5 Electrons in solids—an introduction to band theory

5.1 Introduction

The conduction theory presented in the previous chapter assumes that many free electrons are available within the body of the material, which behave as classical particles. In a metal the free valence electrons are shared by all atoms in the solid; hence there is a tendency for the periodic potential of the crystal lattice as seen by the conduction electrons to be smeared out and to appear almost constant. This accounts for the success of the free-electron model in explaining most, if not all, of the conduction phenomena in metals. However, for materials with different crystal structures, for example in the important case of covalent bonded solids such as some semiconductors, valence electrons are located much nearer to the parent atoms and cannot be associated with the entire collection of atoms as in a metal. The free-electron model fails for such materials since the potential seen by valence electrons can no longer be regarded as constant since it varies rapidly, particularly near to ion cores in the lattice.

A quantum-mechanical model that overcomes this difficulty assumes that the conduction electrons, as well as being subject to the restriction of the exclusion principle as before, are not entirely free but move in the perfectly periodic potential of a crystal. Such a distribution of potential arises because of the regular spacing of ion cores in the lattice and its perodicity is equal to the lattice constant. We shall see that in this situation the energy of electrons can be situated only in allowed bands, which are separated by forbidden energy regions. Within a particular allowed band, electrons behave in much the same way as free electrons; they can again interact with externally applied fields to produce conduction effects but the interaction parameters have to be modified to account for the presence of the lattice.

The so-called band theory of solids, which is developed from the periodic potential model, has been most successful in explaining some of the anomalies predicted by the free-electron model and also can account for the differing electrical properties of conductors, semiconductors and insulators. What

determines the conduction properties of a particular material is whether the electronic states within an allowed energy band are empty or full.

The more complete model also accounts for apparent changes in effective electron mass with position in an energy band. Further, it will be shown that the properties of a material with an almost filled band are identical to those of a material containing a few positive charge carriers in an otherwise empty band; this is a quantum-mechanical justification of the concept of a *hole*, which will be used extensively when discussing semiconductors later.

Finally, departures from the assumed perfect periodicity of potential will be shown to account for resistive effects in a practical material.

5.2 Allowed energy bands of electrons in solids

5.2.1 General concepts

It was shown in Chapter 2 that the electrons in an isolated atom are only allowed to possess discrete values of energy. The exclusion principle also stipulates that each energy level, which is defined by a set of three quantum numbers, can only be occupied by at most two electrons, provided they have opposite spins.

When atoms are packed closely together in a solid such that the electronic orbitals of neighbouring atoms tend to overlap, the allowed electron energy levels are modified from those of the individual constituent atoms. Consider, for example, two identical atoms that are gradually brought together. As the outermost orbitals overlap, electrons that originally had the same energy in the isolated atoms have their energies slightly modified so that the exclusion principle is not violated for the two-atom system; each allowed energy level is split into two closely spaced levels. If the atoms are brought still closer together such that the electrons in inner orbitals of each atom interact, the lower energy levels split in a similar manner.

Energy level splitting of atoms in close proximity can be explained in terms of a simple quantum-mechanical model, as follows. Consider, for example, two hydrogen atoms separated by an initially large distance, r. The electronic energy structure of each atom will be as shown in Fig. 2.3(a) and for the two atoms is as shown in Fig. 5.1(a). To a first approximation each atom can be represented by a one-dimensional rectangular potential well of width δ, as shown in Fig. 5.1(b). We have seen that the wavefunctions of bound electrons in such a potential well are nearly sinusoidal (or exactly sinusoidal if the well is infinitely deep) and that the corresponding energies are discrete eigenvalues. For example, the possible wavefunctions of the lowest energy state ($n=1$) are as shown in Fig. 5.1(c). Notice that the negative ψ solution, shown as a broken curve, is included, since it is equally possible because it is only the quantity $|\psi|^2$ that has any physical significance. The general expression for the energy eigenvalues of a trapped electron is given in Eq. (2.6); hence the lowest energy

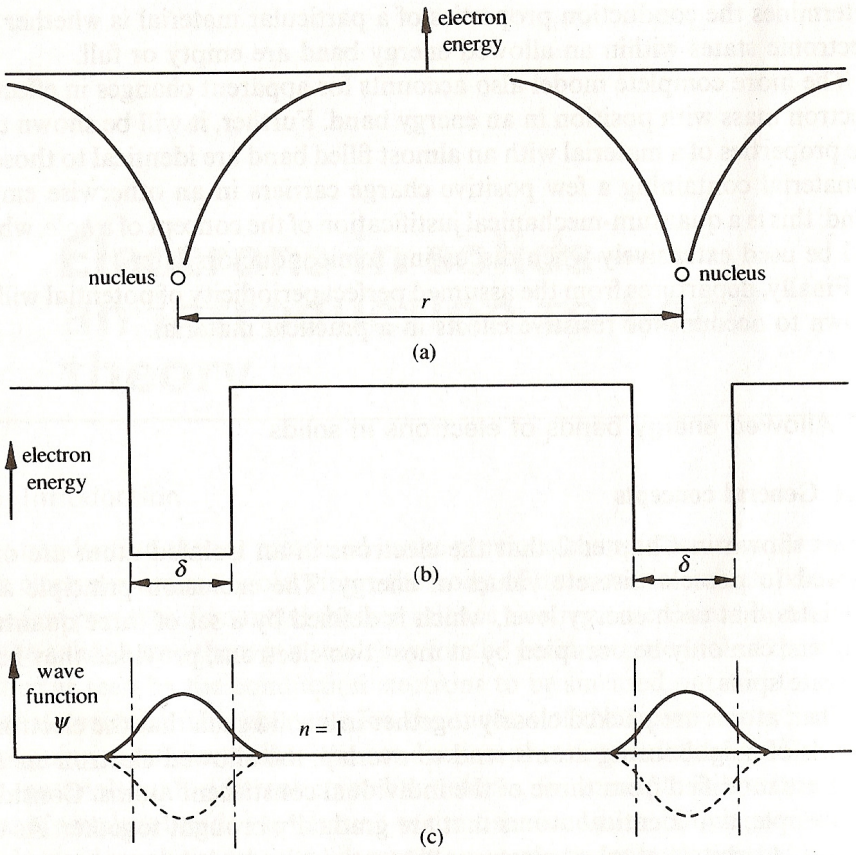


Fig. 5.1 (a) Potential energy of electrons in two isolated hydrogen atoms; (b) one-dimensional potential well equivalent; (c) wavefunction of trapped electrons in the lowest energy state.

state for an individual well is

$$E_1 = h^2/(8m\delta^2) \tag{5.1}$$

Now, consider the atoms when they are brought more closely together such that $r > \delta$, as shown in Fig. 5.2(a). The boundary conditions stipulate that both the wavefunction, ψ, and its gradient, $\partial\psi/\partial x$, must always remain continuous. It would appear that there are two possible configurations for the wavefunction of the complete system, as shown, one in which ψ is symmetrical about the centre line dividing the atoms and one in which ψ is antisymmetrical. When the atoms are brought even nearer, such that $r = \delta$, the wavefunctions of the lowest-energy electrons merge into the symmetrical and antisymmetrical forms shown in Fig. 5.2(b) The electron energies corresponding to each wavefunction of the complete system can again be obtained from Eq. (2.6) and are

