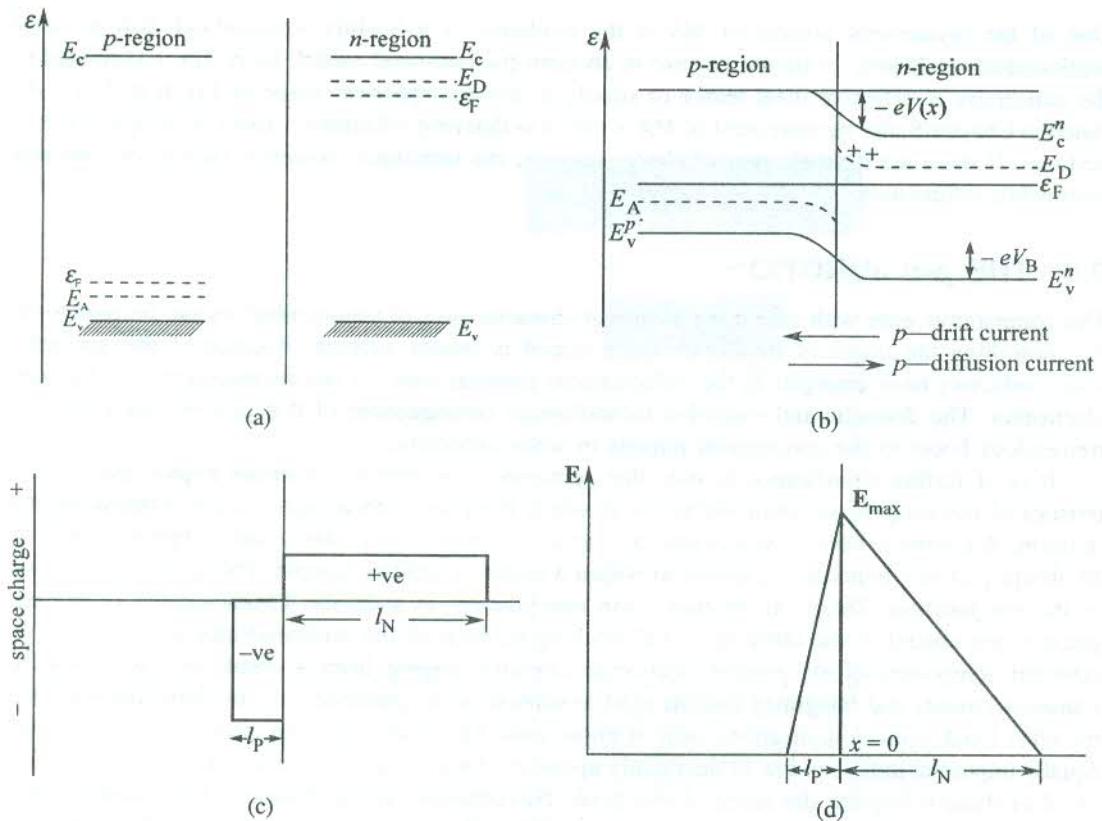


FIG. 9.15 (a) Relative positions of the conduction and valence band edges, the donor and acceptor levels; and the Fermi level in a *p*-*n* junction in the state of thermal equilibrium when *p*- and *n*-type regions are treated as isolated. (b) All the above positions when the *p*- and *n*-type regions are brought in contact. (c) The space charge distribution at a *p*-*n* junction. When the effect of the mobile carriers is neglected, i.e. $l_N N_D = l_p N_A$. (d) Variation of the space charge electric field E across the junction plane ($x = 0$).

current stops as soon as the field of ionized impurities becomes big enough just to keep the diffusion in complete check. The electric field in question is basically interpreted as the consequence of a macroscopic potential $V(x)$ varying over the transition region. In a one-dimensional model with an abrupt change from *p*-type to *n*-type at $x = 0$, the potential $V(x)$ is related to the space charge by the Poisson equation,

$$\frac{d^2V}{dx^2} = \frac{-e}{\epsilon_0 \epsilon_s} (N_D^+ + p_n - n_n) \quad \text{for } 0 < x < l_N \text{ (n-type region)} \quad (9.74)$$

$$\frac{d^2V}{dx^2} = \frac{e}{\epsilon_0 \epsilon_s} (N_A^- + n_p - p_p) \quad \text{for } -l_p < x < 0 \text{ (p-type region)}$$

where the subscripts *n* and *p* refer to concentrations in the *n*- and *p*-type regions, respectively, and $-l_p < x < l_N$ defines the space charge region.

The exact solution of (9.74) is almost impossible, since the carrier concentrations depend on position. To make the solution feasible, we simplify (9.74) by assuming that the electric field is strong enough to keep all mobile free carriers away from the space charge region. The variation of electric field \mathbf{E} across the junction is shown in Fig. 9.15(d). It is maximum at $x = 0$. Thus the density of space charge is given by simply N_D^+ in the n -region and by N_A^- in the p -region. This approximation is good except at the boundary of the space charge region for a small current across the junction. In this assumption the space charge region is depleted of free carriers and, therefore, is also identified as the *depletion layer*. The relation (9.74) is solved under the boundary conditions:

$$\frac{dV}{dx} = 0 \text{ at } x = l_N \quad \text{and} \quad x = -l_P$$

with $\frac{dV}{dx}$ continuous at $x = 0$ and $V(l_N) - V(-l_P) = V_B$, where V_B is the height of the potential step, known as the *barrier height*. It is also called the *diffusion voltage*.

Well outside the space charge zone, N_D^+ and N_A^- are compensated by equally large free carrier concentrations n_n and p_p , respectively. In accordance with the type of doping, electrons serve as majority carriers in the n -type region and holes as majority carriers in the p -type region. There may always be present small concentration of holes in the n -region (p_n) and of electrons in p -region (n_p), contributed by the naturally occurring impurities in the host crystal.

An estimate of V_B is made in terms of carrier concentrations in the state of thermal equilibrium. These concentrations are given by (9.39a) and (9.40a) which we rewrite below:

$$n_n = N(c) \exp\left[\frac{-(E_c^n - \epsilon_F)}{k_B T}\right] \quad (9.75a)$$

$$p_p = N(v) \exp\left[\frac{-(\epsilon_F - E_v^p)}{k_B T}\right] \quad (9.75b)$$

Their relationship with the intrinsic concentration n_i is then given by

$$n_i^2 = n_n p_n = N(v) N(c) \exp\left[\frac{-(E_c^n - E_v^n)}{k_B T}\right] \quad (9.76)$$

Here E_c^n , E_v^n , E_v^p are conduction band edge (in the n -region) and valence band edges (in the n - and p -regions), respectively [see Fig. 9.15(b)].

From (9.75b), we have

$$E_v^p - \epsilon_F = k_B T \ln\left(\frac{p_p}{N(v)}\right) \quad (9.77)$$

Analogically,

$$E_v^n - \epsilon_F = k_B T \ln\left(\frac{p_n}{N(v)}\right) \quad (9.78)$$

Subtracting (9.78) from (9.77) and using (9.76), we get

$$E_v^P - E_v^n = k_B T \ln \left(\frac{p_p n_n}{n_i^2} \right) \quad (9.79)$$

But V_B is equal to the difference between the maximum and minimum of the macroscopic potential $V(x)$. Hence,

$$eV_B = -(E_v^n - E_v^P) = k_B T \ln \left(\frac{p_p n_n}{n_i^2} \right) \quad [\text{using (9.79)}]$$

or

$$V_B = \frac{k_B T}{e} \ln \left(\frac{p_p n_n}{n_i^2} \right) \quad (9.80)$$

Other important properties of a $p-n$ junction are the maximum value of the electric field E_x and the space charge width $l (= l_N + l_P)$. The expressions for them can be obtained from (9.74).

When an external steady voltage is applied across a $p-n$ junction, the barrier step height changes. In a forward-biased configuration[†] it decreases to value,

$$V_{BF} = V_B - V_{ext} \quad (9.81)$$

where V_{ext} denotes the applied voltage.

The ‘flow of current changes’ in the band edges and the Fermi level are depicted in Fig. 9.16(a). On the other hand, for a reverse-biased connection in contrast to the forward-biased case, the

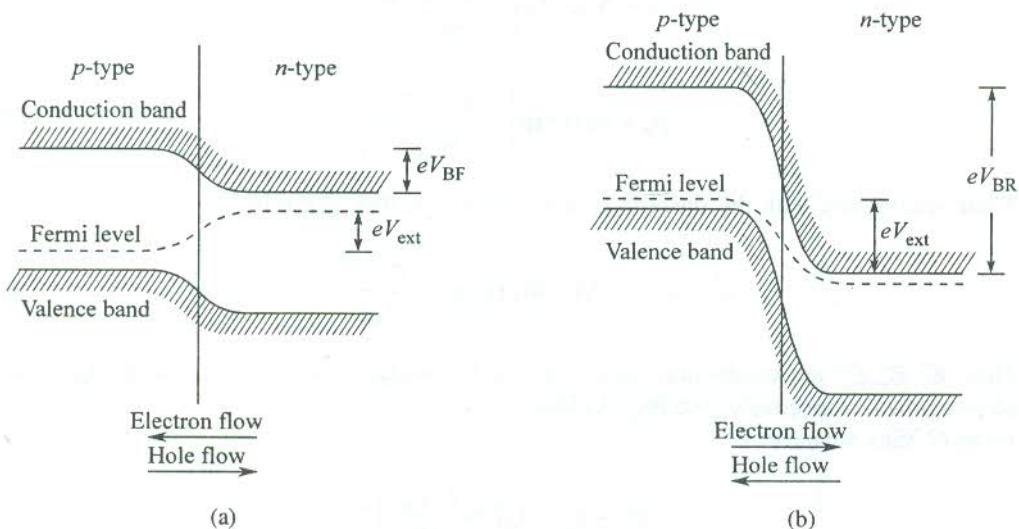


FIG. 9.16 Qualitative behaviour of the Fermi level and the electron potential energy at points near the $p-n$ junction when (a) a forward bias of V_{ext} volt is applied, (b) a reverse bias of V_{ext} volt is applied. Note the direction of electron and hole currents in the two cases.

[†] *Forward biased.* The free end of the p -region is connected to the positive terminal and that of the n -region to the negative terminal of the voltage source.

Reverse biased. The polarities of the p - and n -free ends are opposite to those in the forward-biased arrangement.

direction of the applied field does not favour the flow of majority carriers across the junction, thus increasing the barrier height to

$$V_{BR} = V_B + V_{ext} \quad (9.82)$$

The electronic conduction and changes in the band scheme structure are illustrated in Fig. 9.16(b).

In the forward-biased configuration the current is carried by majority carriers and is of the order of a few mA which is much higher than that in the reverse-biased case ($\sim \mu\text{A}$). The current in a reverse-biased junction is low as it is contributed by the flow of minority carriers that are relatively very small in number and are required to cross a higher potential barrier. The property that a *p-n* junction favours the flow of current in one direction across the junction forms the basis for using it as a rectifier. The knowledge of voltage-current characteristics of a *p-n* junction is crucial in order to understand the behaviour of the *p-n* junction for the desired application. The case of a lightly doped *p-n* junction is the subject matter of an elementary course in basic electronics. Hence it is omitted from here for brevity. An important device employing a heavily doped *p-n* junction is, however, treated below in Section 9.10.

9.10 EXAMPLES OF *p-n* JUNCTION-BASED DEVICES

Several excellent books are available on semiconductor devices. A couple of them considered suitable to meet our requirement are cited at the end of this chapter. It may again be reminded that there are numerous devices that make use of homogeneous semiconductors. Perhaps, the simplest of them is a thermistor which is used as a standard temperature measuring device. On the other hand, the *p-n* junction based devices have acquired such a tremendous technological potential that even their listing is quite a formidable task. As examples for discussion, our choice falls on (i) tunnel diode and (ii) injection laser, on the merit of their role in the development of modern electronics and photonics.

9.10.1 The Tunnel Diode

Tunnelling is a well-founded concept in quantum mechanics. The electron tunnelling at a *p-n* junction implies that electrons reach the other side of the junction by penetrating through the potential wall at the junction. It should not be confused with the act of crossing or jumping over the potential hill. It is purely a quantum statistical phenomenon based on the concept that the electron has a finite probability of being found on either side of the junction.

