CHAPTER 5 METALS II: ENERGY BANDS IN SOLIDS

On the surface there is infinite variety
of things; at base a simplicity of cause.

Ralph Waldo Emerson

5.1 INTRODUCTION

In Chapter 4 we talked about the motion of electrons in solids, using the free-electron model. This model is oversimplified, however, because the crystal potential is neglected. But this potential cannot be entirely disregarded if one is to explain the experimental results quantitatively. In addition, some effects cannot be explained at all without taking this potential into account, as we pointed out at the end of Chapter 4. The present chapter therefore treats the influence of the crystal potential on the electronic properties of solids.

In the first part of the chapter we shall consider the energy spectrum of an electron in a crystal. We shall see that the spectrum is composed of continuous *bands*, unlike the case for atoms, in which the spectrum is a set of discrete levels. We shall discuss the properties and the corresponding wave functions of these bands in detail, and develop a useful criterion for distinguishing metals from insulators in this band model. Then we shall deal with the density of states and the Fermi surface, which serve as useful characteristics of a solid.

The electrons in a crystal are in a constant state of motion. Formulas are developed for calculating the velocity of an electron, and its effective mass. We shall study the effects of an electric field on the motion of an electron, and then derive an expression for the electron's electrical conductivity. Although this expression reduces to the one derived previously in Chapter 4 under the appropriate circumstances, the form we shall develop here is more general, and brings out more clearly the physical factors influencing conductivity.

Cyclotron resonance and the Hall effect will also be discussed again and we shall show how these phenomena may be used to obtain information on a solid.

The last section will deal with the limitations of the energy-band model, and the metal–insulator transition.

5.2 ENERGY SPECTRA IN ATOMS, MOLECULES, AND SOLIDS

The primary purpose of this section is to describe qualitatively the energy spectrum of an electron moving in a crystalline solid. It is helpful, however, to begin the discussion by considering the spectrum of a free atom, and see how this spectrum is gradually modified as atoms are assembled to form the solid.

Let us take lithium as a concrete example. Consider a free lithium atom: The electron moves in a potential well, as shown in Fig. 5.1(a). When we solve the Schrödinger equation, we obtain a series of discrete energy levels, as shown. As in the case of the hydrogen atom, these levels are denoted by 1s, 2s, 2p, etc. The lithium atom contains three electrons, two of which occupy the 1s shell (completely full), and the third the 2s subshell.

Now consider the situation in which two lithium atoms assemble to form the lithium molecule Li_2. The potential "seen" by the electron is now the double well shown in Fig. 5.1(b). The energy spectrum here is comprised of a set of discrete *doublets*: Each of the atomic levels—that is, the 1s, 2s, 2p, etc.—has split into two closely spaced levels. Because of the close generic relation between the atomic and molecular levels, we may also speak of the 1s, 2s, 2p, etc., molecular

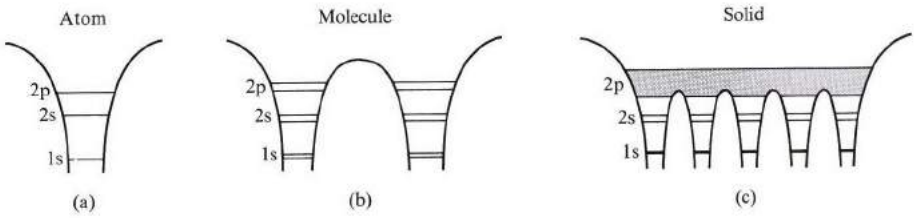


Fig. 5.1 The evolution of the energy spectrum of Li from an atom (a), to a molecule (b), to a solid (c).

energy levels, recognizing that each of these is, in fact, composed of two sublevels.

We can see why the atomic level splits into two, and only two, sublevels in a diatomic molecule from our treatment of the hydrogen molecule ion H_2^+ (Section A.7). The reason is essentially as follows: When the two Li atoms are far apart, the influence of one atom on an electron in the other atom is very small, and may be treated as a perturbation. In this approximation, the unperturbed levels 1s, 2s, etc., are each doubly degenerate, because an electron in a 1s level, for instance, may occupy that level in either atom; and since there are two atoms, the energy is thus doubly degenerate. This degeneracy is strictly valid only if the interaction between the atoms is neglected entirely. When this interaction is included, the double degeneracy is lifted, and each level is split into its two sublevels. The molecular orbitals corresponding to these sublevels are usually taken to be the symmetric and antisymmetric combinations of the corresponding atomic orbitals, as in the case of H_2^+ (Section A.7).

Each molecular level can accommodate at most two electrons, of opposite spins, according to the exclusion principle. The Li_2 molecule has six electrons; four occupy the 1s molecular doublet, and the other two the lower level of the 2s doublet.

According to this discussion, the amount of splitting depends strongly on the internuclear distance of the two atoms in the molecule. The closer the two nuclei, the stronger the perturbation and the larger the splitting. The splitting also depends on the atomic orbital: The splitting of the 2p level is larger than that of the 2s level, which is larger still than that of the 1s level. The reason is that the radius of the ls orbital, for instance, is very small, and the orbital is therefore tightly bound to its own nucleus. It is not greatly affected by the perturbation. The same is not true for the 2s and 2p orbitals, which have larger radii and are only loosely bound to their own nuclei. It follows that, generally speaking, the higher the energy, the greater the splitting incurred.

The above considerations may be generalized to a polyatomic Li molecule of an arbitrary number of atoms. Thus in a 3-atom molecule, each atomic level is split into a triplet, in a 4-atom molecule into a quadruplet, and so forth. The lithium solid may then be viewed as the limiting case in which the number of atoms has

become very large, resulting in a gigantic lithium molecule. What has happened to the shape of the energy spectrum? We can answer this on the basis of the above discussion: Each of the atomic levels is split into N closely spaced sublevels, where N is the number of atoms in the solid. But since N is so very large, about 10^{23}, the sublevels are so extremely close to each other that they coalesce, and form an *energy band.* Thus the 1s, 2s, 2p levels give rise, respectively, to the 1s, 2s, and 2p bands, as shown in Fig. 5.1(c).

To illustrate how close to each other the sublevels lie within the bands, consider the following numerical example. Suppose that the width of the band is 5 eV (a typical value). The energy interval between two adjacent levels is therefore of the order $5/10^{23} = 5 \times 10^{-23}$ eV. Since this is an extremely small value, the individual sublevels are indistinguishable, so we can consider their distribution as a continuous energy band.

To recapitulate, the spectrum in a solid is composed of a set of energy bands. The intervening regions separating these bands are energy *gaps*—i.e., regions of forbidden energy—which cannot be occupied by electrons. Contrast this situation with that of a free atom or a molecule, in which the allowed energies form a set of discrete levels. This broadening of discrete levels into bands is one of the most fundamental properties of a solid, and one we shall use often throughout this book.

The width of the band varies, but in general the higher the band the greater its width, because, as we recall from the case of molecules, a high energy state corresponds to a large atomic radius, and hence a strong perturbation, which is the cause of the level broadening in the first place. By contrast, low energy states correspond to tightly bound orbitals, which are affected but slightly by the perturbation.

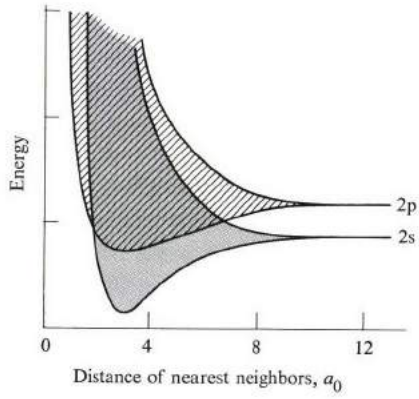


Fig. 5.2 The broadening of the 2s and 2p levels into energy bands in a lithium crystal (a_0 is the Bohr radius, 0.53 Å).

Figure 5.2 shows 2s and 2p bands for metallic lithium plotted as functions of the lattice constants a. Note that the band widths increase as a decreases, as is to be expected, since the smaller the interatomic distance the greater the perturbation. Note also that, for $a < 6a_0$, the 2s and 2p bands broaden to the point at which they begin to overlap, and the gap between them vanishes entirely.

The *crystal orbitals*—i.e., the wave functions describing the electronic states in the bands—extend throughout the solid, unlike the atomic orbitals, which are *localized* around particular atoms, and decay exponentially away from those atoms. In this sense, we refer to solid wave functions as *delocalized orbitals*. We shall see shortly that these orbitals actually describe electron waves traveling in the solid. The concept of delocalization is a basic one. It is responsible for all electronic transport phenomena in solids, e.g., electrical conduction.

We have already presented many concepts related to electronic states in a crystalline solid. In the following sections we shall place these concepts on a firmer, more mathematical basis by writing the Schrödinger equation and discussing the properties of its solution. This will also lead to many interesting and novel concepts which we shall discuss as we go along.

5.3 ENERGY BANDS IN SOLIDS; THE BLOCH THEOREM

The Bloch function

The behavior of an electron in a crystalline solid is determined by studying the appropriate Schrödinger equation. This may be written as (Section A.2),

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}), \tag{5.1}$$

where $V(\mathbf{r})$ is the crystal potential "seen" by the electron, and $\psi(\mathbf{r})$ and E are, respectively, the state function and energy of this electron. The potential $V(\mathbf{r})$ includes the interaction of the electron with all atoms in the solid, as well as its interaction with other electrons (we will get back to this later). At this point we make the important observation that the potential $V(\mathbf{r})$ is periodic. It has the same

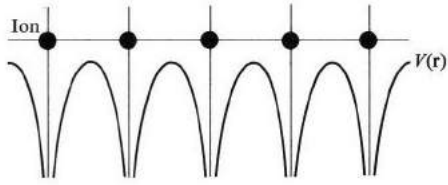


Fig. 5.3 The crystal potential seen by the electron.

translational symmetry as the lattice, that is,

$$V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r}), \tag{5.2}$$

where $\mathbf{R}$ is a lattice vector. Such a potential is shown schematically in Fig. 5.3.

According to the *Bloch theorem*, the solution of (5.1) for a periodic potential $V(\mathbf{r})$ has the form

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \tag{5.3}$$

where the function $u_{\mathbf{k}}(\mathbf{r})$ has the same translational symmetry as the lattice, that is,

$$u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{\mathbf{k}}(\mathbf{r}). \tag{5.4}$$

The vector $\mathbf{k}$ is a quantity related to the momentum of the particle, as we shall see.

We shall now give a physical proof of the Bloch theorem. Anyone interested may pursue the more rigorous treatment in the references cited in the bibliography, e.g., Seitz (1940). The proof presented here is chosen to bring out the physical concepts with a minimum of mathematical detail. Returning to Eq. (5.1), it is always possible to write its solution as

$$\psi(\mathbf{r}) = f(\mathbf{r})\, u(\mathbf{r}),$$

where $u(\mathbf{r})$ is periodic, as in (5.4), and where the function $f(\mathbf{r})$ is to be determined. However, since the potential $V(\mathbf{r})$ is periodic, one requires that all observable quantities associated with the electron also be periodic. In particular, the quantity $|\psi(\mathbf{r})|^2$, which gives the electron probability, must also be periodic.† This imposes the following condition on $f(\mathbf{r})$:

$$|f(\mathbf{r} + \mathbf{R})|^2 = |f(\mathbf{r})|^2.$$

The only function which satisfies this requirement for all $\mathbf{R}$'s is one of the exponential form $e^{i\mathbf{k}\cdot\mathbf{r}}$. This demonstrates that the solution of the Schrödinger equation has the Bloch form (5.3), as we set out to prove.

The state function $\psi_{\mathbf{k}}$ of the form (5.3), known as the *Bloch function*, has several interesting properties.

a) It has the form of a traveling plane wave, as represented by the factor $e^{i\mathbf{k}\cdot\mathbf{r}}$, which implies that the electron propagates through the crystal like a free particle. The effect of the function $u_{\mathbf{k}}(\mathbf{r})$ is to modulate this wave so that the amplitude oscillates periodically from one cell to the next, as shown in Fig. 5.4, but this does not affect the basic character of the state function, which is that of a traveling wave.

†It is well known in quantum mechanics that the quantity $|\psi(\mathbf{r})|^2$ is the probability density, and as such is physically measurable. However, the wave function $\psi(\mathbf{r})$ itself is *not* physically measurable.

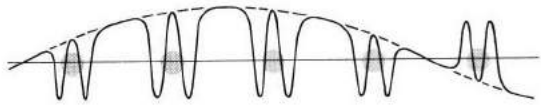


Fig. 5.4 The Bloch function or wave. The smooth curve represents the wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ which is modulated by the atomic-like "wiggly" function $u_{\mathbf{k}}(\mathbf{r})$.

If the electron were indeed entirely free, the state function $\psi_{\mathbf{k}}$ would be given by $(1/V^{1/2})\, e^{i\mathbf{k}\cdot\mathbf{r}}$, that is, the function $u_{\mathbf{k}}(\mathbf{r})$ is a constant. But the electron is not free, since it interacts with the lattice, and this interaction determines the special character of the periodic function $u_{\mathbf{k}}$.

b) Because the electron behaves as a wave of vector $\mathbf{k}$, it has a deBroglie wavelength $\lambda = 2\pi/k$, and hence a *momentum*

$$\mathbf{p} = \hbar\mathbf{k}, \tag{5.5}$$

according to the deBroglie relation. We shall call the vector the *crystal momentum* of the electron, and discuss its properties in later sections.

c) The Bloch function $\psi_{\mathbf{k}}$ is a crystal orbital, as it is delocalized throughout the solid, and not localized around any particular atom. Thus the electron is shared by the whole crystal. This is, of course, consistent with property (a) above, in which we described the electron as a traveling wave. Note also that the function $\psi_{\mathbf{k}}$ is so chosen that the electron probability distribution $|\psi_{\mathbf{k}}|^2$ is periodic in the crystal.

In the above discussion, we have stressed the analogy between a *crystalline* electron and a *free* one; this is very helpful in understanding the properties of electrons in crystals. One should not, however, jump to the conclusion that the two are identical in their behavior. The Bloch-function electron exhibits many intriguing properties not shared by a free electron, properties which result from the interaction of the electron with the lattice.

Energy bands

The discussion has thus far centered on the state function; nothing has been said about energy. We now turn to the energy spectrum which results from solving the Schrödinger equation (5.1). Toward this end, we rewrite this equation in a different form. Substituting for $\psi_{\mathbf{k}}$ from the Bloch form (5.3), and eliminating the factor $e^{i\mathbf{k}\cdot\mathbf{r}}$, after performing the necessary operations, we arrive at

$$\left[-\frac{\hbar^2}{2m} (\nabla + i\mathbf{k})^2 + V(\mathbf{r}) \right] u_{\mathbf{k}}(\mathbf{r}) = E_{\mathbf{k}}\, u_{\mathbf{k}}(\mathbf{r}), \tag{5.6}$$

which is actually the wave equation for the periodic function $u_{\mathbf{k}}(\mathbf{r})$. This is an eigenvalue equation, like the Schrödinger equation, and can therefore be solved in a

similar manner. Note that the operator in the brackets is an explicit function of $\mathbf{k}$, and hence both the eigenfunctions and eigenvalues depend on $\mathbf{k}$, a fact we have already used explicitly by labeling them with the vector $\mathbf{k}$. An eigenvalue equation leads, however, not to one but to many solutions. For each value of $\mathbf{k}$, therefore, we find a large number of solutions, giving a set of *discrete* energies $E_{1,\mathbf{k}}, E_{2,\mathbf{k}}, \ldots$, as shown in Fig. 5.5.† Since these energies depend on $\mathbf{k}$, they vary continuously as $\mathbf{k}$ is varied over its range of values. Each level leads to an energy band, as shown in the figure. We shall henceforth write the energy eigenvalue as $E_n(\mathbf{k})$, and refer to the subscript n as the *band index*, for obvious reasons.

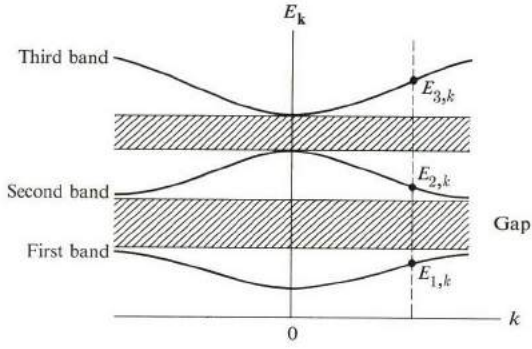


Fig. 5.5 Energy bands and gaps. The cross-hatched regions indicate energy gaps.

The number of bands is large—usually infinite—but only the lowest ones are occupied by electrons. Each band covers a certain energy range, extending from the lowest to the highest value it takes when plotted in $\mathbf{k}$-space. The energy intervals interspersed between the bands constitute the energy gaps, which are forbidden energies that cannot be occupied by electrons.

Note also that, since $\mathbf{k}$ is a vector quantity, a diagram such as Fig. 5.5 is a plot of the energy bands in only one particular direction in $\mathbf{k}$-space. If these bands were plotted in a different $\mathbf{k}$-direction, their appearance would change, in general. A complete representation of the bands therefore requires one to specify the energy values throughout the $\mathbf{k}$-space. Often this is accomplished, at least partially, by drawing the energy contours in $\mathbf{k}$-space for the various bands, as we shall do in the following sections. We shall also show that the bands satisfy certain important symmetry relations that enable us to restrict our considerations to relatively small regions in $\mathbf{k}$-space.

The energy bands which have emerged from this analysis are the same as those discussed in the previous section, and in fact we can establish a one-to-one correspondence between the energy bands and the atomic levels from which they arise. The particular significance of the present results is that here we can classify

† In other words, the energy is a multivalued function of $\mathbf{k}$.

the electron states within the band according to their momentum as given by $\mathbf{k}$. Such a classification, which we shall find extremely useful, was not evident from the last section.

The crystal potential

We turn now to the crystal potential $V(\mathbf{r})$ which acts on the electron. This potential is composed of two parts: the interaction of the electron with the ion cores, forming the lattice, and its interaction with other Bloch electrons moving through the lattice. In metallic sodium, for example, an electron in the 3s band interacts with the Na^+ ions forming the bcc structure, as well as with other electrons in this band. We may therefore write $V(\mathbf{r})$ as the sum

$$V(\mathbf{r}) = V_i(\mathbf{r}) + V_e(\mathbf{r}), \tag{5.7}$$

where the first term on the right represents the interaction with the ion cores and the second the interaction with the electrons.

The ionic part may be written as

$$V_i(\mathbf{r}) = \sum_j v_i(\mathbf{r} - \mathbf{R}_j), \tag{5.8}$$

where $v_i(\mathbf{r} - \mathbf{R}_j)$ is the potential of an ion located at the lattice vector $\mathbf{R}_j$, as in Fig. 5.6(a). and the summation is over all the ions. The potential $V_i(\mathbf{r})$ obviously has the same periodicity as that of the lattice.

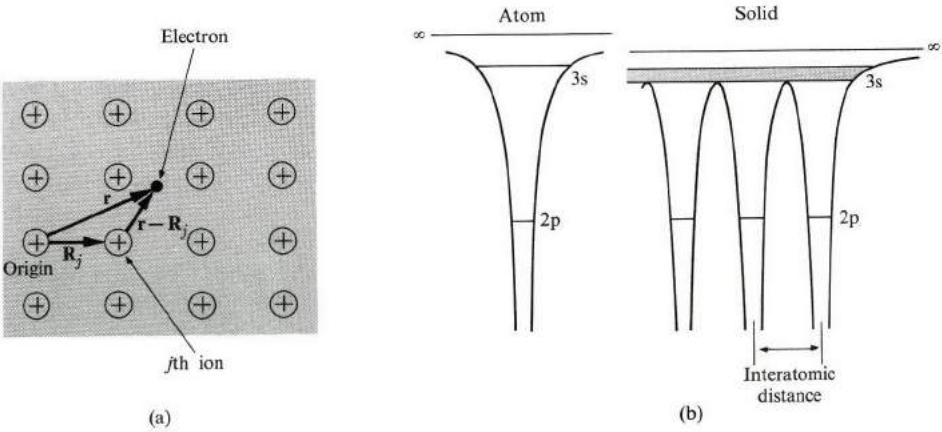


Fig. 5.6 (a) The interaction of an electron with ion cores. The small dots represent electrons. (The spatial distribution of the electrons is not shown accurately. They actually tend to be positioned primarily around the ions.) (b) The spectrum of an Na atom (left), and an Na solid (right). [After J. C. Slater, *Physics Today* **21**, 43 (1968)]. Note the broadening of the 3s level into a 3s band in the solid, and that this band lies almost entirely above the potential barriers of the atoms, which facilitates the delocalization of the electrons in this band. By contrast, electrons in the 2p level or band are so highly constrained by the barriers that they are localized.

The electronic potential $V_e(\mathbf{r})$, the so-called *electron–electron interaction*, presents several hurdles which make its treatment very difficult. First, we can evaluate this term only if we know the states for all other electrons, but these states are not given in advance. In fact, they are the very states we are trying to find. Second, the potential $V_e(\mathbf{r})$ is not strictly periodic, since the electrons are in constant motion through the lattice. Third, a proper treatment should really consider the dynamics of all the electrons simultaneously, not one electron at a time, as we have done above. This is a typical example of the *many-body problems* which are often encountered in solid-state physics.

In view of these difficulties, it is fortunate that the electron–electron interaction turns out to be quite weak, for the reason given in Section 4.3, because this fact makes the above difficulties far less serious than they could otherwise be. The major effect of this interaction is that the electrons distribute themselves primarily around the ions, so that they *screen* these ions from other electrons. This has the additional effect of making the electron–ion interaction weak even at long range, which is another fortunate circumstance.

So we can write an approximate expression for the potential as

$$V(\mathbf{r}) = \sum_j v_s(\mathbf{r} - \mathbf{R}_j), \tag{5.9}$$

where $v_s(\mathbf{r} - \mathbf{R}_j)$ is the potential of the screened ion located at the lattice point $\mathbf{R}_j$. And precisely because this potential *is* once again periodic, it satisfies the requirements of the Bloch theorem. Figure 5.6(b) shows the crystal potential for Na.

In discussing the crystal potential, we have so far tacitly assumed that the atoms are at rest at their lattice sites. However, they are not in fact stationary. They are in a constant state of oscillation as a result of their thermal excitation, as discussed in Chapter 3. Clearly, then, our assumption of a stationary lattice is an approximation, and the question now is: How good is our approximation? One may answer this pragmatically by pointing out that band structures calculated on the basis of a stationary lattice are usually in good agreement with experiment, except at temperatures close to the melting point of the solid. The reason the stationary-lattice approximation seems to hold so well is that amplitudes of lattice vibrations are much smaller than the interatomic distance at all temperatures, even up to the melting point.† Therefore the distortion of the lattice, as seen by the electron, is not appreciable.

5.4 BAND SYMMETRY IN k-SPACE; BRILLOUIN ZONES

The energy eigenvalues $E_n(\mathbf{k})$ for the bands have many useful symmetry properties when these bands are plotted in $\mathbf{k}$-space. Before broaching this subject, however, let us say a few words about the Brillouin zones.

† The average amplitude of the atomic oscillation due to thermal excitation at the melting point is typically about 5% of the interatomic distance.

Brillouin zones

We first encountered Brillouin zones in our discussion of Bragg diffraction of x-rays in Section 2.6. When one draws the normal planes which bisect the reciprocal lattice vectors, the regions enclosed between these planes form the various Brillouin zones.