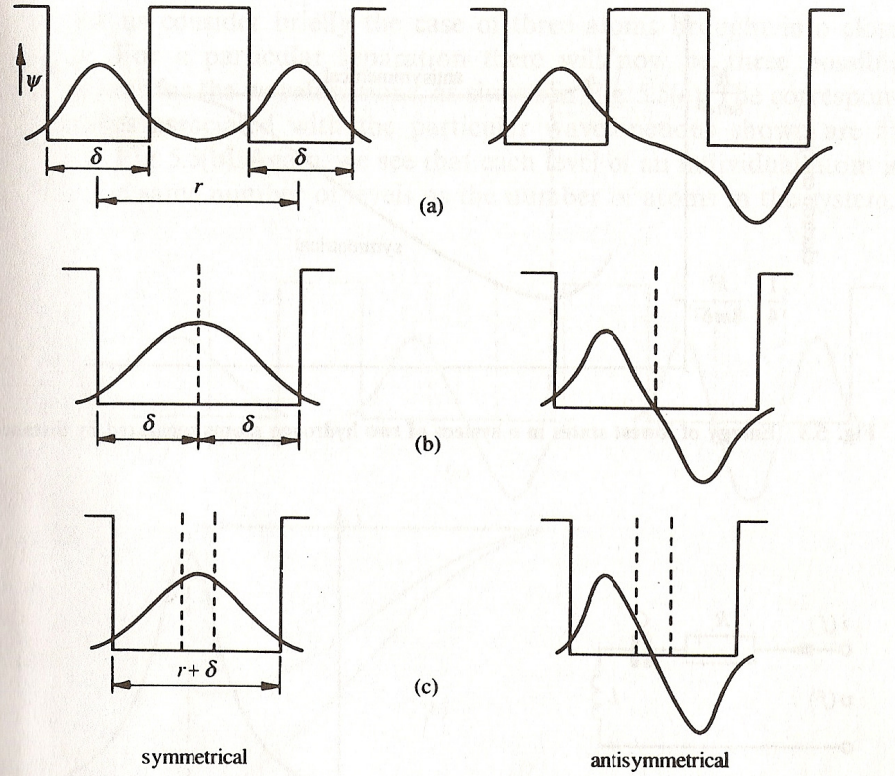


Fig. 5.2 Possible wavefunctions for the lowest energy state of a two-hydrogen-atom system as the separation is changed.

and

$$E_{1,\text{antisym}} = \frac{2^2 h^2}{8m(2\delta)^2} = \frac{h^2}{8m\delta^2} \tag{5.3}$$

For even closer separation such that $r < \delta$, the possible wavefunctions for the system are shown in Fig. 5.2(c) and the corresponding electron energy values are

$$E_{1,\text{sym}} = \frac{h^2}{8m(r+\delta)^2} \tag{5.4}$$

and

$$E_{1,\text{antisym}} = \frac{4h^2}{8m(r+\delta)^2} \tag{5.5}$$

These possible energy states of the system for various spacings r are collected together in Fig. 5.3. It will be seen that as soon as the spacing between the atoms is such that $r \approx \delta$, two definite energy states exist for the system where for the individual atoms there was only one.

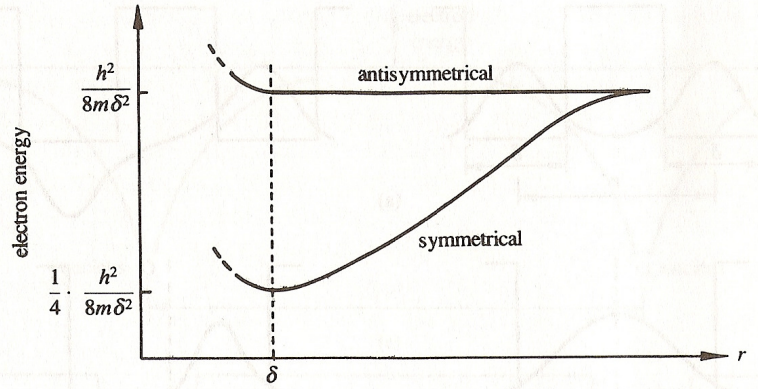


Fig. 5.3 Energy of lowest states in a system of two hydrogen atoms separated by distance r.

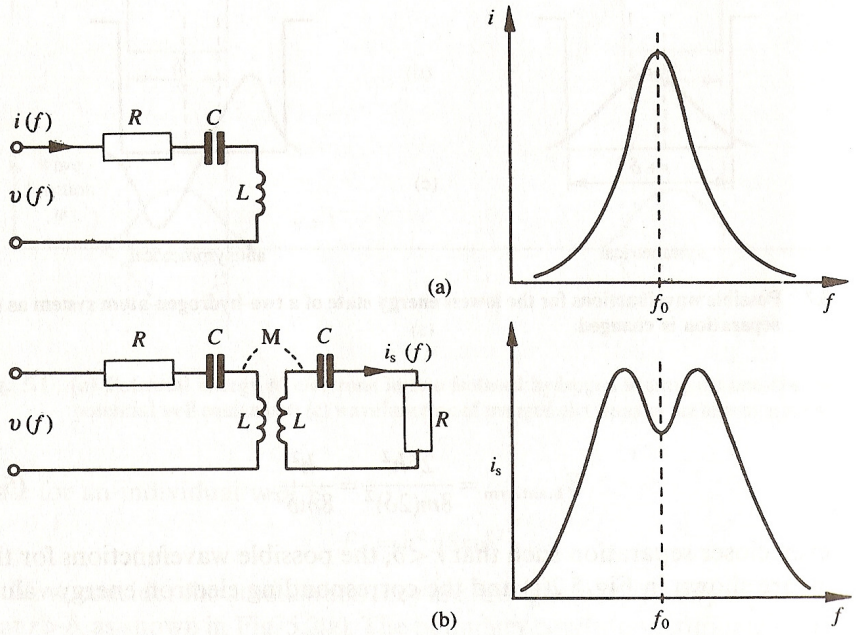


Fig. 5.4 (a) Series resonant circuit; (b) coupled resonant circuits.

coupled resonant circuits. The simple series *LRC* circuit depicted in Fig. 5.4(a) has a single current maximum when the frequency of the applied sinusoidal voltage is equal to the resonant frequency, f_0, as shown. When two such circuits are tightly coupled via mutual inductance as shown in Fig. 5.4(b), two current maxima occur at frequencies displaced slightly to either side of f_0. Such frequency splitting can be compared to energy level splitting in atomic systems, particularly when it is remembered that energy and frequency are related quantities in quantum-mechanical systems.

Now let us consider briefly the case of three atoms brought into close proximity. For a particular separation there will now be three possible configurations for the wavefunctions, as shown in Fig. 5.5(a). The corresponding energies associated with the particular wavefunctions shown are as depicted in Fig. 5.5(b). Again, we see that each level of an individual atom is split into the same number of levels as the number of atoms in the system.

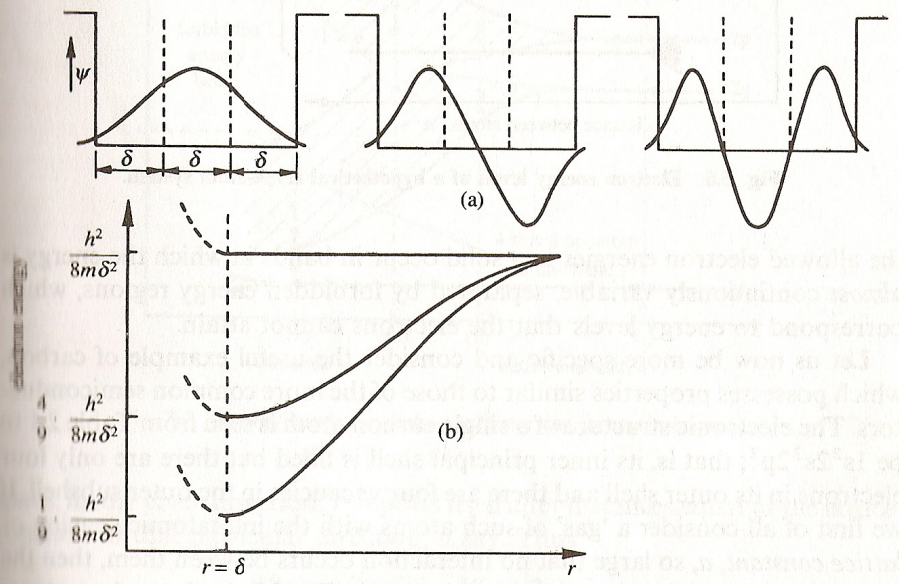


Fig. 5.5 Three-atom system: (a) possible wavefunctions for the lowest energy states; (b) electron energy as a function of atomic separation.