For tunnelling to occur, the width of the space charge layer (or the depletion layer) should be of the order of, or less than, the electron mean free path. The width of the space charge can be reduced to this value by doping both regions of the *p-n* junction heavily. A *p-n* junction that is doped heavily enough to make the tunnel current greater than the usual diffusion current under certain conditions is called a *tunnel diode*. The width of the space charge layer in these diodes is usually less than 100 Å. The impurity concentrations are in the range $10^{19}\text{--}10^{20} \text{ cm}^{-3}$, whereas in ordinary *p-n* junctions they vary between 10^{14} and 10^{18} cm^{-3} .

The material of a tunnel diode behaves as a degenerate semiconductor. The Fermi level ϵ_F no more lies in the gap. Instead, it lies within the valence band of the *p*-region and the conduction band of the *n*-region. We say that the Fermi level lies in a hybrid impurity-intrinsic band. This follows from arguments advanced in Section 9.6. Though a small width for the space charge layer is essential, it is not a sufficient condition for tunnelling. There must be an unoccupied state on the other side of the junction into which an electron could tunnel at constant energy.

Figure 9.17(a) shows the state of an unbiased junction where the Fermi level has the same energy in both the regions of the junction. Let us first examine tunnelling in a reverse-biased junction as shown in Fig. 9.17(b). When a small reverse bias voltage V_{ext} is applied, the height of the barrier step becomes much more than V_B . On account of this, the Fermi level has different values in the regions and at large values of V_{ext} a large number of occupied states of the valence band on the p -side of the depletion region lie opposite even to a larger number of empty states of the conduction band on n -side. The band picture is similar to that in a metal. When l_T is of the order of the electron wavelength or less, the junction has ideal conditions for electron tunnelling. At higher values of V_{ext} the tunnel current can be quite large because of the enhanced level of overlap between the occupied and empty states as referred above. The probability of tunnelling is given as

$$P_t \sim \exp(-kl_T) \quad (9.83)$$

where k is the electron wavevector.

For a simple band structure of geometry as shown in Fig. 9.17(b),

$$l_T \simeq \frac{E_g}{eE} \quad (9.84)$$

Substituting this value of l_T in (9.83), we get

$$P_t \simeq \exp \left[\frac{-E_g(2m^*\epsilon)^{1/2}}{eE\hbar} \right] \quad (9.85)$$

Here E is an average value of the electric field at the junction which generally does not deviate much from E_{max} . Owing to the tunnel current, there is a quick onset of a critical reverse voltage. This property enables a reverse-biased tunnel diode to be used as a voltage regulator. These tunnel diodes are known as Zener diodes. The relation (9.85) provides a good description of tunnelling in a Zener diode.

Now, we proceed to describe the other important tunnel diode that is used in the forward-biased configuration [Fig. 9.17(c)]. This is again a heavily doped p - n junction, known as the Esaki diode. If one approaches the forward-biased state from the reverse-biased condition, by decreasing V_{ext} , the tunnel current is on decrease and becomes zero at $V_{\text{ext}} = 0$. The decrease in current occurs because of the reducing level of overlap at lower values of V_{ext} between the regions of occupied states in the valence band of the p -side and the unoccupied states in the conduction band of the n -side [Fig. 9.17(b)]. When unbiased ($V_{\text{ext}} = 0$) there is no current at $T = 0$. At non-zero temperatures in this condition almost an equal small number of electrons tunnel from both sides of the junction keeping the current at zero value. The V - I characteristics of a tunnel diode are drawn in Fig. 9.17(e). The portion AO refers to a reverse biased junction.

The characteristic curve beyond the point O describes the V - I relationship in an Esaki diode. As the forward voltage grows from zero to a certain value, the current keeps on increasing and approaches a maximum value. During this rise in the applied voltage, the Fermi level ϵ_F^p dips towards the conduction band edge of the n -side (E_c^n) and ϵ_F^n rises towards the valence band edge on p -side (E_v^p) as shown in Fig. 9.17(c). In the state of the maximum overlap of interest, the maximum tunnel current results. With a further increase in the applied voltage, the Fermi levels maintain the trend of dipping and rising in p - and n -regions, respectively. For the present band structure, the above change

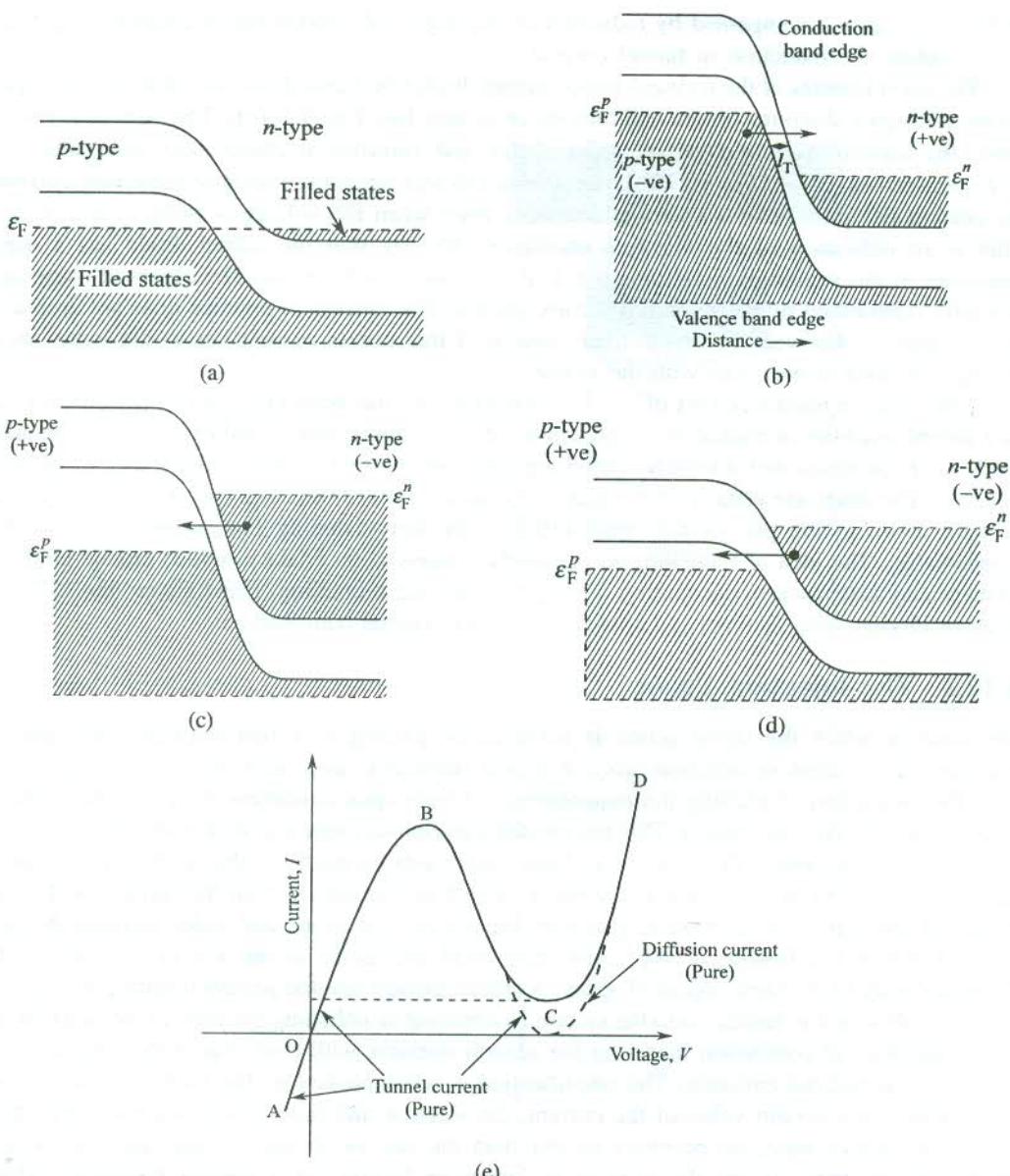


FIG. 9.17 (a) An unbiased tunnel diode in the state of thermal equilibrium. Because of heavy doping the Fermi level ϵ_F lies in the conduction band of the n-type region and in the valence band of the p-type region with a common value of energy. There is no current through the junction. (b) A reverse-biased tunnel diode with different values of Fermi energy on the two sides of the junction. A tunnel current flows across the junction because of the substantial overlap between the unoccupied states of the conduction band in the n-region and the occupied states of the valence band in the p-region. A tunnel diode in this configuration functions as a voltage regulator (the Zener diode). (c) A forward-biased tunnel diode (Esaki diode). The increase in tunnel current to a maximum on increasing the bias voltage is explained on the basis of the increasing degree of overlap between the occupied and unoccupied states as referred in case (b). (d) The decrease in tunnel current (Esaki current) on increasing the bias voltage further refers to the region of the negative resistance of the Esaki diode. The decrease in current is due to the decreasing degree of overlap of occupied and unoccupied states affected by the increasing applied voltage. (e) Voltage-current characteristics of a tunnel diode. The portion AO represents the reversed-biased state; OB—current increases when forward biased; BC—negative resistance region when forward biased; CD—purely diffusion current.

in Fermi levels is accompanied by reduction in the degree of overlap under question [Fig. 9.17(d)] and therefore the reduction in tunnel current.

The main features of the forward-biased tunnel diode (the Esaki diode) are that the tunnel current passes through a sharp maximum and then drops to zero [see Fig. 9.17(e)]. The current is maximum when the forward voltage is about $(\varepsilon_F^p + \varepsilon_F^n)/2e$ and vanishes at about twice this value. In the V - I characteristic curve [Fig. 9.17(e)], the portion OB represents the region of increasing current and the portion BC shows that the current decreases even when the voltage is being further increased. This is an indication of the negative resistance. We say that the tunnel diode has a negative resistance in this region. Beyond the point C there is no tunnel current. The observed rising current is totally contributed by the normal diffusion current. The current in practice does not drop to zero but a valley is observed. The most likely source of this excess current lies in the losses incurred during collisions of electrons with the lattice.

The negative resistance part of the V - I characteristics has been exploited to use an Esaki diode as a power amplifier or source. In conjunction with a capacitor and an inductance it can be made to work as an oscillator and a switch. Switching times are limited by high capacitance C_j of the thin junction. The times are usually of the order of nanosecond (10^{-9} s). The switching time RC_j can be brought down to the picosecond range (10^{-12} s) by heavy doping that lowers the resistance R considerably. The idea of tunnelling as originally conceived by Esaki provided crucial guidance to Giaever and Josephson in accounting for superconducting tunnelling. Esaki shared the 1973 Nobel prize in Physics with Giaever and Josephson for his original contribution.

9.10.2 The Injection Laser

The laser in which the lasing action is achieved by passing a current through a forward-biased p - n junction is called an *injection laser*. It is also referred to as a laser diode. The injection lasers have the distinction of meeting the requirements of fibre optic communication systems in the most convenient and effective manner. This has created tremendous interest in their study and development.

The principle behind the emission of light from a semiconductor is that of the recombination of electrons and holes at a p - n junction when a current is injected through the diode. As the current is passed through a forward-biased junction, the injected electrons and holes increase the density of electrons in the bottom of the conduction band and holes at the top of the valence band, simultaneously in the same region of space. A spontaneously emitted photon resulting from electron-hole recombination is feedback into the system by cleaving or polishing the ends of the junction diode. Under the state of population inversion the photon interacts with electrons in the conduction band to produce stimulated emission. The amplification is accomplished by the multiple reflection at the diode ends. At a certain value of the current, the electron and hole densities are so large that the rate of stimulated emission becomes greater than the rate of absorption and thus the process of amplification starts. When the current is increased further, at a certain threshold value the amplification exceeds the cavity losses and a coherent radiation is obtained in the output.

The threshold current density in the p - n junction based lasers is inconveniently high ($\sim 50,000$ A cm^{-2}). This is drastically reduced in practical lasers that employ a double heterojunction. A heterojunction has two different semiconductor materials on its two sides. In a double heterojunction a lasing semiconductor is sandwiched between two wider-gap semiconductors of opposite doping. In a heterojunction laser developed by Kressel and Butler*, the active layer is an undoped

* H. Kressel and J.K. Butler, *Semiconductor Lasers and Heterojunction LEDs* (Academic Press, 1977).