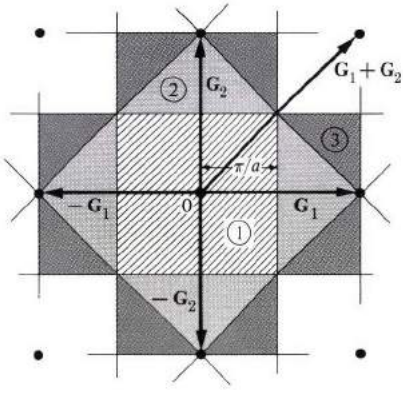


Fig. 5.7 The first three Brillouin zones of the square lattice: First zone (cross-hatched), second zone (shaded) and third zone (screened). Numbers indicate indices of zones.

Consider, for instance, the square lattice whose reciprocal—also a square lattice of edge equal to $2\pi/a$—is shown, in Fig. 5.7, which also shows the reciprocal vectors $\mathbf{G}_1$, $-\mathbf{G}_1$, $\mathbf{G}_2$, and $-\mathbf{G}_2$, etc., as well as the corresponding normal bisectors. The smallest enclosed region centered around the origin (the cross-hatched area) is the first zone. The shaded area (composed of four separate half-

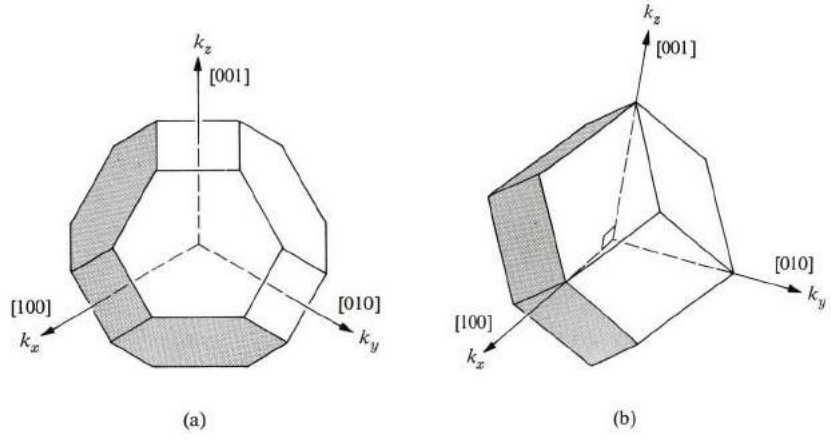


Fig. 5.8 The first Brillouin zone for (a) an fcc lattice, and (b) a bcc lattice.

diamond-shaped pieces enclosed between the normal bisectors to $\mathbf{G}_1$, $\mathbf{G}_2$, and $\mathbf{G}_1 + \mathbf{G}_2$, etc.) forms the second zone. Similarly, the screened area (eight parts) forms the third zone. As higher-order bisectors are included, higher-order zones are also formed, which may have quite complicated shapes.

However, *all the zones have the same area*, regardless of the complexity of the zone. Thus we can see in the figure that the second zone has the same area as the first, that is, $(2\pi/a)^2$. The same is true for the third zone, and this can also be shown to hold true for all zones. This equality of the areas of the Brillouin zones holds true for all plane lattices, not just for square lattices.

In three dimensions, the zones are three-dimensional volumes. Figure 5.8 shows the first zone for fcc (a truncated octahedron) and bcc (a regular rhombic dodecahedron) lattices. Higher-order zones in these lattices are somewhat complicated in appearance and difficult to visualize; they will not concern us further here.

Let us now discuss the relation of the Brillouin zones to the band structure.

Symmetry properties

It can be shown that each energy band $E_n(\mathbf{k})$ satisfies the following symmetry properties.

i) $$E_n(\mathbf{k} + \mathbf{G}) = E_n(\mathbf{k}) \tag{5.10}$$

ii) $$E_n(-\mathbf{k}) = E_n(\mathbf{k}) \tag{5.11}$$

iii) $E_n(\mathbf{k})$ has the same rotational symmetry as the real lattice.

Note that these properties are the same as those obeyed by the dispersion relations of lattice vibrations (Section 3.6), and can be proved in a similar manner—i.e., by invoking the symmetry properties of the real lattice—as will be discussed later in this section.

Property (i) indicates that $E_n(\mathbf{k})$ is periodic, with a period equal to the reciprocal lattice vector. In other words, any two points in $\mathbf{k}$-space related to each other by a displacement equal to a reciprocal lattice vector have the same energy. For instance, in Fig. 5.9(a), the energy is the same at points P_1, P_2, and P_3, because P_2 is related to P_1 by a translation equal to $-\mathbf{G}_2$, P_3 is related to P_1 by a translation $-\mathbf{G}_1$, and both $-\mathbf{G}_1$ and $-\mathbf{G}_2$ are reciprocal lattice vectors.

Figure 5.9(b) illustrates how, by using this translational symmetry, the various pieces of the second zones may be translated by reciprocal lattice vectors to fit precisely over the first zone. Each two areas connected by an arrow are *equivalent*. The first and second zones are equivalent. Similarly, higher-order zones can be appropriately translated to fit over the first zone. It follows, therefore, that we may confine our attention to the first zone only, since this contains all the necessary information.

The inversion property (ii) shows that the band is symmetric with respect to inversion around the origin $\mathbf{k} = 0$. Thus, in Fig. 5.9(a), the energy at point P_1' is equal to that at P_1.

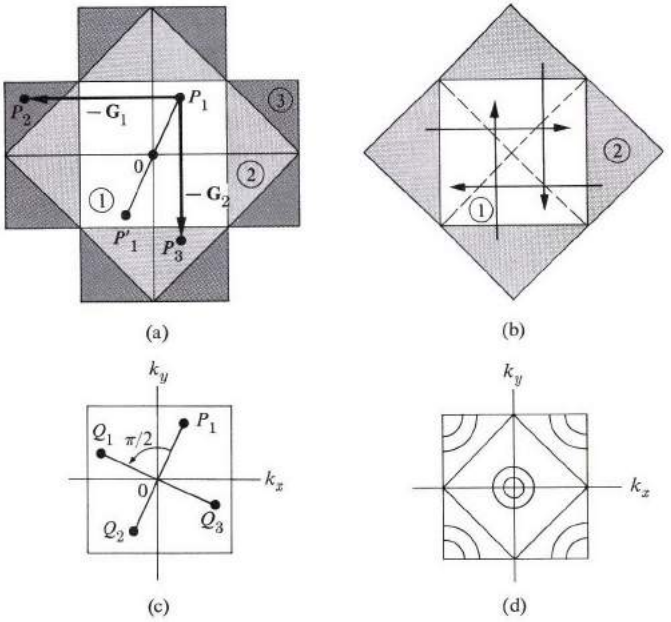


Fig. 5.9 (a) Translational symmetry of the energy $E(\mathbf{k})$ in $\mathbf{k}$-space for a square lattice. (b) Mapping of the second zone into the first. (c) Rotational symmetry of $E(\mathbf{k})$ in $\mathbf{k}$-space for a square lattice. (d) Energy contours in the first zone.

Property (iii) asserts that the band has the same rotational symmetry as the real lattice. For instance, in a square lattice, the energy should exhibit the rotational symmetry of the square. Since this is symmetric with respect to a rotation by $\pi/2$ (and its multiples), it follows that in Fig. 5.9(c) the energies at points Q_1, Q_2, and Q_3 are equal to that at point P_1, because these points may be obtained from P_1 by symmetry rotations. [Note that Q_2 is the same as P'_1 of Fig. 5.9(a); this is so for a square lattice, but it does not hold good for other lattices.]

In Fig. 5.9(d) energy contours are sketched for a band in the first zone of a square lattice. This figure satisfies the various symmetry properties described above.

The symmetry properties are particularly important because we can use them to reduce the labor involved in determining energy bands. For example, with inversion symmetry, we need to know the band in only half of the first zone, and rotational symmetry usually enables us to reduce this even further. In the case of a square lattice, for example, only one-eighth of the zone need be specified independently, as you may see, and the remainder of the zone can then be completed by using symmetry properties.

The labor-saving is even greater in three-dimensional cases. Thus, in the case of a cubic lattice, the band need be specified independently in only 1/48th of the first zone.

Note that the symmetry properties discussed above refer to the same band. They hold for every band separately, but do not relate one band to another.

Let us turn now to the proofs of the above properties. We shall only outline these proofs here, leaving you to pursue the details in some of the advanced references listed at the end of the chapter. Consider first the translational property (i): The Bloch function at the point $\mathbf{k} + \mathbf{G}$ may be written as

$$\psi_{\mathbf{k}+\mathbf{G}} = e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} u_{\mathbf{k}+\mathbf{G}} = e^{i\mathbf{k}\cdot\mathbf{r}}(e^{i\mathbf{G}\cdot\mathbf{r}} u_{\mathbf{k}+\mathbf{G}}). \tag{5.12}$$

Note that the factor inside the brackets of the last expression, which may be denoted by $v(\mathbf{r})$, is periodic in the $\mathbf{r}$-space with a period equal to the lattice vector. That is,

$$v(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{G}\cdot(\mathbf{r}+\mathbf{R})} u_{\mathbf{k}+\mathbf{G}}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{G}\cdot\mathbf{r}} u_{\mathbf{k}+\mathbf{G}}(\mathbf{r}) = (v)\mathbf{r}.$$

This follows from the fact that $u_{\mathbf{k}+\mathbf{G}}$ is periodic, and $e^{i\mathbf{G}\cdot\mathbf{R}} = 1$, since $\mathbf{G}\cdot\mathbf{R} = n2\pi$, where n is some integer. The expression in the brackets in (5.12) has, therefore, the same behavior as $u_{\mathbf{k}}(\mathbf{r})$ in Eq. (5.3). We have thus shown that the state function $\psi_{\mathbf{k}+\mathbf{G}}$ has the same form as $\psi_{\mathbf{k}}$, and consequently the two functions have the same energy, since there is no physical basis for distinguishing between them.

Property (ii) may be established by noting that the Schrödinger equation analogous to (5.6), which corresponds to the point $-\mathbf{k}$, is the same as the equation obtained by writing the complex conjugate equation of (5.6). This means that the corresponding eigenvalues are equal, that is, that $E_n(-\mathbf{k}) = E_n^*(\mathbf{k})$. Since the energy $E_n(\mathbf{k})$ is a real number, however, it follows that $E_n(-\mathbf{k}) = E_n(\mathbf{k})$, which is property (ii).

Property (iii) is derived by noting that if the real lattice is rotated by a symmetry operation, the potential $V(\mathbf{r})$ remains unchanged, and hence the new state function obtained must have the same energy as the original state function. One may show further that these new states correspond to rotations in $\mathbf{k}$-space, and this leads to the desired property.

5.5 NUMBER OF STATES IN THE BAND

We denoted the Bloch function by $\psi_{n,\mathbf{k}}$, which indicates that each value of the band index n and the vector $\mathbf{k}$ specifies an electron state, or orbital. We shall now show that *the number of orbitals in a band inside the first zone is equal to the number of unit cells in the crystal.* This is much the same as the statement made in connection with the number of lattice vibrational modes (Section 3.3), and is proved in a like manner, by appealing to the boundary conditions.

Consider first the one-dimensional case, in which the Bloch function has the form

$$\psi_k(x) = e^{ikx} u_k(x). \tag{5.13}$$

If we impose the periodic boundary condition on this function, it follows that the only allowed values of k are given by

$$k = n\frac{2\pi}{L}, \tag{5.14}$$

where $n = 0, \pm 1, \pm 2$, etc. [Note that $u_k(x)$ is intrinsically periodic, so the condition $u_k(x + L) = u_k(x)$ is automatically satisfied.] As in Section 3.3, the allowed values of k form a uniform mesh whose unit spacing is $2\pi/L$. The number of states inside the first zone, whose length is $2\pi/a$, is therefore equal to

$$(2\pi/a)/(2\pi/L) = L/a = N,$$

where N is the number of unit cells, in agreement with the assertion made earlier.

A similar argument may be used to establish the validity of the statement in two- and three-dimensional lattices.

It has been shown that each band has N states inside the first zone. Since each such state can accommodate at most two electrons, of opposite spins, in accordance with the Pauli exclusion principle, it follows that the maximum number of electrons that may occupy a single band is $2N$. This result is significant, as it will be used in a later section to establish the criterion for predicting whether a solid is going to behave as a metal or an insulator.

5.6 THE NEARLY-FREE-ELECTRON MODEL

In Section 5.3 and 5.4 we studied the general properties of the state functions, and of the energies of an electron moving in a crystalline solid. To obtain explicit results, however, we must solve the Schrödinger equation (5.1) for the actual potential $V(\mathbf{r})$ in the particular solid of interest. But the process of solving the Schrödinger equation for any but the simplest potentials is an arduous and time-consuming task, inundated with mathematical details. Although this is essential for obtaining results that may be compared with experiments, it is preferable to start the discussion of explicit solutions by using rather simplified potentials. The advantage is that we can solve the Schrödinger equation with only minimal mathematical effort and thus concentrate on the new physical concepts involved.

In the present section we shall treat the *nearly-free-electron* (NFE) *model*, in which it is assumed that the crystal potential is so weak that the electron behaves essentially like a free particle. The effects of the potential are then treated by the use of perturbation methods, which should be valid inasmuch as the potential is weak. This model should serve as a rough approximation to the valence bands in the simple metals, that is, Na, K, Al, etc.

In the following section, we shall treat the *tight-binding model*, in which the atomic potentials are so strong that the electron moves essentially around a single atom, except for a small interaction with neighboring atoms, which may then be treated as a perturbation. This model lies at the opposite end from the NFE model in terms of the strength of crystal potential involved, and should serve as a rough approximation to the narrow, inner bands in solids, e.g., the 3d band in transition metals.

The empty-lattice model

The starting point for the NFE model is the solution of the Schrödinger equation for the case in which the potential is exactly zero, i.e., the electron is entirely free. However, we also require that the solutions satisfy the symmetry properties of Section 5.4, which are imposed by the translational symmetry of the real lattice. This leads to the so-called *empty-lattice model.*

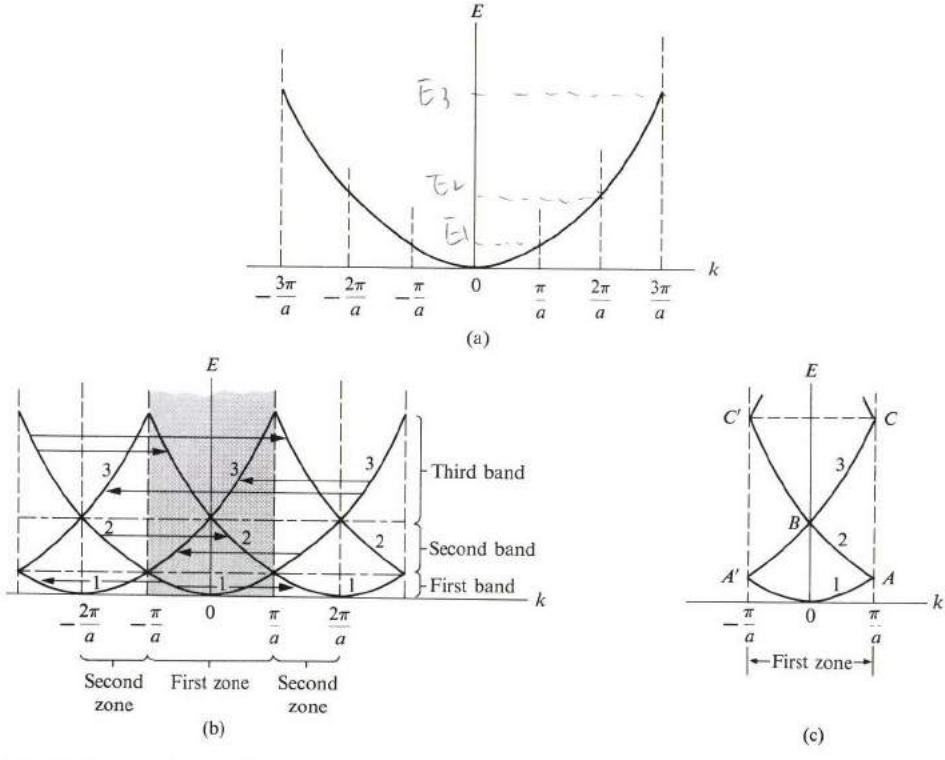


Fig. 5.10 (a) The familiar parabola representing the dispersion curve for a free particle, $E = \hbar^2 k^2 / 2m_0$. (b) The dispersion curves for the same particle in the empty-lattice model, showing translational symmetry and the various bands. (c) Dispersion curves in the empty-lattice model (first zone only).

For a one-dimensional lattice, the state functions and energies for the empty-lattice model are

$$\psi_k^{(0)} = \frac{1}{L^{1/2}} e^{ikx}, \tag{5.15}$$

and

$$E_{(k)}^{(0)} = \frac{\hbar^2 k^2}{2m_0}, \tag{5.16}$$

where the superscript 0 indicates that the solutions refer to the unperturbed state (Section A.7). The energy $E_{(k)}^{(0)}$ which is plotted versus k in Fig. 5.10(a) exhibits a curve in the familiar *parabolic* shape. Figure 5.10(b) shows the result of imposing the symmetry property (i) of Section 5.4. Segments of the parabola of Fig. 5.10(a) are cut at the edges of the various zones, and are translated by multiples of $G = 2\pi/a$ in order to ensure that the energy is the same at any two equivalent points. Figure 5.10(c) displays the shape of the energy spectrum when we confine our consideration to the first Brillouin zone only. [Conversely, Fig. 5.10(b) may be viewed as the result of translating Fig. 5.10(c) by multiples of G.]

The type of representation used in Fig. 5.10(c) is referred to as the *reduced-zone scheme*. Because it specifies all the needed information, it is the one we shall find most convenient. The representation of Fig. 5.10(a), known as the *extended-zone scheme*, is convenient when we wish to emphasize the close connection between a crystalline and a free electron. However, Fig. 5.10(b) employs the *periodic-zone* scheme, and is sometimes useful in topological considerations involving the **k**-space. All these representations are strictly equivalent; the use of any particular one is dictated by convenience, and not by any intrinsic advantages it has over the others.

The nearly-free-electron model

How is the energy spectrum of Fig. 5.10(c) altered when the crystal potential is taken into account, or "turned on?" Figure 5.11(a) shows this. The first and second bands, which previously touched at the point A (and A') in Fig. 5.10(c) are now split, so that an energy gap is created at the boundary of the Brillouin zone. A similar gap is created at the center of the zone, where bands 2 and 3 previously intersected (point B in Fig. 5.10c) and also at point C, where bands 3 and 4 previously intersected. Thus, in general, in the empty-lattice model, energy gaps are created in **k**-space wherever bands intersect, which occurs either at the center or the boundaries of the BZ. At these points the shape of the spectrum is strongly modified by the crystal potential, weak as this may be. (In effect, what the crystal potential has accomplished is to smooth over the sharp "corners" present in the band structure of the empty lattice.)

In the remainder of the zone, however, the shape of the spectrum is affected very little by the crystal potential, since this is assumed to be weak. In that region

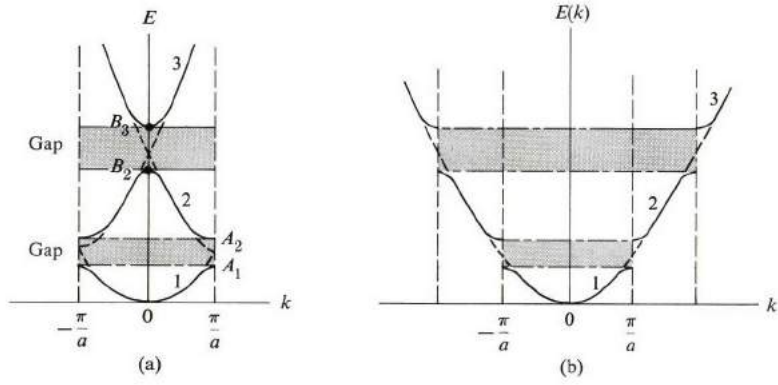


Fig. 5.11 (a) Dispersion curves in the nearly-free-electron model, in the reduced-zone scheme. (b) The same dispersion curves in the extended-zone scheme.

of the **k**-space the bands essentially retain their parabolic shape inherited from the empty-lattice model of Fig. 5.10(c), and the electron there behaves essentially like a free electron.

By comparing Fig. 5.10(c) and Fig. 5.11(a), one notes that a hint of a band structure is almost present even in the empty-lattice model, except that the gaps there vanish, since the bands touch at the zone boundaries. This vanishing is foreseen, of course, since no energy gaps are expected to appear in the spectrum of a free particle. The point is that even a weak potential leads to the creation of gaps, in agreement with the results of Sections 5.2 and 5.3.

Figure 5.11(b) shows the band structure for the NFE model, represented according to the extended-zone scheme, which should be compared with Fig. 5.10(a). Note that, except at the zone boundaries at which gaps are created, the dispersion curve is essentially the same as the free-electron curve.

We made the above assertions without proofs; we shall now outline proofs on the basis of the perturbation method of Section A.7. Suppose, for instance, that we seek to find the influence of the crystal potential on the first band in Fig. 5.10(c). When we treat the potential $V(x)$ as a perturbation, the perturbed energy $E_1(k)$ up to the second order of the potential is given by

$$E_1(k) = E_1^{(0)}(k) + \langle \psi_{1,k}^{(0)} | V | \psi_{1,k}^{(0)} \rangle + \sum_{k',n}{}' \frac{|\langle n, k' | V | 1, k \rangle|^2}{E_1^{(0)}(k) - E_n^{(0)}(k')}. \tag{5.17}$$

Here the subscript 1 refers to the first band, which is the one of interest, and the superscript 0 refers to the empty-lattice model of Eqs. (5.15) and (5.16). The second term on the right side of (5.17), which is the first-order correction, is the average value of the potential. The third term, giving the second-order correction, involves summing over all states n, k, except where these indices are equal to the state 1, k under investigation.

First we note that the first-order correction is equal to

$$\langle \psi^{(0)}_{1,k} \mid V \mid \psi^{(0)}_{1,k} \rangle = \frac{1}{L}\int e^{-ikx} V(x)\, e^{ikx}\, dx = \frac{1}{L}\int V(x)\, dx,$$

which is the average value of the potential over the entire lattice. It is independent of k, and hence it is merely a constant. Its effect on the spectrum of Fig. 5.10(c) is simply to displace it rigidly by a constant amount, without causing any change in the shape of the energy spectrum. Since this term does not lead to anything of interest to us here, it will be set equal to zero, which can be accomplished by shifting the zero energy level.

We must therefore consider the second-order correction in Eq. (5.17). We first assert that the quantity $\langle n, k' \mid V \mid 1, k \rangle$ can be shown to vanish except when $k' = k$, where both k and k' are restricted to the first zone. That is, the only states which are coupled to the $1, k$ state by the perturbation are those lying directly above this state, as shown in Fig. 5.12.

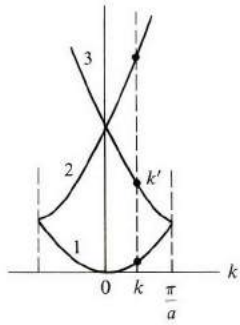


Fig. 5.12 Only those states lying directly above the state $\psi^{(0)}_{1,k}$ in **k**-space are coupled to it by the perturbation.