The extension of this argument is straightforward and it is reasonable to expect that in a system of n interacting atoms each discrete energy level of an individual atom is split into n closely spaced levels as the atoms are brought together. A system containing seven atoms is shown diagrammatically in Fig. 5.6; notice that the higher energy levels split at larger separations. This is because the electrons at these levels are on average further from the nucleus and interact with neighbouring atoms more readily.

Of course, in a more realistic system such as a solid the number of interacting atoms is much higher than seven; a typical figure may be, for example, 10^{22}. Also the total width of each band of allowed energy levels is of order 1 eV and depends not on the number of atoms grouped together but on their interatomic spacing. Since in this instance 10^{22} discrete levels have to be accommodated in an energy range that is only 1 eV in extent, individual levels in a band are of necessity very closely spaced together; the allowed energy levels within a band are therefore said to be *quasi-continuous*. Summarizing,

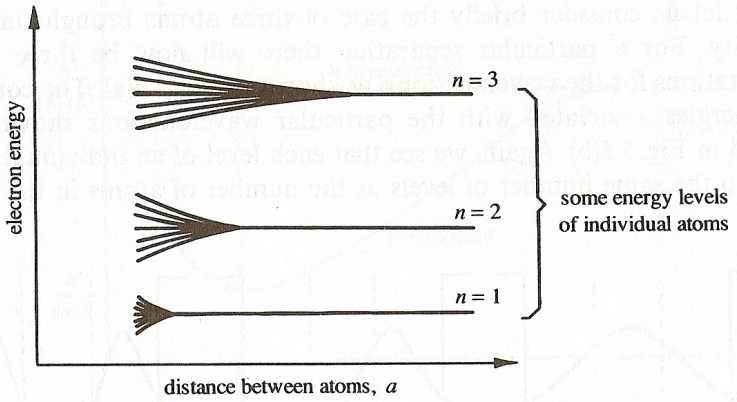


Fig. 5.6 Electron energy levels of a hypothetical seven-atom system.

the allowed electron energies in a solid occur in bands in which the energy is *almost* continuously variable, separated by forbidden energy regions, which correspond to energy levels that the electrons cannot attain.

Let us now be more specific and consider the useful example of carbon, which possesses properties similar to those of the more common semiconductors. The electronic structure of a single carbon atom is seen from Table 2.1 to be $1s^2 2s^2 2p^2$; that is, its inner principal shell is filled but there are only four electrons in its outer shell and there are four vacancies in the outer subshell. If we first of all consider a 'gas' of such atoms with the interatomic spacing or *lattice constant*, a, so large that no interaction occurs between them, then the energy levels of each atom are as shown at (i) in Fig. 5.7. As the carbon atoms are brought into closer proximity (i.e. a is reduced), level splitting occurs as described, which results in bands of allowed energies, as for example at (ii). For even closer spacings as at (iii), the bands can overlap. Eventually, as a is reduced still further, the energies of the outer-shell electrons can lie in one of two bands, separated by a forbidden gap, as at (iv).

Of couse, it is not possible to vary the interatomic spacing, continuously as we have assumed for convenience, since for a particular crystalline solid the lattice constant is fixed. The band structure of a particular allotropic form of carbon will then correspond to a vertical slice through Fig. 5.7; for example, carbon in the diamond form has a band structure similar to that at (iv). We shall see later that it is the magnitude and form of the band structure of a particular solid that completely specify the electrical conducting properties peculiar to it. Before doing so we will consider a more quantitative approach to the investigation of the band structure of solids.

5.2.2 Mathematical model of a solid

In an ideal solid, the ion cores of the crystal are spaced with perfect regularity and the potential experienced by an electron in the solid, V, is periodic in

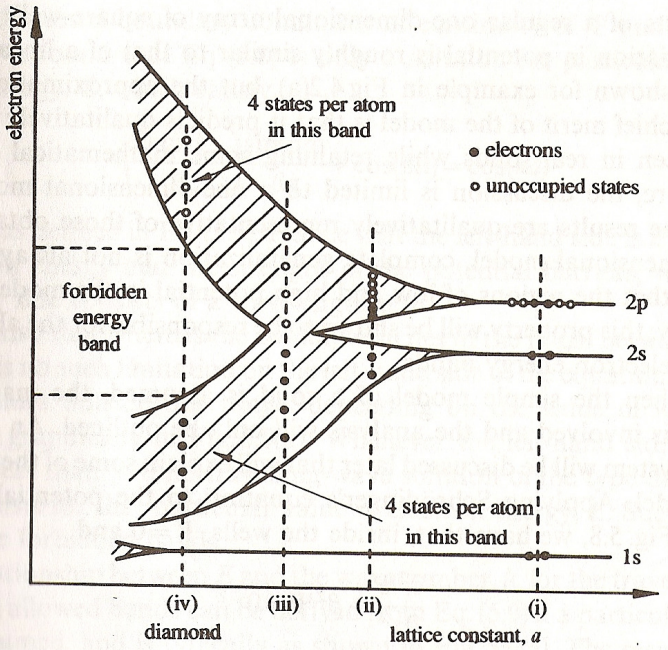
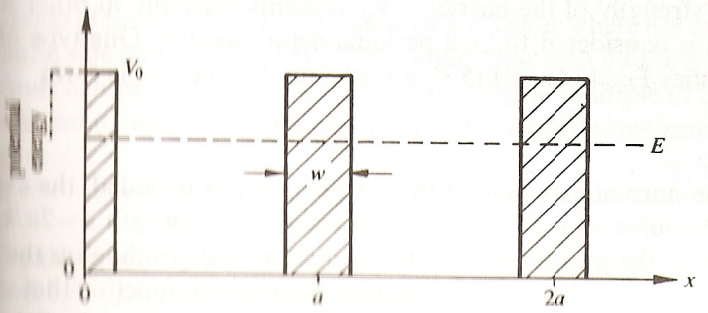


Fig. 5.7 Energy bands for carbon with varying interatomic spacing.

space; in any given direction, V repeats itself after distances equal to the lattice constant in that particular direction, a say. Thus

$$V(x) = V(x+a) = V(x+2a) = \ldots$$

Of course, the precise nature of $V(x)$ is complex and the solution of Schrödinger's equation including such a function is difficult. The problem is considerably eased if a simpler mathematical model of the solid is assumed in which the potential in a given direction, as seen by electrons in higher-lying energy bands, changes abruptly from some value V_0 to zero with a periodicity a as shown in Fig. 5.8. The model, which was proposed by Kronig and Penney,



thus consists of a regular one-dimensional array of square-well potentials. Such a variation in potential is roughly similar to that of a linear array of atoms, as shown for example in Fig 4.2(a), but the approximation is most crude; the chief merit of the model is that it predicts qualitatively the effects that are seen in real solids while retaining some mathematical simplicity. Furthermore, the discussion is limited to a one-dimensional model; while many of the results are qualitatively representative of those obtainable for a three-dimensional model, complete generalization is not always possible. Note also that the regions of low and high potential in the model alternate periodically; this property will be shown to be responsible for the allowed and forbidden electron energy bands.

