

**QUANTUM THEORY OF ELECTRONS - 2**READING ASSIGNMENT

"Quantum Theory of Electrons in Periodic Lattices", from *Solid State and Semiconductor Physics*, J.P. McKelvey

SUPPLEMENTARY REFERENCE

"Energy Bands", from *Physical Properties of Semiconductors*, C.M. Wolfe, N. Holonyak, Jr., G.E. Stillman

Band Structure Handout

LECTURE PROGRAM

1. Tight-Binding Approximation
  - a. Physical description
  - b. Expansion in terms of atomic orbitals
  - c. Overlap of energy bands
2. Dynamics of Electrons in 3D Lattices
  - a. 3D representation of dynamical equations
  - b. Group velocity
  - c. Effective mass tensor
  - d. Successive configurations of the Fermi surface
3. Brillouin Zones (see Sections 3.3 - 3.3.3 of Pierret handout from Lecture 8)
  - a. First Brillouin zone for diamond/zincblendel lattices
  - b. High symmetry points
  - c. E-k diagrams
  - d. Direct and indirect materials
  - e. Light and heavy hole bands; split-off bands
  - f. Constant energy surfaces
4. Insulators, Semiconductors and Metals
  - a. Insulators
  - b. Metals
  - c. Semiconductors
  - d. Positive Hall coefficient
5. Band Gap Energy (see Section 3.3.5 of Pierret)
6. Other Band Models (see Wolfe handout and Band Structure handout)

C. 8.6

usly  
sfied  
The  
veen  
dges  
h aspara-  
from  
it is

5-36)

then

5-37)

5-38)

represents the width of the forbidden energy region at  $k = n\pi/a$ . For points near the bottom of the upper band,  $k'$  is very small and the radical in (8.6-37) may be expanded by the binomial theorem to give, approximately,

$$\varepsilon(k) = \varepsilon_n + \frac{1}{2} \Delta\varepsilon + \frac{\hbar^2 k'^2}{2m} \left( 1 + \frac{4\varepsilon_n}{\Delta\varepsilon} \right), \quad (8.6-39)$$

where  $\varepsilon_n$  is given by (8.6-35). Differentiating twice, and using (8.4-8) it is easily seen that

$$m^* = \frac{\hbar^2}{d^2\varepsilon/dk^2} = \frac{\hbar^2}{d^2\varepsilon/dk'^2} = \frac{m}{1 + \frac{4\varepsilon_n}{\Delta\varepsilon}}. \quad (8.6-40)$$

It can be shown in a somewhat similar fashion that the effective mass for holes near the top of the lower band is the same as the value for electrons given by (8.6-40).

## 8.7 THE TIGHT-BINDING APPROXIMATION

In the free-electron approximation, the potential energy of the electron was assumed to be small compared to its total energy. This led to be an approximate treatment in which allowed and forbidden energy bands were found, the width of the forbidden bands being small compared to that of the allowed bands in the limit in which the initial assumption was well satisfied. The tight binding approximation proceeds from the opposite point of view, namely that the potential energy of the electron accounts for nearly all of the total energy; in this instance the allowed energy bands are narrow in comparison with the forbidden bands. In the free-electron model, the atoms of the crystal are assumed to be so close together that the wave functions for electrons on neighboring atoms overlap to a large extent. There is thus a strong interaction between neighboring atoms, and the allowed energy states of the resulting crystal bear little resemblance to the atomic wave functions of the individual atoms of which the crystal is composed. The tight binding approximation, however, is based on the assumption that the atoms of the crystal are so far apart that the wave functions for electrons associated with neighboring atoms overlap only to a small extent. The interaction between neighboring atoms will in this case be relatively weak, and the wave functions and allowed energy levels of the crystal as a whole will be closely related to the wave functions and energy levels of isolated atoms. The question of which of the two approximations is correct in any given situation, of course, depends upon the particular material at hand. In some substances the free-electron approximation is quite good, while for others the tight binding approximation is more nearly correct; there are also crystals where neither is very good, the situation being intermediate between the two extreme cases.

The crystal wave functions in the tight-binding approximation are based upon the wave functions of isolated atoms. If the potential function associated with an

the value given by (8.6-34) and again to approach the solutions obtained previously when  $k^2$  differs from  $k_n^2$  by a sufficient amount. These conditions will be satisfied if in (8.6-33) the plus sign is chosen for  $k > n\pi/a$  and the minus sign for  $k < n\pi/a$ . The solution (8.6-15), or the better second approximation (8.6-24), in the regions between the band edges, coupled with the solution (8.6-33) for the regions close to band edges (the signs being chosen as described above) lead to a relation between  $\epsilon$  and  $k$  such as that illustrated in Figure 8.7.