This assertion rests on the translational symmetry of the crystal potential $V(x)$.† Furthermore, since the energy difference in the denominator of the third

† An arbitrary potential $V(x)$ can always be expanded as a Fourier series

$$V(x) = \textstyle\sum_k V_k\, e^{ikx},$$

where the summation is over all the allowed k's. The Fourier coefficient V_k is given by

$$V_k = (1/L)\int_0^L V(x)\, e^{-ikx}\, dx.$$

But if $V(x)$ is periodic, as is the case in a crystal, then only the values $k = G$ contribute to the above summation; that is, $V_k = 0$ for $k \neq G$. A periodic potential therefore has the expansion

$$V(x) = \textstyle\sum_G V_G\, e^{iGx}.$$

It can be shown that the bracket in the numerator of (5.17) is the Fourier coefficient $V_{k'-k}$, and hence this bracket vanishes except for $k' - k = G$.

term in (5.17) increases rapidly as the band n rises, the major effect on band 1 arises from its coupling to band 2. We may therefore write

$$E_1(k) \simeq E_1^{(0)}(k) + \frac{|V_{-2\pi/a}|^2}{E_1^{(0)}(k) - E_2^{(0)}(k)}, \tag{5.18}$$

where $V_{-2\pi/a}$ is the Fourier component of the potential, that is,

$$V_{-2\pi/a} = \frac{1}{L}\int V(x)\, e^{+i(2\pi/a)x}\, dx.$$

An explicit expression for $E_1(k)$ can be obtained by substituting the values for $E_1^{(0)}(k)$ and $E_2^{(0)}(k)$: namely $E_1^{(0)}(k) = \hbar^2k^2/2m$, and $E_2^{(0)}(k) = \hbar^2(k - 2\pi/a)^2/2m$. [Note that if $0 < k < \pi/a$, then the second band is obtained by translating that part of the free-electron curve lying in the interval $-2\pi/a < k < -\pi/a$, as seen in Fig. 5.10(b), and hence the above expression for $E_2^{(0)}(k)$.] But this is not really necessary, because if the potential is weak, then $|V_{-2\pi/a}|^2$ is very small, and the second term in (5.18) is negligibly small compared with the first. In other words, $E_1(k) \simeq E_1^{(0)}(k)$, and the effect of the lattice potential is negligible.

There is, however, one point in k-space at which the above conclusion breaks down: the point $k = \pi/a$ at the zone edge. At this point the energies $E_1^{(0)}(k)$ and $E_2^{(0)}(k)$ are equal [recall that bands 1 and 2 touch there; see Fig. 5.10(c)], the denominator of the perturbation term in (5.18) vanishes, and hence the perturbation correction becomes very large. Since the above perturbation theory presumes the smallness of the correction, it follows that this theory cannot hold true in the neighborhood of the zone edge. In this neighborhood, one should instead invoke the *degenerate* perturbation theory, in which both bands 1 and 2 are treated simultaneously, and on an equal footing. The resulting energy values are (Ziman, 1963),

$$E_\pm(k) = \tfrac{1}{2}\{E_1^{(0)}(k) + E_2^{(0)}(k) \pm [(E_2^{(0)}(k) - E_1^{(0)}(k))^2 + 4|V_{-2\pi/a}|^2]^{1/2}\}, \tag{5.19}$$

where the plus sign corresponds to the deformed upper band—i.e., band 2—near the edge of the zone, and the minus sign refers to the deformed lower band—i.e., band 1.

Now let us substitute the values of $E_1^{(0)}(k)$ and $E_2^{(0)}(k)$ into (5.19) and plot $E_+(k)$ and $E_-(k)$ in the neighborhood of the zone edge. We obtain the spectrum shown in Fig. 5.11(a). In particular, the energy gap E_g is equal to the difference $E_+(k) - E_-(k)$ evaluated at the point $k = \pi/a$. Using (5.19), we readily find that

$$E_g = 2|V_{-2\pi/a}|. \tag{5.20}$$

That is, the energy gap is equal to twice the Fourier component of the crystal potential. In effect, band 1 has been depressed by an amount equal to $|V_{-2\pi/a}|$ and band 2 has been raised by the same amount, leading to an energy gap given by (5.20).

The same formula (5.19) may also be used to find the energy gap that arises at the center of the zone, at the intersection between bands 2 and 3, except that we now replace $E_1^{(0)}(k)$, $E_2^{(0)}(k)$ by $E_2^{(0)}(k)$ and $E_3^{(0)}(k)$, respectively. We also replace the potential term by $V_{-4\pi/a}$. This leads to the splitting of bands 2 and 3, as shown in Fig. 5.11(a), with an energy gap of $2 \mid V_{-4\pi/a} \mid$. Obviously the procedure can be used to find both the splitting of the bands and the corresponding gaps at all appropriate points.

In addition to the above results, two qualitative conclusions emerge from the analysis. First, the higher the band, the greater its width; this is evident from referring back to the empty lattice model in Fig. 5.10(a), since the energy there increases as k^2. Second, the higher the energy, the narrower the gap; this follows from the fact that the gap is proportional to a certain Fourier component of the crystal potential, but note that the order of the component increases as the energy rises (from $V_{-2\pi/a}$ to $V_{-4\pi/a}$ in our discussion above). Since the potential is assumed to be well behaved, the components decrease rapidly as the order increases, and this leads to a decrease in the energy gap. It follows therefore that, as we move up the energy scale, the bands become wider and the gaps narrower; i.e., the electron behaves more and more like a free particle. This agrees with the qualitative picture drawn in Section 5.2.

Since the greatest effect of the crystal potential takes place near the points in k-space at which two bands touch, let us examine the behavior there more closely. If one applies the degenerate perturbation formula (5.17) to the splitting of bands 2 and 3 at the center of the zone, one finds that, for small k ($k \ll \pi/a$),

$$E_3(k) = E_B + \mid V_{-4\pi/a} \mid + \frac{\hbar^2}{2m_0} \alpha k^2, \tag{5.21}$$

and

$$E_2(k) = E_B - \mid V_{-4\pi/a} \mid - \frac{\hbar^2}{2m_0} \alpha k^2, \tag{5.22}$$

where the parameter α is given by

$$\alpha = 1 + \frac{4E_B}{E_g} \tag{5.23}$$

and $E_B = \hbar^2(2\pi/a)^2/2m_0$ is the energy of point B in Fig. 5.10(c). These results are very interesting for several reasons.

a) Equation (5.21) shows that, for an electron near the bottom of the third band, $E \sim k^2$ (ignoring the first two terms on the right, since they are simply constants), which is similar to the dispersion relation of a free electron. In other words, the electron there behaves like a free electron, with an effective mass m^* given by

$$m^* = m_0/\alpha,$$

which is different from the free mass. Referring to (5.23), one sees that the effective mass increases as the energy gap E_g increases. Such a relationship between m^* and E_g is familiar in the study of semiconductors.

b) Equation (5.22) shows that, for an electron near the top of the second band, $E \sim -k^2$, which is like a free electron, except for the surprising fact that the effective mass is negative. Such behavior is very unlike that of a free electron, and its cause lies, of course, in the crystal potential. The phenomenon of a negative effective mass near the top of the band is a frequent occurrence in solids, particularly in semiconductors, as we shall see later (Chapter 6).

We have thus far confined ourselves to a one-dimensional lattice, but we may extend this treatment to two- and three-dimensional lattices in a straightforward fashion. We find again, as expected, that starting with the empty-lattice model, the "turning on" of the crystal potential leads to the creation of energy gaps. Furthermore, these gaps occur at the boundaries of the Brillouin zone.

5.7 THE ENERGY GAP AND THE BRAGG REFLECTION

In discussing the NFE model, we focused on energy values. But perturbation also modifies state functions, and we shall now study this modification. If we apply the perturbation theory to the one-dimensional empty lattice, we find that the state function of the first band in Fig. 5.11(a) is given by

$$\psi_{1,k} = \psi_{1,k}^{(0)} + \frac{V_{-2\pi/a}}{E_1^{(0)}(k) - E_2^{(0)}(k)} \psi_{2,k}^{(0)}, \tag{5.24}$$

where—again because of the form of the potential and also the energy difference in the denominator—the perturbation summation has been reduced to one term only, involving the state function of the second band $\psi_{2,k}^{(0)}$.

The state functions $\psi_{1,k}^{(0)}$ and $\psi_{2,k}^{(0)}$ refer to a free electron; $\psi_{2,k}^{(0)} \sim e^{ikx}$ represents a wave traveling to the right, while $\psi_{2,k}^{(c)} \sim e^{i(k-2\pi/a)x}$ represents a wave traveling to the left (note that $|k| < \pi/a$). The effect of the lattice potential is then to introduce a new left-traveling wave in addition to the incident free wave. This new wave is generated by the scattering of the electron by the crystal potential. If k is not close to the zone edge, however, the coefficient of $\psi_{2,k}^{(0)}$ in (5.24) is negligible. That is,

$$\psi_{1,k} \simeq \psi_{1,k}^{(0)} = \frac{1}{L^{1/2}} e^{ikx}, \tag{5.25}$$

and the electron behaves like a free electron. The effects of the potential are negligible there, which is in agreement with the conclusions reached in Section 5.6.

Near the zone edge, however, the energy denominator in the correction term in (5.24) becomes very small, and the perturbation term large, which means that

the form (5.24) becomes invalid. As stated in Section 5.6, one must then use the degenerate perturbation theory, in which the state functions $\psi^{(0)}_{1,k}$ and $\psi^{(0)}_{2,k}$ are treated on an equal footing. One finds that, at the zone edge itself,

$$\psi_{\pm}(x) = \frac{1}{\sqrt{2}} [\psi^{(0)}_{1,\pi/a}(x) \pm \psi^{(0)}_{2,\pi/a}(x)] = \frac{1}{\sqrt{2L}} (e^{i(\pi/a)x} \pm e^{-i(\pi/a)x}). \tag{5.26}$$

The function $\psi_+(x) \sim \cos(\pi/a)x$, and hence the probability is proportional to $|\psi_+(x)|^2 \sim \cos^2(\pi/a)x$. Such a state function distributes the electron so that it is piled predominantly at the nuclei (recall that the origin $x = 0$ is at the center of an ion) [see Fig. 5.13], and since the potential is most negative there, this

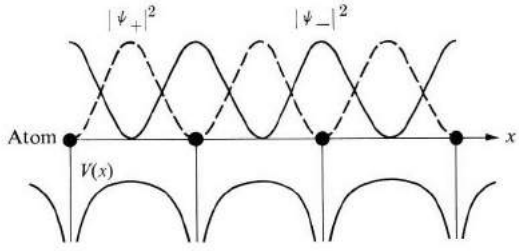


Fig. 5.13 Spatial distributions of electrons described by the functions ψ_+ and ψ_-.

distribution has a low energy. The function $\psi_+(x)$ therefore corresponds to the energy at the top of band 1, that is, point A_1 in Fig. 5.11(a).

By contrast, the function $\psi_-(x) \sim \sin \pi/ax$, depositing its electron mostly between the ions (as shown in Fig. 5.13), corresponds to the bottom of band 2 in Fig. 5.11(a), that is, point A_2. The gap arises, therefore, because of the two different distributions for the same value $k = \pi/a$, the distributions having different energies.

Scrutinizing (5.26) from the viewpoint of scattering, we see that at the zone edge, $k = \pi/a$, the scattering is so strong that the reflected wave has the same amplitude as the incident wave. As found above, the electron is represented there by a *standing* wave, $\cos \pi/ax$ or $\sin \pi/ax$, very unlike a free particle. An interesting result of this is that the electron, as a standing wave, has a zero velocity at $k = \pi/a$. This is a general result which is valid at all zone boundaries, and one which we shall encounter often in the following sections.

We have seen that the periodic potential causes strong scattering at $k = \pi/a$. Recall from Section 3.6 on lattice vibrations that this strong scattering arises as a result of the Bragg diffraction at the zone edge. In the present situation, the wave diffracted is the electron wave, whose wavelength is $\lambda = 2\pi/k$.

In higher-dimension lattices, the Bragg condition is satisfied along all boundaries of the Brillouin zone, as discussed in Section 2.6, and this results in the creation of energy gaps along these boundaries, in agreement with the conclusions of the last section.

5.8 THE TIGHT-BINDING MODEL

In the tight-binding model, it is assumed that the crystal potential is strong, which is the same as saying that the ionic potentials are strong. It follows, therefore, that when an electron is captured by an ion during its motion through the lattice, the electron remains there for a long time before leaking, or tunneling, to the next ion [see Fig. 5.14(a), which also shows that the energy of the electron is appreciably

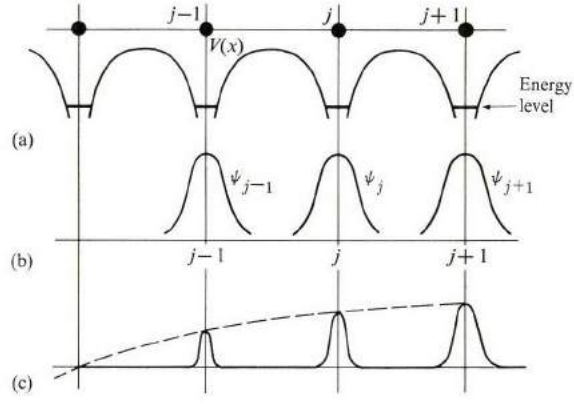


Fig. 5.14 The tight-binding model. (a) The crystal potential. (b) The atomic wave functions. (c) The corresponding Bloch function.

lower than the top of the potential barrier]. During the capture interval, the electron orbits primarily around a single ion, i.e., its state function is essentially that of an atomic orbital, uninfluenced by other atoms. Most of the time the electron is tightly bound to its own atom. The mathematical analysis to be developed must reflect this important fact.

As we said in Section 5.6, the TB (tight-binding) model is primarily suited to the description of low-lying narrow bands for which the shell radius is much smaller than the lattice constant. Here the atomic orbital is modified only slightly by the other atoms in the solid. An example is the 3d band, so important in transition metals.

Let us begin, then, with an atomic orbital, $\phi_\nu(x)$, whose energy in a free atom is E_ν. We wish to examine the effects of the presence of other atoms in the solid. The index ν characterizes the atomic orbital (for the atomic shell of interest).

First, the one-dimensional case: It is necessary to choose a suitable Bloch function, and while the choice is not unique, the following offers a reasonable form.

$$\psi_k(x) = \frac{1}{N^{1/2}} \sum_{j=1}^{N} e^{ikX_j} \phi_v(x - X_j), \tag{5.27}$$

where the summation extends over all the atoms in the lattice. The coordinate X_j specifies the position of the j^{th} atom. That is, $X_j = ja$, where a is the lattice constant. The function $\phi_v(x - X_j)$ is the atomic orbital centered around the j^{th} atom; it is large in the neighborhood of X_j, but decays rapidly away from this point, as shown in Fig. 5.14(b). By the time the neighboring site at X_{j+1} (or X_{j-1}) is reached, the function $\phi_v(x - X_j)$ has decayed so much that it has become almost negligible. In other words, there is only a little overlap between neighboring atomic orbitals. This is the basic assumption of the TB model. The factor $N^{1/2}$ is included in (5.27) to ensure that the function ψ_k is normalized to unity (if the atomic orbital ϕ_v is so normalized).

Let us turn now to the properties of the function $\psi_k(x)$, as defined by (5.27), First, it is necessary to ascertain that this function is a Bloch function, namely, that it can be written in the form (5.3). This can be established by rewriting (5.27) in the form

$$\psi_k(x) = \frac{1}{N^{1/2}} e^{ikx} \sum_{j=1}^{N} e^{-ik(x - X_j)} \phi_v(x - X_j),$$

where it is now readily recognized that the factor defined by the summation is periodic, with a period equal to the lattice constant a. Thus the function $\psi_k(x)$ has indeed the desired Bloch form, i.e., it describes a propagating electron wave, as shown in Fig 5.14(c).

Note also that near the center of the j^{th} ion, the function $\psi_k(x)$ reduces to

$$\psi_k(x) \simeq e^{ikX_j} \phi_v(x - X_j) \sim \phi_v(x - X_j). \tag{5.28}$$

That is, the Bloch function is proportional to the atomic orbital. Thus in the neighborhood of the j^{th} ion, the crystal orbital behaves much like an atomic orbital, in agreement with the basic physical assumption of the TB model.

The function $\psi_k(x)$ therefore satisfies both the mathematical requirement of the Bloch theorem and the basic assumption of the TB model, and as such is a suitable crystal orbital. It will be used now to calculate the energy of the band.

The energy of the electron described by ψ_k is given, according to quantum mechanics, by

$$E(k) = \langle \psi_k \mid H \mid \psi_k \rangle, \tag{5.29}$$

where H is the Hamiltonian of the electron†. Substituting for ψ_k from (5.27), one has

$$E(k) = \frac{1}{N} \sum_{j,j'} e^{ik(X_j - X_{j'})} \langle \phi_v(x - X_{j'}) \mid H \mid \phi_v(x - X_j) \rangle, \tag{5.30}$$

where the double summation over j and j' extends over all the atoms in the lattice. Note that each term in the summation is a function of the difference $X_j - X_{j'}$, and not of X_j and $X_{j'}$ individually. Therefore, for each particular choice of j', the sum over j yields the same result, and since j' can take N different values, one obtains N equal terms, which thus leads to

$$E(k) = \sum_{j=-N/2}^{(N-1)/2} e^{ikX_j} \langle \phi_v(x) \mid H \mid \phi_v(x - X_j) \rangle, \tag{5.31}$$

where we have arbitrarily put $X_{j'} = 0$ in (5.30). By splitting the term $j = 0$ from the others, one may write the above expression as

$$E(k) = \langle \phi_v(x) \mid H \mid \phi_v(x) \rangle + \sum_j{}' e^{ikX_j} \langle \phi_v(x) \mid H \mid \phi_v(x - X_j) \rangle, \tag{5.32}$$

The first term gives the energy the electron would have if it were indeed entirely localized around the atom $j = 0$, while the second term includes the effects of the electron tunneling to the various other atoms. The terms in the summation are expected to be appreciable only for nearest neighbors—that is, $j = 1$ and $j = -1$—because as j increases beyond that point, the overlap between the corresponding functions and the state function at the origin becomes negligible (Fig. 5.14b). Note also that, since the property of electron delocalization is included entirely in the second term of (5.32), it is this term which is responsible for the band structure, and as such is of particular interest to us here.

To proceed with the evaluation of $E(k)$, according to (5.32), we need to examine the Hamiltonian H more closely. The expression for this quantity is given by

$$H = -\frac{\hbar^2}{2m_0} \frac{d^2}{dx^2} + V(x), \tag{5.33}$$

where $V(x)$ is the crystal potential. Writing this potential as a sum of atomic potentials, one has

$$V(x) = \sum_j v(x - X_j). \tag{5.34}$$

† The Hamiltonian H is simply the quantum operator which represents the total energy of the particle. Thus $H = -(\hbar^2/2m_0)\nabla^2 + V(\mathbf{r})$, where the first term on the right represents kinetic energy and the second term potential energy. The expression (5.29) for the energy is very plausible, since the term on the right is the average value of the energy in quantum mechanics.

In using this to evaluate the first term in Eq. (5.32), we shall find it convenient to split $V(x)$ into a sum of two terms

$$V(x) = v(x) + V'(x), \tag{5.35}$$

where $v(x)$ is the atomic potential due to the atom at the origin and $V'(x)$ is that due to all the other atoms. These potentials are plotted in Figs. 5.15(a) and (b),

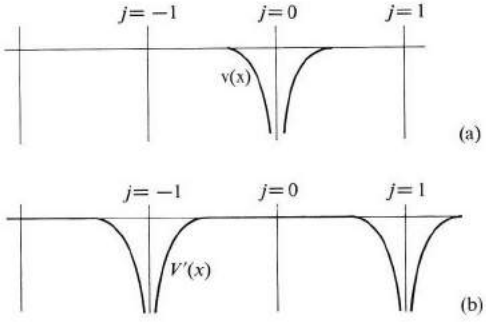


Fig. 5.15 The splitting of the crystal potential into (a) an atomic potential and (b) the remainder of the crystal potential.

respectively. Note in particular that $V'(x)$ is small in the neighborhood of the origin. The first term in (5.32) may now be written as

$$\langle \phi_\nu(x) \mid H \mid \phi_\nu(x) \rangle = \left\langle \phi_\nu(x) \left| \left[-\frac{h^2}{2m}\frac{d^2}{dx^2} + v(x) \right] \right| \phi_\nu(x) \right\rangle + \langle \phi_\nu(x) \mid V'(x) \mid \phi_\nu(x) \rangle. \tag{5.36}$$

The first term on the right is equal to E_ν, the atomic energy, since the operator involved is the Hamiltonian for a free atom. The second term is an integral which can be evaluated, and will be denoted by the constant $-\beta$. Explicitly,

$$\beta = -\int \phi_\nu^*(x)\, V'(x) \phi_\nu(x)\, dx, \tag{5.37}$$

where the minus sign is introduced so that β is a positive number.† Note that β is a small quantity, since the function $\phi_\nu(x)$ is appreciable only near the origin, whereas $V'(x)$ is small there. Collecting the two terms above, we have

$$\langle \phi_\nu(x) \mid H \mid \phi_\nu(x) \rangle = E_\nu - \beta. \tag{5.38}$$

† The integral in (5.37) is negative because $V'(x)$ is negative (Fig. 5.15b).

Let us now turn to the interaction term, i.e., the summation in (5.32). The term involving interaction with the nearest neighbor at $X_1 = a$ involves an integral which may be written as

$$\langle \phi_v(x) \mid H \mid \phi_v(x-a) \rangle = \langle \phi_v(x) \mid -\frac{h^2}{2m_0}\frac{d^2}{dx^2} + v(x-a) \mid \phi_v(x-a) \rangle + \langle \phi_v(x) \mid V'(x-a) \mid \phi_v(x-a) \rangle. \quad (5.39)$$

The first term on the right is equal to $E_v \langle \phi_v(x) \mid \phi_v(x-a) \rangle$, which is a negligible quantity, since the two functions $\phi_v(x)$ and $\phi_v(x-a)$, being centered at two different atoms, do not overlap appreciably. The second term on the right of (5.39) is a constant which we shall call $-\gamma$, that is,

$$\gamma = -\int \phi_v^*(x)\, V'(x-a)\, \phi_v(x-a)\, dx. \quad (5.40)$$

Note that γ, though small, is still nonvanishing because $V'(x-a)$ is appreciable near the origin, that is, $x = 0$ (although not at $x = a$). The parameter γ is called the *overlap integral*, since it is dependent on the overlap between orbitals centered at two neighboring atoms.

The integral arising from the term $j = -1$ in the sum in (5.32), which is due to the atom on the left side of the origin, yields the same result as (5.39) because the atomic functions are symmetric.

Substituting the above results into (5.32), and restricting the sum to nearest neighbors only, one finds

$$E(k) = E_v - \beta - \gamma \sum_{j=-1}^{1}{}' e^{ikX_j}, \quad (5.41)$$

which may thus be written as

$$E(k) = E_v - \beta - 2\gamma \cos ka. \quad (5.42)$$

This is the expression we have been seeking. It gives band energy as a function of k in terms of well-defined parameters which we can evaluate from our knowledge of atomic energy and atomic orbitals.

Equation (5.42) may be rewritten more conveniently as

$$E(k) = E_0 + 4\gamma \sin^2\left(\frac{ka}{2}\right), \quad (5.43)$$

where

$$E_0 = E_v - \beta - 2\gamma. \quad (5.44)$$

The energy $E(k)$ is plotted versus k in Fig. 5.16, where k is restricted to the first zone [although $E(k)$ is obviously periodic in k, in agreement with property (i)

of Section 5.4]. We see, as expected, that the original atomic level E_v has broadened into an energy band. The bottom of the band, located at $k = 0$, is equal to E_0, and its *width* is equal to 4γ.