Even when the simple model of a solid is assumed, the mathematical treatment is involved and the analysis will only be outlined. An analogous electrical system will be discussed later that will explain some of the properties of the model. Applying Schrödinger's equation to the potential variation shown in Fig. 5.8, we have that, inside the wells, $V=0$ and

$$\frac{\partial^2\Psi}{\partial x^2}+\beta^2\Psi=0 \tag{5.6}$$

where

$$\beta^2=2mE/\hbar^2$$

and in the barriers, where $V_0>E$,

$$\frac{\partial^2\Psi}{\partial x^2}-\alpha^2\Psi=0 \tag{5.7}$$

where

$$\alpha^2=2m(V_0-E)/\hbar^2$$

Equations (5.6) and (5.7) can be solved using the appropriate boundary conditions. It is usual at this stage, however, to make a further simplification to the model so that the problem becomes more tractable. This consists of letting the width of the barriers, w, go to zero and their height to infinity, in such a way that the 'strength' of the barrier, wV_0, remains constant; in other words, the potential is considered to be a periodic delta function. One type of solution that satisfies Eqs (5.6) and (5.7) is then found to be of the form

$$\Psi=U_k(x)e^{jkx} \tag{5.8}$$

When the normal exponential time dependence is included, the exponential part of the solution represents a plane wave of wavelength $\lambda=2\pi/k$, which is travelling in the positive or negative x direction, depending on the sign of k. The factor $U_k(x)$, called a Bloch function, is a periodic function that varies with the same periodicity as the lattice, a. Thus the solution to Schrödinger's

When the usual boundary conditions of continuity of Ψ and $\partial\Psi/\partial x$ are applied, it is found that Eq. (5.8) is only a solution for particular values of electron energy, E, which satisfy

$$\left(\frac{mawV_0}{\hbar^2}\right)\frac{\sin(\beta a)}{\beta a}+\cos(\beta a)=\cos(ka) \tag{5.9}$$

where β is as defined in Eq. (5.6). Notice that the left-hand side is a function of the electron energy and the strength of the potential barriers, whereas the right-hand side consists of a wavelength term.

Notice also that whereas the $\cos(ka)$ term lies in the range between -1 and $+1$, there is no such limitation on the left-hand side of the equation, which can assume values outside this range, depending on the value of wV_0; this is illustrated graphically in Fig. 5.9(a). Whenever the left-hand side is greater than 1 or less than -1, no travelling-wave solution of the type described by Eq. (5.8) exists for that particular value of electron energy, E. Such values of E lie in the forbidden bands.

The relationship between E and the wavenumber, k, for the travelling-wave solution in allowed bands can be derived from Eq. (5.9) if a particular value of wV_0 is assumed, and is typically as shown in Fig. 5.9(a). The resulting band structure of allowed electron energy bands separated by forbidden energy gaps is also included in the diagram.

It will be noticed from Eq. (5.9) and Fig. 5.9(a) that if the barrier strength is increased, i.e. wV_0 is made larger, the allowed bands become much narrower and the forbidden bands are correspondingly widened. At the other extreme, as wV_0 is reduced to zero, Eq. (5.9) reduces to $\beta=k$ and hence from Eq. (5.6)

$$\beta=k=2\pi/\lambda=(2mE)^{1/2}/\hbar$$

$$E=h^2/(2m\lambda^2) \tag{5.10}$$

This energy expression is identical to that for the free electron, Eq. (4.11); the result is not surprising, since as wV_0 becomes zero the potential barriers are removed and electrons can move freely inside the solid.

Considering Eq. (5.9) again, and the solution in the form of an $E-k$ diagram, Fig. 5.9(b), the value of k is not uniquely determined, e.g. if on the right-hand side of the equation k is replaced by $k+2\pi n/a$, where n is an integer, the right-hand side remains the same. This implies that each plot of $E-k$ in the various bands can be shifted left or right by an integral multiple of $2\pi/a$. It is often convenient to use this concept and to define a *reduced wavevector* limited to the region

$$\pi/a\geq k\geq-\pi/a$$

as shown by the broken curves in Fig. 5.9(b), in which the curves marked a and b are shifted by $-2\pi/a$ and $2\pi/a$ along the axis to appear as a′ and b′ in the reduced zone; similar shifts of $2\pi n/a$ in k values transform curves c, d, e and f to c′, d′, e′ and f′ in the reduced zone.

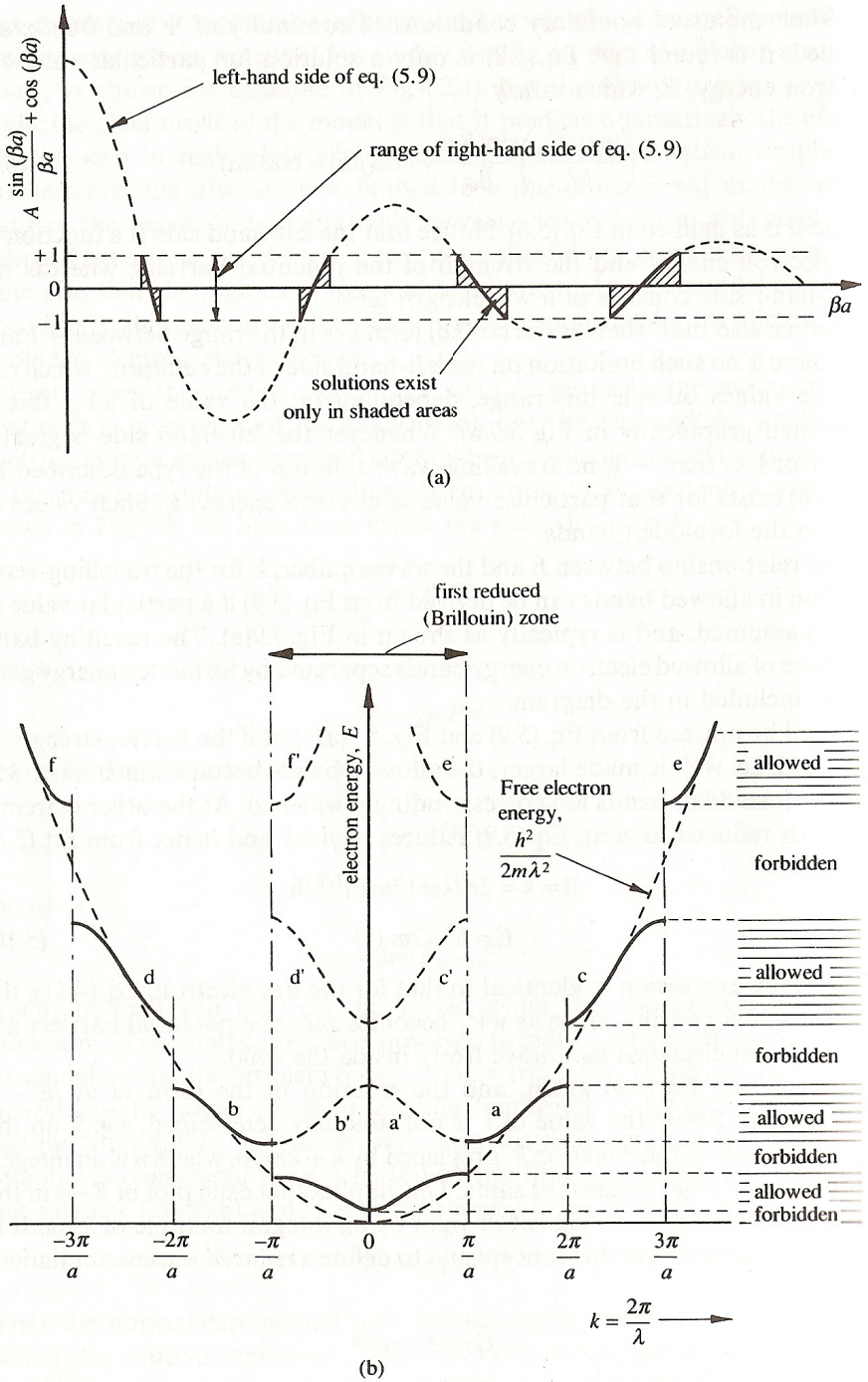


Fig. 5.9 (a) Range of possible solutions of Eq. (5.9). (b) Electronic energy as a function of wavelength and the energy band structure of a hypothetical solid. ——reduced zone representation.