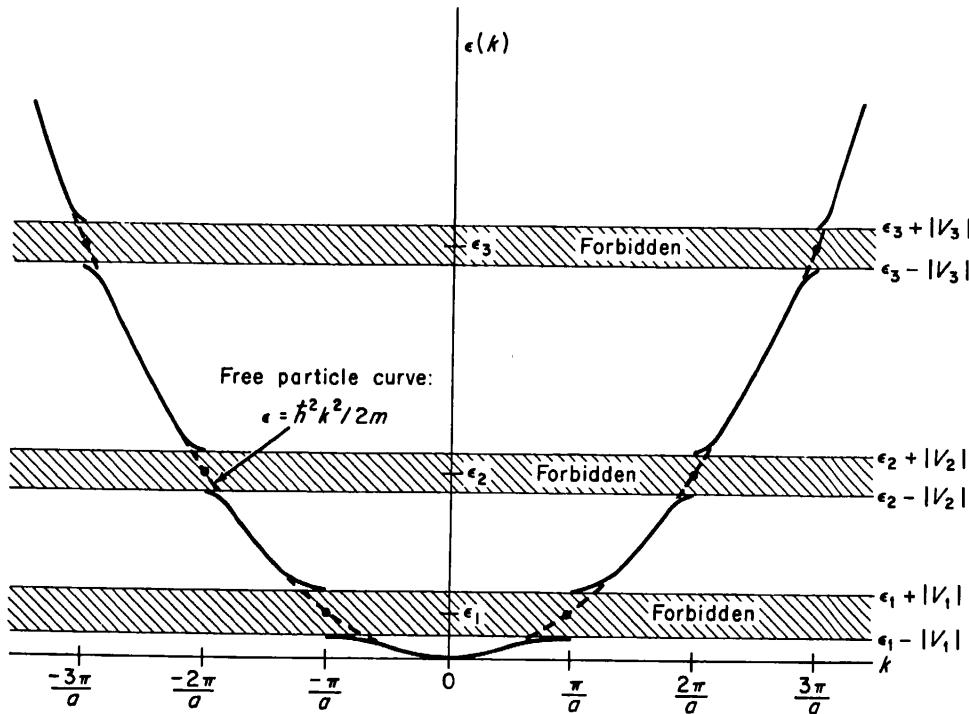


FIGURE 8.7. Schematic representation of the  $\epsilon$  versus  $k$  relation for the free electron approximation as obtained from (8.6-24) and (8.6-33).

It is apparent from this figure that the  $\epsilon$  versus  $k$  curves are approximately parabolic in form near the band edge points. The effective mass can be established from (8.4-8), by direct differentiation of (8.6-33). This calculation is very involved, and it is somewhat easier to proceed by setting

$$k = \frac{n\pi}{a} + k', \quad (8.6-36)$$

where  $k'$  is small compared with  $\pi/a$ , in (8.6-33). After a little algebra, one may then write (8.6-36) in the form

$$\epsilon(k) = \frac{\hbar^2}{2m} \left[ \left( \frac{n\pi}{a} \right)^2 + k'^2 + \frac{m\Delta\epsilon}{\hbar^2} \sqrt{1 + 4k'^2 \left( \frac{n\pi}{a} \right)^2 \left( \frac{\hbar^2}{m\Delta\epsilon} \right)^2} \right] \quad (8.6-37)$$

where

$$\Delta\epsilon = 2|V_n| \quad (8.6-38)$$

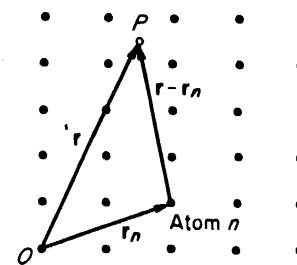
isolated atom is  $V_0(\mathbf{r})$ , then the solutions of the Schrödinger equation

$$\mathcal{H}_0 \psi_0 = -\frac{\hbar^2}{2m} \nabla^2 \psi_0 + V_0(\mathbf{r}) \psi_0(\mathbf{r}) = \varepsilon_0 \psi_0(\mathbf{r}) \quad (8.7-1)$$

represent the electronic wave functions of the atom. Let us assume that the ground state wave function  $\psi_0$  is nondegenerate and corresponds to a ground state energy  $\varepsilon_0$ . If a large number of atoms of this type are combined into a periodic lattice in such a way that the value of the potential in the regions about each individual atom where the ground-state wave function is large is not much affected by the presence of the neighboring atoms, then the crystal wave function (for these most tightly bound states, at any rate) can be written as a linear superposition of atomic wave functions of the form

$$\psi(\mathbf{r}) = \sum_n a_n \psi_0(\mathbf{r} - \mathbf{r}_n), \quad (8.7-2)$$

FIGURE 8.8. Vector geometry used in the calculation of Section 8.7.



the vectors  $r$  and  $r_n$  being related as shown in Figure 8.8. The sum is taken over all atoms of the crystal, which for simplicity we shall assume to be infinite in extent. Since all atoms of the crystal are equivalent, all the coefficients  $a_n$  must have the same absolute value; they must therefore be expressible in the form  $ae^{i\phi_n}$ , where  $a$  and  $\phi_n$  are real quantities. For the same reason, the phase difference  $\phi$  between each pair of neighboring atoms along any given crystal axis must be the same (although it is *not* necessarily the same along all three axes). This leads to the choice of the phase factor  $\phi_n = \mathbf{k} \cdot \mathbf{r}_n$  with  $\mathbf{k}$  a constant vector. Finally, if the functions  $\psi_0$  are normalized such that their integral taken over all space is unity, and if we wish to preserve this normalization, we must take  $a = 1$ . We may then write (8.7-2) in the form<sup>8</sup>

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_n e^{i\mathbf{k} \cdot \mathbf{r}_n} \psi_0(\mathbf{r} - \mathbf{r}_n). \quad (8.7-3)$$

These wave functions satisfy the Schrödinger equation for the entire crystal, with the periodic potential function of Figure 8.9, namely

$$\mathcal{H} \psi_{\mathbf{k}} = \left( -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \right) \psi_{\mathbf{k}} = \varepsilon \psi_{\mathbf{k}}. \quad (8.7-4)$$

<sup>8</sup> Equation (8.7-3) can be shown to have the form of a Bloch function such as (8.2-14). The details of the proof are assigned as an exercise.

The total Hamiltonian  $\mathcal{H}$  can be written as the sum of two parts,

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}' \quad (8.7-5)$$

with

$$\mathcal{H}_0 = -\frac{\hbar^2}{2m} \nabla^2 + V_0(\mathbf{r} - \mathbf{r}_n) \quad (8.7-6)$$

and

$$\mathcal{H}' = V(\mathbf{r}) - V_0(\mathbf{r} - \mathbf{r}_n). \quad (8.7-7)$$

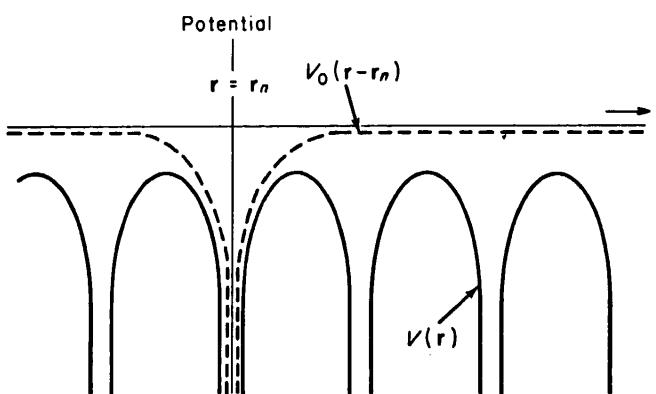


FIGURE 8.9. Potential functions used in the calculations of Section 8.7.