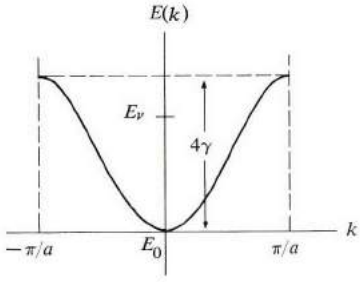


Fig. 5.16 The dispersion curve in the tight-binding model.

Note that the bottom of the band E_0 is lower than the atomic energy E_v, which is to be expected, since one effect of the presence of the other atom is to depress the potential throughout the system (refer to Fig. 5.14a). In addition to E_0, the electron has an amount of energy given by the second term in (5.43). This is a *kinetic energy*, arising from the fact that the electron is now able to move through the crystal.

Note also that the bandwidth, 4γ, is proportional to the overlap integral. This is reasonable, because, as we saw in Section 5.2, the greater the overlap the stronger the interaction, and consequently the wider the band.

When the electron is near the bottom of the band, where k is small, one may make the approximation $\sin(ka/2) \sim ka/2$, and hence

$$E(k) - E_0 = \gamma a^2 k^2, \tag{5.45}$$

which is of the same form as the dispersion relation of a free electron. An electron in that region of k-space behaves like a free electron with an effective mass

$$m^* = \frac{h^2}{2a^2}\frac{1}{\gamma}. \tag{5.46}$$

It is seen that the effective mass is inversely proportional to the overlap integral γ. This is intuitively reasonable, since the greater the overlap the easier it is for the electron to tunnel from one atomic site to another, and hence the smaller is the inertia (or mass) of the electron. Conversely, a small overlap leads to a large mass, i.e., a sluggish electron. Of course, in the TB model, the overlap is supposed to be small, implying a large effective mass.

Note, however, that an electron near the top of the band shows unusual behavior. If we define $k' = \pi/a - k$, and expand the energy $E(k)$ near the

maximum point, using (5.43), we arrive at

$$E(k') - E_{\max} = -\frac{a^2}{2}\gamma k'^2, \tag{5.47}$$

which shows that the electron behaves like a particle of *negative* effective mass

$$m^* = -\frac{\hbar^2}{a^2\gamma}. \tag{5.48}$$

This, you recall, is in agreement with the results obtained on the basis of the NFE model.

The above treatment can be extended to three dimensions in a straightforward manner. Thus for a sc lattice, the band energy is given by

$$E(k) = E_0' + 4\gamma\left[\sin^2\left(\frac{k_x a}{2}\right) + \sin^2\left(\frac{k_y a}{2}\right) + \sin^2\left(\frac{k_z a}{2}\right)\right]. \tag{5.49}$$

where E_0' is the energy at the bottom of the band. The energy contours for this band, in the $k_x - k_y$ plane, are shown in Fig. 5.17(a), and the dispersion curves along the [100] and [111] directions are shown in Fig. 5.17(b). The bottom of the band is at the origin $k = 0$, and the electron there behaves as a free particle with an effective mass given by (5.46). The top of the band is located at the corner of the zone along the [111] direction, that is, at $[\pi/a, \pi/a, \pi/a]$; the electron there has a negative effective mass given by (5.48). The width of the band is equal to 12γ.

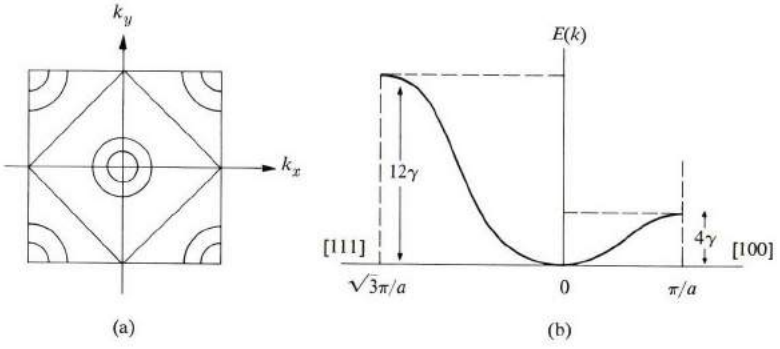


Fig. 5.17 (a) Energy contours for an sc lattice in the tight-binding model. (b) Dispersion curves along the [100] and [111] directions for an sc lattice in the TB model.

In this treatment of the TB model, we have seen how an atomic level *broadens* into a band as a result of the interaction between atoms in the solid. In this manner, each atomic level leads to its own corresponding band, and each band reflects the character of the atomic level from which it has originated.

In conclusion, we see that both the NFE and TB models lead to the same qualitative results, although the models start from opposite points of view. The principal results arrived at in both models are: (a) Energy gaps appear at zone boundaries. (b) An electron near the bottom of the band behaves like a free particle with a positive effective mass. (c) An electron near the top of the band behaves like a free particle with a negative effective mass.

5.9 CALCULATIONS OF ENERGY BANDS

In the last few sections we have discussed some methods of calculating energy bands. However, these methods—the NFE and TB models—are too crude to be useful in calculations of actual bands which are to be compared with experimental results. In this section we shall consider therefore some of the common methods employed in calculations of actual bands. Because this subject is an advanced one, requiring a considerable background in quantum mechanics, as well as meticulous attention to almost endless mathematical details, our discussion will be brief, primarily qualitative, and somewhat superficial. We shall nevertheless try to give the reader a glimpse of this fundamental subject in the hope that he may pursue it further, if he so desires, by referring to books listed in the bibliography at the end of the chapter.

Several different schemes for calculating energy bands have been used. Let us now discuss them individually.

The cellular method

The cellular method was the earliest method employed in band calculations (Wigner and Seitz, 1935). It was applied with success to the alkali metals, particularly Na and K; we shall use Na as an example.

The Schrödinger equation whose solution we seek is

$$\left[-\frac{\hbar^2}{2m_0} \nabla^2 + V(\mathbf{r}) \right] \psi_{\mathbf{k}} = E(\mathbf{k}) \psi_{\mathbf{k}}, \tag{5.50}$$

where $V(\mathbf{r})$ is the crystal potential and $\psi_{\mathbf{k}}$ the Bloch function. Here we are interested only in the 3s band. It is at once evident that this equation cannot be solved analytically. We must therefore use an approximation procedure.

When we use the cellular method, we divide the crystal into unit cells; each atom is centered at the middle of its cell, as shown in Fig. 5.18. Such a cell, known as the *Wigner–Seitz* (WS) *cell*, is constructed by drawing bisecting planes normal to the lines connecting an atom A, say, to its neighbors, and "picking out" the volume enclosed by these planes. (The procedure for constructing the WS cell, you may note, is analogous to that used in constructing the Brillouin zone in $\mathbf{k}$-space.) For Na, which has a bcc structure, the WS cell has the shape of a regular dodecahedron (similar to Fig. 5.8b, but in real space).

In order to solve (5.50), we now assume that the electron, when in a particular cell, say A, is influenced by the potential of the ion in that cell only. The ions in other cells have a negligible effect on the electron in cell A because each of these cells is occupied, on the average, by another conduction electron which tends to screen the ion, thereby reducing its potential drastically. To ensure that the function $\psi_{\mathbf{k}}$ satisfies the Bloch form, it is necessary that $u_{\mathbf{k}}$—where $\psi_{\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}$—be periodic, that is, $u_{\mathbf{k}}$ has the same points on opposite faces of the cell, e.g., points P_1 and P_2 in Fig. 5.18(a).

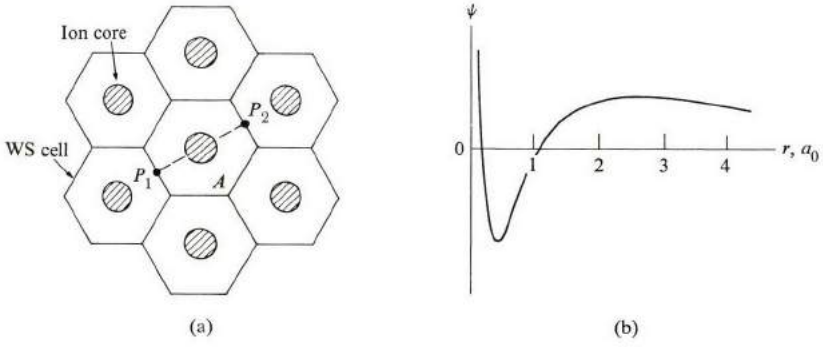


Fig. 5.18 (a) The WS cell. (b) The wave function ψ_0 at the bottom of the 3s band in Na versus the radial distance, in units of the Bohr radius.

The procedure is now clear in principle: We attempt to solve (5.50) in a single cell, using for $V(\mathbf{r})$ the potential of a *free* ion, which can be found from atomic physics. In Na, for instance, $V(\mathbf{r})$ is the potential of the ion core Na^+. It is still very difficult, however, to impose the requirements of periodicity on the function for the actual shape of the cell (the truncated octahedron), and to overcome this hurdle Wigner and Seitz replaced the cell by a WS *sphere* of the same volume as the actual cell, i.e., one employs a *WS sphere*. Using these simplifying assumptions concerning the potential and the periodic conditions, one then solves the Schrödinger equation numerically, since an analytical solution cannot usually be found. The resulting wave function ψ_0 at the bottom of the band, $k = 0$, is shown in Fig. 5.18(b). The wave functions at other values of $\mathbf{k}$ near the bottom of the band may then be approximated by

$$\psi_{\mathbf{k}} \simeq \frac{1}{V^{1/2}} e^{i\mathbf{k}\cdot\mathbf{r}}\psi_0, \tag{5.51}$$

which has the Bloch form.

The procedure is also capable of yielding the energy $E(\mathbf{k})$. The energy E_0 of the bottom of the band is obtained from the same calculations which give ψ_0,

and the energy at any other point $\mathbf{k}$ is obtained by using

$$E(\mathbf{k}) = \left\langle \psi_{\mathbf{k}} \left| -\frac{\hbar^2}{2m_0} \nabla^2 + V(\mathbf{r}) \right| \psi_{\mathbf{k}} \right\rangle, \tag{5.52}$$

where the wave function ψ_k is substituted from (5.51). The energy found in this manner was used by Wigner and Seitz to evaluate the cohesive energy, and the results are in satisfactory agreement with experiment.

One noteworthy feature of these results is the shape of the wave function in Fig. 5.18(b). The wave function oscillates at the ion core, but once outside the core the function is essentially a constant. This constancy of the wave function holds true for almost 90% of the cell volume. Thus the wave function behaves like a plane wave, as seen from (5.51), over most of the cell, and hence over most of the crystal. Looking at this in terms of the potential, we see that where the function is a plane wave, the potential must be a constant. Thus the *effective* potential acting on the electron is essentially a constant, except in the region at the ion core itself. Viewing the motion of the electron in the crystal as a whole, we conclude that the electron moves in a region of constant potential throughout most of the crystal; only at the cores themselves does the electron experience any appreciable potential. This surprising result explains why the conduction electrons in Na, for example, may be regarded as essentially free electrons. Mathematically, it is a consequence of the periodic conditions imposed on the wave function in the cell, and this is particularly apparent when one realizes that the wave function for the 3s electron in a free Na atom is very unlike ψ_0 outside the ion core. The flatness of ψ_0 is thus due to the imposition of the periodic conditions, and not to any special property of the ionic potential.† The effect of the periodic condition is to cancel out the ionic potential outside the core, and thus render the potential a constant. We shall find this result very useful in the development of other methods of band calculation.

Despite its usefulness, the cellular method is greatly oversimplified, and is not currently much in use. One of its chief disadvantages is that when one replaces the WS cell by a sphere, one ignores the crystal structure entirely. All anisotropic effects, for instance, are completely masked out.

The augmented-plane wave (APW) method

The APW method (Slater, 1937) uses the results of the cellular method, but is so formulated as to avoid its shortcomings. Since the effective crystal potential was found to be constant in most of the open spaces between the cores, the APW method begins by assuming such a potential (Fig. 5.19), which is referred to as the

† The boundary conditions require that the derivative of the function ψ_0 vanish at the surface of the WS sphere (why?). Thus the function is flat near the surface of this sphere, as shown in Fig. 5.18(b).

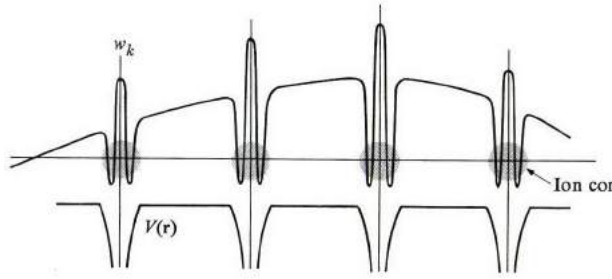


Fig. 5.19 The potential and wave function in the APW method.

muffin-tin potential. The potential is that of a free ion at the core, and is strictly constant outside the core. The wave function for the wave vector **k** is now taken to be

$$w_{\mathbf{k}} = \begin{cases} \dfrac{1}{V^{1/2}} e^{i\mathbf{k}\cdot\mathbf{r}}, & r > r_s, \\ \text{atomic function}, & r < r_s, \end{cases} \tag{5.53}$$

where r_s is the core radius. Outside the core the function is a plane wave because the potential is constant there. Inside the core the function is atom-like, and is found by solving the appropriate free-atom Schrödinger equation. Also, the atomic function in (5.53) is chosen such that it joins continuously to the plane wave at the surface of the sphere forming the core; this is the boundary condition here.

The function $w_{\mathbf{k}}$ does not have the Bloch form, but this can be remedied by forming the linear combination

$$\psi_{\mathbf{k}} = \sum_{\mathbf{G}} a_{\mathbf{k}+\mathbf{G}}\, w_{\mathbf{k}+\mathbf{G}}, \tag{5.54}$$

where the sum is over the reciprocal lattice vectors, which has the proper form. The coefficients $a_{\mathbf{k}+\mathbf{G}}$ are determined by requiring that $\psi_{\mathbf{k}}$ minimize the energy.† In practice the series in (5.54) converges quite rapidly, and only four or five terms—or even less—suffice to give the desired accuracy.

The APW method is a sound one for calculating the band structure in metals, and has been used a great deal in the past few years. It incorporates the essential features of the problem in a straightforward and natural fashion.

The pseudopotential method

Yet another method popular among solid-state physicists for calculating band structure in solids is the pseudopotential method, which is distinguished by the manner

† The "best" linear combination (5.54) is that which makes the energy as low as possible.

in which the wave function is chosen. We seek a function which oscillates rapidly inside the core, but runs smoothly as a plane wave in the remainder of the open space of the WS cell. Such a function was chosen in the APW method according to (5.53), but this is not the only choice possible. Suppose we take

$$w_{\mathbf{k}} = \phi_{\mathbf{k}} - \sum_i a_i v_i, \tag{5.55}$$

where $\phi_{\mathbf{k}}$ is a plane wave and v_i an atomic function. The sum over i extends over all the atomic shells which are occupied. For example, in Na, the sum extends over the 1s, 2s, and 2p shells. The coefficients a_i are chosen such that the function $w_{\mathbf{k}}$, representing a 3s electron, is orthogonal to the core function v_i.† By requiring this orthogonality, we ensure that the 3s electron, when at the core, does not occupy the other atomic orbitals already occupied. Thus we avoid violating the Pauli exclusion principle.

The function $w_{\mathbf{k}}$ has the features we are seeking: Away from the core, the atomic functions v_i are negligible, and thus $w_{\mathbf{k}} \simeq \phi_{\mathbf{k}}$, a plane wave. At the core, the atomic functions are appreciable, and act so as to induce rapid oscillations, as shown in Fig. 5.20.

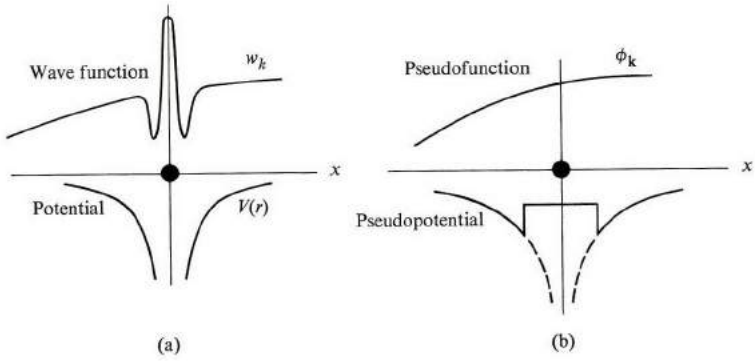


Fig. 5.20 The pseudopotential concept. (a) The actual potential and the corresponding wave function, as seen by the electron. (b) The corresponding pseudopotential and pseudofunction.

If one now substitutes $w_{\mathbf{k}}$ into the Schrödinger equation

$$\left[-\frac{\hbar^2}{2m_0}\nabla^2 + V\right] w_{\mathbf{k}} = E(\mathbf{k}) w_{\mathbf{k}}, \tag{5.56}$$

† Two functions ψ_1 and ψ_2 are said to be *orthogonal* if the integral $\int \psi_1^* \psi_2 \, d^3r = 0$. This concept of orthogonality is very useful in quantum mechanics. The atomic functions in the various atomic shells are all mutually orthogonal.

and rearranges the terms, one finds that the equation may be written in the form

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V'\right]\phi_{\mathbf{k}} = E(\mathbf{k})\,\phi_{\mathbf{k}}, \tag{5.57}$$

where

$$V' = V - \sum_i b_i \langle v_i | V | v_i \rangle. \tag{5.58}$$

These results are very interesting: Equation (5.57) shows that the effective potential is given by V, while (5.58) shows that V' is weaker than V, because the second term on the right of (5.58) tends to cancel the first term. This cancellation of the crystal potential by the atomic functions is usually appreciable, often leading to a very weak potential V'. This is known as the *pseudopotential.* Since V' is so weak, the wave function as seen from (5.57) is almost a plane wave, given by $\phi_{\mathbf{k}}$, and is called the *pseudofunction.*

The pseudopotential and pseudofunction are illustrated graphically in Fig. 5.20(b). Note that the potential is quite weak, and, in particular, the singularity at the ion core is entirely removed. Correspondingly, the rapid "wiggles" in the wave function have been erased, so that there is a smooth plane-wave-like function.

Now we can understand one point which has troubled us for some time: why the electrons in Na, for instance, seem to behave as free particles despite the fact that the crystal potential is very strong at the ionic cores. Now we see that, when the exclusion principle is properly taken into account, the effective potential is indeed quite weak. The free-particle behavior, long taken to be an empirical fact, is now borne out by quantum-mechanical calculations. The explanation of this basic paradox is one of the major achievements of the pseudopotential method. This method has also been used to calculate band structure in many metals and semiconductors (Be, Na, K, Ge, Si, etc.) with considerable success.

The APW and pseudopotential methods, as well as other related systems, require much numerical work which can feasibly be carried out only by modern electronic computers. It often takes a whole year or more to develop the necessary program and perform the calculations for one substance on a large computer!

5.10 METALS, INSULATORS, AND SEMICONDUCTORS

Solids are divided into two major classes: *Metals* and *insulators.* A metal—or conductor—is a solid in which an electric current flows under the application of an electric field. By contrast, application of an electric field produces no current in an insulator. There is a simple criterion for distinguishing between the two classes on the basis of the energy-band theory. This criterion rests on the following statement: *A band which is completely full carries no electric current, even in the*

presence of an electric field. It follows therefore that a solid behaves as a metal only when some of the bands are partially occupied. The proof of this statement will be supplied later (Section 5.13), but we shall accept it for the time being as an established fact.

Let us now apply this statement to Na, for example. Since the inner bands 1s, 2s, 2p are all fully occupied, they do not contribute to the current. We may therefore concern ourselves only with the topmost occupied band, the *valence band.* In Na, this is the 3s band. As we saw in Section 5.5, this band can accommodate $2N_c$ electrons, where N_c is the total number of primitive unit cells. Now in Na, a Bravais bcc lattice, each cell has one atom, which contributes one valence (or 3s) electron. Therefore the total number of valence electrons is N_c, and as these electrons occupy the band, only half of it is filled, as shown in Fig. 5.21(a). Thus sodium behaves like a metal because its valence band is only partially filled.

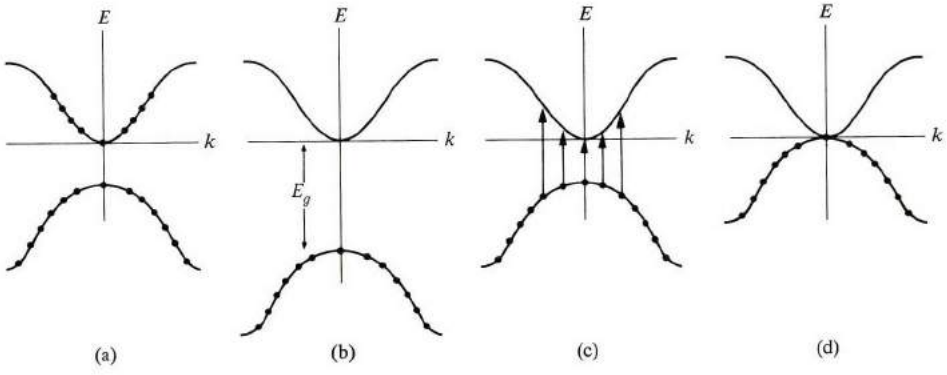


Fig. 5.21 The distribution of electrons in the bands of (a) a metal, (b) an insulator, (c) a semiconductor, and (d) a semimetal.

In a similar fashion, we conclude that the other alkalis, Li, K, etc., are also metals because their valence bands—the 2s, 4s, etc., respectively—are only partially full. The noble metals, Cu, Ag, Au, are likewise conductors for the same reason. Thus in Cu the valence band (the 4s band) is only half full, because each cell in its fcc structure contributes only one valence electron.

As an example of a good insulator, we mention diamond (carbon). Here the top band originates from a hybridization of the 2s and 2p atomic states (Section A.8), which gives rise to two bands split by an energy gap (Fig. 5.21b.) Since these bands arise from s and p states, and since the unit cell here contains two atoms, each of these bands can accommodate $8N_c$ electrons. Now in diamond each atom contributes 4 electrons, resulting in 8 valence electrons per cell. Thus the

valence band here is completely full, and the substance is an insulator, as stated above.†

There are substances which fall in an intermediate position between metals and insulators. If the gap between the valence band and the band immediately above it is small, then electrons are readily excitable thermally from the former to the latter band. Both bands become only partially filled and both contribute to the electric condition. Such a substance is known as a *semiconductor*. Examples are Si and Ge, in which the gaps are about 1 and 0.7 eV, respectively. By contrast, the gap in diamond is about 7 eV. Roughly speaking, a substance behaves as a semiconductor at room temperature whenever the gap is less than 2 eV.

The conductivity of a typical semiconductor is very small compared to that of a metal, but it is still many orders of magnitude larger than that of an insulator. It is justifiable, therefore, to classify semiconductors as a new class of substance, although they are, strictly speaking, insulators at very low temperatures.

In some substances the gap vanishes entirely, or the two bands even overlap slightly, and we speak of *semimetals* (Fig. 5.21d). The best-known example is Bi, but other such substances are As, Sb, and white Sn.

An interesting problem is presented in this connection by the divalent elements, for example, Be, Mg, Zn, etc. For instance, Be crystallizes in the hcp structure, with one atom per cell. Since there are two valence electrons per cell, the 2s band should completely fill up, resulting in an insulator. In fact, however, Be is a metal—although a poor one, in that its conductivity is small. The reason for the apparent paradox is that the 2s and 2p bands in Be overlap somewhat, so that electrons are transferred from the former to the latter, resulting in incompletely filled bands, and hence a metal. The same condition accounts for the metallicity of Mg, Ca, Zn, and other divalent metals.