Let us now try to gain some physical insight into the striking behaviour shown in the E–k diagram whenever $k=n\pi/a$, by employing the following electric circuit analogy. Consider the coaxial transmission line shown in Fig. 5.10(a); the voltage, V, at any point x on the line is given by

$$\frac{d^2V}{dx^2}+\omega^2\mu_0\epsilon_r\epsilon_0 V=0 \tag{5.11}$$

If the line is air-filled but periodically loaded at intervals of a with dielectric discs of relative permittivity ϵ_r and thickness w, as shown, then the voltage equation becomes similar in form to Eqs (5.6) and (5.7), provided $\omega^2\mu_0\epsilon_0\equiv\beta^2$

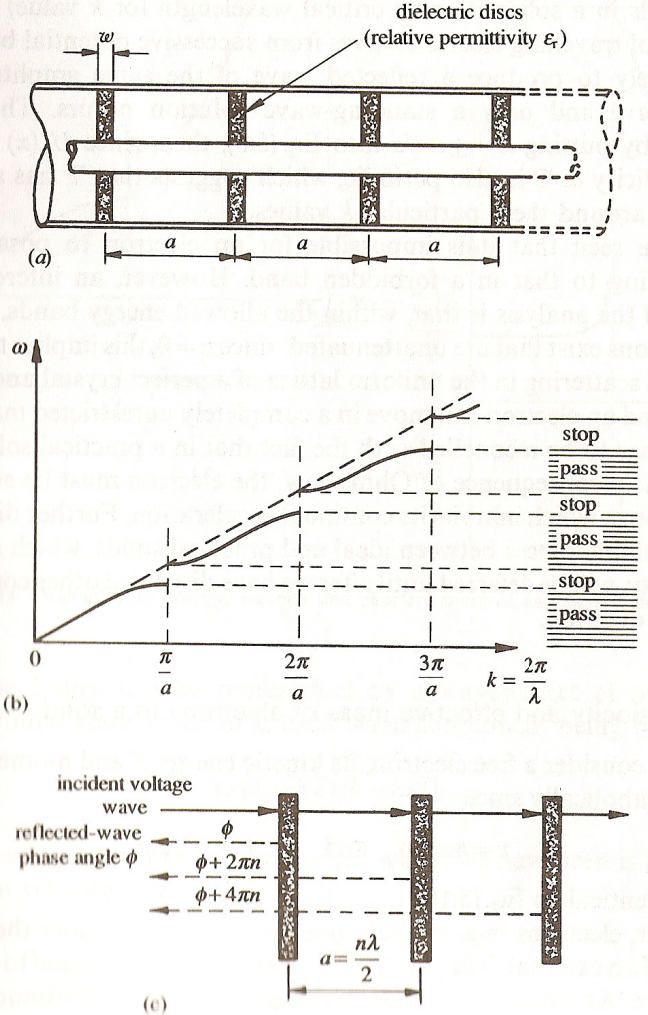


Fig. 5.10 (a) A periodically loaded transmission line; (b) ω–k diagram of the loaded line; (c) total reflection occurring when $ka=n\pi$.

and $\omega^2\mu_0\epsilon_r\epsilon_0 \equiv -\alpha^2$. Hence, in the circuit analogy, voltage V is equivalent to electron energy E and frequency bands are analogous to electron energy bands. The ω–k diagram for the loaded line can be obtained by solving Eq. (5.11) using the appropriate boundary conditions and results in series of pass bands of frequency and stop bands that occur when $ka = \pi, 2\pi, 3\pi, n\pi$, as shown in Fig. 5.10(b). The stop bands arise because when $ka = n\pi$ the discs are spaced an integral number of half-wavelengths apart and reflections from successive discs add in phase as shown diagrammatically in Fig. 5.10(c). Then even if individual reflections are weak, their combined effect is to produce total reflection; hence, for this condition, no travelling-wave solution exists for the voltage and only standing-wave solutions are possible.

The situation in the stop bands is analogous to what occurs at the forbidden energy levels in a solid; at some critical wavelength (or k value) the partial reflections of travelling electron waves from successive potential barriers add constructively to produce a reflected wave of the same amplitude as the incident wave and only a standing-wave solution occurs. This may be confirmed by putting $k = \pm n\pi/a$ into Eq. (5.8); then, since $U_k(x)$ is periodic with periodicity a, Ψ is also periodic, which suggests that Ψ has a standing-wave form around these particular k values.

We have seen that it is impossible for an electron to possess energy corresponding to that in a forbidden band. However, an interesting consequence of the analysis is that, within the allowed energy bands, travelling-wave solutions exist that are unattenuated, since $\alpha = 0$; this implies that there is no electron scattering in the uniform lattice of a perfect crystal and within an allowed band an electron can move in a completely unrestricted manner. This statement has to be reconciled with the fact that in a practical solid we have seen that, as a consequence of Ohm's law, the electron must be subjected to a viscous force, which inhibits its continual acceleration. Further discussion of the essential differences between ideal and practical solids, which can resolve this difficulty, will be deferred until after we have discussed other consequences of the E–k diagram.

5.3 The velocity and effective mass of electrons in a solid

Let us first consider a free electron; its kinetic energy E and momentum p are related parabolically since

$$p = mv \qquad \text{and} \qquad E = p^2/2m \tag{5.12}$$

which is identical to Eq. (5.10).

However, electrons in a solid are not free; they move under the combined influence of an external field plus that of a periodic potential due to atom cores in the lattice. As a result, the electron energy is no longer continuous and the energy–momentum relationship, since $p = \hbar k$, will be similar in shape to that shown in Fig. 5.9(b), as depicted in Fig. 5.11(a). Now, an electron moving

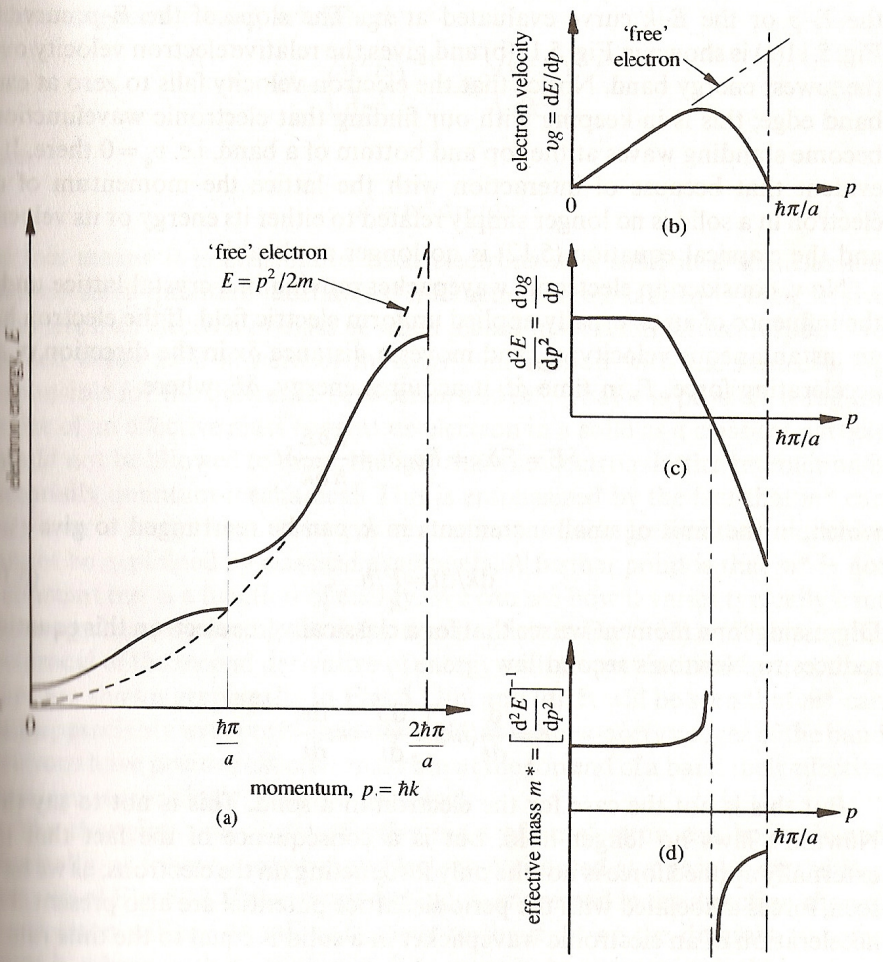


Fig. 5.11 Variation of energy, velocity and effective mass of an electron in a solid.