$\text{Ga}_{1-x}\text{Al}_x\text{As}$ ($x = 0.05 - 0.1$) which is embedded between *n*- and *p*-type layers of $\text{Ga}_{1-y}\text{Al}_y\text{As}$ ($y = 0.3 - 0.4$). The *n*- and *p*-type layers have the same band gap and the same refractive index. The band gap of the active layer is lower and characterized by a higher value of the refractive index. This ensures that the radiation emitted in the active layer is confined within this layer because of the total internal reflection occurring at its interfaces with the *n*- and *p*-type materials. Furthermore, the electrons from the *n*-side and the holes from the *p*-side that enter the active layer are not allowed by the potential barriers at the junctions to escape into the *p*- and *n*-sides (regions), respectively. This confinement of the light and the carriers results in amplification and wave guidance. For the high efficiency, materials only with the direct band gap are selected for fabrication. One of the biggest advantages with this heterostructure is that the frequency of the laser output can be finely monitored by varying the composition of the compound. The reader interested in complete details may consult the book authored by Chai Yeh,* and that by Ghatak and Thyagarajan.**

For the present heterostructure there are two quasi-Fermi levels ε_F^c and ε_F^v referred to the conduction and valence band edges. The ε_F^c in the optically active layer is on level with the ε_F^n in the *n*-region and ε_F^v with ε_F^p in the *p*-region. The population inversion is achieved by applying a forward voltage that exceeds the voltage equivalent of the band gap of the active layer. Once the electrons and holes enter the active layer from *n*- and *p*-type regions respectively, they remain confined in the active layer as explained in the preceding paragraph. The emission of radiation occurs in the plane of the active layer when an electron from the conduction band combines with a hole in the valence band as shown in Fig. 9.18. The process of stimulated emission and amplification proceeds the same way as in a *p-n* junction based device. The frequency of the emitted radiation is temperature dependent. A typical GaAs laser emits in the near infrared around 8380 Å.

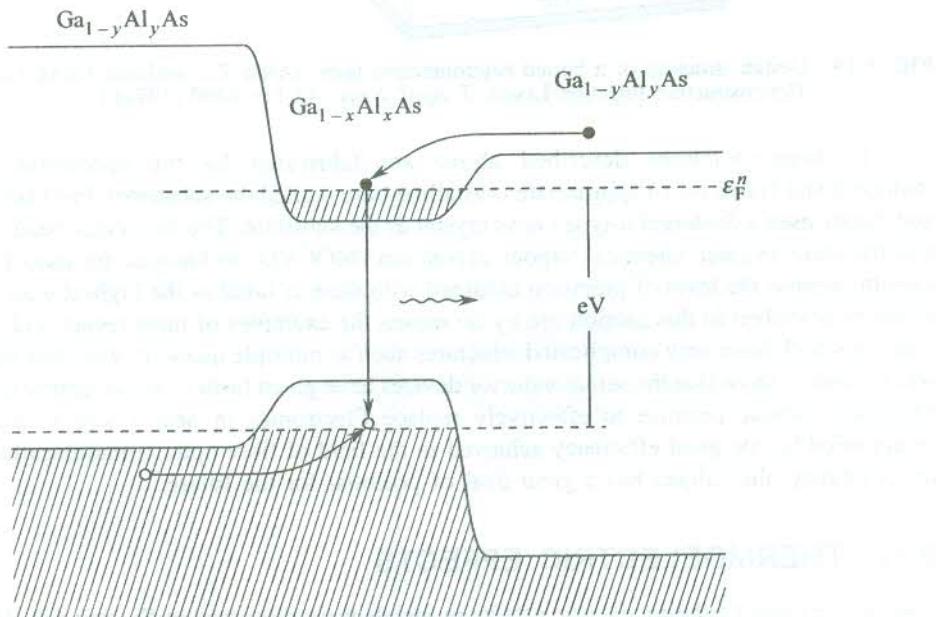


FIG. 9.18 Double heterojunction GaAs-Al laser. The optically active layer has a higher refractive index (lower band gap) with $x = 0.05 - 0.1$. The material of *p*- and *n*-type regions has a wider band gap (lower refractive index) with $y = 0.3 - 0.4$. The confinement to carriers in the active layer is provided by the potential barriers at the two junctions.

* Chai Yeh, *Handbook of Fiber Optics: Theory and Applications* (Academic Press, 1990).

** A.K. Ghatak and K. Thyagarajan, *Optical Electronics* (Cambridge, 1989).

In the above device structure the carriers and the light are confined only along one direction. The confinement can also be provided in the lateral direction by surrounding the optically active layer with higher band gap materials on all sides, leaving a window for the output. Such a heterostructure laser is known as the *buried heterostructure laser* and shown in Fig. 9.19. The required threshold current densities ($2000\text{--}4000 \text{ A cm}^{-2}$) are achieved in this structure for a significantly reduced value of current (< 50 mA), whereas for a stripe-geometry laser with a standard size cavity the threshold current shoots a little over one ampere.

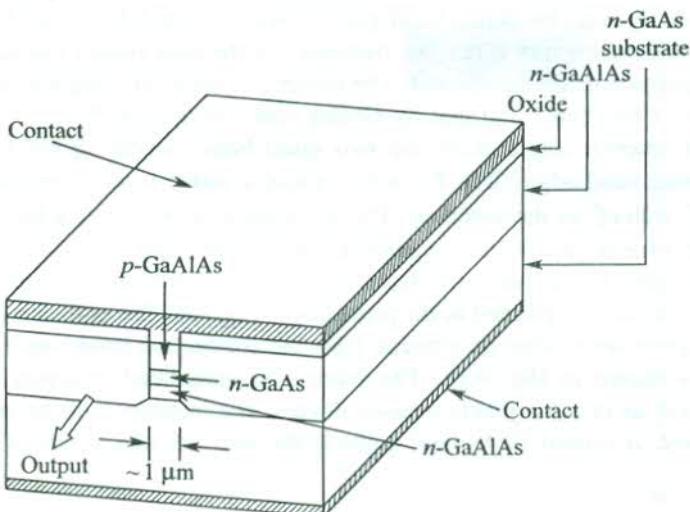


FIG. 9.19 Design structure of a buried heterojunction laser. [After T. Tasukada, GaAs-Ga_{1-x}Al_xAs Buried Heterostructure Injection Lasers, *J. Appl. Phys.*, **45**(11), 4899 (1974).]

The heterostructures described above are fabricated by the successive deposition of multiepitaxial layers on an appropriate crystalline base (called the substrate). For GaAs laser, Kressel and Butler used a Si-doped *n*-type GaAs crystal as the substrate. The molecular beam epitaxy (MBE) and the metal organic chemical vapour deposition (MOCVD) techniques are used for the epitaxial growth because the level of precision obtained with them is rated as the highest to day. The injection lasers as described in this section are by no means the examples of most recent and versatile lasers, some of which have very complicated structures such as multiple quantum well structures. Our object was limited to show that the semiconductor devices have given birth to an altogether new field (called *Photonics*) whose promise to effectively replace Electronics in near future to a large extent is exemplified by the great efficiency achieved in the field of fibre optic communication. Though still in its infancy, the subject has a great deal of potential for the future.

9.11 THERMOELECTRIC EFFECTS

Essential features of thermoelectric effects in metals were discussed in Section 6.8. The treatment is being extended here to semiconductors whose thermoelectric response is many times stronger. The three main thermoelectric parameters, namely the thermoelectric power S , the Thomson coefficient μ_T and the Peltier coefficient Π in semiconductors may be positive or negative, depending on whether

the semiconductor is of *p*-type or *n*-type. We may recall that these parameters can have only negative sign in ideal conductors. Consider a semiconductor rod whose one end is hot and the other cold. As far as the algebraic sign is concerned, the case of an *n*-type semiconductor is similar to that of ideal conductors. But electrons at the hot end in a *p*-type semiconductor occupy acceptor levels on being excited from the valence band. This enables electrons at the cold end near the top of the valence band to lower their energy by moving into holes created in the valence band at the hot end. Thus the hot end becomes negatively charged and the cold end positively charged. This situation is opposite to that described in Section 6.8. Hence in a *p*-type semiconductor the parameters S , μ_T and Π are all positive. These arguments lead to a quick method of determining whether a specimen is of *n*-type or *p*-type. All that is needed is to measure the sign of the voltage across the given specimen whose one end is hot.

Now, we first consider the above-mentioned semiconductor rod in an open circuit, ensuring that a constant temperature difference is maintained between its ends. In this condition a constant difference of potential is observed between its end. This potential difference can be interpreted to be associated with an electric field,

$$E = \frac{dV}{dx} = \frac{dV}{dT} \frac{dT}{dx} = S \frac{dT}{dx} \quad (9.86)$$

Next, we consider a semiconductor rod kept at a constant temperature in a closed circuit. If an electric current is forced into it at one end from a certain external source, the current density in an *n*-type semiconductor may be written as

$$j_e = n(-e)(-\mu_n) E = ne\mu_n E \quad (9.87)$$

where μ_n is the electron mobility in the direction opposite to that of the electric field E from the external source.

It is appropriate to measure the electron energies relative to the Fermi energy ε_F because the two materials in contact have a common Fermi level. Therefore, the average energy transported by electrons is equal to

$$(E_c - \varepsilon_F) + \frac{3}{2} k_B T \quad (9.88)$$

where E_c denotes the energy at the conduction band edge.

The heat current density j_Q associated with the electric current density j_e is given by

$$j_Q = -n \left(E_c - \varepsilon_F + \frac{3}{2} k_B T \right) \mu_n E \quad (9.89)$$

But we know from (6.94) that

$$j_Q = \Pi_e j_e$$

Therefore, the Peltier coefficient of an *n*-type semiconductor is

$$\Pi_e = - \frac{E_c - \varepsilon_F + \frac{3}{2} k_B T}{e} \quad (9.90)$$

Similarly, for a *p*-type semiconductor, we have

$$j_h = pe\mu_p E \quad (9.91)$$

$$j_Q = p \left(\epsilon_F - E_v + \frac{3}{2} k_B T \right) \mu_p E \quad (9.92)$$

and

$$\Pi_h = \frac{\epsilon_F - E_v + \frac{3}{2} k_B T}{e} \quad (9.93)$$

The Peltier coefficient can be easily determined by measuring the absolute thermoelectric power to which it is linked by the Kelvin relation (6.96), $\Pi = ST$. The Π versus T plots for *n*-type and *p*-type specimens of silicon are displayed in Fig. 9.20. The significant observation is that the specimens behave as intrinsic semiconductors above 600 K.

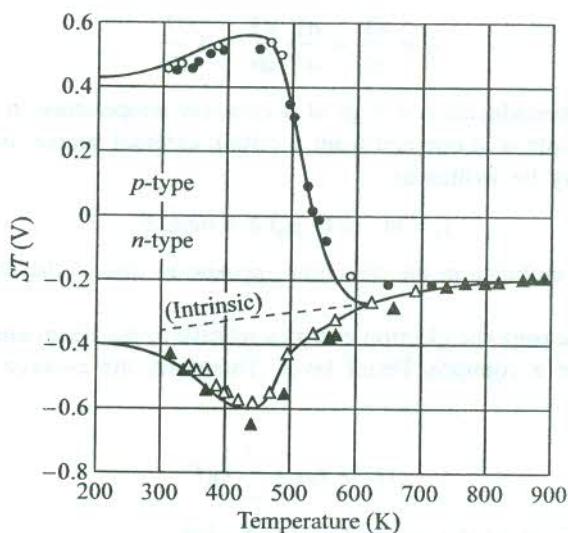


FIG. 9.20 Variation of the Peltier coefficient ($= ST$, with S as the absolute thermoelectric power) of *n*-type and *p*-type specimens of silicon as a function of temperature. The two specimens show intrinsic behaviour above 600 K. The curves are based on theoretical estimates. The points show the experimental data. [After T.H. Geballe and G.W. Hall, *Phys. Rev.*, **98**, 940 (1955).]