From (8.7-1) it is clear that  $\mathcal{H}_0\psi_0 = \varepsilon_0\psi_0$ , whereby, using (8.7-3),

$$\mathcal{H}_0\psi_{\mathbf{k}} = \sum_n e^{i\mathbf{k}\cdot\mathbf{r}_n} \mathcal{H}_0\psi_0(\mathbf{r} - \mathbf{r}_n) = \varepsilon_0 \sum_n e^{i\mathbf{k}\cdot\mathbf{r}_n} \psi_0(\mathbf{r} - \mathbf{r}_n) = \varepsilon_0\psi_{\mathbf{k}}. \quad (8.7-8)$$

Now, according to (4.14-5) the energy  $\varepsilon$  can be found by evaluating the expectation value of the Hamiltonian operator, whence, using (8.7-5) and (8.7-8),

$$\varepsilon = \frac{\int_v \psi_{\mathbf{k}}^*(\mathcal{H}_0 + \mathcal{H}')\psi_{\mathbf{k}} dv}{\int_v \psi_{\mathbf{k}}^*\psi_{\mathbf{k}} dv} = \varepsilon_0 + \frac{\int_v \psi_{\mathbf{k}}^* \sum_n e^{i\mathbf{k}\cdot\mathbf{r}_n} [V(\mathbf{r}) - V_0(\mathbf{r} - \mathbf{r}_n)] \psi_0(\mathbf{r} - \mathbf{r}_n) dv}{\int_v \psi_{\mathbf{k}}^*\psi_{\mathbf{k}} dv}. \quad (8.7-9)$$

Provided that the small overlapping of atomic wave functions centered on different lattice sites is neglected, the normalization properties of the functions  $\psi_0$  and the definition (8.7-3) require that  $\int_v \psi_{\mathbf{k}}^*\psi_{\mathbf{k}} dv$  be equal to  $N$ , the number of atoms in the crystal. Utilizing this fact, and substituting the value given by (8.7-3) for  $\psi_{\mathbf{k}}^*$  in the numerator of (8.7-9), we obtain

$$\varepsilon = \varepsilon_0 + \frac{1}{N} \sum_n \left\{ \sum_m e^{i\mathbf{k}\cdot(\mathbf{r}_n - \mathbf{r}_m)} \int_v \psi_0^*(\mathbf{r} - \mathbf{r}_m) [V(\mathbf{r}) - V_0(\mathbf{r} - \mathbf{r}_n)] \psi_0(\mathbf{r} - \mathbf{r}_n) dv \right\}. \quad (8.7-10)$$

It is not difficult to see that due to the periodicity of the crystal, and due to the summation over all values of  $m$ , each term of the summation over  $n$  yields the same value. The value of the sum over  $n$  is then just the value of any term of the summation (say, for simplicity, that for which  $n = 0$ ) times the number of terms in the sum, which is  $N$ . We may then write (8.7-10) as

$$\epsilon(\mathbf{k}) = \epsilon_0 + \sum_m e^{-i\mathbf{k}\cdot\mathbf{r}_m} \int_v \psi_0^*(\mathbf{r} - \mathbf{r}_m)[V(\mathbf{r}) - V_0(\mathbf{r})]\psi_0(\mathbf{r}) dv, \quad (8.7-11)$$

where the sum is taken over all atoms of the crystal. However, since the wave function  $\psi_0$  ordinarily falls off very rapidly with distance, and since the magnitudes of the integrals in (8.7-11) are governed essentially by the amount of overlap between two wave functions centered on atoms separated by distances  $r_m$ , the contributions of the terms in the summation decrease very rapidly as  $r_m$  increases. It is therefore usually quite a good approximation to consider nearest neighbor terms only. Accordingly we shall neglect all terms beyond nearest neighbor contributions in (8.7-11). We shall also assume that the wave functions  $\psi_0$  are *spherically symmetric*; under these conditions all nearest neighbor contributions will be the same. Our treatment is thus restricted to the case where the ground state electronic configuration in the isolated atom is that of an *s*-state, as it is, for example, for the alkali metals, although the extension to *p*-states is quite straightforward and can be accomplished without much difficulty.<sup>9</sup>

For the case  $m = 0$ , the integral in (8.7-11) becomes

$$\int_v \psi_0^*(\mathbf{r})[V(\mathbf{r}) - V_0(\mathbf{r})]\psi_0(\mathbf{r}) dv = -\alpha, \quad (8.7-12)$$

while for the nearest neighbor atoms

$$\int_v \psi_0^*(\mathbf{r} - \mathbf{r}_m)[V(\mathbf{r}) - V_0(\mathbf{r})]\psi_0(\mathbf{r}) dv = -\beta, \quad (8.7-13)$$

where  $\mathbf{r}_m$  is a vector connecting the atom at the origin with a nearest neighbor atom. Since the actual calculation of the overlap integrals is quite tedious we shall be content with expressing our results in terms of these quantities, which are defined above as  $-\alpha$  and  $-\beta$ . Under these circumstances, equation (8.7-11) may be written

$$\epsilon(\mathbf{k}) = \epsilon_0 - \alpha - \beta \sum_m e^{-i\mathbf{k}\cdot\mathbf{r}_m}, \quad (8.7-14)$$

where the summation is taken only over nearest neighbor atoms. In the case of a *simple cubic* crystal, the components of the vector  $\mathbf{r}_m$  may be expressed as

$$\mathbf{r}_m = (\pm a, 0, 0), (0, \pm a, 0), (0, 0, \pm a), \quad (8.7-15)$$

---

<sup>9</sup> See for instance N. F. Mott and H. Jones, *Theory of Metals and Alloys*, Dover Publications, New York (1958), p. 70.