through the lattice can be represented by a wavepacket of plane waves grouped around some value of k, each wave component being of the form

$$\exp[-j(Et/\hbar - kx)]$$

The group velocity, v_g, of the wavepacket, which we have seen is identical to the electron velocity is, from Eq. (1.30)

$$v_g = \frac{\partial\omega}{\partial k} = \frac{\partial(E/\hbar)}{\partial k} = \frac{1}{\hbar}\frac{\partial E}{\partial k} = \frac{\partial E}{\partial p} \tag{5.13}$$

Thus, the velocity of an electron, which is represented by a packet of waves centred near to a particular value of $k = k_0$, say, is proportional to the slope of

the E–p or the E–k curve evaluated at k_0. The slope of the E–p curve in Fig. 5.11(a) is shown in Fig. 5.11(b) and gives the relative electron velocity over the lowest energy band. Notice that the electron velocity falls to zero at each band edge; this is in keeping with our finding that electronic wavefunctions become standing waves at the top and bottom of a band, i.e. $v_g = 0$ there. It is evident that because of interaction with the lattice the momentum of an electron in a solid is no longer simply related to either its energy or its velocity and the classical equation (5.12) is no longer applicable.

Now, consider an electronic wavepacket moving in a crystal lattice under the influence of an externally applied uniform electric field. If the electron has an instantaneous velocity, v_g, and moves a distance δx in the direction of an accelerating force, F, in time δt, it acquires energy, δE, where

$$\delta E = F\delta x = Fv_g\delta t = \frac{F\delta E}{\hbar\,\delta k}\delta t$$

which, in the limit of small increments in k, can be rearranged to give

$$\mathrm{d}k/\mathrm{d}t = F/\hbar \tag{5.14}$$

Digressing for a moment we see that for a classically free electron this equation reduces to Newton's second law

$$F = \frac{\mathrm{d}}{\mathrm{d}t}(\hbar k) = \frac{\mathrm{d}p}{\mathrm{d}t} = m\frac{\mathrm{d}v}{\mathrm{d}t} \tag{5.15}$$

But this is not the case for the electron in a solid. This is not to say that Newton's laws no longer hold, but is a consequence of the fact that the externally applied force is not the only force acting on the electrons; as we have seen, forces associated with the periodic lattice potential are also present. The acceleration of an electronic wavepacket in a solid is equal to the time rate of change of its velocity, thus

$$\text{acceleration} = \frac{\mathrm{d}v_g}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}E}{\mathrm{d}p}\right) = \frac{\mathrm{d}^2E}{\mathrm{d}t\mathrm{d}p}$$

or, using Eq. (5.14)

$$\frac{\mathrm{d}v_g}{\mathrm{d}t} = \frac{\mathrm{d}p}{\mathrm{d}t}\frac{\mathrm{d}^2E}{\mathrm{d}p^2} = F\frac{\mathrm{d}^2E}{\mathrm{d}p^2}$$

which can be rearranged to give

$$F = \left(\frac{\mathrm{d}^2E}{\mathrm{d}p^2}\right)^{-1}\frac{\mathrm{d}v_g}{\mathrm{d}t} \tag{5.16}$$

Comparing this equation with the classical equation of motion for a particle, Eq. (5.15), we see that the quantity $(\mathrm{d}^2E/\mathrm{d}p^2)^{-1}$ is equivalent to the mass of the free electron. Thus if, for an electron moving in the periodic lattice of a solid, we

define an *effective mass*, m^*, where

$$m^* = \left(\frac{\mathrm{d}^2E}{\mathrm{d}p^2}\right)^{-1} = \hbar^2\left(\frac{\mathrm{d}^2E}{\mathrm{d}k^2}\right)^{-1} \tag{5.17}$$

then

$$F = m^*\mathrm{d}v_g/\mathrm{d}t \tag{5.18}$$

By this means it is possible to treat electrons in a solid in a semiclassical manner since quantum-mechanical interactions are included in the effective mass term; an electron of mass m, when placed in a crystal lattice, responds to applied fields as if it were of mass m^*, interaction with the lattice being responsible for the difference between m and m^*. That it is possible using the device of an effective mass to treat an electron in a solid as a classical particle should not be allowed to mask the fact that the electron–lattice interaction is essentially quantum-mechanical. This is emphasized by the fact that m^* can vary over a range from a few per cent of m to much greater than m, which cannot be explained by classical arguments. A further point is that m^* is not a constant but is a function of energy. We can see how it varies typically over an energy band by noting the definition implicit in Eq. (5.17) and forming the reciprocal of the second derivative of energy with respect to momentum; these steps are shown graphically in Figs 5.11(c) and (d). It will be seen that m^* can vary appreciably with position in the band; at the low-energy edges of the band electrons have positive effective mass but at the top end of a band their effective mass can, surprisingly, become negative!

The changing sign of the effective mass of an electron can be explained physically, as follows. Suppose an electron is situated at a point a on the E–p diagram of Fig. 5.12. If an electric field, $\mathscr{E}$, is impressed in the direction shown, the electron will accelerate and moves to the right on the diagram to some point b, where both its energy and its velocity have increased; this conventional behaviour corresponds to a positive effective mass. Now, consider an

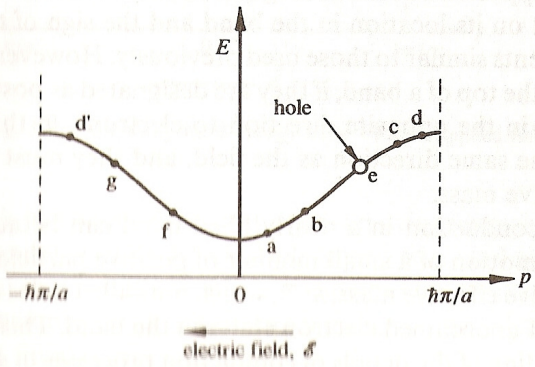


Fig. 5.12 E–p diagram of a hypothetical solid.

electron at the upper end of a band, at c say. When the field is applied again the electron moves to d, where its energy is increased but its velocity has decreased (see for example Fig. 5.11(b)). The electron appears to have been decelerated by the previously accelerating force and since Eq. (5.18) applies always this can be accounted for by the electron having a negative effective mass. It will be apparent from Fig. 5.11 that negative effective masses occur whenever the E–p curve is concave downwards and that the electron mass is positive whenever the curve is concave upwards. Now, the direction of acceleration of an electron is determined by the sign of both its effective mass and its charge, so an alternative way of accounting for the properties of an electron with negative mass is to consider it as a particle with positive mass but having a *positive* charge e, since acceleration $= -e\mathscr{E}/(-m^*) = +e\mathscr{E}/(+m^*)$. Thus, when electrons at the top of a band are acted on by an applied field the resulting currents correspond to the movement of electrons with a positive charge $+e$ and a positive effective mass m^*.

In most materials, at least one band is only partly filled. For example, we have seen that, in a metal, conduction electrons occupy all levels up to around the Fermi energy, E_F. If, as is often the case with other materials, a band is almost entirely filled, it is often convenient to discuss its properties in terms of the relatively few unfilled states rather than those occupied by the many remaining electrons. An unfilled energy state providing a vacancy that can be occupied by an electron wavepacket is called a *hole*. Consider the hole shown diagrammatically in Fig. 5.12 at e. Because of the symmetry of the E–p diagram an electron with positive momentum, say at b, will be cancelled by another electron, in this case at f, with the same magnitude of momentum but oppositely directed. This cancellation of momentum by pairs applies to all electrons except the one at g, which is at the corresponding energy level to the hole; this electron has a negative velocity, v_g (from Eq. (5.13)), and so the hole at e must have a corresponding positive velocity, which can be accounted for by attributing a positive charge to the hole. Hence an electron moving in one direction results in a current flow that is equivalent to a hole or vacancy moving in the opposite direction. In general, the effective mass of a hole is again dependent on its location in the band and the sign of the mass can be found by arguments similar to those used previously. However, since holes are often located at the top of a band, if they are designated as positive charge and moreover move in the opposite direction to electrons, in this case they are accelerated in the same direction as the field, and they must usually possess a positive effective mass.