A substance in which the number of valence electrons per unit cell is *odd* is necessarily a metal, since it takes an even number of electrons to fill a band completely. But when the number is even, the substance may be either an insulator or a metal, depending on whether the bands are disparate or overlapping.

---

† The case of hydrogen is of special interest. Although it is gaseous at atmospheric pressure, hydrogen solidifies at high pressure. But the familiar solid hydrogen is an insulator, having two atoms per unit cell, which causes the complete filling of the ls band. Theory predicts, however, that at very high pressure ($\simeq 2$ megabars), solid hydrogen undergoes a crystal structure transformation and a concomitant change to a metallic state. Many experimenters are currently attempting to observe this transformation, and tentative successes have been reported, but definitive results are still lacking at the time of writing. Even diamond has been reported to undergo transition to the metallic state at high pressure ($\simeq 1.5$ megabars). Simultaneously a structural phase transformation to a body-centered tetragonal structure occurs. The decrease in the lattice constant caused by the pressure is about 17%.

5.11 DENSITY OF STATES

The *density of states* for electrons in a band yields the number of states in a certain energy range. This function is important in electronic processes, particularly in transport phenomena. When we denote the density-of-states function by $g(E)$, it is defined by the relation

$$g(E)\,dE = \text{number of electron states per unit volume in the energy range } (E, E + dE). \tag{5.59}$$

This definition of $g(E)$ is analogous to that of the phonon density of states $g(\omega)$, so our discussion here parallels that presented in connection with $g(\omega)$. (See Sections 3.3 and 3.7; particularly 3.7.) To evaluate $g(E)$ one applies the definition (5.59): One draws a shell in **k**-space whose inner and outer surfaces are determined by the energy contours $E(\mathbf{k}) = E$ and $E(\mathbf{k}) = E + dE$, respectively, as shown in Fig. 5.22. The number of allowed **k** values lying inside this shell then gives the number of states which, when divided by the thickness of the shell dE, yields the desired function $g(E)$.

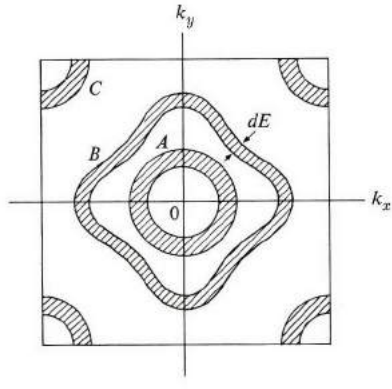


Fig. 5.22 Concentric shells in **k**-space used to evaluate the density of states $g(E)$.

It is evident that $g(E)$ is intimately related to the shape of the energy contours, and hence the band structure. The complexities of this structure are reflected in the form taken by $g(E)$. Let us first evaluate $g(E)$ for the case in which the dispersion relation for electron energy has the standard form

$$E = \frac{\hbar^2 k^2}{2m^*}. \tag{5.60}$$

As we have seen earlier, such a dispersion relation often holds true for those states lying close to the bottom of the band near the origin of the Brillouin zone. The energy contours corresponding to (5.60) are clearly concentric spheres surrounding the origin. The resulting density-of-states shell is then spherical in shape, as illustrated by shell A in Fig. 5.22, and since this is spherical, its volume is given by $4\pi k^2\,dk$, where k is the radius and dk the thickness of the shell. Recalling from Section 3.3 that the number of allowed $\mathbf{k}$ values per unit volume of $\mathbf{k}$-space is $1/(2\pi)^3$, it follows that the number of states lying in the shell—i.e., in the energy range $(E, E + dE)$—is

$$\text{Number of states} = \frac{1}{(2\pi)^3}\,4\pi k^2\,dk. \tag{5.61}$$

We may convert the right side by writing it in terms of E, the energy, rather than in terms of k, by using (5.60). We then find that

$$\text{Number of states} = \frac{1}{4\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} E^{1/2}dE.$$

Comparing this result with the definition (5.59), we infer that

$$g(E) = \frac{1}{4\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} E^{1/2}. \tag{5.62}$$

In order to take into account the spin degeneracy—i.e., the fact that each $\mathbf{k}$ state may accommodate two electrons of opposite spins—we multiply this expression by 2, which yields

$$g(E) = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} E^{1/2}. \tag{5.63}$$

This shows that $g(E) \sim E^{1/2}$, which means that the curve $g(E)$ has a parabolic shape (Fig. 5.23). The function $g(E)$ increases with E because, as we see from Fig. 5.22, the larger the energy the greater the radius, and hence the volume of the shell,

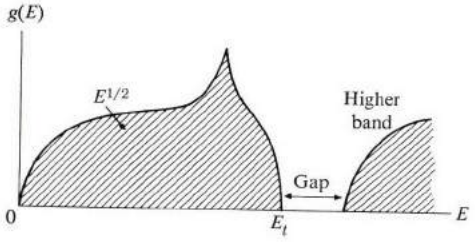


Fig. 5.23 The density of states.

and consequently the larger the number of states lying within it. Also note that $g(E) \sim m^{*3/2}$. That is, the larger the mass the greater the density of states.

The result (5.63) is very useful, and will be used repeatedly in subsequent discussions, but note that its validity is restricted to that region in **k**-space in which the standard dispersion relation (5.60) is satisfied. As the energy increases, a point is reached at which the energy contours become nonspherical—e.g., shell B in Fig. 5.22, in which region Eq. (5.63) no longer holds. One must then resort to a more complicated formula to evaluate $g(E)$. As a result, the shape of $g(E)$ is no longer parabolic at large energy, as shown in Fig. 5.23, the actual shape being determined by the dispersion relation $E = E(\mathbf{k})$ of the band. Note also that, at sufficiently large energies, the shell begins to intersect the boundaries of the zone, e.g., shell C in Fig. 5.22, in which case the volume of the shell begins to shrink, with a concomitant decrease in the number of states. The density of states of the shell plummets, and continues to decrease as the energy increases, until it vanishes completely when the shell lies entirely outside the zone, as shown in Fig. 5.23. The energy at which $g(E)$ vanishes marks the top of the valence band. The density of states remains zero for a certain energy range beyond that, this range marking the energy gap, until a new energy band appears, with its own density of states.

In simple metals, such as alkalis and noble metals, the standard form (5.60) holds true for most of the zone until the energy contours come close to the boundaries of the zone. It follows therefore that for these substances the expression (5.63) applies throughout most of the energy band, except close to the top of the band.

It is sometimes useful to have an expression for the density of states in the energy range lying close to the top of the band. This can be derived readily if the band there can be represented by a negative effective mass, as is usually the case (see Section 3.6). We may then show, by following a procedure analogous to that in deriving (5.63), that

$$g(E) = \frac{1}{2\pi^2}\left(\frac{2\,|m^*|}{\hbar^2}\right)^{3/2}(E_t - E)^{1/2}, \tag{5.64}$$

where E_t is the top of the valence band (note that here $E < E_t$). Thus the density function $g(E)$ has an inverted parabolic shape, where the parabola is at the top of the band. (See Fig. 5.23.).

Figure 5.24 illustrates situations in which bands overlap each other. Figure 5.24(a) represents a circumstance typical of divalent metals, in which the top of a band is at higher energy than the bottom of the next-higher band. Figure 5.24(b) shows the overlap of the 4s and 3d bands in transition metals. The 3d band, narrow and high, lies in the midst of the wide and flat 4s band.

According to definition, the quantity $g(E)\,dE$ gives the number of states lying in the energy range $(E, E + dE)$. The number of electrons actually occupying this

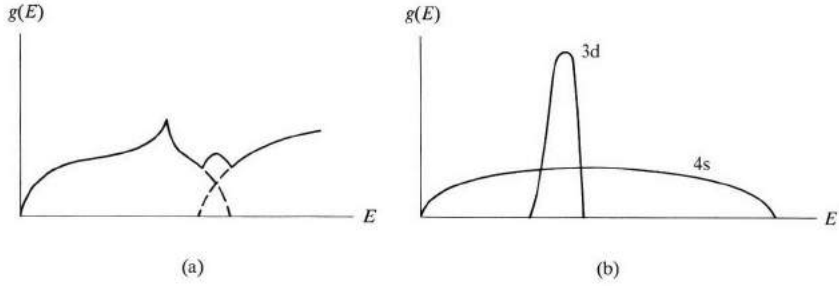


Fig. 5.24 (a) The shape of the density of states when two bands overlap each other as, e.g., in divalent metals. (b) The overlap of the 3d and 4s bands in transition metals.

range of energy is then given by

$$dn(E) = f(E)\, g(E)\, dE, \tag{5.65}$$

where $f(E)$ is the Fermi–Dirac distribution function, $f(E) = (1 + e^{(E - E_F)/k_B T})^{-1}$, discussed in Section 4.6. Expression (5.65) follows from the fact that since $g(E)\, dE$ gives the number of available states, and $f(E)$ the probability that each of these is occupied by an electron, then the product $f(E)\, g(E)\, dE$ must give the number of electrons present in that energy range.

5.12 THE FERMI SURFACE

In Section 4.7 we discussed the *Fermi surface* (FS) in connection with the free-electron model. There we saw that the significance of this surface in solid-state physics derives from the fact that only those electrons lying near it participate in thermal excitations or transport processes. Here we shall consider the Fermi surface again, and now we shall incorporate the effects of the crystal potential. The significance of the FS remains unchanged, but its shape, in some cases, may be considerably more complicated than the spherical shape of the free-electron model. We shall now consider the effects of the crystal potential on the shape of the FS, while in later sections we shall see how this change may influence the physical properties of the crystal. Experimental determination of the FS will be considered in Section 5.19.

As we recall, the FS is defined as the surface in **k**-space inside which all the states are occupied by valence electrons.† All the states lying outside the surface are

† In Section 4.7 we discussed the FS in velocity space. However, for a free-like electron, the velocity is given by $\mathbf{v} = \hbar\mathbf{k}/m^*$. Thus $\mathbf{v}$ and $\mathbf{k}$ are proportional to each other, and one could equally well speak of the FS in **k**-space, provided an appropriate change in scale were made.

empty. The definition is strictly valid only at absolute zero, $T = 0°K$, but, as we saw in Section 4.6, the effect of temperature on the FS is very slight, and the surface remains sharp even at room temperature or higher. The shape of the FS is determined by the geometry of the energy contours in the band, since the FS is itself an energy contour, where $E(\mathbf{k}) = E_F$, E_F being the Fermi energy. (Because of this, the FS should display the same rotational symmetry as the lattice.)

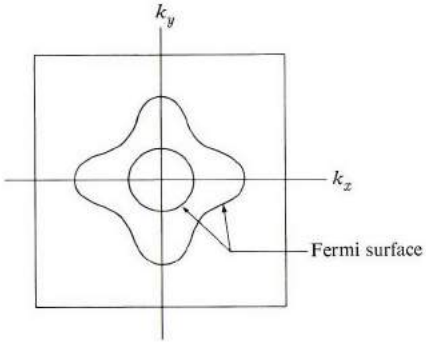


Fig. 5.25 The evolution of the shape of the FS as the concentration of valence or conduction electrons increases.

Figure 5.25 illustrates the evolution of the shape of the FS as the concentration of valence electrons increases. For small n, only those states lying near the bottom of the band at the center of the zone are populated, and the occupied volume is a sphere in **k**-space, which is therefore bounded by a spherical FS. As n increases and more states are populated, the "Fermi volume" expands, and so does the FS. This surface, which is spherical near the origin, begins to deform gradually as n increases, following the distortion in the contours at large energies (as discussed previously) as seen in Fig. 5.25. The distortion in the shape of the FS may become quite pronounced, particularly as the FS approaches the boundaries of the zone. The distortion is even greater when the surface intersects the boundaries, as will be discussed later in this section.

The alkali metals Li, Na, and K crystallize in the bcc structure, whose Brillouin zone is a regular rhombic dodecahedron (Fig. 5.8b). As we saw in Section 5.6, the valence band is half filled. The FS is still far from the boundaries, and since the standard dispersion relation holds well throughout most of the zone, it follows that the FS in these substances is essentially spherical in shape. Experiments confirm this, showing that in Na and K the distortion of the FS from sphericity is of the order of 10^{-3}.

The noble metals Cu, Ag, and Au crystallize in the fcc structure. The shape of the BZ here is that of a truncated octahedron (Fig. 5.26). Here again the valence band is only half-filled, and consequently the FS, being far from the zone

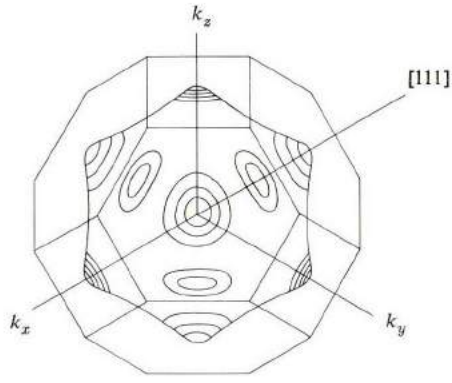


Fig. 5.26 The FS in noble metals. The surface protrudes toward the zone faces in the [111] directions.

boundaries, should be essentially spherical, which is substantially true for most of the FS. However, along the ⟨111⟩ directions, the FS comes close to the zone boundaries, because of the shape of the zone, and as a result the surface suffers strong distortion in that region. As seen in Fig. 5.26, the FS protrudes along the ⟨111⟩ directions so much as to touch the zone face. In effect the zone boundaries have "pulled" the FS, giving it the shape shown in the figure—a sphere with eight "necks" protruding in the ⟨111⟩ directions. In this respect the FS in the noble metals is quite different from that in the alkali metals.

The position of the Fermi level E_F for various classes of solids is illustrated in Fig. 5.27. Figure 5.27(a) illustrates the density of states and the position of E_F

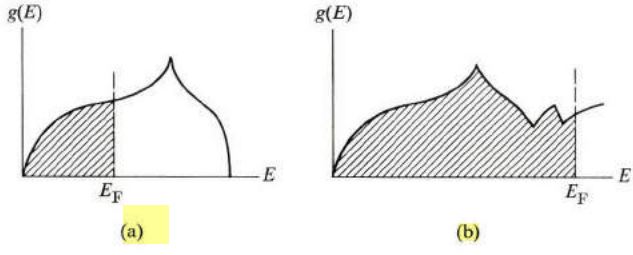


Fig. 5.27 The position of the Fermi energy in (a) a monovalent metal, and (b) a divalent metal.

for a typical monovalent metal, where only half the band is filled, and the substance acts as a conductor. Figure 5.27(b) shows a divalent metal. Here the bands overlap to some extent, and the number of valence electrons is so large that the FS spills over into the higher band. Figure 5.28 shows an insulator, in which the

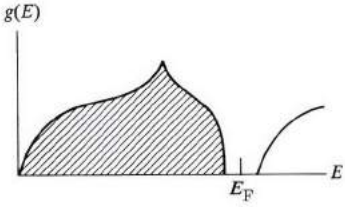


Fig. 5.28 The position of the Fermi energy in an insulator.

valence band is completely filled and the Fermi level lies somewhere in the energy gap.

We shall now determine the Fermi energy E_F for the case in which the standard form (5.60) holds. As seen above, this applies to the alkali metals and, to a lesser extent, to the noble metals as well. By its very definition, the Fermi energy satisfies the relation (at $T = 0°K$)

$$\int_0^{E_F} g(E)\,dE = n, \tag{5.66}$$

because the integral on the left gives the number of states from the bottom of the band, $E = 0$, right up to the Fermi level. This number must be equal to the number of electrons, which is the meaning of (5.66). If we substitute for $g(E)$ from (5.63), perform the necessary integration (which can be readily accomplished), and solve for E_F, we find that

$$E_F = \frac{\hbar^2}{2m^*}(3\pi^2 n)^{2/3}, \tag{5.67}$$

which is the result quoted previously in the case of the free-electron model (Section 4.7). Refer to Table 4.1 for a list of Fermi levels, and note that E_F is typically of the order of a few electron volts.

Let us now turn to the FS in polyvalent metals. Suppose that the number of valence electrons is sufficiently large so that the FS intersects the boundaries of the zone, as shown in Fig. 5.29(a). In constructing the FS here, we used the empty-lattice model, so the crystal potential is set equal to zero. The FS is now seen to extend over two zones. The part of the FS lying in the first zone is repeated in Fig. 5.29(b). Note that it is composed of the four sides of a diamond-shaped figure. Figure 5.29(c) replots the part of the FS lying in the second zone using the reduced-zone scheme. We see that it is composed of the sides of four half-bubble-shaped figures. When viewed in the various individual zones, the shape of the FS appears quite complicated, even for a free electron, belying its original simplicity. Of course, if one uses the extended-zone scheme, the original spherical shape of Fig. 5.29(a) may be reconstructed, but this is not immediately apparent from Fig. 5.29(b) or (c) individually. If we now turn on a weak crystal potential, the

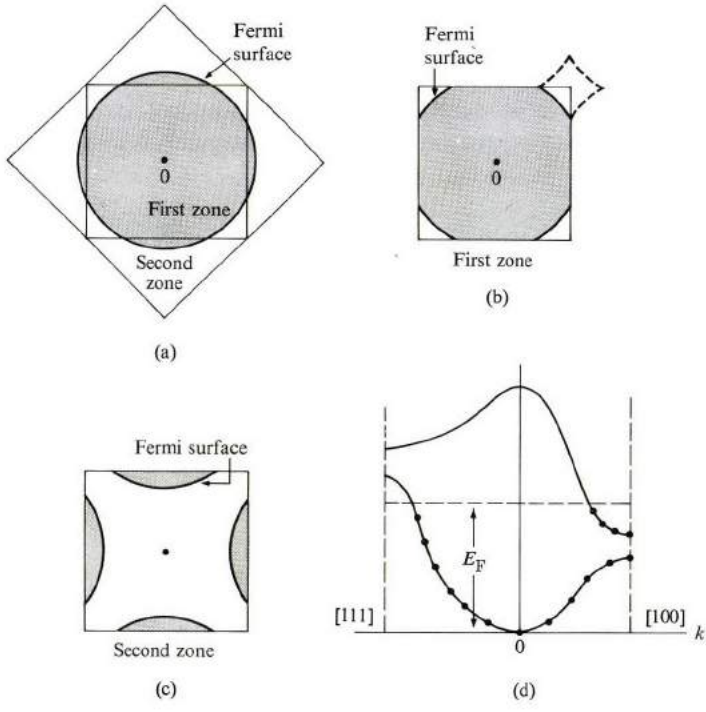


Fig. 5.29 The Harrison construction. (a) The FS in the empty-lattice model using the extended-zone scheme. (b) The FS in the first zone. (c) The FS in the second zone. (d) Band overlap.

shape of the FS in the two zones is affected only slightly, the effect being primarily to round off the sharp corners. The point here is that the complicated FS's usually observed in polyvalent metals are not necessarily the result of strong crystal potentials (as was once thought to be the case). They may be due largely to the crossing of the zone and the piecing together of the various parts of the FS. (The procedure for reconstructing Fermi surfaces on the basis of the empty-lattice model is known as the *Harrison construction.*)

Figure 5.29(d) shows the energy bands in the two zones plotted in two different directions. The two bands overlap. The top of the first band along the [111] direction is higher than the bottom of the second band in the [100] direction. The Fermi level crosses both bands, and both contribute to the conduction process.

It is important to note here that the Fermi level crosses the lower band (on the left in Fig. 5.29d) in a region in which the curvature of the band is downward, i.e., a region of negative effective mass. As we shall see in Section 5.17, such a situation is best described in terms of *holes.*

Figure 5.29(d) illustrates what is known as the *two-band model* for a metal.

The electric current is transported by carriers in two bands: Electrons in the higher band, holes in the lower. We shall exploit this model to full advantage in Section 5.18.

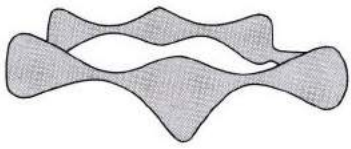


Fig. 5.30 The Fermi surface of beryllium.

Finally, Fig. 5.30 shows the FS for Be (known also as the Be *coronet*). Complicated as this appears to be, the surface is quite similar to the shape obtained using the Harrison construction. Note the hexagonal symmetry, expected as a consequence of the hexagonal crystal structure of Be.

5.13 VELOCITY OF THE BLOCH ELECTRON

Now let us study the motion of the Bloch electrons in solids. An electron in a state $\psi_{\mathbf{k}}$ moves through the crystal with a velocity directly related to the energy of that state. Consider first the case of a free particle. The velocity is given by $\mathbf{v} = \mathbf{p}/m_0$, where $\mathbf{p}$ is the momentum. Since $\mathbf{p} = \hbar\mathbf{k}$, it follows that, for a free electron, the velocity is given by

$$\mathbf{v} = \frac{\hbar\mathbf{k}}{m_0}, \tag{5.68}$$

i.e., the velocity is proportional to and parallel to the wave vector $\mathbf{k}$, as shown in Fig. 5.31(a).

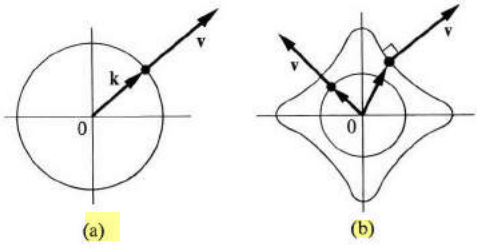


Fig. 5.31 The velocity of (a) a free electron, and (b) a Bloch electron.

For a Bloch electron, the velocity is also a function of $\mathbf{k}$, but the functional relationship is not as simple as (5.68). To derive this relationship, we use a well-known

formula in wave propagation. That is, the group velocity of a wave packet is given by

$$\mathbf{v} = \nabla_{\mathbf{k}}\, \omega(\mathbf{k}), \tag{5.69}$$

where ω is the frequency and $\mathbf{k}$ the wave vector of the wave packet. Applying this equation to the electron wave in the crystal, and noting the Einstein relation $\omega = E/\hbar$, we may write for the velocity of the Bloch electron

$$\mathbf{v} = \frac{1}{\hbar}\, \nabla_{\mathbf{k}}\, E(\mathbf{k}), \tag{5.70}$$

which states that the velocity of an electron in state $\mathbf{k}$ is proportional to the gradient of the energy in $\mathbf{k}$-space. [Equation (5.70) can also be derived more rigorously by writing the quantum expression for the velocity of the probability wave associated with the Bloch electron and finding the quantum expectation value; see Mott (1936).] We assume implicitly that we are dealing here with the valence band, and hence the band index has been suppressed, although it should be clear from the derivation that (5.70) is valid in any band.

Since the gradient vector is perpendicular to the contour lines, a fact well known from vector analysis, it follows that the velocity $\mathbf{v}$ at every point in $\mathbf{k}$-space is normal to the energy contour passing through that point, as shown in Fig. 5.31(b). Because these contours are in general nonspherical, it follows that the velocity is not necessarily parallel to the wave vector $\mathbf{k}$, unlike the situation of a free particle.

Note, however, that near the center of the zone, where the standard dispersion relation $E = \hbar^2 k^2/2m^*$ is expected to hold true, the relation (5.70) leads to

$$\mathbf{v} = \frac{\hbar \mathbf{k}}{m^*}, \tag{5.71}$$

which is of the same form as the relation for a free particle, (5.68), except that m_0 has been replaced by m^*, the effective mass. This is to be expected, of course, since we have often stated that a Bloch electron behaves in many respects like a free electron, except for the difference in mass. It follows that near the center of the zone $\mathbf{v}$ is parallel to $\mathbf{k}$, and points radially outward, as shown in Fig. 5.31(b). It is near the zone boundaries at which the energy contours are so distorted that this simple relationship between $\mathbf{v}$ and $\mathbf{k}$ is destroyed, and so one must resort to the more general result (5.70).