Being many times intense than in metals, the thermoelectric effects in semiconductors are of added interest. For example, the thermoelectric powers of semiconductors are two orders of magnitude larger than those for non-ferromagnetic metals. The reasons ascribed to this huge quantitative difference in response are mainly two-fold. Firstly, the carrier density of semiconductors is sensitive to temperature owing to which the hot end has more conduction electrons or holes per unit volume, depending on whether the material is of *p*-type or *n*-type. Secondly, and more importantly the presence of a forbidden energy gap in semiconductors proves to be a major cause of the observed behaviour.

To elaborate on the second reason we examine the quantity of Peltier heat evolved or absorbed at a metal semiconductor junction. Let a current be flown from a metal to an *n*-type semiconductor in contact forming a junction as shown in Fig. 9.21(a). In an *n*-type semiconductor the majority carriers (electrons) are in the conduction band placed well above the Fermi level ϵ_F . On the other hand, most of the electrons in a metal are near the Fermi level. Because of being in contact, the Fermi energy would be at the same level in the metal and the semiconductor. Figure 9.21(a) clearly indicates that the average energy of conduction electrons in the electric current decreases as the same electric

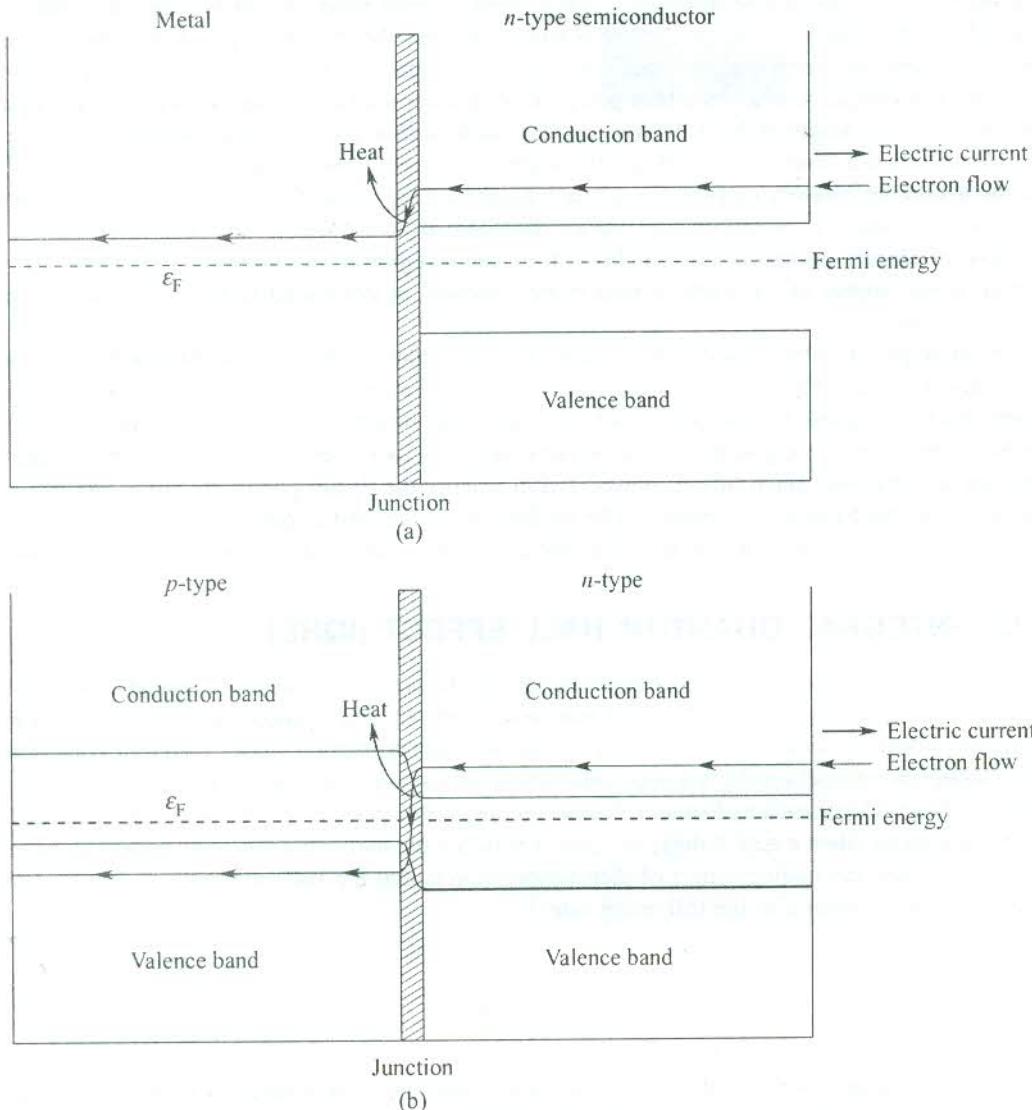


FIG. 9.21 (a) Peltier effect at a metal–semiconductor junction. Heat is evolved at the junction when an electric current flows from a metal to an *n*-type semiconductor. (b) Peltier effect at a *p-n* junction. Relatively large amount of heat is evolved at the junction (compared to that in (a)) in the case of an electric current flowing from the *p*-region to the *n*-region.

current enters the semiconductor. The difference of electron energies in the two regions is lost in the form of heat at the junction. Similarly, when a *p*-type material is in contact with a metal, heat would be absorbed for the shown direction of electric current because the majority carriers (holes) in the semiconductor would occupy energy states in the valence band that are much below the Fermi level (the approximate level of electric current in metal).

Application

It is obvious from the above discussion that a thermocouple junction can function as a heat pump or a refrigerator when an electric current is forced through the thermocouple. Under the flow of an electric current the thermocouple pumps heat from one junction to the other. We show below that a *p-n* junction can serve as a better heat pump or refrigerator than any common thermocouple junction consisting of two semimetals or even a junction such as that shown in Fig. 9.21(a).

Consider a *p-n* junction through which an electric current flows as demonstrated in Fig. 9.21(b). On the *n*-side the majority carriers are at much higher energy levels than the Fermi level whereas on the *p*-side the majority carriers occupy energy states that are much below the Fermi level. Hence, with the flow of electric current in this condition from the *n*-region to the *p*-region a relatively far more energy would appear in the form of heat at the junction, as compared to that in a junction shown in Fig. 9.21(a).

In principle, a Peltier heater can be more efficient than an electrical resistance heater. For the requirement of each kW of heat, an electrical resistance heater must consume one kW of electrical power that is dissipated. On the other hand, an ideal Peltier heater needs electrical power only to pump the heat energy from one junction (cold) to the other (hot) like a refrigerator or heat engine and thus operates at a relatively much smaller power. But in practice the Peltier pumps are not as efficient as the conventional mechanical heat pumps. The inefficiency is attributed partly to the heat loss through the thermocouples from the hot to the cold side and partly to the Joule heating of the thermocouples.

9.12 INTEGRAL QUANTUM HALL EFFECT (IQHE)

Although the treatment of Hall effect given in Section 9.8 is based on purely classical considerations, it gives a good account of the electrical transport in metals and semiconductors. But the classical magnetoconducting scenario undergoes a spectacular transformation under quantum conditions of temperature and magnetic field in a two-dimensional conductivity channel. K. von Klitzing, Dorda and Pepper* observed that such a channel is formed at the oxide interface in a metal-oxide-semiconductor (MOS) transistor when a gate voltage is applied between the metal and the semiconductor, as shown in Fig. 9.22. The fascinating aspect of their observation is that the Hall resistance ρ_H varies with the magnetic field according to the following rule:

$$\rho_H = \frac{h}{ie^2} \quad (9.94)$$

where i is an integer ($= 1, 2, 3, \dots$).

The phenomenon expressed by this rule, where the Hall conductance is quantized in units of e^2/h , is called the *Integral Quantum Hall Effect* (IQHE).

* K. von Klitzing, G. Dorda and M. Pepper, *Phys. Rev. Lett.*, **45**, 494 (1980).

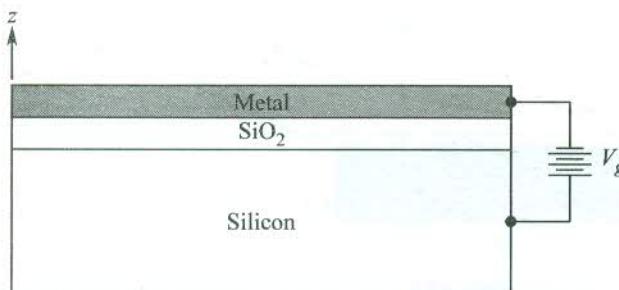


FIG. 9.22 A metal-oxide-semiconductor (MOS) transistor. An accumulation of charge carriers occurs at the oxide interface on the application of a gate voltage between the metal and the semiconductor. The oxide interface (the xy -plane) behaves as a two-dimensional conductivity channel.

In the original experiment a constant current of $1 \mu\text{A}$ was forced to flow between the source and the drain in the presence of a magnetic field of 18 tesla at 1.5 K . The results of this experiment are graphically displayed in Fig. 9.23 and the experimental geometry is illustrated in Fig. 9.24(a). The Hall field versus magnetic field plot in this case is not a straight line (as in the classical Hall effect). Instead, it shows plateaus with steps in between at certain values of the magnetic field [see Fig. 9.24(b)].

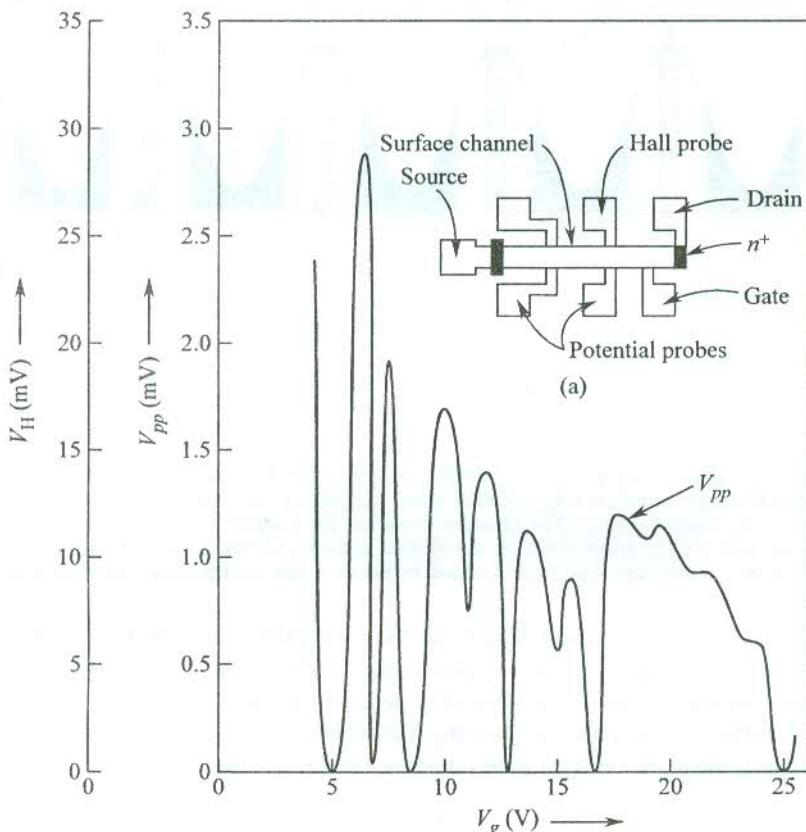


FIG. 9.23 The IQHE voltage V_{pp} and the Hall voltage V_H as functions of the gate voltage V_g . The MOS transistor set-up for the IQHE measurements is shown above the plots.