We see that conduction in a nearly filled band can be accounted for by considering the motion of a small number of positive particles of charge $+e$, possessing positive effective mass, m_h^*, which are called holes and correspond to the number of unoccupied electron states in the band. This concept is vital to an understanding of the details of conduction processes in semiconductors,

5.4 Conductors, semiconductors and insulators

The band structure of a solid is a convenient method of classifying its conduction properties. Of course, electrical engineering materials can readily be characterized experimentally by means of their conductivities, but the band theory explains the essential differences between materials with widely differing conductivities.

The conduction process in any material is dependent on the availability of charge carriers. Clearly, if a given energy band is unoccupied it can make no contribution to electronic conduction. What is not quite so obvious is the fact that there can be no net conduction effects if all the bands are completely full either. Consider, for example, a completely full band with E–p diagram as in Fig. 5.12 (disregarding the hole now). We have seen that because of symmetry of the graph there can be no net electron momentum when no external field is applied; obviously, no current flows when the field is zero. When the field, $\mathscr{E}$, is applied, electrons at a or b, say, are accelerated by the field and their momentum is increased. However, an electron at d, say, can have its momentum so increased that it reaches the boundary of the band edge, is reflected and reappears at d′ with oppositely directed momentum. Thus, since all levels remain filled before and after the application of the field, and since the distribution of electron momenta is unaltered, there is still no net flow of current.

In order for conduction to occur there must be empty available states in a particular band. Then, when an electric field is applied, as shown in Fig. 5.13,

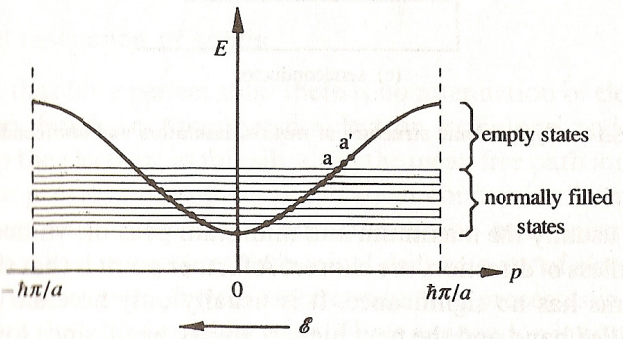


Fig. 5.13 Partially filled energy band with an applied field.

electrons at levels just below the empty ones can gain energy from the field and move into the available levels, for example from a to a′, and all other electrons move to the right in momentum space, as shown. This results in a net electron momentum in the opposite direction to the field that is no longer zero and an electron current flows.

In a good conductor, the essential requirement of many carriers being

available in a partially filled energy band is achieved by the two outermost bands, one of which is completely empty and one full, overlapping; this situation is demonstrated diagrammatically in Fig. 5.7, where the section at (iii) represents the band structure of a metal. It is conventional to draw the band structure of a particular material with fixed lattice constant as shown in Fig. 5.14. While the extremities of each band are, in fact, dependent on crystal

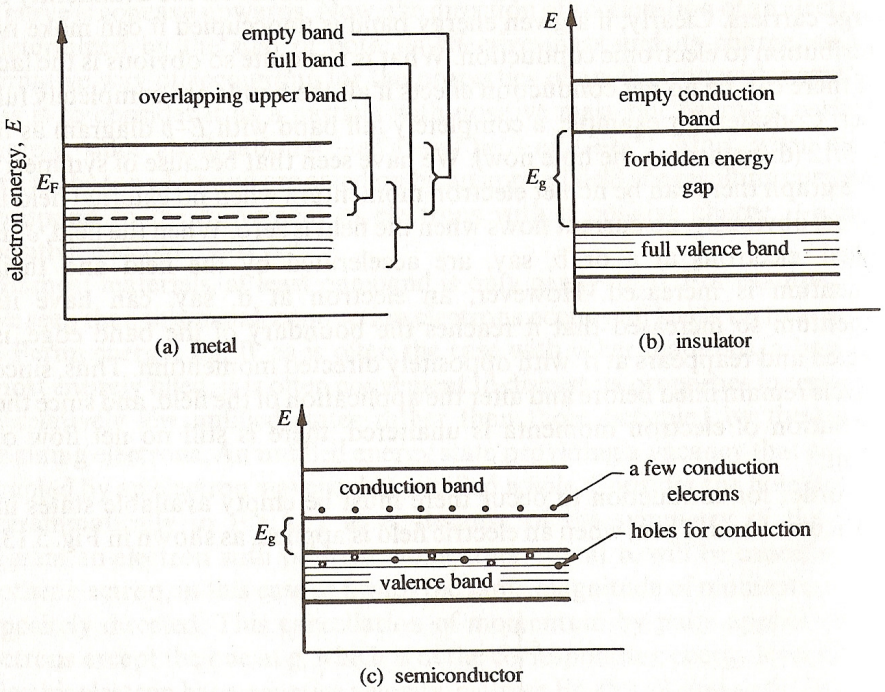


Fig. 5.14 Typical band structure of metals, insulators and semiconductors.

orientation, usually the maximum and minimum possible values for the band edges, regardless of direction, are chosen. A further point is that the abscissa on such diagrams has no significance. It is usually only necessary to show the outermost *filled* band and the next highest empty band, since lower bands are usually completely occupied and play no part in the conduction process. The band diagram of a metal is as shown in Fig. 5.14(a), then. All levels are filled in the band up to some level approaching E_F, above which there are many empty states. Only electrons near the Fermi level participate in conduction and they behave as if they had an effective mass, m^*, evaluated at $E = E_F$. Since E is not located near a band edge, m^* is nearly equal to m, as shown in Fig. 5.9(b); this accounts for the success of the free-electron model of a conductor.

At the other conductivity extreme, insulators are characterized by a band structure consisting of a completely full band, the *valence* band, separated from

an empty band, the *conduction* band, by a wide forbidden energy gap of several electronvolts, as shown in Fig. 5.14(b). At all ordinary temperatures, the statistical probability of electrons from the full band gaining sufficient energy to surmount the energy gap and becoming available for conduction in the conduction band is slight. This very limited number of free conduction electrons at all but very elevated temperatures accounts for the high resistivity of insulators.

Semiconductors, as their name suggests, have conduction properties that are intermediate between those of metals and insulators. They have a band structure as shown in Fig. 5.14(c), which is basically similar to that of insulators, except that the gap energies are very much smaller, being typically of order 1 eV. Since the gap is appreciably smaller than for insulators, it is statistically more probable at ordinary temperatures for electrons in the otherwise full valence band to be elevated across the forbidden gap to the empty conduction band, where they are available for conduction. An essential difference between conduction in metals and conduction in semiconductors is that when electrons in a semiconductor gain sufficient energy to occupy the conduction band they automatically create vacancies in the valence band due to their absence. Thus, additional current flow is possible due to charge motion in the now partially empty valence band; such currents can be described in terms of the motion of holes, as discussed in the previous section. The relatively low conductivities of semiconductors compared with metals is, of course, a consequence of the relatively small number of charge carriers, both electrons and holes, available for conduction.