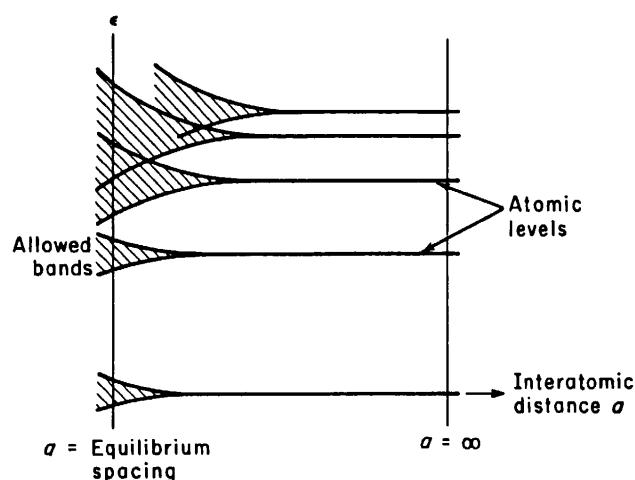
where  $a$  is the lattice constant. Equation (8.7-14) then gives

$$\epsilon(\mathbf{k}) = \epsilon_0 - \alpha - 2\beta(\cos k_x a + \cos k_y a + \cos k_z a). \quad (8.7-16)$$

From this equation it is clear that there is a *range* of allowed energy values, corresponding to the various values which  $k_x$ ,  $k_y$ , and  $k_z$  may have. The allowed states thus form an *energy band* in much the same way as did those which were discussed in connection with the free electron approximation.

The minimum value of the energy, according to (8.7-16), is at the point  $k_x = k_y = k_z = 0$ , that is, at the origin in a plot involving orthogonal coordinates  $(k_x, k_y, k_z)$  in  $\mathbf{k}$ -space. The maximum values occur at the corners of a cube in  $\mathbf{k}$  space whose coordinates are  $\left(\pm \frac{\pi}{a}, \pm \frac{\pi}{a}, \pm \frac{\pi}{a}\right)$ , all permutations of + and - signs between the three coordinates being taken to generate the eight corner points. At these sites each cosine term above takes on the value -1. The difference in energy between the maximum energy points and the minimum energy is, from (8.7-16),  $12\beta$ , and this is the width of the energy band corresponding to the related *s*-state of the isolated atom. If the nearest neighbor atoms are quite far apart, then the overlap of the wave functions  $\psi_0(\mathbf{r})$  and  $\psi_0(\mathbf{r} - \mathbf{r}_m)$  as expressed by the product  $\psi_0(\mathbf{r})\psi_0(\mathbf{r} - \mathbf{r}_m)$  in the integral (8.7-13) becomes very small, as does the value of  $\beta$  as given by that equation. In the limit where the atomic distances become large (and it is in this limit that the approximations upon which the tight binding calculation are based become really good), the allowed energy bands become narrow, approaching a single discrete energy value, corresponding to the atomic *s*-state energy level, as the separation becomes infinite. This variation of band width with atomic spacing is shown schematically in Figure 8.10.

FIGURE 8.10. Schematic representation of the splitting of discrete atomic energy levels into bands as isolated atoms are assembled into a crystal lattice.



For an electron moving in the  $x$ -direction,  $k_y = k_z = 0$ . If the momentum is sufficiently small that  $k_x \ll \pi/a$  the electron will be near the bottom of the energy band. Under these conditions, approximating  $\cos k_x a$  in (8.7-16) by the first two terms of its series expansion, (8.7-16) becomes

$$\epsilon(k_x) = \epsilon_0 - \alpha - 2\beta\left(3 - \frac{k_x^2 a^2}{2}\right). \quad (8.7-17)$$

From this it is clear that the  $\epsilon$  versus  $k_x$  relation is parabolic, and the electron therefore behaves essentially as a free electron. The quantity  $\hbar k_x$  may, as usual, be interpreted as the  $x$ -component of a crystal momentum vector  $\mathbf{p} = \hbar \mathbf{k}$ . The effective mass is given by

$$m^* = \frac{\hbar^2}{d^2\epsilon/dk_x^2} = \frac{\hbar^2}{2\beta a^2}. \quad (8.7-18)$$

From the symmetry of equation (8.7-16) it is evident that the effective mass associated with motion in the  $y$ - and  $z$ -directions for small momenta will be the *same* as that given by (8.7-18), and the effective mass will therefore be isotropic or independent of direction. This is not always the case, however, especially in noncubic crystals; the effective mass is then often anisotropic and must be regarded as a tensor. This point will be discussed more fully in the next section. From (8.7-18) it would appear that as the lattice distance  $a$  increases the effective mass decreases; exactly the *opposite* is true, however, because as  $a$  increases the overlap integral  $\beta$  decreases much more rapidly, so that the product  $\beta a^2$  decreases rapidly with increasing lattice distances. In the extreme tight binding limit, then, according to (8.7-18) an electron becomes very "heavy," which is merely an expression of the fact that it is not easily transferred from one atom to another under these circumstances.