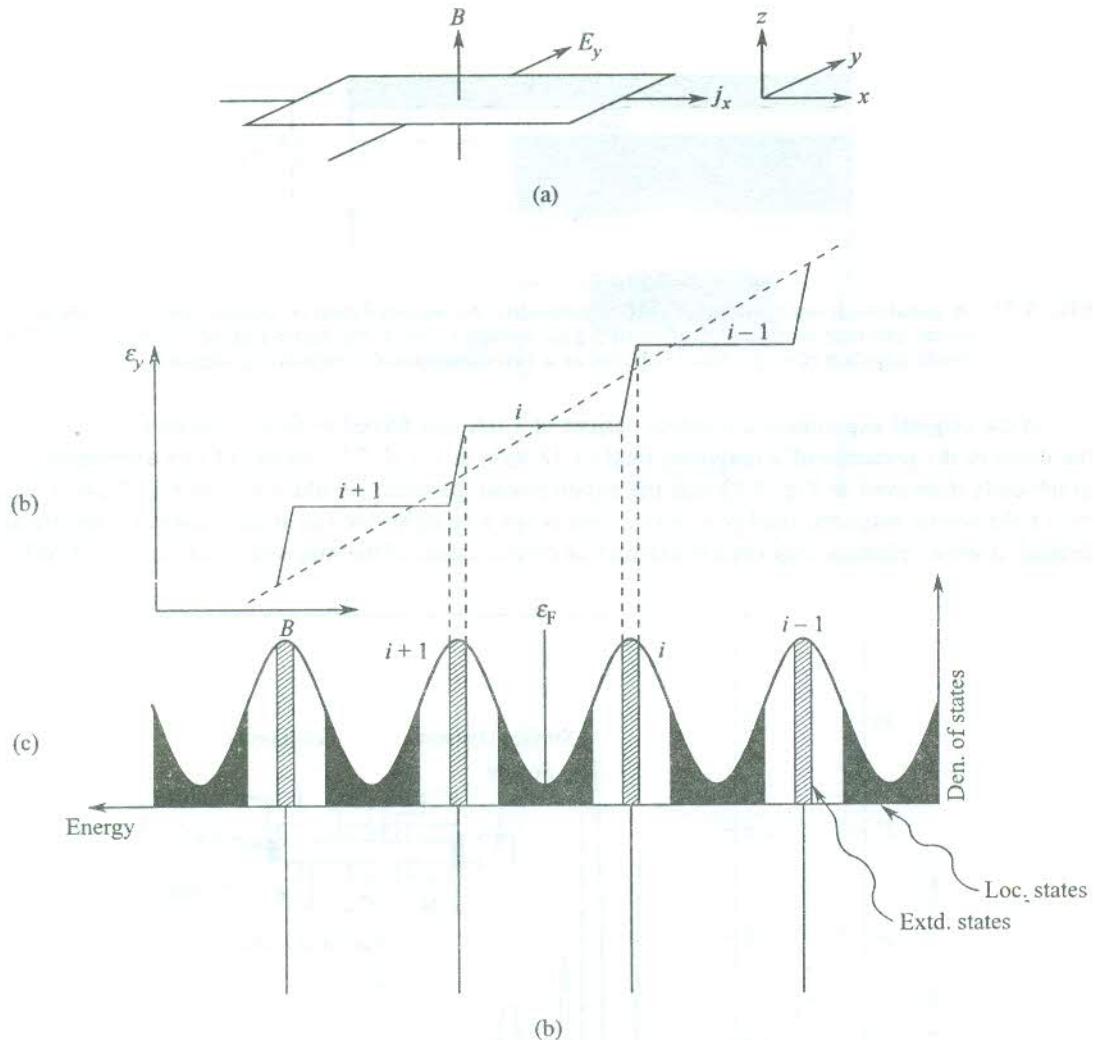


FIG. 9.24 (a) The IQHE geometry. (b) The IQHE plateaus in the Hall field for a fixed system current j_x . The dotted line represents the classical Hall effect. (c) Density of states in a two-dimensional real system in a strong magnetic field. The disorder broadens the Landau levels with the extended states in the middle and the localized states in the shaded regions, known as the mobility gaps. Vertical lines shown below indicate the sharp Landau levels in a two-dimensional ideal system.

First we show how the observed minima in the longitudinal voltage V_{pp} can be explained and relation (9.94) be derived, using only a crude model. Then, we give a qualitative analysis of real two-dimensional systems in the framework of a general theory.

Consider a surface current density \mathbf{j}_x in the x -direction defined as the current crossing a line of unit length in the y -direction on the oxide interface (the xy -plane):

$$I_x = j_x L_y \quad (9.95)$$

If n is the total electron density per unit area of the xy -plane,

$$j_x = nev_d \quad (9.96)$$

with v_d as the drift velocity of electrons in the x -direction.

In an alternative description of the Hall effect, the current in a specimen with mobile charge carriers is produced when the specimen is placed in a region of mutually crossed (perpendicular) electric and magnetic fields. The flow of current in a direction orthogonal to both the fields is detected on closing the circuit. If the electric field (E_y) and the magnetic field \mathbf{B} act along y - and z -directions, respectively

$$\mathbf{j}_x = \sigma_{xx}\mathbf{E}_x + \sigma_{xy}\mathbf{E}_y = \sigma_{xy}\mathbf{E}_y \quad (9.97)$$

(since $\mathbf{E}_x = 0$)

and

$$v_d = \frac{E_y}{B} \quad (9.98)$$

Here σ_{xy} denotes the conductivity tensor in the plane of the two-dimensional channel. Interpreting the resistivity ρ_{xy} as the Hall resistance ρ_H ,

$$\rho_H = \frac{V_H}{I_x} = \frac{E_y L_y}{j_x L_y} = \frac{B}{ne} \quad (9.99)$$

As follows from relation (9.95), ρ_H represents the resistance of a channel of unit thickness. In the commonly used geometry for the Hall effect, a current is flown along the x -direction in the presence of a magnetic field along the z -direction and the Hall voltage V_H measured along the y -direction in the specimen. The description contained in relation (9.99) is consistent with this experimental geometry.

It is now required to exploit relation (8.38) which states that the areas of the successive electron orbits in the k -space in the presence of a magnetic field B differ by an amount $2\pi eB/\hbar$. Considering a square of side L on the xy -plane, the number of states in this area are estimated as

$$\left(\frac{2\pi eB}{\hbar}\right) \left(\frac{L}{2\pi}\right)^2 = \frac{eL^2 B}{h} \quad (9.100)$$

It gives the number of electron levels that coalesce into a single magnetic level as soon as even a small magnetic field is switched on. This magnetic level defines a Landau level whose energies in the xy -plane are quantized as $(i + 1/2)\hbar\omega_c$ with $\omega_c = eB/m^*$ (the cyclotron frequency). Relation (9.100) gives essentially the measure of degeneracy of a Landau level. In the present context we define the degeneracy per unit area of the xy -plane as

$$D(B) = \frac{eB}{h} \quad (9.101)$$

Let the applied magnetic field be so strong that $\hbar\omega_c \gg k_B T$. It is then reasonable to talk in terms of completely filled or completely empty Landau levels. Suppose B_i is the critical field at which no level is partly filled and i is the magnetic quantum number of the highest occupied level. When the

electron density on the oxide interface is adjusted by varying the gate voltage so that the Fermi level coincides with level i ,

$$n = i \cdot D(B_i) \quad (9.102)$$

Under these conditions electrons can undergo neither elastic nor inelastic collisions. The elastic collisions would involve scattering of electrons from one state to the other in the same Landau level. But this is not permitted by the Pauli principle, since all possible final states of equal energy are occupied. The inelastic collisions can be possible with the scattering of electrons to a vacant Landau level by absorbing the required energy from some source, most likely phonons. But in the experimental conditions of low temperature and $\hbar\omega_c \gg k_B T$ as established here, there are hardly any phonons whose energy could compare with the large energy interval $\hbar\omega_c$. Therefore, the inelastic collisions are too ruled out and the electron mean free paths are greatly enhanced. This results in the occurrence of the voltage minima in V_{pp} (or the longitudinal resistance minima).

Placing the value of $D(B)$ from (9.101) in (9.102), we obtain (9.94):

$$\rho_H = \frac{h}{ie^2} \text{ (in ohms)}$$

Analysis in real systems

The experimental evidence for the Hall resistance being accurately quantized at h/e^2 ohms might apparently suggest that IQHE is independent of purity and crystallinity, simply because the theory predicting this quantization does not take these aspects into consideration. But the presence of impurities or microcrystallinity produces disorder, rendering the crystal potential irregular as a result of which the sharp levels in ideal systems are broadened into bands in real systems [Fig. 9.24(c)]. This affects the Hall resistance to such an extent that its linear variation changes to develop plateaus. In a two-dimensional system that concerns the present discussion all the electron states are predicted as localized at any disorder.* To the credit of this prediction the IQHE actually approaches this limit as the magnetic field goes to zero. Therefore, it is logical to believe that there can exist both the extended and localized types of carrier states in a band. As per the latest concept of localization, the extended and localized states cannot coexist at the same energy. The localized states occupy the region of the lowest density of states forming the mobility gap and do not contribute to the flow of electronic current. The extended states, on the other hand, appear around the peaks of the density of states [see Fig. 9.24(c)].

In light of his thought experiment on a two-dimensional system, Laughlin** has interpreted the IQHE in real systems as a consequence of the principle of gauge invariance. By analogy with the flux quantization in a superconductor (where the unit of charge is $2e$), the flux quantization in the IQHE (with e as the unit of charge) is discussed. For a certain increase δB in the magnetic field, there is an addition of one flux quantum that enhances the degeneracy of each Landau level by one. Suppose i denotes the magnetic quantum number of the highest completely filled Landau level. If all electrons cannot be accommodated up to this level at absolute zero, the Fermi level ϵ_F will coincide with the level $(i + 1)$ which is only partially filled.[†]

* E. Abraham, P.W. Anderson, D.C. Licciardello and T.V. Ramakrishnan, *Phys. Rev. Lett.*, **42**, 673 (1979).

** R.B. Laughlin, *Phys. Rev.*, **B23**, 5632 (1981).

† The statement is valid for temperatures (~ 1 K) at which the IQHE experiment is conducted.

With an increase of one flux quantum, each of the Landau levels will have one additional level in its subband. Consequently, the level $(i + 1)$ will be vacated, since the electrons in this level can now be accommodated in the newly created lower energy levels (i in total). Over a fixed range δB (equivalent to one flux quantum), during which the Fermi level remains in a level, a plateau in the Hall resistance is observed. The increase δB equals a magnitude for which at the stage under consideration all the states in the level $(i + 1)$ are vacated and the Fermi level drops to coincide with the level i . Thus for each increase of δB , a jump to the next plateau takes place. The next plateau refers to a level with the next lower magnetic quantum number [see Fig. 9.24(c)].

It is important to assert the role of disorder and localization in the IQHE for real systems. Given next is a brief discussion conducted in this approach.

The Hall electric field in an ideal system is

$$E_y(B) = j_x B_z R_H = v_d B_z = \frac{D(B) v_d h}{e} \quad (9.103)$$

where R_H denotes the Hall coefficient.

In a disordered system the degeneracy of states and the electron densities may be split as

$$D(B) = D^E(B) + D^L(B) \quad (9.104)$$

$$n = n^E(B) + n^L(B) \quad (9.105)$$

where E and L refer to the extended and localized states, respectively.

Then, the Hall field in a disordered system may be expressed as

$$\varepsilon_y(B) = \frac{D^E(B) v_d(B) h}{e} \quad (9.106)$$

where $v_d(B)$ stands for the drift velocity in the disordered system.

As observed in a typical IQHE experiment, the current density j_x carried by electrons in the extended states must remain unchanged at the value $n e v_d$. Hence,

$$v_d(B) = \left(\frac{n}{n^E(B)} \right) v_d \quad (9.107)$$

It implies that n^E electrons per unit area carry the current (of density $n e v_d$) with a higher drift velocity to compensate for the loss of current because of the localization of n^L electrons per unit area.

The $D^E(B)$ in a particular band always increases with B , though non-monotonically. It increases by one, only when a δB -increase of B creates an extended state in this band. We can check that this occurs with probability $1/(v + 1)$, where

$$v = \frac{D^L(B)}{D^E(B)}$$

From (9.98) we see that $\varepsilon_y(B)$ can remain unchanged (forming a plateau) with B whenever $v_d(B)$ decreases, provided

$$D^E(B) v_d(B) = \text{a constant} \quad (9.108)$$

But it should be observed that $v_d(B)$ increases as well as decreases with B depending on where the Fermi level ϵ_F is located. When ϵ_F lies in the mobility gap, we will have either (i) $v_d(B + \delta B) < v_d(B)$, when an extended state is produced and D^E increases in the subbands below ϵ_F or (ii) $v_d(B + \delta B) = v_d(B)$, in between the above events where the Hall field $\epsilon_y(B)$ remains unchanged and the plateau occurs.