Note also that when an electron is in a certain state $\psi_{\mathbf{k}}$, it remains in that state forever, provided only that the lattice remains periodic. Thus as long as this situation persists, the electron will continue to move through the crystal with the same velocity $\mathbf{v}$, unhampered by any scattering from the lattice.† In other words,

† See the remarks about the propagation of waves in periodic lattices (Section 4.5).

the velocity of the electron is a constant. Any effect the lattice may exert on the propagation velocity has already been included in (5.70) through the energy $E(\mathbf{k})$.

Deviations in the periodicity of the lattice would, of course, cause a scattering of the electron, and hence a change in its velocity. For example, an electron moving in a vibrating lattice suffers numerous collisions with phonons, resulting in a profound influence being exerted on the velocity. Also, external fields—electric or magnetic—lead to change in the velocity of the electron. We shall discuss these effects in the following sections.

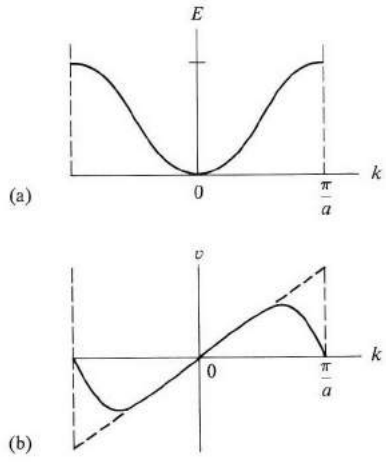


Fig. 5.32 (a) The band structure, and (b) the corresponding electron velocity in a one-dimensional lattice. The dashed line in (b) represents the free-electron velocity.

Figure 5.32(a) shows a typical one-dimensional band structure, and Fig. 5.32(b) shows the corresponding velocity, which in this case reduces to

$$v = \frac{1}{\hbar}\frac{\partial E}{\partial k}, \tag{5.72}$$

that is, the velocity is proportional to the slope of the energy curve. We see that as k varies from the origin to the edge of the zone, the velocity increases at first linearly, reaches a maximum, and then decreases to zero at the edge of the zone. We wish now to explain this behavior on the basis of the NFE model, particularly the seemingly anomalous decreases in the velocity near the edge of the zone. The following discussion is closely related to the discussion in Section 5.7.

Near the zone center, the electron may be adequately represented by a single plane wave $\psi_k \sim e^{ikx}$, and hence $\mathbf{v} = \hbar k/m_0$, explaining the linear region of Fig. 5.32(b). However, as $\mathbf{k}$ increases, the scattering of the free wave by the lattice introduces a new left-traveling wave whose wave vector $k' = k - 2\pi/a$, and which

is to be superimposed on the original right-traveling wave k. Therefore the electron is now represented by the wave mixture

$$\psi_k \simeq e^{ikx} + be^{-i(2\pi/a-k)x}, \tag{5.73}$$

where the coefficient b is found from perturbation theory (Eq. 5.24). The velocity of this wave, according to quantum mechanics, is given by

$$v = \frac{\hbar k}{m_0} - |b|^2 \frac{\hbar}{m_0}\left(\frac{2\pi}{a} - k\right), \tag{5.74}$$

where the first term on the right is the contribution of the right-traveling wave, while the second term is the contribution of the left-traveling wave. At small k, the coefficient b is small, and v is given essentially by $\hbar k/m_0$, as stated above. As k increases, however, the coefficient of the scattered wave increases, and so the second term in (5.74) becomes appreciable. Since the second term is negative ($k < 2\pi/a$), its effect tends to cancel the first term. Near the zone boundaries, the coefficient b is so large that the resulting cancellation is greater than the increase in the first term, which leads to a net decrease in the velocity, as we have seen.

At the zone boundary itself ($k = \pi/a$), the scattered wave becomes equal to the incident wave as a result of the strong Bragg reflection, that is, $b = 1$, which, when substituted into (5.74), yields $v = 0$, in agreement with Fig. 5.32(b). We anticipated this result in Section 5.7, in which we found that at the zone edge the electron is represented by a standing wave.

Similar applications of the NFE model in two and three dimensions explain why the relationship between $\mathbf{v}$ and $\mathbf{k}$ near the zone boundaries differs considerably from that for a free particle (see the problem section at the end of this chapter).

Now we shall derive a result which was used earlier in Section 5.10, namely, that a completely filled band carries no electric current. To establish this, we note that according to (5.70)

$$\mathbf{v}(-\mathbf{k}) = -\mathbf{v}(\mathbf{k}), \tag{5.75}$$

where $\mathbf{v}(\mathbf{k})$ and $\mathbf{v}(-\mathbf{k})$ are the velocities of electrons in the Bloch states $\mathbf{k}$ and $-\mathbf{k}$, respectively (see Fig. 5.33). This equation follows from the symmetry relation

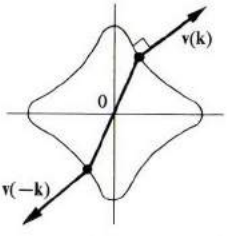


Fig. 5.33 $\mathbf{v}(-\mathbf{k}) = -\mathbf{v}(\mathbf{k})$.

$E(-\mathbf{k}) = E(\mathbf{k})$, which was established in Section 5.4. The current density due to all electrons in the band is given by

$$\mathbf{J} = \frac{1}{V}(-e)\sum_{\mathbf{k}} \mathbf{v}(\mathbf{k}), \tag{5.76}$$

where V is the volume, $-e$ the electronic charge, and the sum is over all states in the band. But as a consequence of (5.75), the sum over a whole band is seen to vanish, that is, $\mathbf{J} = 0$, with the electrons' velocities canceling each other out in pairs.

5.14 ELECTRON DYNAMICS IN AN ELECTRIC FIELD

When an electric field is applied to the solid, the electrons in the solid are accelerated. We can study their motions most easily in k-space. Suppose that an electric field $\mathscr{E}$ is applied to a given crystal. As a result, an electron in the crystal experiences a force $\mathbf{F} = -e\mathscr{E}$, and hence a change in its energy. The rate of absorption of energy by the electron is

$$\frac{dE(\mathbf{k})}{dt} = -e\mathscr{E}\cdot\mathbf{v}, \tag{5.77}$$

where the term on the right is clearly the expression for the power absorbed by a moving object. If we write

$$\frac{dE(\mathbf{k})}{dt} = \nabla_{\mathbf{k}}E(\mathbf{k})\cdot\frac{d\mathbf{k}}{dt},$$

and use the expression (5.70) for $\mathbf{v}$, then substitute these into (5.77), we find the surprisingly simple relation

$$\hbar\frac{d\mathbf{k}}{dt} = -e\mathscr{E} = \mathbf{F}. \tag{5.78}$$

This shows that the rate of change of $\mathbf{k}$ is proportional to—and lies in the same direction as—the electric force $\mathbf{F}$ (i.e., opposite to the field $\mathscr{E}$, by virtue of the negative electron charge). This relation is a very important one in the dynamics of Bloch electrons, and is known as the *acceleration theorem*.

Equation (5.78) is not totally unexpected. We have already noted the fact that the vector $\hbar\mathbf{k}$ behaves like the momentum of the Bloch electron (Section 5.3). In that context, Eq. (5.78) simply states that the time rate of change of the momentum is equal to the force, which is Newton's second law.

Let us now consider the consequences of the acceleration theorem, starting

with the one-dimensional case. Equation (5.78) may be written in the form

$$\frac{dk}{dt} = \frac{F}{\hbar}, \tag{5.79}$$

showing that the wave vector k increases uniformly with time. Thus, as t increases, the electron traverses the k-space at a uniform rate, as shown in Fig. 5.34. If we

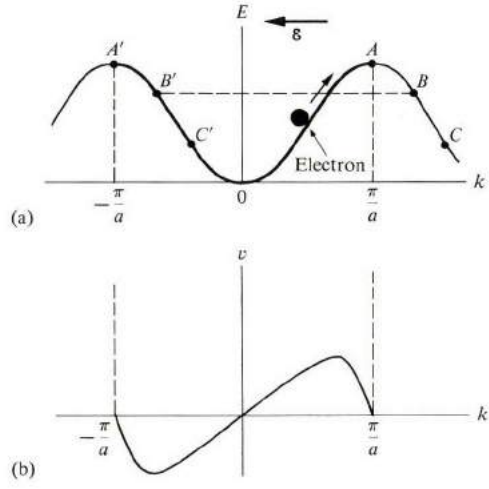


Fig. 5.34 (a) The motion of an electron in k-space in the presence of an electric field (directed to the left). (b) The corresponding velocity.

use the repeated-zone scheme, the electron, starting from $k = 0$, for example, moves up the band until it reaches the top (point A) and then starts to descend along the path BC. If we use the reduced-zone scheme, then once the electron passes the zone edge at A, it immediately reappears at the equivalent point A', then continues to descend along the path $A'B'C'$. Recall that, according to the translational-symmetry property of Section 5.4, the points B', C' are respectively equivalent to the points B, C, so that we may use either of the two schemes.

Note that, in the presence of an electric field, the electron is in constant motion in k-space; it is never at rest.

Also note that the motion in k-space is periodic in the reduced-zone scheme, since after traversing the zone once, the electron repeats the motion. The *period* of the motion is readily found, on the basis of (5.79), to be

$$T = \frac{2\pi\hbar}{Fa} = \frac{2\pi\hbar}{e\mathscr{E}a}. \tag{5.80}$$

Figure 5.34(b) shows the velocity of the electron as it traverses the k-axis. Starting at $k = 0$, as time passes, the velocity increases, reaches a maximum,

decreases. and then vanishes at the zone edge. The electron then turns around and acquires a negative velocity, and so forth. The velocity we are discussing is the velocity in real space, i.e., the usual physical velocity. It follows that a Bloch electron, in the presence of a static electric field, executes an oscillatory periodic motion in real space, very much unlike a free electron. This is one of the surprising conclusions of electron dynamics in a crystal.

Yet the oscillatory motion described above has not been observed, and the reason is not hard to come by. The period T of (5.80) is about 10^{-5} s for usual values of the parameters, compared with a typical electron collision time $\tau = 10^{-14}$ s at room temperature. Thus the electron undergoes an enormous number of collisions, about 10^9, in the time of one cycle. Consequently the oscillatory motion is completely "washed out."†

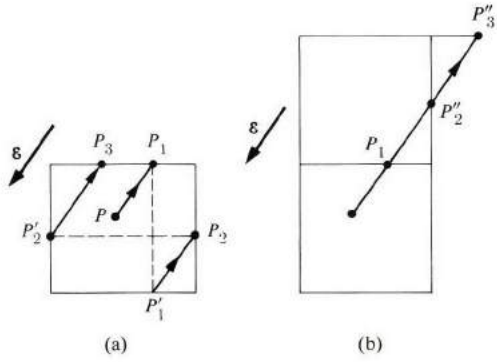


Fig. 5.35 The motion of an electron in a two-dimensional lattice in the presence of an electric field (a) according to the reduced-zone scheme, (b) according to the repeated-zone scheme.

Let us now consider the situation in two dimensions (Fig. 5.35). When an electric force **F** is applied, the electron, starting at some arbitrary point P, moves in a straight line in **k**-space, according to (5.78). As it reaches the zone edge at point P_1, it reappears at P'_1, continues on to P_2, and reappears at P'_2. It follows the crisscross path shown in Fig. 5.35(a). If we used the repreated-zone scheme instead (Fig. 5.35b), then the path of the electron in **k**-space would simply be the straight line $P\,P_1\,P''_2\,P''_3$ (note that P''_2 is equivalent to P_2, P''_3 to P_3, etc.). This is one situation in which the repeated-zone scheme proves to be more convenient than the extended-zone scheme.

5.15 THE DYNAMICAL EFFECTIVE MASS

When an electric field is applied to a crystal, the Bloch electron undergoes an

†Leo Esaki and his collaborators are currently attempting to build a device for which $T \leqslant \tau$, by growing highly pure superlattices for which $a \simeq 50 - 100$Å. Such a *Bloch oscillator* may be used as an oscillator or amplifier.

acceleration. This can be calculated as follows: Since acceleration is the time derivative of velocity, we have

$$a = \frac{dv}{dt}, \tag{5.81}$$

where we have chosen to treat the one-dimensional case first. But velocity is a function of the wave vector k, and consequently the above equation may be rewritten as

$$a = \frac{dv}{dk}\frac{dk}{dt},$$

which, when we substitute for the velocity from (5.72), and for dk/dt from (5.78), yields

$$a = \frac{1}{\hbar^2}\frac{d^2E}{dk^2}F. \tag{5.82}$$

This has the same form as Newton's second law, provided we define a *dynamical effective mass* m^* by the relation

$$m^* = \hbar^2 \Big/ \left(\frac{d^2E}{dk^2}\right). \tag{5.83}$$

Thus, insofar as the motion in an electric field is concerned, the Bloch electron behaves like a free electron whose effective mass is given by (5.83).

The mass m^* is inversely proportional to the curvature of the band; where the curvature is large—that is, d^2E/dk^2 is large—the mass is small; a small curvature implies a large mass (Fig. 5.36).

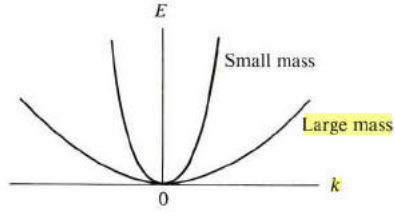


Fig. 5.36 The inverse relationship between the mass and the curvature of the energy band.

We have previously used the concept of effective mass (Sections 5.6 and 5.8). Those situations are now superseded by—and are in fact special cases of—the

general relation (5.83). Thus, if the energy is quadratic in k,

$$E = \alpha k^2, \tag{5.84}$$

where α is a constant. Then Eq. (5.83) yields

$$m^* = \hbar^2/2\alpha, \tag{5.85}$$

which is equivalent to rewriting (5.84) as $E = \hbar^2 k^2/2m^*$, the standard form.

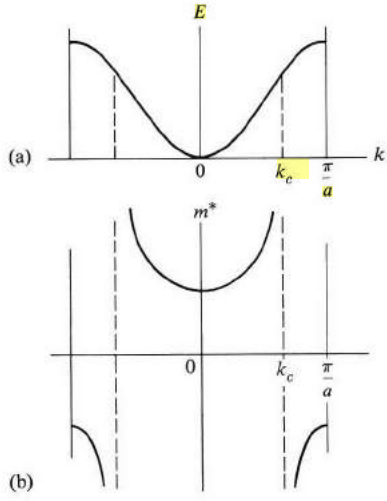


Fig. 5.37 (a) The band structure, and (b) the effective mass m^* versus k.

Figures 5.37(a) and (b) show, respectively, the band structure and the effective mass m^*, the latter calculated according to (5.83). Near the bottom of the band, the effective mass m^* has a constant value which is positive, because the quadratic relation (5.84) is satisfied near the bottom of the band. But as k increases, m^* is no longer a strict constant, being now a function of k, because the quadratic relation (5.84) is no longer valid.

Note also that beyond the *inflection point* k_c the mass m^* becomes negative, since the region is now close to the top of the band, and a negative mass is to be expected (Sections 5.6 and 5.8).

The negative mass can be seen dynamically by noting that, according to Fig. 5.34, the velocity decreases for $k > k_c$. Thus the acceleration is negative, i.e., opposite to the applied force, implying a negative mass. This means that in this region of k-space the lattice exerts such a large retarding (or braking) force on the electron that it overcomes the applied force and produces a negative acceleration.

The above results may be extended to three dimensions. The acceleration is

now

$$\mathbf{a} = \frac{d\mathbf{v}}{dt}.$$

If we write this in cartesian coordinates, and use (5.70) and (5.78), we find that

$$a_i = \sum_j \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k_i \partial k_j} F_j, \qquad i, j = x, y, z,$$

which leads to the definition of effective mass as

$$\left(\frac{1}{m^*}\right)_{ij} = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k_i \partial k_j}, \qquad i, j = x, y, z. \tag{5.86}$$

The effective mass is now a second-order tensor which has nine components.

When the dispersion relation can be written as†

$$E(\mathbf{k}) = (\alpha_1 k_x^2 + \alpha_2 k_y^2 + \alpha_3 k_z^2), \tag{5.87}$$

then using (5.86) leads to an effective mass with three components: $m_{xx}^* = \hbar^2/2\alpha_1$, $m_{yy}^* = \hbar^2/2\alpha_2$, and $m_{zz}^* = \hbar^2/2\alpha_3$. In this case the mass of the electron is *anisotropic*, and depends on the direction of the external force. When the force is along the k_x-axis, the electron responds with a mass m_{xx}^*, while a force in the k_y-direction elicits an effective mass m_{yy}^*. A relation of the type (5.87), corresponding to ellipsoidal contours, is a common occurrence in semiconductors, e.g., Si and Ge. Note that in this case, unlike the free-electron case, the acceleration is not, in general, in the same direction as the applied force.

It may also happen that one of the α_i's in (5.87) is negative. This means that the mass in the corresponding direction is negative, while the other directions exhibit positive masses. This again is vastly different from the behavior of the free electron.

The concept of effective mass is very useful, in that it often enables us to treat the Bloch electron in a manner analogous to a free electron. Nonetheless, the Bloch electron exhibits many unusual properties which are alien to those of a free electron.

5.16 MOMENTUM, CRYSTAL MOMENTUM, AND PHYSICAL ORIGIN OF THE EFFECTIVE MASS

We have said on several occasions that a Bloch electron in the state $\psi_{\mathbf{k}}$ behaves as if it had a momentum $\hbar\mathbf{k}$. Basically, there are three different reasons to support this statement.

† This is possible near a point at which the energy has a minimum, a maximum, or a saddle point.

a) The Bloch function has the form

$$\psi_{\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}, \tag{5.89}$$

which, since $u_{\mathbf{k}}$ is periodic, appears essentially as a plane wave of wavelength $\lambda = 2\pi/k$. This, combined with the deBroglie relation, leads to a momentum $\hbar\mathbf{k}$.

b) When an electric field is applied, the wave vector varies with time according to

$$\frac{d(\hbar\mathbf{k})}{dt} = \mathbf{F}_{\text{ext}}, \tag{5.90}$$

again indicating that $\hbar\mathbf{k}$ acts as a momentum. Here $\mathbf{F}_{\text{ext}}$ refers to the external force applied to the crystal.

c) In collision processes involving a Bloch electron, the electron contributes a momentum equal to $\hbar\mathbf{k}$.

These reasons are sufficiently important to warrant identification of $\hbar\mathbf{k}$ with the momentum. The fact is, nevertheless, that $\hbar\mathbf{k}$ is *not* equal to the actual momentum of the Bloch electron. To make the distinction clear, let us denote the vector $\hbar\mathbf{k}$ by $\mathbf{p}_c$. That is,

$$\mathbf{p}_c = \hbar\mathbf{k}. \tag{5.91}$$

We shall refer to this as the *crystal momentum.*

The actual momentum of the electron $\mathbf{p}$ can be evaluated using quantum methods. According to quantum mechanics the average momentum is given by

$$\mathbf{p} = \langle\psi_{\mathbf{k}}| -i\hbar\nabla|\psi_{\mathbf{k}}\rangle, \tag{5.92}$$

where $-i\hbar\nabla$ is the momentum operator and $\psi_{\mathbf{k}}$ is the Bloch function. If one evaluates this integral, using the properties of the wave function $\psi_{\mathbf{k}}$ (see the problem section at the end of this chapter), one finds that

$$\mathbf{p} = m_0\mathbf{v}, \tag{5.93}$$

where m is the mass of the *free* electron and $\mathbf{v}$ is the velocity as given by (5.70). Thus the true momentum of the electron is equal to the true mass m times the actual velocity $\mathbf{v}$, which seems to be a plausible result.

In retrospect, one may have suspected the original identification of $\mathbf{p}_c$ with the actual momentum from the outset. Since the function $u_{\mathbf{k}}$ in (5.89) is not a constant, the Bloch function ψ_k is not quite a plane wave, and correspondingly the vector $\hbar\mathbf{k}$ is not quite equal to the momentum. Also, if $\mathbf{p}_c = \hbar\mathbf{k}$ were the true momentum, then the force appearing on the right of (5.90) should have been the *total* force, and not just the *external* force. As we shall see, there is a force exerted by the lattice, yet this force does not appear to influence $\mathbf{p}_c$.

The above ideas may now be assembled to give a physical interpretation of the effective mass. Since the vector $\mathbf{p} = m_0\mathbf{v}$ is equal to the true momentum, one may write

$$m_0 \frac{dv}{dt} = F_{\text{tot}} = F_{\text{ext}} + F_L, \tag{5.94}$$

where F_{tot} and F_L are, respectively, the total force and the lattice force acting on the electron. By lattice force, we mean the force exerted by the lattice on the electron as a result of its interaction with the crystal potential. The left side in (5.94) can be readily expressed in terms of the effective mass, namely

$$m_0 \frac{dv}{dt} = m_0 \frac{F_{\text{ext}}}{m^*}, \tag{5.95}$$

as we can see by referring to Eqs. (5.81) through (5.83). Substituting this into (5.94), and solving for m^*, one finds

$$m^* = m_0 \frac{F_{\text{ext}}}{F_{\text{ext}} + F_L}. \tag{5.96}$$

Now we see that the reason why m^* is different from m_0, the free mass, lies in the presence of the lattice force F_L. If F_L were to vanish, the effective mass would become equal to the true mass.

The effective mass m^* may be smaller or larger than m_0, or even negative, depending on the lattice force. Suppose that the electron is "piled up" primarily near the top of the crystal potential, as shown in Fig. 5.38(a). When an

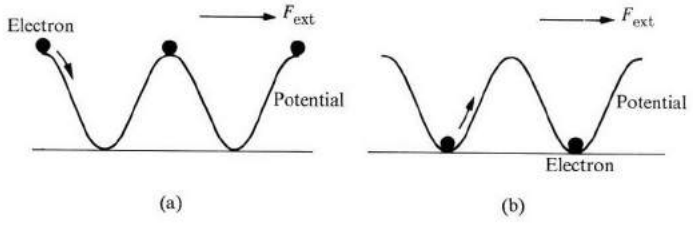


Fig. 5.38 (a) Electron spatial distribution leading to an effective mass m^* smaller than m_0. (b) A distribution leading to $m^* > m_0$.

external force is applied, it causes the electron to "roll downhill" along the potential curve. As a result, a positive lattice force becomes operative and hence, according to (5.96), $m^* < m_0$. This is what happens in alkali metals, for instance, and in the conduction band in semiconductors. Here m^* is less than m_0 because the lattice force assists the external force.

On the other hand, when the electron is piled mainly near the bottom of the potential curve (Fig. 5.38b), then clearly the lattice force tends to oppose the external force, resulting in $m^* > m_0$. This is the situation in the alkali halides, for instance. If the potential wave is sufficiently steep, then F_L becomes larger than F_{ext}, and m^* becomes negative.

Note that the lattice force F_L, which appears in (5.94), is a force induced by the external force. Thus if $F_{ext} = 0$, then the velocity is constant (Section 5.13), and hence $F_L = 0$, according to (5.94). It is true that the lattice also exerts a force on an otherwise-free electron even in the absence of F_{ext}, but that force has already been included in the solution of the Schrödinger equation, and hence in the properties of the state $\psi_{\mathbf{k}}$. That force (as we stated in Sections 5.13 and 4.4) does not scatter the wave $\psi_{\mathbf{k}}$.