5.5 Electrical resistance of solids

We have seen that for a perfect solid there is no attenuation of electrons in an allowed energy band, i.e. there are no electron 'collisions' and the solid is transparent to the electron; in this situation the mean free path for collisions is infinite. In the previous discussion of electrical conduction in metals it was found necessary to postulate some sort of collision mechanism to provide a frictional force to account for the terminal drift velocity of electrons, which ensures that Ohm's law is obeyed. Since conduction processes in other solids are similar, with the added complication of hole conduction in some materials, it is again clearly necessary for some kind of collisional damping to be present. An electron in an allowed band can then gain energy from an applied field and move higher in the band, but then can suffer a collision (more usually it is said to be *scattered*), give up its energy in the form of Joule heating of the lattice, and return to lower down in the energy band. The conductivity of the solid can again be expressed using equations similar to Eq. (4.35), except that, since the carriers are no longer necessarily free, their effective mass must be included to account for interaction of carrier and lattice.

But what are electrons colliding with? What are they scattered by? It was

shown in Sec. 4.7 that electrons colliding with ion cores cannot account for the electrical resistance of a solid. It is the interaction of electrons with the slightly aperiodic potential fields experienced in real solids that can cause scattering effects and hence account for electrical resistance effects. Such deviations from the perfect periodicity in potential, which was assumed for an ideal solid, can be due to thermal lattice vibrations, lattice defects, or the presence of impurity atoms and boundaries, some or all of which are normally present in practical materials.

The most important scattering process at ordinary temperatures can occur in crystals where impurity atom or structural imperfections are negligible. The departure from periodicity necessary to produce scattering is in this case brought about by thermal vibration of the lattice atoms about their equilibrium position. Such a displacement alters the local potential, hence its regular periodicity, and an electron travelling in this field can have both the magnitude and direction of its momentum altered. Such an event constitutes what we have thought of as a collision and is often called *lattice scattering*. The description is something of a misnomer since it is not the lattice that produces the scattering so much as its thermally induced vibrations. Scattering interactions between carriers and lattice vibrations become more probable the higher the temperature because of the larger amplitude of vibrations. Thus, we see qualitatively that the average time between collisions, or relaxation time, τ_r, and hence the carrier mobility decreases with increasing temperature in materials in which lattice scattering is the dominant mechanism, such as relatively pure or structurally perfect crystals.

A further scattering mechanism is attributable to the presence of impurity atoms in the lattice, which, may be ionized or otherwise, although the former are more important. Such atoms alter the local electrostatic potential and create the necessary aperiodicity in the field to cause *impurity scattering* of the electrons. The effectiveness of the deflection of an electron by an ionized impurity is greater the lower the velocity of the electron; hence impurity scattering tends to dominate in purer crystals at lower temperatures when thermal scattering is weak, as well as being important when the impurity concentration is high.

Other possible scattering mechanisms are due to vacancies, dislocations and other lattice imperfections. It is also conceivable that conduction electrons could be scattered by holes and vice versa but the probability that these processes occur is slight.

Now let us try to estimate more quantitatively the effects that departures from periodicity have on the resistivity of a real solid. Again, it is convenient to revert to the dielectric disc-loaded coaxial line model of a solid discussed in Sec. 5.2.2. For a uniform disc spacing, a, we saw that there was no attenuation in the line, i.e. all reflected waves cancel, unless $ka = n\pi$. If we now assume a small deviation from periodicity δ, such that $\delta \ll a$, it can be shown that the reflection coefficient will be proportional to δ and the back-scattered power to

$\overline{\delta^2}$. Thus the fraction of power reflected, dP, in a length of line dx is given by

$$dP/P = K\overline{\delta^2}\,dx$$

where K is some constant. Integrating we have

$$P(x) = P_0 \exp(-K\overline{\delta^2}x) \tag{5.19}$$

where P_0 is the power at $x = 0$.

Applying this result to the analogous case of electron waves travelling in the periodic potential of a solid, then if the mean-square departure from periodicity in potential is $\overline{\delta^2}$, it follows by comparison with Eq. (5.19) that

$$\psi^2 = \psi_0^2 \exp(-K\overline{\delta^2}x) \tag{5.20}$$

In other words, electrons are scattered by the aperiodic potential and the probable electron density falls off with distance as $\exp(-K\overline{\delta^2}x)$. Notice that if $\overline{\delta^2} = 0$ there is no scattering. When δ is finite, however, the electron density falls to $1/e$ of its initial value in a mean free path, $\bar{l}$, and so

$$\bar{l} \propto 1/(K\overline{\delta^2}) \tag{5.21}$$

Since the principal scattering mechanism is due to thermal vibrations we shall consider the consequences of Eq. (5.21) applied to this case. Ion cores in a realistic solid have a natural frequency of vibration about their equilibrium position; also, as one is displaced, the position of its neighbour is affected and acoustic waves propagate. A simple mechanical analogy is that of a linear chain of masses, M, joined by springs of stiffness C. Each simple oscillator has kinetic energy $Mx^2/2$ and potential energy $Cx^2/2$. Now, the equipartition of energy condition requires that the *mean* energy in each energy state at any particular temperature, T, is $kT/2$. Hence

$$C\overline{x^2}/2 = kT/2 \tag{5.22}$$

and the mean-square deviation from the equilibrium is proportional to T. Thus, for a practical solid,

$$\overline{\delta^2} \propto T$$

and from Eq. (5.21)

$$\bar{l} \propto T^{-1}$$

It follows from (4.37) that the conductivity, σ, is proportional to $\bar{l}$ and hence

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

or the resistivity, ρ, is proportional to temperature. This result is in accordance with practice, where

$$\rho = \rho_0(1 + \alpha_t T) \tag{5.23}$$

where α_T is the temperature coefficient of resistance. The residual resistivity term in the expression is due to lattice defects, which are present even at very low temperatures. Such defects are particularly large in disordered alloys such as nichrome and their effect is to tend to swamp the increase in ρ due to lattice vibrations. Hence such alloys are useful where high resistance combined with a low temperature coefficient of resistance are required.

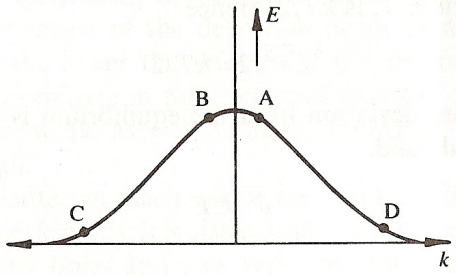
Now, the energy of the acoustic wave due to thermal vibrations of the lattice is quantized and can only change in units of hf. The quantum of acoustic energy is called a *phonon* (cf. photon of electromagnetic energy). Hence, at very low temperatures, Eq. (5.22) is no longer valid since kT becomes comparable to hf and a phonon description of the interaction must be used. This can account for the departure from linearity of the ρ–T curve of a practical solid at very low temperatures.

Although a phonon is actually an acoustic wave propagating through a solid, it is often convenient to think of the associated quantum of energy, the phonon, as a particle in the solid capable of interacting with other particles. Thus, the mechanism we have discussed can be considered as the collision between electrons and phonons and is indeed sometimes referred to as *electron–phonon collisional scattering.*

It is now evident that the classical collisions that have been postulated to account for resistive effects are realized in a solid by the interaction of electron waves with other waves due to lattice vibrations.

Problems

1. The E–k diagram for an energy band in a particular material is as shown. If



an electric field is applied to the material in the negative k direction (force in the positive direction), find (a) the polarity of the effective masses of the four wavepackets made up of groups of states near A, B, C and D, (b) the direction of the velocity of each of the four wavepackets and (c) the direction of the acceleration of each. What are the physical consequences of these results?

Ans (a) −, −, +, +, (b) −, +, +, −, (c) −, −, +, +

2. The conductivity of a metal having n free electrons per unit volume is given by Eq. (4.37) and the Fermi energy by Eq. (4.22). Consider a metal with a simple-cubic lattice structure of side 0.2 nm and one free conduction electron per atom. Assuming that the mean free path for electron collisions with the lattice is 100 lattice constants, find the relaxation time for an electron with the Fermi energy.

Ans. 2.2×10^{-14} s