On the other hand, when ϵ_F falls in the band of extended states we always have $v_d(B + \delta B) > v_d(B)$ because D^E decreases on account of the downward movement of ϵ_F . For a complete treatment, the reader is referred to calculations made elsewhere.* The calculations show that a plateau in the Hall field $\epsilon_y(B)$ is formed within an accuracy of a few parts in $10^6 D(B)/(v + 1)$.

In the extreme quantum conditions, i.e. at extremely low temperatures and extremely high magnetic fields a QHE has been observed** for fractional values of i in relation (9.94). In this limit the lowest Landau level is only partially occupied and the IQHE is not expected. Some of the fractional values of i for which the Hall resistance has been observed to be quantized are $\frac{1}{3}, \frac{2}{3}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}$ and $\frac{2}{7}$. At these occupancies the longitudinal Hall resistance ρ_{xx} is found to vanish. This phenomenon is called the *Fractional Quantum Hall Effect* (FQHE).

SUMMARY

1. The electron concentration n_i and the hole concentration p_i in an intrinsic semiconductor are given by

$$n_i = p_i = 2 \left(\frac{k_B T}{2\pi\hbar^2} \right)^{3/2} (m_e^* m_h^*)^{3/4} \exp\left(-\frac{E_g}{2k_B T}\right)$$

2. The electrical conductivity of intrinsic semiconductors depends on temperature according to

$$\sigma_i \propto \exp(-E_g/2k_B T)$$

3. The Fermi energy ϵ_F of a semiconductor is expressed as

$$\epsilon_F = E_v + \frac{1}{2} E_g + \frac{3}{4} k_B T \ln \left(\frac{m_h^*}{m_e^*} \right)$$

The Fermi level is centred in the forbidden energy gap, when $m_e^* = m_h^*$.

4. In an extrinsic semiconductor,

$$n = 2 \left(\frac{m_e^* k_B T}{2\pi\hbar^2} \right)^{3/2} \exp \left[-\frac{(E_c - \epsilon_F)}{k_B T} \right]$$

* Vipin Srivastava, *Found. Phys Lett.*, **11**(6), 561 (1998).

** D.C. Tsui, H.L. Stormer and A.C. Gossard, *Phys Rev. Lett.*, **48**, 1559 (1982).

$$p = 2 \left(\frac{m_h^* k_B T}{2\pi\hbar^2} \right)^{3/2} \exp \left[-\frac{(E_F - E_v)}{k_B T} \right]$$

where E_v and E_c denote the valence band and conduction band edges, respectively.

5. The electrical conductivity of a semiconductor is

$$\sigma = |e| (n\mu_n + p\mu_p)$$

where μ_n and μ_p are the electron and hole mobilities, respectively.

6. The magnetoresistance is defined as

$$\frac{\Delta \rho}{\rho(0)} = \frac{\rho(B) - \rho(0)}{\rho(0)}$$

Here $\rho(B)$ is the resistivity in the presence of a magnetic field \mathbf{B} and $\rho(0)$ is the zero-field resistivity. In the case of intrinsic materials

$$\frac{\Delta \rho}{\rho(0)} = \omega_c^2 \tau^2 \quad \text{with } \omega_c = \frac{eB}{m^*}$$

where τ is the relaxation time of carriers.

7. In a heterojunction ($p-n$), p and n regions are derived from different semiconductor materials.
 8. The Hall resistance in IQHE is

$$\rho_H = \frac{h}{ie^2}, \text{ where } i = 1, 2, 3, \dots$$

PROBLEMS

- 9.1 Assuming the resistivity of an intrinsic Ge crystal as 47 ohm cm and electron and hole mobilities as 0.39 and $0.19 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ respectively, calculate its intrinsic carrier density at room temperature.
- 9.2 A p -type semiconductor has an acceptor density of 10^{18} cm^{-3} . If $E_D - E_v = 0.2 \text{ eV}$ and $m_e^* = m_h^* = m$ (the free electron mass),
 (a) show that the intrinsic conduction in the crystal is negligible at room temperature;
 (b) estimate the conductivity of the crystal at room temperature, assuming the hole mobility to be equal to $0.01 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ at room temperature.
- 9.3 Plot the intrinsic carrier concentration as a function of $1/T$ on a semilog graph paper over 10–500 K for:
 (a) GaAs ($E_g = 1.5 \text{ eV}$, $m_e^* = 0.1m$, $m_h^* = 0.4m$)
 (b) InSb ($E_g = 0.22 \text{ eV}$, $m_e^* = 0.013m$, $m_h^* = 0.18m$)

For which electron and hole concentrations is the conductivity of a semiconductor minimum? To which value of the net impurity content $|N_D - N_A|$ does this correspond?

- 9.4 In an InSb crystal, $E_g = 0.23$ eV: the dielectric constant (ϵ_s) = 18 and $m_e^* = 0.015m$. Calculate (i) E_d and (ii) the radius of the ground state orbit. What should be the minimum donor concentration so as to produce an appreciable overlap between the orbits of adjacent impurity atoms?
- 9.5 The effective mass of electrons at the lower conduction band edge of a semiconductor is three times higher than that of holes at the upper valence band edge. How far is the Fermi level located from the middle of the forbidden energy gap, assuming that the semiconductor is intrinsic? Explain why E_g should be greater than $8k_B T$ for your calculation.
- 9.6 Calculate $N(c)$ and $N(v)$ in silicon at room temperature. In the conduction band of silicon there are six minima with $m_1^* = 0.97m$, $m_2^* = 0.19m$. The valence band can be approximated by two independent bands coincident at the band edge with $m_1^* = 0.5m$ and $m_2^* = 0.16m$.
- 9.7 In a two-dimensional intrinsic semiconductor the density of states is defined as

$$D_c(\epsilon) = \frac{4\pi m_e^*}{h^2}, \quad \epsilon > E_c$$

$$D(\epsilon) = 0, \quad E_v < \epsilon < E_c$$

$$D_v(\epsilon) = \frac{4\pi m_h^*}{h^2}, \quad \epsilon < E_v$$

Use Fermi-Dirac statistics without making any approximation to show that

$$\epsilon_F = E_c - \frac{1}{2}E_g + \frac{k_B T}{2} \ln \frac{8}{3} - k_B T \ln \cos \frac{\phi}{3}$$

$$\text{with } \phi = \tan^{-1} \left[\left(\frac{32}{27} \right) \exp(-E_g/k_B T) - 1 \right]^{1/2}$$

- 9.8 A semiconductor cuboid crystal (1 cm \times 5 mm \times 1 mm) has a resistivity of 12.5 ohm cm. A Hall voltage of 5 mV across the 5 mm width is measured when the applied magnetic field is 2000 gauss and a current of 1 mA flows along the length of the crystal. Calculate the carrier concentration and Hall mobility in the crystal.
- 9.9 Hall measurements are made on a *p*-type semiconductor bar 500 μm wide and 20 μm thick. The Hall contacts A and B are displaced 2 μm with respect to each other in the direction of current flow of 3 mA. The voltage between A and B with a magnetic field of 10 kG pointing out of the plane of the sample is 3.2 mV. When the magnetic field direction is reversed the voltage changes to -2.8 mV. What is the hole concentration and the mobility?
- 9.10 In an acceptorless *n*-type semiconductor with 10^{13} donors cm^{-3} , the donor ionization energy is 1 meV. Taking the effective mass as 0.01 m and assuming that $E_g \gg k_B T$, calculate (a) the density of conduction electrons at 4 K and (b) the Hall coefficient.
- 9.11 Show that the Hall coefficient of a *p*-type semiconductor is zero when the excess density of acceptors over donors is

$$N_A - N_D = n_i \left(\frac{b^2 - 1}{b} \right)$$

where n_i is the intrinsic density and $b = \mu_n/\mu_p$

- 9.12 Can the longitudinal magnetoresistance of n -type germanium be zero for any orientation of the magnetic field? If yes, find the direction.
- 9.13 Prove that the minimum conductivity of an extrinsic semiconductor is given by

$$\sigma = 2n_i e(\mu_n \mu_p)^{1/2}$$

Show that the conductivity minimum occurs when

$$N_A - N_D = n_i \left[\left(\frac{\mu_n}{\mu_p} \right)^{1/2} - \left(\frac{\mu_p}{\mu_n} \right)^{1/2} \right]$$

- 9.14 Prove that the barrier step height of an unbiased $p-n$ junction is given by

$$eV_B = k_B T \ln \left(\frac{n_n}{n_p} \right)$$

where all of the symbols have their usual meaning.

- 9.15 Calculate the built-in voltage V_B for a $p-n$ junction formed by diffusing boron ($n_p = 10^{15} \text{ cm}^{-3}$) into one end of an n -type silicon chip ($n_n = 3.87 \times 10^{16} T^{3/2} \exp(-E_g/2k_B T) \text{ cm}^{-3}$, $E_g = 1.1 \text{ eV}$) at room temperature and 127°C .
- 9.16 Explain why Esaki diodes do not show high electrical conductivity in spite of having very large carrier concentrations.

SUGGESTED FURTHER READING

- Dalven, R., *Introduction to Applied Solid State Physics* (Plenum Press, 1981).
- Elliott, R.J. and A.F. Gibson, *Introduction to Solid State Physics and its Applications* (Macmillan, 1978).
- Long, D., *Energy Bands in Semiconductors* (Wiley-Interscience, 1968).
- Lüth, H., *Surfaces and Interfaces of Solid Materials* (Springer, 1995).
- Milnes, A.G., *Deep Impurities in Semiconductors* (Wiley-Interscience, 1973).
- Phillips, J.C., *Bands and Bonds in Semiconductors* (Academic Press, 1973).
- Seeger, R., *Semiconductor Physics*, 2nd ed. (Springer, 1982).
- Streetman, B.G., *Solid State Electronic Devices* (Prentice-Hall, New Delhi, 1997).
- Sze, S.M., *Physics of Semiconductor Devices* (Wiley-Interscience, 1981).
- Wolf, H.F., *Semiconductor* (Wiley-Interscience, 1971).

Theory of Dielectrics: Applications to Plasmons, Polaritons and Polarons

In accordance with their behaviour in the presence of electric fields, the structure of any insulating solid can be viewed as a body of bound positively and negatively charged particles (atomic/molecular ions). In this context, insulators are referred to as *dielectric materials* or simply *dielectrics*. We begin this chapter with the study of response of insulators to the external electric fields. Both the static and alternating fields produce interesting effects that are exemplified by a good number of physical situations. The effects are described under two approaches: the microscopic and the macroscopic. The interaction of the electromagnetic radiation with matter is perhaps the best example where the difference between the two approaches can be made out easily. In the microscopic approach the properties are interpreted on the basis of the absorption of a photon with the creation of a phonon or an electron-hole pair. On the other hand, the macroscopic approach concerns the use of the Maxwell's equations, given below (in the SI system):

$$\text{div } \mathbf{D} = \rho \quad (10.1a)$$

$$\text{curl } \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t} \quad (10.1b)$$

$$\text{div } \mathbf{B} = 0 \quad (10.1c)$$

$$\text{curl } \mathbf{H} = \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t} \quad (10.1d)$$

These classical equations are of central importance in the macroscopic description of electromagnetic effects in solids. Vectors \mathbf{E} and \mathbf{D} denote the electric field and the electric displacement vectors, respectively. Vectors \mathbf{B} and \mathbf{H} are the analogously defined quantities for the description of magnetic effects: the magnetic induction and the intensity of magnetic field in that order. ρ represents the volume density of quasi-free charge carriers (or of charges from an external source) and \mathbf{j} is the electric current density.

10.1 POLARIZATION

When a dielectric is placed in the region of an electric field, the centres of positive and negative charges within its constituent entities (atoms/molecules) separate out if they had coincided in the

absence of the field. The typical order of the displacement is, however, less than the atomic diameter. Thus the individual constituent entities begin to behave as electric dipoles, giving a net induced dipole moment to the dielectric medium. Nevertheless, we must be aware that asymmetrical molecules such as HCl and H₂O have permanent electric dipole moment by virtue of an appreciable difference in electronegativities of the atoms they are composed of. The dipole moments of a single HCl and a single H₂O molecule are 1.0 and 1.9 debye, respectively (1 debye = 10⁻¹⁸ esu-cm = 3.3 × 10⁻³⁰ C m). The effect of the electric field in these materials is to orient the dipoles, tending to align them along the field direction.