## 8.8 DYNAMICS OF ELECTRONS IN TWO- AND THREE-DIMENSIONAL LATTICES: CONSTANT ENERGY SURFACES AND BRILLOUIN ZONES

In this section we shall discuss in detail the behavior of electrons in two- and three-dimensional periodic lattices, using wherever possible the example of the tight binding approximation calculation for the simple cubic structure as an illustration of the general principles which are involved. It is easily seen that the dynamical equations of Section 8.4 can be extended in a straightforward manner to a three-dimensional situation, whereupon one obtains

$$\mathbf{v}_g = \frac{1}{\hbar} \nabla_{\mathbf{k}} \epsilon \quad (8.8-1)$$

for the electron velocity, where  $\nabla_{\mathbf{k}}$  refers to the gradient operator in the orthogonal  $\mathbf{k}$ -space whose coordinates are  $(k_x, k_y, k_z)$ . One may then write

$$\frac{d\mathbf{v}_g}{dt} = \frac{1}{\hbar} \frac{d}{dt} \nabla_{\mathbf{k}} \epsilon. \quad (8.8-2)$$

If  $\mathbf{A}$  is a vector whose components are functions of  $k_x, k_y$ , and  $k_z$ , then each component  $A_{\alpha}$  ( $\alpha = x, y, z$ ) must satisfy the relation

$$\frac{dA_{\alpha}}{dt} = \frac{\partial A_{\alpha}}{\partial k_x} \frac{dk_x}{dt} + \frac{\partial A_{\alpha}}{\partial k_y} \frac{dk_y}{dt} + \frac{\partial A_{\alpha}}{\partial k_z} \frac{dk_z}{dt} = (\nabla_{\mathbf{k}} A_{\alpha}) \cdot \frac{d\mathbf{k}}{dt}. \quad (8.8-3)$$

The vector  $d\mathbf{A}/dt$  may thus be expressed as the scalar product of the tensor<sup>10</sup>  $\nabla_k \mathbf{A}$  and the vector  $d\mathbf{k}/dt$ , thus

$$\frac{d\mathbf{A}}{dt} = (\nabla_k \mathbf{A}) \cdot \frac{d\mathbf{k}}{dt}, \quad (8.8-4)$$

each component of the tensor product above leading to an equation of the form of (8.8-3). Applying this formula to (8.8-3), letting  $\mathbf{k} = \mathbf{p}/\hbar$ , one obtains

$$\frac{d\mathbf{v}_g}{dt} = \frac{1}{\hbar^2} \nabla_k (\nabla_k \epsilon) \cdot \frac{d\mathbf{p}}{dt} = \frac{1}{\hbar^2} \nabla_k (\nabla_k \epsilon) \cdot \mathbf{F}. \quad (8.8-5)$$

This equation is essentially Newton's force equation,  $d\mathbf{v}_g/dt = \mathbf{F}/m^*$ . It is evident, however, that the reciprocal of the effective mass is a *tensor* of the form

$$\left( \frac{1}{m^*} \right) = \frac{1}{\hbar^2} \nabla_k (\nabla_k \epsilon). \quad (8.8-6)$$

The elements of this tensor are easily seen to be

$$\left( \frac{1}{m^*} \right)_{\alpha\beta} = \frac{1}{\hbar^2} \frac{\partial^2 \epsilon}{\partial k_\alpha \partial k_\beta}, \quad (8.8-7)$$

and from this it is apparent that the reciprocal effective mass tensor is symmetric, that is, that

$$(1/m^*)_{\beta\alpha} = (1/m^*)_{\alpha\beta}. \quad (8.8-8)$$

The effective mass must in general be regarded as a tensor because the curvature of the  $\epsilon(\mathbf{k})$  relation is not necessarily the same along all possible directions in  $\mathbf{k}$ -space; a different effective mass may thus be required for each possible direction.

For the simple cubic crystal of Section 8.7, for values  $ka \ll 1$ , equation (8.7-16) may be written in the form

$$\epsilon(\mathbf{k}) = \epsilon_0 - \alpha - 6\beta + \beta a^2(k_x^2 + k_y^2 + k_z^2). \quad (8.8-9)$$

It is evident from this that all the off-diagonal elements of the effective mass tensor (8.8-7) are zero, and that each of the elements on the diagonal is just equal to  $\hbar^2/2\beta a^2$ , whereby the tensor reduces to  $\hbar^2/2\beta a^2$  times the unit tensor  $[\delta_{\alpha\beta}]$ . The effective mass is clearly isotropic under these circumstances, and may be represented as the *scalar* quantity  $\hbar^2/2\beta^2$  in agreement with the conclusions of the previous section.

Let us now investigate the form of the surfaces of constant energy in  $\mathbf{k}$ -space for the simple cubic structure discussed in the tight-binding approximation in the preceding section. The equation of these surfaces is simply (8.7-16) with  $\epsilon(\mathbf{k})$  being regarded as a

---

<sup>10</sup> A brief outline of the aspects of tensor analysis which are essential for an understanding of the concepts discussed in this book is given in Appendix B.

parametric constant. In connection with (8.7-16), it was mentioned that the *minimum* value of  $\epsilon$  was obtained at the point  $k_x = k_y = k_z = 0$  where all the cosine terms attain the value unity. For values of  $\epsilon$  slightly greater than this minimum, the cosines must be *near* unity, and thus the arguments  $k_x a$ ,  $k_y a$ , and  $k_z a$  must all be very small. In this case (8.7-16) reduces to the form written above as (8.8-9), which can be slightly rearranged to read

$$k_x^2 + k_y^2 + k_z^2 = \frac{\epsilon(\mathbf{k})}{\beta a^2} - \frac{\epsilon_0 - \alpha - 6\beta}{\beta a^2} = \text{const.} \quad (8.8-10)$$