However, the crystal momentum $\mathbf{p}_c = \hbar\mathbf{k}$ is still a very useful quantity. In problems of electron dynamics in external fields, crystal momentum is much more useful than true momentum, since it is easier to follow motion in k-space than in real space. Therefore we shall continue to use $\mathbf{p}_c$ and refer to it as the momentum, when there is no ambiguity, and even drop the subscript c.

In other words, the effective mass m^* and the crystal momentum $\hbar\mathbf{k}$ are artifices which allow us—formally at least—to ignore the lattice force and concentrate on the external force only. This is very useful, because lattice force is not known *a priori*, nor is it easily found and manipulated as is the external force.

5.17 THE HOLE

A *hole* occurs in a band that is completely filled except for one vacant state. Figure 5.39 shows such a hole. When we consider the dynamics of the hole in an

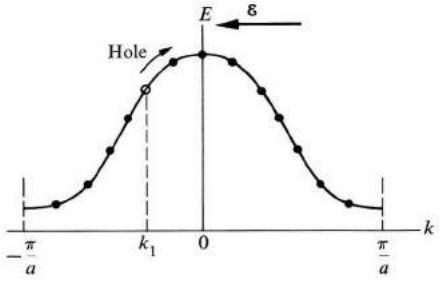


Fig. 5.39 The hole and its motion in the presence of an electric field.

external field, we find it far more convenient to focus on the motion of the vacant site than on the motion of the enormous number of electrons filling the band. The concept of the hole is an important one in band theory, particularly in semi-

conductors, in which it is essential to the operation of many valuable devices, e.g., the transistor.

Suppose the hole is located at the wave vector k_1, as shown in Fig. 5.39. The current density of the whole system is

$$J_h = \frac{-e}{V} \sum_k{}' v_e(k), \tag{5.97}$$

where the sum is over all the electrons in the band, with the prime over the summation sign, indicating that the state k_1 is to be excluded, since that state is vacant. Since the sum over the filled band is zero, the current density (5.97) is also equal to

$$J_h = \frac{e}{V} v_e(k_1). \tag{5.98}$$

That is, the current is the same as if the band were empty, except for an electron of *positive* charge $+e$ located at k_1.

When an electric field is now applied to the system, and directed to the left (Fig. 5.39), all the electrons move uniformly to the right, in k-space, and at the same rate (Section 5.14). Consequently the vacant site also moves to the right, together with the rest of the system. The change in the hole current in a time interval δt can be found from (5.98):

$$\delta J_h = \frac{e}{V} \left(\frac{dv_e}{dk} \right)_{k_1} \frac{dk}{dt} \delta t,$$

which, when we use (5.70), (5.83), and (5.78), can be transformed into

$$\delta J_h = \frac{e}{V} \frac{1}{m^*(k_1)} F \delta t = \frac{1}{V} \left(\frac{-e^2}{m^*(k_1)} \right) \mathscr{E} \delta t, \tag{5.99}$$

where $m^*(k_1)$ is the mass of an electron occupying state k_1.

This equation gives the electric current of the hole, induced by the electric field, which is the observed current.† Since the hole usually occurs near the top of the band—due to thermal excitation of the electron to the next-higher band, where the mass $m^*(k_1)$ is negative—it is convenient to define the mass of a hole as

$$m_h^* = -m^*(k_1), \tag{5.100}$$

† In practice a band contains not a single hole but a large number of holes, and in the absence of an electric field the net current of these holes is zero because of the mutual cancelation of the contributions of the various holes, i.e., the sum of the expression (5.98) over the holes vanishes. When a field is applied, however, induced currents are created, and since these are additive, as seen from (5.99), a nonvanishing net current is established.

which is a positive quantity, and write (5.99) as

$$\delta J_h = \frac{1}{V}\frac{e^2}{m_h^*}\mathscr{E}\,\delta t. \tag{5.101}$$

Note that the hole current, like the electron current, is in the same direction as the electric field.

By examining (5.98) and (5.101), we can see that the motion of the hole, both with and without an electric field, is the same as that of a *particle with a positive charge e and a positive mass* m_h^*. Viewing the hole in this manner results in a great simplification, in that the motion of all the electrons in the band has been reduced to that of a single "particle." This representation will be used frequently in the following discussions.

We may note, incidentally, that according to (5.99), if the hole were to lie near the bottom of the band, where $m^*(k_1) > 0$, then the current would be opposite to the field. This means that the system would act as an amplifier, with the field absorbing energy from the system. This situation is not likely to occur, however, because the hole usually lies near the top of the band.†

5.18 ELECTRICAL CONDUCTIVITY

We discussed electrical conductivity previously in connection with the free-electron model (Sections 4.4 and 4.8), in which we obtained the result

$$\sigma = \frac{ne^2\tau_F}{m^*}. \tag{5.102}$$

The quantity n is the concentration of the conduction—or valence—electrons and τ_F is the collision time for an electron at the Fermi surface. Now let us derive the corresponding expression for electrical conductivity within the framework of band theory.

When the system is at equilibrium—i.e., when there is no electric field—the FS is centered exactly at the origin, as shown in Fig. 5.40(a). Consequently the net current is zero, because the velocities of the electrons cancel in pairs. That is, for every electron in state $\mathbf{k}$ whose velocity is $\mathbf{v}(\mathbf{k})$, another electron exists in state $-\mathbf{k}$ whose velocity $\mathbf{v}(-\mathbf{k}) = -\mathbf{v}(\mathbf{k})$ is simply the reverse of the former. This result, found in the free-electron model, also holds good in band theory, and accounts for the vanishing of the current at equilibrium.

When an electric field is applied, each electron travels through $\mathbf{k}$-space at a

† A proposal for an amplifier operating on essentially the same principle was advanced by H. Kroemer, *Phys. Rev.* **109**, 1856 (1955).

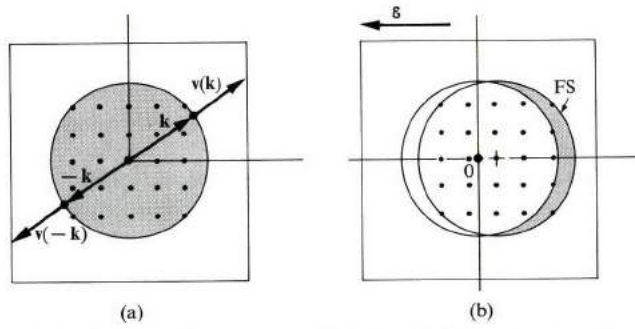


Fig. 5.40 (a) In the absence of an electric field the FS is centered at the origin, and the electron currents cancel in pairs. (b) In the presence of an electric field, the FS is displaced and a net current results.

uniform rate, as discussed in Section 5.14. That is,

$$\delta k_x = -\frac{e\mathscr{E}}{\hbar}\delta t,$$

where δk_x is the displacement in a time interval δt. Since an electron usually "lives" for an interval equal to the collision time τ, the average displacement is

$$\delta k_x = -\frac{e\mathscr{E}}{\hbar}\tau. \tag{5.103}$$

Consequently the FS is displaced rigidly by this amount, as shown in Fig. 5.40(b). There are now some electrons which are *not* compensated—i.e., canceled—by other electrons, and which are indicated by the cross-hatched crescent-shaped region. They contribute a *net* current.

The density of this current can be calculated as follows: It is given by

$$\begin{aligned} J_x &= -e\,\bar{v}_{\mathrm{F},x} \times \text{concentration of uncompensated electrons} \\ &= -e\bar{v}_{\mathrm{F},x} g(E_\mathrm{F})\,\delta E \\ &= -e\bar{v}_{\mathrm{F},x} g(E_\mathrm{F}) \left(\frac{\partial E}{\partial k_x}\right)_{E_\mathrm{F}} \delta k_x, \end{aligned} \tag{5.104}$$

where $\bar{v}_{\mathrm{F},x}$ is the component of the Fermi velocity in the x-direction and the bar indicates an average value.

Note that $g(E_\mathrm{F})\,\delta E$ gives the concentration of uncompensated electrons, $g(E_\mathrm{F})$ being the density of states at the FS and δE the energy absorbed by the electron from the field. Noting that $\partial E/\partial k_x = \hbar v_{\mathrm{F},x}$, and substituting for δk_x from (5.103). one obtains

$$J_x = e^2 \bar{v}_{\mathrm{F},x}^2 \tau_\mathrm{F} g(E_\mathrm{F}) \mathscr{E}, \tag{5.105}$$

where the collision time has been designated as τ_F, inasmuch as we are clearly dealing with electrons lying at the FS. Note that the current is in the same direction as the field.

For a spherical FS, there is a spherical symmetry, and hence one may write $\bar{v}_{F,x}^2 = \frac{1}{3}v_F^2$ which, when substituted into (5.105), leads finally to the following expression for the electrical conductivity:

$$\sigma = \tfrac{1}{3}\, e^2 v_F^2 \tau_F g(E_F), \tag{5.106}$$

which is the expression we have been seeking.

Note that σ depends on the Fermi velocity and the collision time, but also note the dependence on the density of states at the FS, $g(E_F)$. Often this is the predominant factor in determining the conductivity, as we shall see shortly.

Expression (5.106) is more general than the free-electron formula (5.102), and far more meaningful. Equation (5.102) implies that conductivity is controlled primarily by n, the electron concentration. However, conductivity is, in fact, controlled primarily by the density of states $g(E_F)$ instead. In the appropriate limit, expression (5.106) reduces to (5.102) as a special case, as it must. To establish this, we use the relation $g(E_F) = \frac{1}{2}\pi^2(2m^*/\hbar^2)^{3/2}E_F^{1/2}$ [see (5.63)], $E_F = \frac{1}{2}m^*v_F^2$, and $E_F = (\hbar^2/2m^*)(3\pi^2 n)^{2/3}$ [from (5.67)], which we find reduce (5.106) to (5.102).

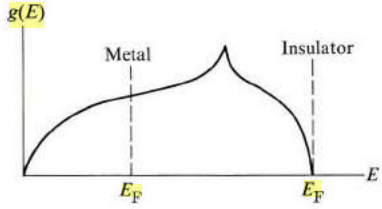


Fig. 5.41 Position of the Fermi energy level in a monovalent metal and in an insulator. In the former, $g(E_F)$ is large, while in the latter, $g(E_F) = 0$.

Figure 5.41 shows the density of states for a typical solid, indicating the position of the Fermi level for a monovalent metal, and also for an insulator. In the metal, the level E_F is located near the middle of the band where $g(E_F)$ is large, leading to a large conductivity, according to (5.106). In the insulator, the level E_F is right at the top of the band, where $g(E_F) = 0$. Thus the conductivity is zero, despite the fact that the Fermi velocity, which also appears in (5.106), is very large.

The expression (5.106), though restricted to the case in which the FS is spherical, is useful in unraveling the important role played by the density of states. The results may be generalized to include the effects of more complex FS shapes

(as you will find by referring to the bibliography), which often lead to unwieldy expressions.

Another important aspect of the electrical conduction process—and of transport phenomena in general—is that they enable us to calculate the collision time τ_F. We discussed this subject in a semiclassical fashion in Section 4.4 for the free-electron model, but a more rigorous treatment involves the use of quantum methods (see Appendix A), and perturbation theory in particular. The scattering mechanisms are the same as those discussed in connection with the free-election model (Section 4.5)—scattering by lattice vibrations, impurities, and other lattice defects—but the details of the calculation are highly complicated (Ziman, 1960), and will not be given here.

5.19 ELECTRON DYNAMICS IN A MAGNETIC FIELD: CYCLOTRON RESONANCE AND THE HALL EFFECT

We discussed electron dynamics in a magnetic field in Section 4.10 with respect to the free-electron model, where we also treated cyclotron resonance and the Hall effect. Here we shall discuss the way in which this is modified for a Bloch electron, taking into account the interaction with the crystal potential. This subject is more useful in practice, as the magnetic field is often used in studies of band structure.

Cyclotron resonance

The basic equation of motion describing the dynamics in a magnetic field is

$$\hbar \frac{d\mathbf{k}}{dt} = -e\,[\mathbf{v}(\mathbf{k}) \times \mathbf{B}], \tag{5.107}$$

where the left side is the time derivative of the crystal momentum, and the right side the well-known Lorentz force due to the magnetic field. This equation is a plausible one in light of the discussion in Sections 5.14 and 5.16, in which we concluded that the momentum of the crystal usually acts as the familiar momentum, provided only the external force is included. [The equation (5.107) may also be derived from detailed quantum calculations.]

According to (5.107), the change in $\mathbf{k}$ in a time interval δt is given by

$$\delta\mathbf{k} = -(e/\hbar)\,[\mathbf{v}(\mathbf{k}) \times \mathbf{B}]\,\delta t, \tag{5.108}$$

which shows that the electron moves in $\mathbf{k}$-space in such a manner that its displacement $\delta\mathbf{k}$ is perpendicular to the plane defined by $\mathbf{v}$ and $\mathbf{B}$. Since $\delta\mathbf{k}$ is perpendicular to $\mathbf{B}$, this means that the electron trajectory lies in a plane normal to the magnetic field. In addition, $\delta\mathbf{k}$ is perpendicular to $\mathbf{v}$ which, inasmuch as $\mathbf{v}$ is normal to the energy contour in $\mathbf{k}$-space, means that $\delta\mathbf{k}$ lies along such a contour. Putting these two bits of information together, we conclude that the electron rotates along

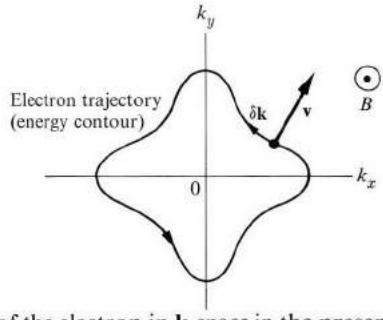


Fig. 5.42 Trajectory of the electron in **k**-space in the presence of a magnetic field.

an energy contour normal to the magnetic field (Fig. 5.42), and in a counterclockwise fashion.

Also note that, because the electron moves along an energy contour, no energy is absorbed from, or delivered to, the magnetic field, in agreement with the well-known facts concerning the interaction of electric charges with a magnetic field.

As Fig. 5.42 shows, the motion of the electron in **k**-space is *cyclic*, since, after a certain time, the electron returns to the point from which it started. The period T for the motion is, according to (5.108), given by

$$T = \oint \delta t = \frac{\hbar}{eB} \oint \frac{\delta k}{v(\mathbf{k})}, \tag{5.109}$$

where the circle on the integration sign denotes that this integration is to be carried out over the complete cycle in **k**-space, i.e., a closed orbit. In (5.109), the differential δk is taken along the perimeter of the orbit, while $v(\mathbf{k})$ is the magnitude of the electron velocity normal to the orbit. Also note that in deriving (5.109) from (5.108), we have used the fact that **v** is normal to **B**, since the electron trajectory lies in a plane normal to **B**.

The angular frequency ω_c associated with the motion is $\omega_c = 2\pi/T$, which, in light of (5.109), is given by

$$\omega_c = (2\pi eB/\hbar) \Big/ \oint \frac{\delta k}{v(\mathbf{k})}. \tag{5.110}$$

This is the *cyclotron frequency* for the Bloch electron. It is the generalization of the cyclotron frequency (4.38) derived for the free-electron model.

We conclude that the motion of a Bloch electron in a magnetic field is a natural generalization of the motion of a free electron in the same field. A free electron executes circular motion in velocity space along an energy contour with a frequency $\omega_c = eB/m^*$. A Bloch electron executes a cyclotron motion along an energy contour with a frequency given by (5.110). The energy contour in this latter case may, of course, be very complicated.

When the standard form $E = \hbar^2 k^2/2m^*$ is applicable, the frequency ω_c in (5.110) may be readily calculated. The cyclotron orbit is circular in this case, and in evaluating the integral we note that $v(\mathbf{k}) = \hbar k/m^*$, which is a constant along the orbit, since the magnitude k of the wave vector is constant along this contour trajectory. Thus

$$\oint \frac{\delta k}{v(\mathbf{k})} = \frac{1}{(\hbar k/m^*)} \oint \delta k = \frac{2\pi k}{(\hbar k/m^*)} = \frac{2\pi m^*}{\hbar},$$

which, when substituted into (5.110), produces

$$\omega_c = eB/m^*.$$

This, as expected, agrees with the result for the free-electron model.

But, of course, Eq. (5.110) is more general than the free-electron result, and applies to a contour of arbitrary shape, although evaluating the integral may become very tedious. In the problem section at the end of this chapter, you will be asked to evaluate ω_c for contours which, although more complicated than those in the free-electron model, are still simple enough to render the integral in (5.110) tractable.

In discussing the above cyclotron motion, we have disregarded the effects of collision. Of course, if this cyclotron motion is to be observed at all, the electron must complete a substantial fraction of its orbit during one collision time; that is, $\omega_c \tau \gtrsim 1$. This necessitates the use of very pure samples at low temperature under a very strong magnetic field.

The Hall effect

When we were discussing the Hall effect in the free-electron model (Section 4.10), we found that the Hall constant is given by

$$R_e = -\frac{1}{n_e e}, \tag{5.111}$$

where n_e is the electron concentration. The negative sign is due to the negative charge of the electron. The general treatment of the Hall effect for Bloch electrons becomes quite complicated for arbitrary FS, requiring considerable mathematical effort (Ziman, 1960). However, we can obtain some important results quite readily.

Suppose that only holes were present in the sample. Then we could apply to the holes the same treatment used for electrons in Section 4.10, and would obtain a Hall constant

$$R_h = \frac{1}{n_h e}, \tag{5.112}$$

where R is now positive because of the positive charge on the hole (n_h is the hole concentration).

Actually, in metals, holes are not present by themselves; there are always some electrons present. Thus when two bands overlap with each other, electrons are present in the upper band and holes in the lower. The expression for the Hall constant when both electrons and holes exist simultaneously is given by (see the problem section)

$$R = \frac{R_e\sigma_e^2 + R_h\sigma_h^2}{(\sigma_e + \sigma_h)^2}, \tag{5.113}$$

where R_e and R_h are the contributions of the individual electrons and holes, as given above, and σ_e and σ_h are the conductivities of the electrons and holes ($\sigma_e = n_e e^2\tau_e/m_e^*$ and $\sigma_h = n_h e^2\tau_h/m_h^*$).

Equation (5.113) shows that the sign of the Hall constant R may be either negative or positive depending on whether the contribution of the electrons or the holes dominates. If we take $n_e = n_h$, which is the case in metals, then $|R_e| = |R_h|$ and the sign of R is determined entirely by the relative magnitudes of the conductivities σ_e and σ_h. Thus if $\sigma_e > \sigma_h$—that is, if the electrons have small mass and long lifetime—the electrons' contribution dominates and R is negative. And when the opposite condition prevails, the holes' contribution dominates, and R is positive. *We* can now understand why some polyvalent metals—e.g., Zn and Cd—exhibit positive Hall constants (see Table 4.3).

5.20 EXPERIMENTAL METHODS IN DETERMINATION OF BAND STRUCTURE

Now let us discuss some of the experimental techniques used to determine the band structure in metals. For example, how did physicists determine the Fermi energies in Table 4.1, or the Fermi surfaces shown in Fig. 5.26 for Cu and Fig. 5.30 for Be? This field of solid-state physics is a wide one, and has been expanding at a rapid pace. Our discussion here will therefore be rather sketchy, leaving it to the reader to pursue the subject in greater detail by referring to the entries in the bibliography.

One can determine the Fermi energy by the method of *soft x-ray emission*. When a metal is bombarded by a beam of high-energy electrons, electrons from the inner K shell† are knocked out, leaving empty states behind. Electrons in the valence band now move to fill these vacancies, undergoing downward transitions, as shown in Fig. 5.43(a). The photons emitted in the transition, usually lying in the soft x-ray region—about 200 eV—are recorded and their energies measured. Figure 5.43(b) shows the intensity of the x-ray spectrum as well as the energy

† The atomic shells $n = 0$, 1, 2, etc., are usually referred to as the K, L, M, etc., shells, respectively.

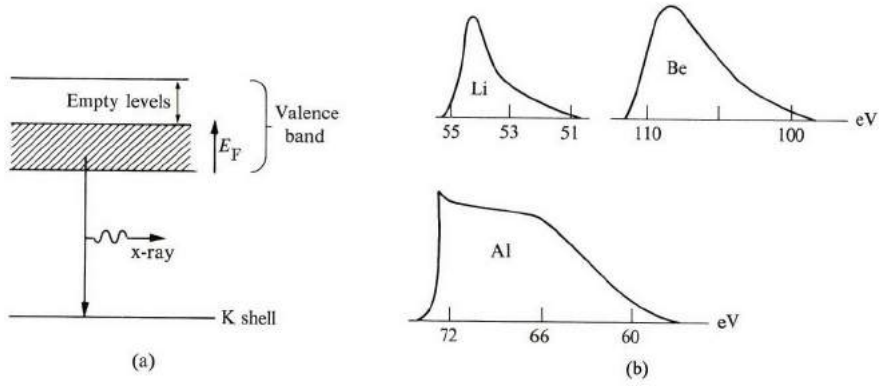


Fig. 5.43 (a) Emission of soft x-rays. (b) Intensity of the spectrum of x-ray emission versus energy for Li, Be, and Al.

range for several metals. Since the K shell is very narrow, almost to the point of being a discrete level, the width of the range shown in Fig. 5.43(b) is due entirely to the spread of the occupied states in the valence band, i.e., the width is equal to the Fermi level. One can also extract information from Fig. 5.43(b) on the shape of the density of states. In fact, the shape of the curve is determined primarily by the density of states of the valence band.

Let us now turn to the determination of the FS, and discuss one of the many methods in common use: the *Azbel–Kaner cyclotron resonance* (AKCR) technique. A semi-infinite metallic slab is placed in a strong static magnetic field $\mathbf{B}_0$, which is parallel to the surface (Fig. 5.44). As a result, electrons in the

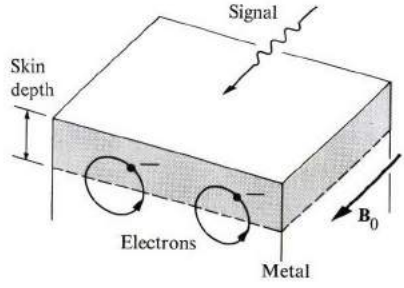


Fig. 5.44 Physical setup for Azbel–Kaner cyclotron resonance.

metal begin to execute a cyclotron motion, with a cyclotron frequency ω_c. Now an alternating electromagnetic signal of frequency ω, circularly polarized in a counterclockwise direction, is allowed to travel parallel to the surface and along the direction of the static field $\mathbf{B}_0$. This signal penetrates the metal only to a

small extent, equal to the skin depth (see Section 4.11), and so is confined to a short distance from the surface. Only electrons in this region are affected by the signal.

The electrons near the surface feel the field of the signal and absorb energy from it. This absorption is greatest when the condition

$$\omega = \omega_c \tag{5.114}$$

is satisfied, because the electron then remains in phase with the signal field throughout the cycle. This is the *resonance* condition.

During a part of its cycle, the electron actually penetrates the metal beyond the skin depth, where the signal field vanishes. A resonance condition is still satisfied, provided only that, when the electron returns to the region at the surface, it is again in phase with the field. In general, therefore, the condition for resonance is

$$\omega = l\omega_c, \tag{5.115}$$

where $l = 1, 2, 3$, etc., at all harmonics of the cyclotron frequency ω_c. The AKCR for Cu is shown in Fig. 5.45. (Usually the frequency ω is held fixed and the field is varied until the resonance condition is satisfied.)