The significance of electric dipoles may be emphasized in two ways. Firstly, the magnitude of the electric field inside a dielectric cannot be accounted without taking the electric field owing to dipoles into consideration. Secondly, the orientation of electric dipoles under the action of an external electric field creates an induced charge density on the surface of the dielectric.

The electric field owing to the internal sources of charge is entirely contributed by electric dipoles that store the bound charges. The field is calculated by replacing the point charges in Coulomb's law with the electric dipoles since there are no free point charges in a dielectric. In a standard problem of electrostatics we show that the electric field owing to a dipole with moment \mathbf{p} at distance \mathbf{r} from its centre is given as

$$\mathbf{E}(\mathbf{r}) = \frac{3(\mathbf{p} \cdot \mathbf{r})\mathbf{r} - r^2 \mathbf{p}}{4\pi \epsilon_0 r^5} \quad (10.2)$$

with

$$\mathbf{p} = q\mathbf{a} \quad (10.3)$$

where q is the magnitude of the positive and negative charges forming the dipole and \mathbf{a} is the distance between the centres of the positive and negative charges (or simply the size of the dipole).

For a certain value of the electric field the dipoles align themselves parallel to the field direction (Fig. 10.1) as a result of which the charges surface on the opposite faces along this direction with

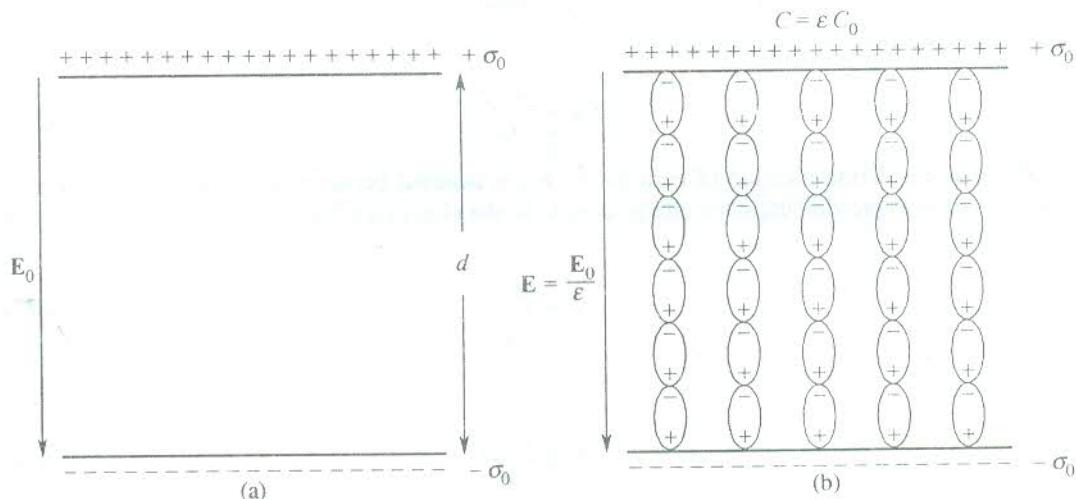


FIG. 10.1 An illustration of the effect of tiny electric dipoles (of atomic dimensions) on the macroscopic electric field: (a) A parallel plate capacitor charged to the surface charge density $σ_0$ with free space between the plates. The field in the region between the plates is E_0 . (b) The capacitor filled with a dielectric and charged to the same charge density $σ_0$. The applied field polarizes the dielectric, creating tiny dipoles which align themselves along the field direction.

the positive charges on one face and the negative on the other. These are essentially the induced charges as the dielectric on the whole continues to be neutral. In this situation the dielectric is said to be polarized and the phenomenon is termed *polarization*. The extent of polarization is given by the electric dipole moment per unit volume (\mathbf{P}):

$$\mathbf{P} = \sum_n \mathbf{p}_i = \sum_n q_i \cdot \mathbf{a}_i \quad (10.4)$$

Here \mathbf{P} is the total dipole moment (including the induced and permanent) and n is the number of dipoles per unit volume.

Polarization \mathbf{P} has the same units as the surface charge density ($C m^{-2}$). This equivalence is substantiated by the fact that electric field induces charges on the surface of the dielectric and the density of charges is a measure of the extent of polarization.

10.2 DIELECTRIC CONSTANT

Consider a parallel plate capacitor with plate area A and plate separation d . When the capacitor is charged at a constant difference of d.c. potential, the electric field in the region between the plates can be treated uniform if the size of the plates is considerably larger than their separation. When the space enclosed within the capacitor is evacuated and the two plates acquire the surface charge densities $+\sigma_0$ and $-\sigma_0$, the magnitude of the electric field over the region within the capacitor is given by

$$E_0 = \frac{\sigma_0}{\epsilon_0} \quad (10.5a)$$

where ϵ_0 denotes the permittivity of the vacuum space, and E_0 is related to the difference of potential V_0 between the plates as

$$V_0 = E_0 d \quad (10.5b)$$

The capacitance may be expressed as

$$C_0 = \frac{A \sigma_0}{V_0} \quad (10.6)$$

When an insulating medium of permittivity ϵ_m is inserted between the plates and the capacitor is charged so as to provide the same charge density to the plates [see Fig. 10.1(b)], the above relations transform to

$$E = \frac{\sigma_0}{\epsilon_m} \quad (10.7a)$$

$$V = Ed \quad (10.7b)$$

$$C = \frac{A \sigma_0}{V} \quad (10.7c)$$

From the above two sets of relations, we get

$$\frac{C}{C_0} = \frac{V_0}{V} = \frac{E_0}{E} = \frac{\epsilon_m}{\epsilon_0} = \epsilon \quad (10.8)$$

where ϵ is a constant known as the *relative permittivity* or the *macroscopic dielectric constant* of the medium between the plates. It is one of the most valuable constants of a solid since it is involved in the description of all of its dielectric and optical properties.

We may rewrite (10.8) as

$$C = \epsilon C_0 \quad (10.9a)$$

$$E = \frac{E_0}{\epsilon} \quad (10.9b)$$

$$V = \frac{V_0}{\epsilon} \quad (10.9c)$$

The relations (10.9) state that on introducing an insulator or a dielectric medium between the plates the capacitance increases to ϵ times the original value; and both the electric field and the potential difference across the plates are reduced to $1/\epsilon$ th of their respective original values. The reduced value of electric field E in the dielectric medium would occur in vacuum for a lower value of the charge density, say, $|\sigma'|$. With this interpretation coupled to the polarization \mathbf{P} being perceived as the reduction in the surface charge density, we have

$$|\mathbf{P}| = |\sigma_0| - |\sigma'| \quad (10.10)$$

And in the SI system,

$$\begin{aligned} \mathbf{P} &= \epsilon_0 \mathbf{E}_0 - \epsilon_0 \mathbf{E} \\ &= \epsilon_0 \epsilon \mathbf{E} - \epsilon_0 \mathbf{E} \end{aligned}$$

or

$$\mathbf{P} = (\epsilon - 1) \epsilon_0 \mathbf{E} \quad (10.11)$$

We rewrite (10.11) as

$$\frac{\mathbf{P}}{\epsilon_0 \mathbf{E}} = \epsilon - 1 = \chi_E \quad (10.12)$$

or

$$\epsilon = 1 + \chi_E \quad (10.13)$$

where χ_E is called the *electric susceptibility* of the dielectric medium.

Relation (10.11) is more often written in the following form,

$$\begin{aligned} \epsilon_0 \mathbf{E} + \mathbf{P} &= \epsilon_0 \epsilon \mathbf{E} \\ &= \epsilon_m \mathbf{E} \\ &= \epsilon_0 \mathbf{E}_0 \end{aligned} \quad (10.14)$$

The combination $(\epsilon_0 \mathbf{E} + \mathbf{P})$ is customarily given a special name—the electric displacement vector—and denoted by \mathbf{D} . That is,

$$\epsilon_0 \mathbf{E} + \mathbf{P} = \mathbf{D} \quad (10.15)$$

Relations (10.11)–(10.15) are written in the form appropriate for an isotropic medium since \mathbf{P} , \mathbf{E} and \mathbf{D} can be in different directions for an anisotropic medium. The susceptibility and the dielectric constant in that case are tensors interpreted with the help of the following relations:

$$\begin{aligned} P_\mu &= (\chi_E)_{\mu\nu} E \\ \epsilon_{\mu\nu} &= 1 + (\chi_E)_{\mu\nu} \end{aligned} \quad (10.16)$$

The displacement vector \mathbf{D} , defined above, appears in Maxwell's equations (10.1a) and (10.1d). This highlights the importance of this vector in the description of the electrostatic and the electromagnetic properties of matter.

10.3 LOCAL ELECTRIC FIELD

It is often required to know the effective electric field at an atomic site for giving a proper account of dielectric properties. This electric field is termed *local electric field*. The effective electric field at a site where an electric dipole may be located equals the externally applied field only in the case of a single electric dipole. In the presence of several dipoles in a system, the neighbouring dipoles also contribute to the electric field at the atomic site of interest. The effect can be ignored only in dilute systems but surely not in solids where the dipole density is considerably large.

We see below that the local field \mathbf{E}_{loc} is different from the macroscopic field \mathbf{E} . The method normally employed to calculate the local electric field is due to Lorentz, according to which the atomic site of concern is considered to be located at the centre of a fictitious cavity. In this model, the local electric field is written as

$$\mathbf{E}_{\text{loc}} = \mathbf{E}_0 + \mathbf{E}_{\text{dep}} + \mathbf{E}_L + \mathbf{E}_{\text{dip}} \quad (10.17)$$

where

\mathbf{E}_0 is the field owing to fixed charges external to the dielectric body.

\mathbf{E}_{dep} is the depolarization field caused by the uniform polarization of the dielectric. This field, having its origin in induced surface charges, opposes \mathbf{E}_0 .

\mathbf{E}_L is the field at the centre of the cavity created by charges induced on the imaginary surface bounding the fictitious cavity. This field is also called the *Lorentz field*.

\mathbf{E}_{dip} is the field at the centre of the cavity contributed by dipoles within the cavity.

The value of \mathbf{E}_{dep} calculated in terms of \mathbf{P} depends on the shape of the dielectric specimen. For symmetrical specimens, say in the form of sphere, cylinder, disc, etc. the medium is uniformly polarized and \mathbf{E}_{dep} in these cases has been calculated by Osborn* and Stoner.** For a sphere placed in a uniform electric field, it is found to be given by

$$\mathbf{E}_{\text{dep}} = - \frac{1}{3\epsilon_0} \mathbf{P} \quad (10.18)$$

The macroscopic field \mathbf{E} is customarily interpreted as

$$\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_{\text{dep}} \quad (10.19)$$

Hence \mathbf{E}_{loc} may be expressed as

$$\mathbf{E}_{\text{loc}} = \mathbf{E} + \mathbf{E}_L + \mathbf{E}_{\text{dip}} \quad (10.20)$$

* J.A. Osborn, *Phys. Rev.*, **67**, 351 (1945).