This is clearly the equation of a sphere in  $k$ -space. The spherical form of the surfaces of constant energy is indicative of the fact that electrons in this region of  $k$ -space behave like free electrons; it will be noted from (5.2-18) that for free particles the surface of constant energy in momentum space is spherical. As the value of  $\epsilon$  increases, the radius of the spherical constant energy surface increases, until the values of  $k$  associated with points on the surface becomes so great that the approximation  $ka \ll 1$  is no longer valid. The more general equation (8.7-16) must then be used to plot the constant energy surfaces, and as  $\epsilon$  increases the surfaces which are so obtained are illustrated in Figure 8.11 (a-f). With increasing  $\epsilon$  the sphere expands, becomes somewhat distorted, as at (b), and at  $\epsilon = \epsilon_1 + 4\beta$  (where  $\epsilon_1$  is the value of  $\epsilon$  at the minimum point, or  $\epsilon_0 - \alpha - 6\beta$ ) assumes the quasi-octahedral form shown at (c). At this point the energy surface touches the limiting cube of side  $2\pi/a$ , about which we shall say more later. As  $\epsilon$  increases further, the surface assumes a shape like that shown in (d), and when  $\epsilon = \epsilon_1 + 8\beta$  takes on the form shown in (e). At this point the surface splits up into eight sheets, each of which approaches a corner of the cube as  $\epsilon$  approaches its maximum value for the band,  $\epsilon_1 + 12\beta$ . These eight branches are shaped approximately like octants of a sphere in the limit where  $\epsilon$  approaches this maximum value. The surface of constant energy always has cubic symmetry in this example. The intersections of the constant energy surface with the  $k_x k_y$ -plane are shown in Figure 8.12; these curves can also be regarded as the curves of constant energy for a *two-dimensional* crystal with a square lattice whose interatomic spacing is  $a$ . It is clear from (8.7-16) that all states within the band must have energy between  $\epsilon_1$  and  $\epsilon_1 + 12\beta$ , and that all possible states within this energy range are represented by values of  $(k_x, k_y, k_z)$  lying between the limits  $\left( \pm \frac{\pi}{a}, \pm \frac{\pi}{a}, \pm \frac{\pi}{a} \right)$ .

If there are a given number of electrons belonging to this energy band, and if the temperature is at absolute zero, the electronic energy states will all be occupied up to a given Fermi energy, in accord with the Pauli exclusion principle. The electron distribution will therefore lie within a volume in  $k$ -space which is enclosed by a surface of constant energy, the value of the energy on the surface being the Fermi energy  $\epsilon_f$ . This surface is often referred to as the *Fermi surface*. From the discussion of Section 5.5, we have already seen that the Fermi surface for free electrons is a sphere. For electrons in a periodic crystal potential, the Fermi surface may also be approximately spherical *if the band is nearly empty or nearly full*; this fact is clearly illustrated in Figure 8.11 (a) and (f) for the example which has been discussed in detail. At temperatures in excess of absolute zero, some of the electrons will be excited to states of energy higher than  $\epsilon_f$ , lying outside the Fermi surface, and there will be some unoccupied states within the Fermi surface. The transition between full and empty states will then no longer be a sharp one. Nevertheless, if the temperature is small compared to the

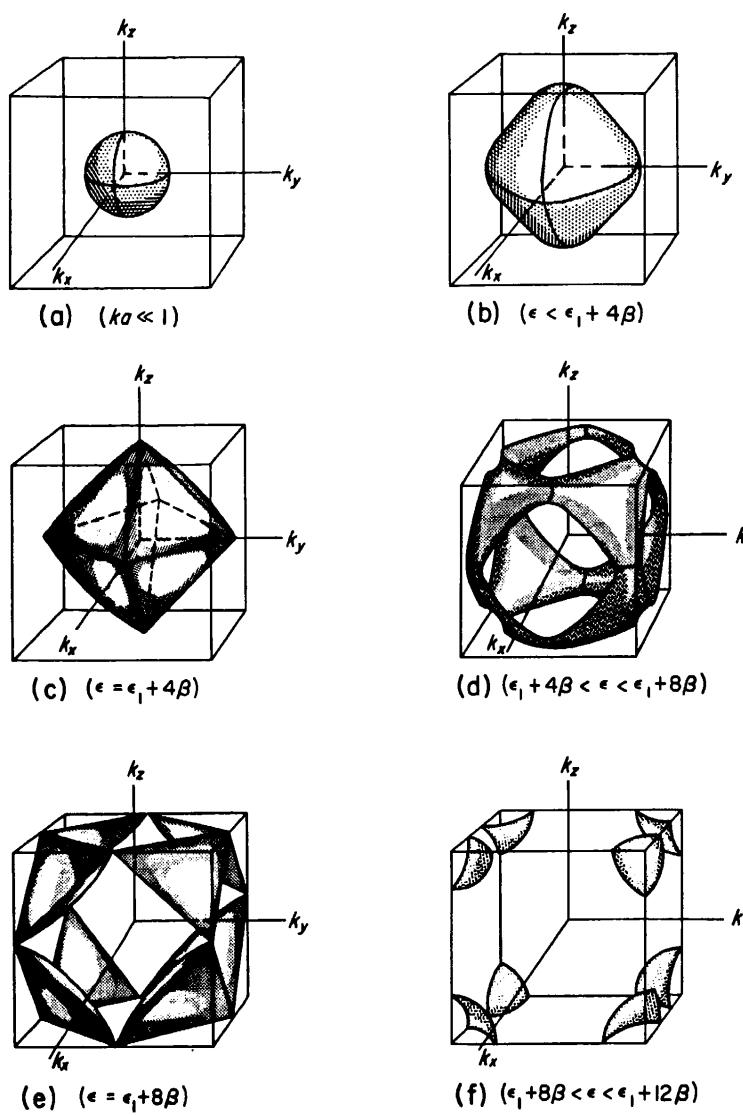


FIGURE 8.11. Successive configurations of the Fermi surface for a simple cubic crystal with spherically symmetric wave functions in the approximation where only nearest neighbor interactions are considered. As the electron population in the band increases, thus raising the Fermi level from the bottom of the band to the top, the Fermi surface goes from the nearly spherical form shown at (a) through the complex intermediate shapes (b)–(e) to the cube-minus-sphere configuration in the nearly filled Brillouin zone shown at (f).

Fermi temperature, the region of transition will be quite thin and the essential validity of the Fermi surface picture will be unimpaired.