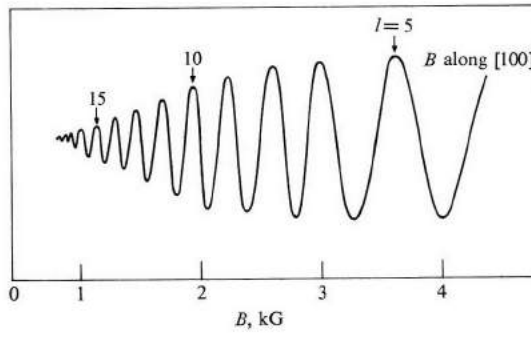


Fig. 5.45 AKCR spectrum in Cu at $T = 4.2°K$. The crystal surface (upper surface) is cut along the (100) plane. The ordinate of the curve represents the derivative of the surface resistivity with respect to the field. [After Haüssler and Wells, *Phys. Rev.*, **152,** 675, 1966]

Not only is the method capable of determining ω_c (and hence the effective mass m^*), but also the actual shape of the FS. In general, electrons in different regions of the surface have different cyclotron frequencies, but the frequency which is most pronounced in the absorption is the frequency appropriate to the *extremal* orbit, i.e., where the FS cross section perpendicular to $\mathbf{B}_0$ is greatest, or smallest. Therefore, by varying the orientation of $\mathbf{B}_0$, one can measure the extremal sections in various directions, and reconstruct the FS.

The experiment is usually performed at very low temperatures, that is, $T \simeq 4°K$, on very pure samples, and at very strong fields—about 100 kG. Under these conditions, the collision time τ is long enough, and the cyclotron frequency ω_c high enough, so that the high-field condition $\omega_c\tau \gg 1$ is satisfied. In this limit, the electron executes many cycles in a single collision time, leading to a sharp, well-resolved resonance. The frequency ω_c usually falls in the microwave range.

Optical ultraviolet techniques are also used in determining band structure. Figure 5.46 shows the principle of the method. When a light beam impinges on a

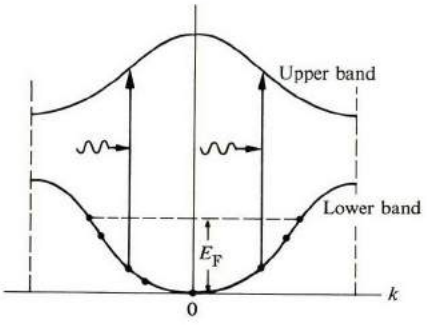


Fig. 5.46 Interband optical absorption.

metal, electrons are excited from below the Fermi level into the next-higher band. This *interband absorption* may be observed by optical means—i.e., reflectance and absorption techniques, which give information concerning the shape of the energy bands. In this case, two bands are involved simultaneously, and the results cannot be expressed in terms of the individual bands separately. But if the shape of one of these is known, the shape of the other may be determined. For further discussion of the optical properties of metals in the ultraviolet region—which is where the frequencies happen to lie in the case of most metals—refer to Section 8.9.

5.21 LIMIT OF THE BAND THEORY; METAL–INSULATOR TRANSITION

So far in this chapter we have based our discussion entirely on the so-called *band model* of solids. This model has been of immense value to us; it is capable of explaining all the observed properties of metals, and is the basis of the semiconductor properties to be discussed in Chapters 6 and 7. Yet this model has a limitation which we now wish to probe.

Consider, for example, the case of Na. This substance is a conductor because the 3s band is only partially filled—half filled, to be exact. Suppose that we cause the Na to expand by some means, so that the lattice constant a can be increased

arbitrarily. Would the material then remain a conductor for any arbitrary value of a? The answer must be yes, if one is to believe the band model, because, regardless of the value of a, the 3s band would always be half full. It is true (the model predicts further) that the conductivity σ decreases as a increases, but the decrease is gradual, as shown in Fig. 5.47.

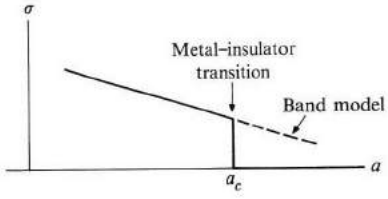


Fig. 5.47 Electrical conductivity σ versus lattice constant a.

In fact, however, this is not correct. As a increases, a critical value a_c is reached at which the conductivity drops to zero abruptly, rendering the solid an *insulator*, and it remains so for all values $a > a_c$. Thus for a sufficiently large lattice constant, the metal is transformed into an insulator, and we speak of the *metal–insulator transition* (also known as the *Mott transition*).

To explain this transition, we need to recall some of the fundamental concepts underlying band theory. In this theory, Bloch electrons are assumed to be *delocalized*, extending throughout the crystal, and it is this delocalization which is responsible for metallic conductivity. As a delocalized particle, the Bloch electron spends a fraction of its time ($1/N$, to be exact), at each atom. The interaction between the various Bloch electrons is taken into account only in an average manner, i.e., the interaction between individual electrons is neglected.

However, as a increases, the bandwidth decreases (recall the TB model, Section 5.8), until it becomes quite small at sufficiently large a. In that case, the band model breaks down because it allows the presence of two or more electrons at the same lattice site, which cannot happen because of the Coulomb repulsion between electrons. When the band is wide, this is not serious, because electrons can readjust their kinetic energies to compensate for the increase in the coulomb potential energy. But for a narrow band the kinetic energy is, at best, quite small, and this readjustment is not possible.

In effect, for very large a, the proper electronic orbitals in a crystal are not of the Bloch type. They are localized orbitals centered around their respective sites, which mitigates the large coulomb energy. Since the orbitals are localized, as in the case of free atoms, conductivity vanishes, as depicted in Fig. 5.47.

Note that the above conclusion holds true even though the energy levels still form a band, and even though the band is only half full. The point is that electronic orbitals become localized, and hence nonconducting.

The metal–insulator transition has been observed in VO_2 (vanadium oxide)

and other oxide materials. Although VO_2 is normally an insulator, it is transformed into a metallic material at sufficiently high pressure.

SUMMARY

The Bloch theorem and energy bands in solids

The wave function for an electron moving in a periodic potential, as in the case of a crystal, may be written in the *Bloch form*,

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}),$$

where the function $u_{\mathbf{k}}(\mathbf{r})$ has the same periodicity as the potential. The function $\psi_{\mathbf{k}}$ has the form of a plane wave of vector $\mathbf{k}$, which is modulated by the periodic function $u_{\mathbf{k}}$. Although the function $\psi_{\mathbf{k}}$ itself is nonperiodic, the electron probability density $|\psi_{\mathbf{k}}|^2$ is periodic; i.e., the electron is delocalized, and is deposited periodically throughout the crystal.

The energy spectrum of the electron is comprised of a set of continuous *bands*, separated by regions of forbidden energies which are called *energy gaps*. The electron energy is commonly denoted by $E_n(\mathbf{k})$, where n is the band index.

Regarded as a function of the vector $\mathbf{k}$, the energy $E(\mathbf{k})$ satisfies several symmetry properties. First, it has translational symmetry

$$E(\mathbf{k} + \mathbf{G}) = E(\mathbf{k}),$$

which enables us to restrict our consideration to the first Brillouin zone only. The energy function $E(\mathbf{k})$ also has inversion symmetry, $E(-\mathbf{k}) = E(\mathbf{k})$, and rotational symmetry in $\mathbf{k}$-space.

The NFE and TB models

In the NFE model the crystal potential is taken to be very weak. Solving the Schrödinger equation shows that the electron behaves essentially as a free particle, except when the wave vector $\mathbf{k}$ is very close to, or at, the boundaries of the zone. In these latter regions, the potential leads to the creation of energy gaps. The first gap is given by

$$E_g = 2\,|V_{-2\pi/a}|,$$

where $V_{-2\pi/a}$ is a *Fourier component* of the crystal potential.

The wave functions at the zone boundaries are described by standing waves, which result from strong Bragg reflection of the electron wave by the lattice.

The TB model, in which the crystal potential is taken to be strong, leads to the same general conclusions as the NFE model, i.e., the energy spectrum is composed of a set of continuous bands. The TB model shows that the width of the band increases and the mobility of the electron becomes greater (the mass lighter) as the overlap between neighboring atomic functions increases.

Metals versus insulators

If the valence band of a given substance is only partially full, the substance acts like a metal or conductor because an electric field produces an electric current in the material. If the valence band is completely full, however, no current is produced, regardless of the field, and the substance is an insulator.

When the gap between the valence band and the band immediately above it is small, electrons may be thermally excited across the gap. This gives rise to a small conductivity, and the metal is called a *semiconductor*.

Velocity of the Bloch electron

An electron in the Bloch state $\psi_{\mathbf{k}}$ moves through the crystal with a velocity

$$\mathbf{v} = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}).$$

This velocity remains constant so long as the lattice remains perfectly periodic.

Electron dynamics in an electric field

In the presence of an electric field, an electron moves in $\mathbf{k}$-space according to the relation

$$\dot{\mathbf{k}} = -(e/\hbar)\mathcal{E}.$$

The motion is uniform, and its rate proportional to the field. One obtains this relation at once if one regards the electron as having a momentum $\hbar\mathbf{k}$.

Effective mass

The effective mass of a Bloch electron is given by

$$m^* = \hbar^2/(d^2E/dk^2).$$

The mass is positive near the bottom of the band, where the curvature is positive. But near the top, where the band curvature is negative, the effective mass is also negative. The fact that the effective mass is different from the free mass is due to the effect of the lattice force on the electron.

The hole

A hole exists in a band which is completely full, with one vacant state. The hole acts as a particle of positive charge $+e$. When the hole lies near the top of the band, which is the usual situation, the hole also behaves as if it has a positive effective mass.

Electrical conductivity

Electrical conductivity is given by

$$\sigma = \tfrac{1}{3} e^2 v_F^2 \tau_F g(E_F).$$

This expression is a particularly sensitive function of $g(E_F)$, the density of states at the Fermi energy. In monovalent metals, σ is large because $g(E_F)$ is large, while the opposite is true for polyvalent metals. In insulators, the electrical conductivity vanishes because $g(E_F) = 0$.

Under appropriate circumstances, the above expression for σ reduces to the familiar form $\sigma = ne^2\tau_F/m^*$ of the free-electron model.

Cyclotron resonance and the Hall effect

The motion of a Bloch electron in a magnetic field is governed by

$$\hbar \frac{d\mathbf{k}}{dt} = -e(\mathbf{v} \times \mathbf{B}).$$

The electron moves along an energy contour in a trajectory perpendicular to the field **B**, and the motion is referred to as cyclotron motion.

The *cyclotron frequency* is found to be

$$\omega_c = (2\pi eB/h) \Big/ \oint \frac{\delta k}{v},$$

where the integral in the denominator is taken over a closed contour. Measuring this frequency gives information about the shape of the contour, and hence about the shape of the band. The above expression reduces to the familiar form $\omega_c = eB/m^*$ for the case of a standard band.

When both electrons and holes are present in the metal, they both contribute to the Hall constant. The resulting expression is

$$R = \frac{R_e\sigma_e^2 + R_h\sigma_h^2}{(\sigma_e + \sigma_h)^2}.$$

When the electron term dominates, the Hall constant R is negative; when the hole term dominates, the Hall constant R is positive.

REFERENCES

J. Callaway, 1963, *Energy Band Theory*, New York: Academic Press

J. F. Cochran and R. R. Haering, editors, 1968, *Electrons in Metals*, London: Gordon and Breach

W. A. Harrison and M. B. Webb, editors, 1968, *The Fermi Surface*, New York: Wiley

W. A. Harrison, 1970, *Solid State Theory*, New York: McGraw-Hill

N. F. Mott and H. Jones, 1936, *Theory of the Properties of Metals and Alloys*, Oxford: Oxford University Press; also Dover Press (reprint)

A. B. Pippard, 1965, *Dynamics of Conduction Electrons*, London: Gordon and Breach

F. Seitz, 1940, *Modern Theory of Solids*, New York: McGraw-Hill

D. Schoenberg, "Metallic Electrons in Magnetic Fields," *Contemp. Phys.* **13**, 321, 1972

J. C. Slater, 1965, *Quantum Theory of Molecules and Solids*, Volume II, New York: McGraw-Hill

A. H. Wilson, 1953, *Theory of Metals*, second edition, Cambridge: Cambridge University Press

QUESTIONS

1. It was pointed out in Sections 6.3 and 4.3 that an electron spends only a little time near an ion, because of the high speed of the electron there. At the same time it was claimed that the ions are "screened" by the electrons, implying that the electrons are so distributed that most of them are located around the ions. Is there a paradox here? Explain.
2. Figure 5.10(c) is obtained from Fig. 5.10(a) by cutting and displacing various segments of the free-electron dispersion curve. Is this rearrangement justifiable for a truly free electron? How do you differentiate between an empty lattice and free space?
3. Explain why the function ψ_0 in Fig. 5.18(b) is flat throughout the Wigner–Seitz cell except close to the ion, noting that this behavior is different from that of an atomic wave function, which decays rapidly away from the ion. This implies that the coulomb force due to the ion in cell A is much weakened in the flat region. What is the physical reason for this?
4. *Band overlap* is important in the conductivity of polyvalent metals. Do you expect it to take place in a one-dimensional crystal? You may invoke the symmetry properties of the energy band.

PROBLEMS

1. Figure 5.7 shows the first three Brillouin zones of a square lattice.
   a) Show that the area of the third zone is equal to that of the first. Do this by appropriately displacing the various fragments of the third zone until the first zone is covered completely.
   b) Draw the fourth zone, and similarly show that its area is equal to that of the first zone.
2. Draw the first three zones for a two-dimensional rectangular lattice for which the ratio of the lattice vectors $a/b = 2$. Show that the areas of the second and third zones are each equal to the area of the first.
3. Convince yourself that the shapes of the first Brillouin zones for the fcc and bcc lattices are those in Fig. 5.8.
4. Show that the number of allowed **k**-values in a band of a three-dimensional sc lattice is N, the number of unit cells in the crystal. Hint : Count the k values in 1st BZ of SC
5. Repeat Problem 4 for the first zone of an fcc lattice (zone shown in Fig. 5.8a).
6. Derive Eqs. (5.21) and (5.22).
7. Show that the first three bands in the empty-lattice model span the following energy ranges.

$E = \frac{\hbar^2 k^2}{2m}$ (5.16) E_1: 0 to $\pi^2\hbar^2/2m_0a^2$; E_2: $\pi^2\hbar^2/m_0a^2$ to $2\pi^2\hbar^2/m_0a^2$;

見圖 5.10(a)，此入兩帶 E_3: $2\pi^2\hbar^2/m_0a^2$ to $9\pi^2\hbar^2/2m_0a^2$.

8. a) Show that the octahedral faces of the first zone of the fcc lattice (Fig. 5.8a) are due to Bragg reflection from the (111) atomic planes, while the other faces are due to reflection from the (200) planes.
   b) Show similarly that the faces of the zone for the bcc lattice are associated with Bragg reflection from the (110) atomic planes.

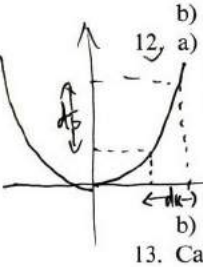
4. The volume of 1st BZ is the same as the unit cell of the reciprocal lattice $(\frac{2\pi}{a})^3$, From (5.14), $k_x = n_1(\frac{2\pi}{L})$, $k_y = n_2(\frac{2\pi}{L})$, $k_z = n_3(\frac{2\pi}{L})$

V reciprocal point is $(\frac{2\pi}{L})^3$. The allowed k values in a band (one BZ) are $(\frac{2\pi}{a})^3 / (\frac{2\pi}{L})^3 \simeq \frac{V}{a^3} = N =$ number of unit cells in the crystal

9. Suppose that the crystal potential in a one-dimensional lattice is composed of a series of rectangular wells which surround the atom. Suppose that the depth of each well is V_0 and its width $a/5$.
   a) Using the NFE model, calculate the values of the first three energy gaps. Compare the magnitudes of these gaps.
   b) Evaluate these gaps for the case in which $V_0 = 5\,\text{eV}$ and $a = 4\,\text{Å}$.
10. Prove that the wave function used in the TB model, Eq. (5.27), is normalized to unity if the atomic function ϕ_v is so normalized. [*Hint*: For the present purpose you may neglect the overlap between the neighboring atomic functions.]
11. The energy of the band in the TB model is given by

$$E(\mathbf{k}) = E_v - \beta - \gamma \sum_j e^{i\mathbf{k}\cdot\mathbf{x}_j},$$

where β and γ are constants, as indicated in the text, and $\mathbf{x}_j$ is the position of the jth atom relative to the atom at the origin.
   a) Find the energy expression for a bcc lattice, using the nearest-neighbor approximation. Plot the energy contours in the $k_x - k_y$ plane. Determine the width of the energy band.
   b) Repeat part (a) for the fcc lattice.
12. a) Using the fact that the allowed values of k in a one-dimensional lattice are given by $k = n(2\pi/L)$, show that the density of electron states in the lattice, for a lattice of unit length, is given by The number of k-values in the segment dk



$$g(E) = \frac{1}{2\pi} \Big/ \left(\frac{dE}{dk}\right) \cdot \frac{dk}{\left(\frac{2\pi}{L}\right)}$$, The corresponding energy level in segment dE is $g(E)dE$

   b) Evaluate this density of states in the TB model, and plot $g(E)$ versus E.
13. Calculate the density of states for the first zone of an sc lattice according to the empty-lattice model. Plot $g(E)$, and determine the energy at which $g(E)$ has its maximum. Explain qualitatively the behavior of this curve.
14. a) Using the free-electron model, and denoting the electron concentration by n, show that the radius of the Fermi sphere in $\mathbf{k}$-space is given by

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

   b) As the electron concentration increases, the Fermi sphere expands. Show that this sphere begins to touch the faces of the first zone in an fcc lattice when the electron-to-atom ratio $n/n_a = 1.36$, where n_a is the atom concentration.
   c) Suppose that some of the atoms in a Cu crystal, which has an fcc lattice, are gradually replaced by Zn atoms. Considering that Zn is divalent while Cu is monovalent, calculate the atomic ratio of Zn to Cu in a CuZn alloy (brass) at which the Fermi sphere touches the zone faces. Use the free-electron model. (This particular mixture is interesting because the solid undergoes a structural phase change at this concentration ratio.)
15. a) Calculate the velocity of the electron for a one-dimensional crystal in the TB model, and prove that the velocity vanishes at the zone edge.
   b) Repeat (a) for a square lattice. Show that the velocity at a zone boundary is parallel to that boundary. Explain this result in terms of the Bragg reflection.

These two numbers must be equal $g(E)\,dE = \dfrac{dk}{\left(\frac{2\pi}{L}\right)}$, since $L = 1$

$$g(E) = \left(\frac{1}{2\pi}\right) \Big/ \frac{dE}{dk}$$

c) Repeat for a three-dimensional sc lattice, and show once more that the electron velocity at a zone face is parallel to that face. Explain this in terms of Bragg reflection. Can you make a general statement about the direction of the velocity at a zone face?

16. Suppose that a static electric field is applied to an electron at time $t = 0$, at which instant the electron is at the bottom of the band. Show that the position of the electron in real space at time t is given by

$$x = x_0 + \frac{1}{F} E(k = Ft/\hbar),$$

where x_0 is the initial position and $F = -e\mathscr{E}$ is the electric force. Assume a one-dimensional crystal, and take the zero-energy level at the bottom of the band. Is the motion in real space periodic? Explain.

17. a) Using the TB model, evaluate the effective mass for an electron in a one-dimensional lattice. Plot the mass m^* versus k, and show that the mass is independent of k only near the origin and near the zone edge.
    b) Calculate the effective mass at the zone center in an sc lattice using the TB model.
    c) Repeat (b) at the zone corner along the [111] direction.
18. Prove Eq. (5.18).
19. a) Calculate the cyclotron frequency ω_c for an energy contour given by

$$E(\mathbf{k}) = \frac{\hbar^2}{2m_1^*} k_x^2 + \frac{\hbar^2}{2m_2^*} k_y^2,$$

where the magnetic field is perpendicular to the plane of the contour.

$$\left[\textit{Answer: } \omega_c = \sqrt{\frac{e^2}{m_1^* m_2^*}} B.\right]$$

    b) Repeat (a) for an ellipsoidal energy surface

$$E(\mathbf{k}) = \frac{\hbar^2}{2m_1^*} (k_x^2 + k_y^2) + \frac{\hbar^2}{2m_3^*} k_z^2,$$

where the field $\mathbf{B}$ makes an angle θ with the k_z-axis of symmetry of the ellipsoid.

$$\left[\textit{Answer: } \omega_c = \left[\left(\frac{eB}{m_1^*}\right)^2 \cos^2\theta + \frac{e^2B^2}{m_1^* m_3^*} \sin^2\theta\right]^{1/2}.\right]$$

20. In Section 5.19 we discussed the motion of a Bloch electron in **k**-space in the presence of a magnetic field. The electron also undergoes a simultaneous motion in **r**-space. Discuss this motion, and in particular show that the trajectory in **r**-space lies in a plane parallel to that in **k**-space, that the shapes of the two trajectories are the same except that the one in **r**-space is rotated by an angle of $-\pi/2$ relative to the other, and expanded by a linear scale factor of $(\hbar/eB)$. [*Hint:* Use Eq. (5.108) to relate the electron displacements in **r**- and **k**-space.]
21. Prove Eq. (5.113) for the Hall constant of an electron–hole system.

14. (a) Each point in the momentum space is surrounded by a volume $(\frac{2\pi}{L})^3$ L^3 is the volume of the metal. $2\times$[Total number of points inside the Fermi sphere]

$= 2 \times \frac{\frac{4}{3}\pi k_F^3}{(\frac{2\pi}{L})^3} = \frac{V}{3\pi^2} k_F^3 = N =$ Total number of electrons in the meta

$k_F = (3\pi^2 \frac{N}{V})^{1/3}$ since $n = \frac{N}{V}$ $k_F = (3\pi^2 n)^{1/3}$.

(b) 倒格子中心到8顶点 1^{st} $d = \frac{1}{4}$ 对角线 (FCC → BCC), concentration of atoms in a FCC lattice is $n_a = \frac{4 \text{ atoms}}{a^3}$, $k_F = \frac{d}{4}$, $n = 1.36\, n$

(c) $\frac{n_e^{Cu}\ (\text{concentration of electrons})}{n_{Cu}\ (\text{concentration of atoms})} = 1$ $\frac{n_e^{Zn}}{n_{Zn}} = 2$

$\frac{n}{n_a} = 1.36$ $\frac{n_e^{Cu} + n_e^{Zn}}{n_{Cu} + n_{Zn}} = 1.36 \rightarrow \frac{\frac{n_e^{Cu}}{n_{Cu}} + \frac{n_e^{Zn}}{n_{Cu}}}{1 + \frac{n_{Zn}}{n_{Cu}}} = 1.36 \rightarrow \frac{1 + 2\frac{n_{Zn}}{n_{Cu}}}{1 + \frac{n_{Zn}}{n_{Cu}}}$

$\frac{n_{Zn}}{n_{Cu}} = 0.56$

16. (5.77) $\frac{dE(k)}{dt} = F \cdot V$, $\frac{1}{F} dE(k) = V\, dt = dx \Rightarrow \frac{1}{F} \int_{E(k)} dE(k) = \int_{x_0}^{x} dx$

$\frac{1}{F} E(k) = x - x_0$, $x = x_0 + \frac{1}{F} E(k)$ From (5.78), $k = \frac{Ft}{\hbar}$