** E.C. Stoner, *Phil. Mag.*, **36**, 803 (1945).

Calculation of E_L

We consider the case of a fictitious cavity of spherical shape for simplicity (Fig. 10.2). For regions external to the cavity, the distances to the cavity centre are so large that we can treat the dipole distribution as continuous, resulting in a macroscopic polarization \mathbf{P} . The polarization produces a field that can be described in terms of polarization charges on the surface of the cavity. In view of the

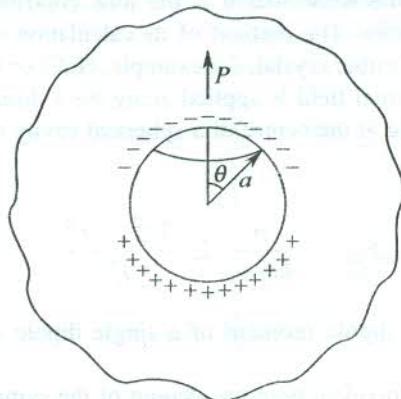


FIG. 10.2 The macroscopic polarization and the induced charges on the surface of a fictitious spherical cavity within the dielectric.

interpretation of polarization \mathbf{P} , given in Section 10.2, the fictitious charge density[†] σ on the surface of the imaginary sphere is written as

$$\sigma = -\mathbf{P}_n \quad (10.21)$$

where \mathbf{P}_n stands for the normal component of \mathbf{P} (see Fig. 10.2). Then, a circular element at polar angle θ bears the charge,

$$dq = -P \cos \theta \cdot 2\pi a \sin \theta \cdot a d\theta \quad (10.22)$$

The electric field produced by this charge at the centre of the cavity is

$$dE_L = -\frac{1}{4\pi\epsilon_0} \frac{dq}{a^2} \cos \theta \quad (10.23)$$

Therefore, the field caused by the total charge on the surface of the cavity is

$$\begin{aligned} E_L &= \frac{\mathbf{P}}{2\epsilon_0} \int_0^\pi \cos^2 \theta \sin \theta d\theta \\ &= \frac{1}{3\epsilon_0} \mathbf{P} \end{aligned} \quad (10.24)$$

[†] Referred to the electric field in vacuum.

Hence, for a specimen of spherical shape [from (10.18) and (10.24)], we get

$$\mathbf{E}_{\text{dep}} + \mathbf{E}_L = 0 \quad (10.25)$$

Calculation of \mathbf{E}_{dip}

The nature of \mathbf{E}_{dip} is special in the sense that it is the lone contribution to the local field which depends on the crystalline symmetry. The method of its calculation in non-cubic crystals is not so straightforward. Let us consider a cubic crystal, for example, NaCl or CsCl in which the atomic sites have cubic symmetry.[†] If the external field is applied along the z -direction, then in accordance with (10.2) the z -component of the field at the centre of a spherical cavity owing to all dipoles within the cavity is written as

$$E_{\text{dip}}^z = \frac{p}{4\pi\epsilon_0} \sum_i \frac{3z_i^2 - r_i^2}{r_i^5} \quad (10.26)$$

where p is the magnitude of the dipole moment of a single dipole and all dipoles are parallel to the z -axis.

Since x, y, z directions are equivalent here on account of the symmetry of the crystal and of the sphere, we have

$$\sum_i \frac{x_i^2}{r_i^5} = \sum_i \frac{y_i^2}{r_i^5} = \sum_i \frac{z_i^2}{r_i^5} \quad (10.27)$$

Therefore, for a cubic crystal with spherical cavity,

$$\mathbf{E}_{\text{dip}} = 0 \quad (10.28)$$

Since the cavity is only imaginary, this result also holds good for a cubic crystal of spherical shape.

At the same time we must appreciate that the net contribution to the local field from all the dipoles in a dielectric can be expressed as

$$\sum_i \frac{3(\mathbf{p}_i \cdot \mathbf{r}_i) \mathbf{r}_i - r_i^2 \mathbf{p}_i}{r_i^5} = \mathbf{E}_{\text{dep}} + \mathbf{E}_L + \mathbf{E}_{\text{dip}} \quad (10.29)$$

Relations (10.25) and (10.28) show that for a cubic crystal of spherical shape, the sum on the RHS of (10.29) is zero.^{††} So for these specimens,

$$\mathbf{E}_{\text{loc}} = \mathbf{E}_0 \quad (10.30)$$

and

$$\mathbf{E} = \mathbf{E}_0 - \frac{1}{3\epsilon_0} \mathbf{P} \quad (10.31)$$

showing different values of the two fields for the same specimen.

[†] Every atomic site in some cubic crystals may not have cubic symmetry (e.g. O²⁻ site in BaTiO₃).

^{††} A proper discussion can be found in *Classical Electrodynamics*, p. 116, J.D. Jackson (Wiley, 1962).

The \mathbf{E}_{loc} at an atomic site having cubic symmetry is generally expressed in the form,

$$\mathbf{E}_{\text{loc}} = \mathbf{E} + \frac{1}{3\epsilon_0} \mathbf{P} \quad (10.32)$$

This is known as the *Lorentz relation*. It is very well satisfied by experimental data on cubic ionic crystals.

The \mathbf{E}_{loc} at an atomic site in a medium of cubic symmetry is expressed in another useful form by substituting the value of \mathbf{P} from (10.11) in the Lorentz relation (10.32):

$$\mathbf{E}_{\text{loc}} = \left(\frac{\epsilon + 2}{3} \right) \mathbf{E} \quad (10.33)$$

10.4 DIELECTRIC POLARIZABILITY

The electric dipole moment of an atom is found to be related to the electric field at its site by

$$\mathbf{p} = \alpha \mathbf{E}_{\text{loc}} \quad (10.34)$$

where α is defined as the *polarizability* of the atom concerned. This relation is valid for both CGS and SI systems with

$$(\alpha)_{\text{SI}} = 4\pi\epsilon_0 (\alpha)_{\text{CGS}} \quad (10.35)$$

We can relate polarizability α , which is an atomic property, to the macroscopic property, polarization \mathbf{P} by

$$\mathbf{P} = \sum_j N_j \alpha_j \mathbf{E}_{\text{loc}}^j \quad (10.36)$$

where

N_j is the number of j th type of atoms in unit volume of the crystal

α_j is the polarizability of an j th type atom

$\mathbf{E}_{\text{loc}}^j$ is the electric field at the site of an j th type atom

Then, using (10.33) the polarization of a cubic crystal is written as

$$\mathbf{P} = \left(\frac{\epsilon + 2}{3} \right) \mathbf{E} \sum_j N_j \alpha_j \quad (10.37)$$

And using (10.11), we get

$$\frac{\epsilon - 1}{\epsilon + 2} = \frac{1}{3\epsilon_0} \sum_j N_j \alpha_j \quad (10.38)$$

This is the famous Clausius–Mossotti relation. It is valid for dielectrics that crystallize in a cubic lattice. It gives the relationship between the dielectric constant and the polarizability of atoms that arises due to the displacement of the electron shell relative to the nucleus in an atom. The polarization of a hydrogen atom under the action of a static electric field is shown in Fig. 10.3. Thus the polarizability appearing in (10.38) is essentially contributed by the electronic polarization. The α_j is a tensor if the respective j th type atom is non-spherical.

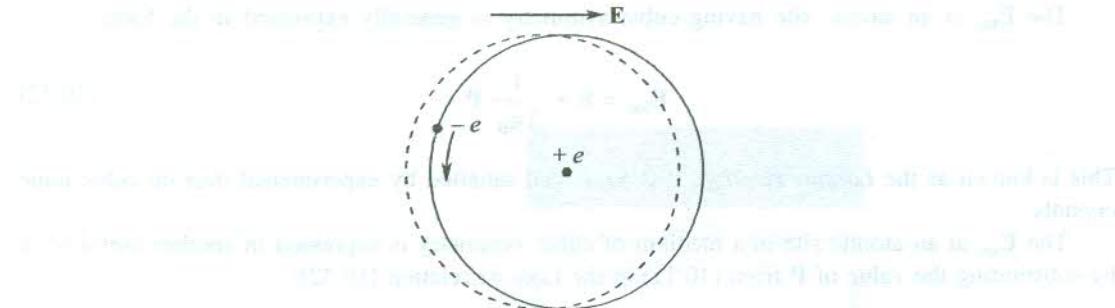


FIG. 10.3 Illustration of the displacement of the electron orbit relative to the nucleus in a hydrogen atom under the influence of an external field \mathbf{E} .

10.5 SOURCES OF POLARIZABILITY

The sources that contribute to the total polarizability of a dielectric crystal are:

- (i) Electronic polarization
- (ii) Ionic polarization
- (iii) Orientational polarization

We discuss below these polarizabilities and the related physical phenomena.

10.5.1 Theory of Electronic Polarizability and Optical Absorption

We were introduced to electronic polarizability in Section 10.4 through the Clausius–Mossotti relation (10.38). In the presence of an electric field the centre of the electron shell and that of the nucleus in an atom remain no more coincided and the atomic system gets polarized as shown for a hydrogen atom in Fig. 10.3. The interaction of optical radiation with matter offers itself, perhaps, as the best example where the consequences of electronic polarization can be grasped in a simple way. We treat this problem below in the framework of a classical theory. When the optical radiation carrying oscillating electric field strikes an atom, the atom experiences an alternating electric field which we consider as the Lorentz local field \mathbf{E}_{loc} expressed as

$$\mathbf{E}_{\text{loc}} = \mathbf{E}_0 \exp(-i\omega t) \quad (10.39)$$

where \mathbf{E}_0 represents the amplitude of the local field.

An electron in an atom polarized by this field can be treated as a harmonic oscillator experiencing a restoring force that originates from the Coulomb attraction between the electron and the nucleus. The complete equation of motion of the electron, taken as an elastically bound particle, has the form

$$m \frac{d^2 \mathbf{x}}{dt^2} + m\gamma \frac{d\mathbf{x}}{dt} + m\omega_0^2 \mathbf{x} = e\mathbf{E}_{\text{loc}} \quad (10.40)$$

where ω_0 is the natural angular frequency of the electron such that $\omega_0 = \left(\frac{f}{m}\right)^{1/2}$, with f denoting the force constant referred to the restoring force. It must be emphasized here that the strength of the internal field \mathbf{E}_{loc} is not the same as that of the field vector of the incident optical radiation. Only in gases where the density is low, the two can be equal in strength.

The second term on the LHS of (10.40) represents the damping force* on the electron characterized by dissipation constant γ with the dimensions of frequency. The dissipation results from the fact that the electron emits radiation as a consequence of its acceleration.

The solution to the equation of motion (10.40) is

$$\mathbf{x} = \frac{e\mathbf{E}_{\text{loc}}}{m(\omega_0^2 - \omega^2 + i\gamma\omega)} \quad (10.41)$$

It varies sinusoidally as dictated by the local field, the driver of these displacements. With this solution the following expression for the electronic polarizability α_e is obtained showing that α_e is a complex quantity:

$$\alpha_e(\omega) = \frac{e\mathbf{x}}{\mathbf{E}_{\text{loc}}} = \frac{e^2}{m(\omega_0^2 - \omega^2 + i\gamma\omega)} \quad (10.42)$$

If we ignore the damping effect, (10.42) would read as

$$\alpha_e(\omega) = \frac{e^2}{m(\omega_0^2 - \omega^2)} \quad (10.43)$$

which shows a singularity in the polarizability at the natural resonance frequency ω_0 . But, practically, the value of the dielectric constant, a corresponding macroscopic physical quantity, never approaches either $+\infty$ or $-\infty$. It indicates to the dissipation of energy or the dielectric loss that prevents the occurrence of singularity. This is perhaps the simplest way we can substantiate the need for having a damping term in the equation of motion of an electron which is a member of the oscillating electronic shell.

The model which we have used to derive (10.42) is distinctly crude. Nevertheless, the utility of this model is underlined by the success of (10.42) in showing that the electronic polarizability is independent of the frequency of optical radiation ω when $\omega \ll \omega_0$ (the natural frequency of the electronic shell) and the polarizability equals its static value,

$$\alpha_e^s = \frac{e^2}{m\omega_0^2} \quad (10.44)$$

This result is supported by more accurate quantum mechanical calculations.

The values of electronic polarizability, also referred to as the atomic polarizability, in some atoms and ions are given in Table 10.1. The typical order of these values is 10^{-24} cm^3 in the CGS system and 10^{-40} F m^2 in the SI system.[†] Substituting this value in (10.44), we get the natural frequency of the electronic shell as $\nu_0 = \frac{\omega_0}{2\pi} \approx 10^{15} \text{ s}^{-1}$ which is about ten times higher than the frequencies in the visible range of the optical radiation. Thus, up to the beginning of the ultraviolet region, the electronic polarizability may be considered to remain constant.

* Proof given by R. Becker, *Theorie der Elektrizität*, 6th ed., Teubner (Leipzig, 1933).

† Multiply the value in CGS units by $\left(\frac{1}{9 \times 10^{15}} \right)$ to obtain the value in SI units.