Let us now examine what happens at points on the surface of the limiting cube enclosing the volume of  $\mathbf{k}$ -space which the electronic states belonging to this band may occupy. For simplicity, we shall initially confine our attention to the  $k_x, k_y$ -plane only; the solutions in this plane are related to a two-dimensional square crystal lattice. Consider a point on the upper (or lower) edge of the square region shown in Figure 8.12; the  $k_y$  coordinate of such a point is  $\pi/a$ , while the  $k_x$  coordinate may have any

arbitrary value, say  $k_{x0}$ . The electron wavelength  $\lambda$  will then be given by

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{\sqrt{k_{x0}^2 + \frac{\pi^2}{a^2}}} \quad (8.8-11)$$

while the sine of the glancing angle  $\theta$  related to a set of crystal planes parallel to the  $x$ -axis will be

$$\sin \theta = \frac{k_y}{\sqrt{k_x^2 + k_y^2}} = \frac{\pi/a}{\sqrt{k_{x0}^2 + \frac{\pi^2}{a^2}}} \quad (8.8-12)$$

From this it is clear that the Bragg condition (2.1-1) with  $n = 1$  is satisfied by the electron wave function at all points on this part of the boundary. In the same way, it is easy to verify that the Bragg condition (related now to a glancing angle referred to an equivalent set of planes parallel to the  $y$ -axis) is satisfied at all points on the right-hand (or left-hand) edge of the bounding square. The electron, regarded as a wave, is internally diffracted by the lattice whenever its momentum vector touches this square boundary. Points in  $k$ -space beyond the limits of this boundary correspond to points *outside* the lowest energy band, belonging to other bands which arise from atomic states of higher energy than the ground state. The limits of the regions of  $k$ -space occupied by the higher bands are determined by the orientation of boundaries along which higher order Bragg reflections ( $n > 1$ ) occur. There is always an energy gap between the bands, as illustrated in Figure 8.10; for example, whenever one crosses the edge of the square region enclosed by the lines  $k_x = \pm\pi/a$ ,  $k_y = \pm\pi/a$ , then a discontinuity in energy accompanied by internal Bragg reflection must occur. The interior of the region contains the *totality* of states belonging to the lowest energy band.

It will be noted that the vector  $\mathbf{k}$  has the same dimensions as the *reciprocal* lattice. In Section 2.6 it has already been shown that the Bragg reflection condition can be written as

$$2\mathbf{k} \cdot \mathbf{G} + \mathbf{G}^2 = 2(k_x G_x + k_y G_y + k_z G_z) + G_x^2 + G_y^2 + G_z^2 = 0 \quad (8.8-13)$$

where  $\mathbf{k}$  is the propagation vector and where  $\mathbf{G}/2\pi$  is a vector from the origin to any point of the *reciprocal* lattice. For a two-dimensional square lattice of points whose lattice spacing is  $a$ , the reciprocal lattice is a square lattice of lattice spacing  $1/a$ . If the origin is taken for simplicity at a lattice point, then the vectors  $\mathbf{G}$  have the form

$$\mathbf{G}_{\mu\nu} = \frac{2\pi}{a} (\mu \mathbf{i}_x + \nu \mathbf{i}_y) \quad (8.8-14)$$

where  $\mu$  and  $\nu$  are *integers* and where  $\mathbf{i}_x$  and  $\mathbf{i}_y$  are unit vectors in the  $k_x$  and  $k_y$  directions, respectively. Using this form for the vector  $\mathbf{G}$ , (8.8-13) becomes

$$\mu k_x + \nu k_y = -\frac{\pi}{a} (\mu^2 + \nu^2). \quad (8.8-15)$$

.8

## SEC. 8.8

## QUANTUM THEORY

241

This equation represents a family of straight lines in the  $k_x k_y$ -plane. Their  $k_x$ -intercepts and  $k_y$  intercepts are

$$k_x \text{ intercepts: } -\frac{\pi \mu^2 + v^2}{a \mu} \quad (8.8-16)$$

$$k_y \text{ intercepts: } -\frac{\pi \mu^2 + v^2}{a v}$$

Each of these lines represents a boundary along which a Bragg reflection takes place, and which may form the boundary between states belonging to different energy bands. If we take  $\mu = 0$ ,  $v = \pm 1$ , and  $\mu = \pm 1$ ,  $v = 0$  in (8.8-15) we obtain the four lines  $k_x = \pm \pi/a$  and  $k_y = \pm \pi/a$  which bound the region shown in Figure 8.12, whose

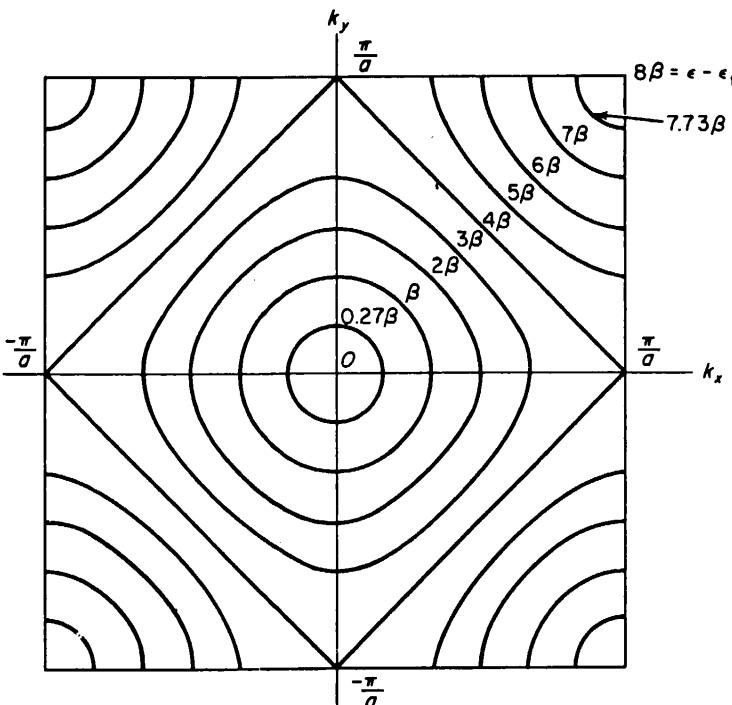


FIGURE 8.12. Constant energy contours in the Brillouin zone for a two-dimensional square lattice using spherically symmetric wave functions and nearest neighbor interactions only. The same pattern is formed by the traces of the Fermi surfaces of Figure 8.11 for the three-dimensional lattice upon the  $k_x k_y$ -plane.

interior contains all the states in the lowest energy band of the crystal. This region of  $\mathbf{k}$ -space is called the *first Brillouin zone* (this terminology has already been introduced in connection with the one-dimensional calculations discussed in Section 3.3). If we now consider  $\mu = \pm 1$ ,  $v = \pm 1$ , we obtain the four lines  $k_x \pm k_y = 2\pi/a$  and  $k_x \pm k_y = -2\pi/a$ . These four lines are plotted in Figure 8.3(a). On these boundaries the Bragg condition (2.1-1) is satisfied with the order of reflection,  $n$ , equal to 1. This can be proved by the same methods which were used previously in connection with the

boundaries of the first Brillouin zone. The region enclosed between the first Brillouin zone and these four intersecting lines contains all the states in the second energy band (which would be related to the second lowest atomic energy level if the tight-binding approximation were valid). It is referred to as the second Brillouin zone. It has the same area as the first zone, and indeed the four separate sections of this zone can be translated into the interior of the first zone by moving the sections parallel to the axes by distances which are integral multiples of  $2\pi/a$ . The four sections then fit together like the pieces of a picture puzzle to form a square region which is an exact replica

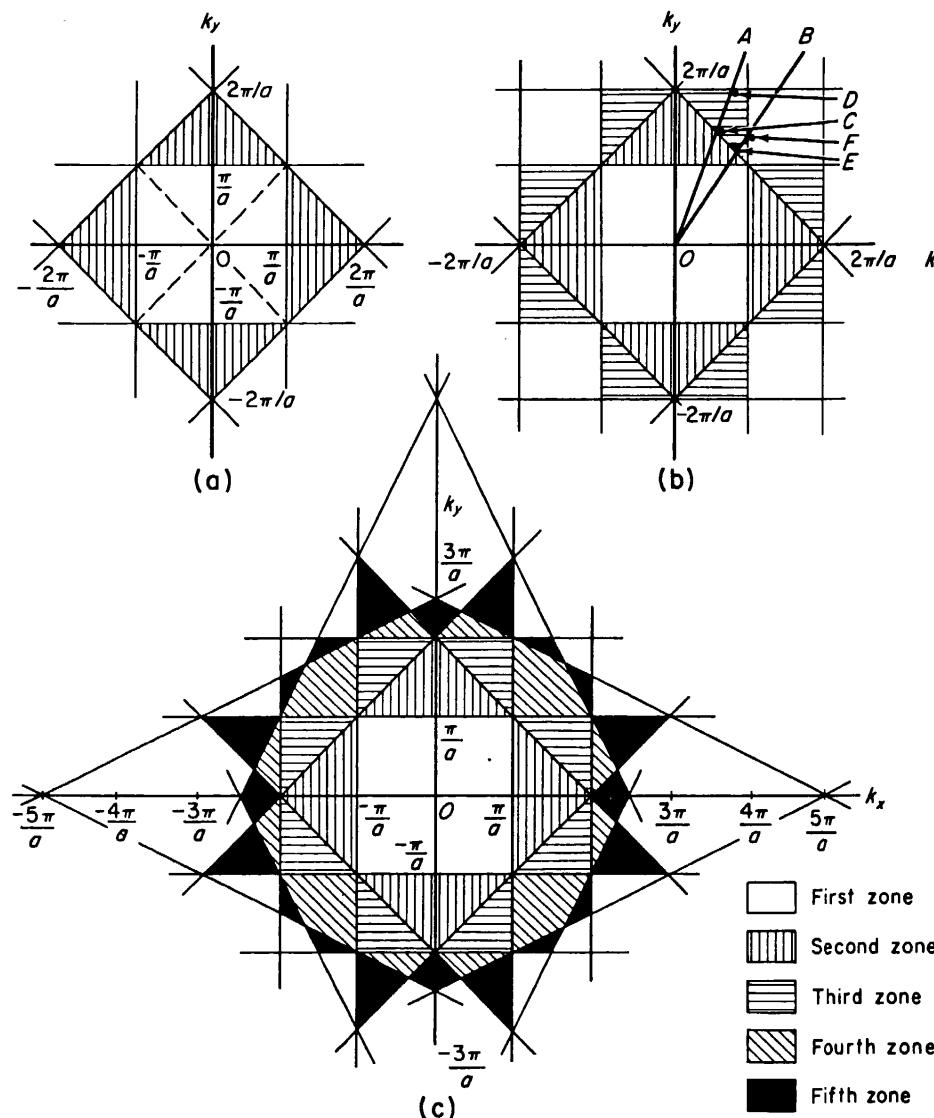


FIGURE 8.13. Construction of Brillouin zones for a two-dimensional square lattice. (a) First and second zones, (b) Third zone construction, (c) First five zones.

of the first zone, as shown by the dotted lines in Figure 8.13(a). The latter representation of the second zone is simply the *reduced zone* representation of Section 8.5. The justification of such a transformation with reference to the Kronig-Penney model has been discussed in that section.

1.8

in  
id  
ig  
ne  
be  
es  
er  
ca

More precisely, if

$$\psi(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} u(\mathbf{r}) = e^{i(k_x x + k_y y + k_z z)} u(\mathbf{r}) \quad (8.8-17)$$

is a wave function of a system, satisfying Schrödinger's equations and all relevant boundary conditions, and if we define

$$\mathbf{k}' = \mathbf{k} - i_x \left( \frac{2n_x \pi}{a} \right) - i_y \left( \frac{2n_y \pi}{a} \right) - i_z \left( \frac{2n_z \pi}{a} \right), \quad (8.8-18)$$

where  $i_x$ ,  $i_y$ , and  $i_z$  are unit vectors along the  $k_x$ ,  $k_y$ ,  $k_z$  directions, and where  $n_x$ ,  $n_y$ ,  $n_z$  are integers, then the *same* wave function can be expressed as

$$\begin{aligned} \psi(\mathbf{r}) &= e^{i\mathbf{k}' \cdot \mathbf{r}} [e^{2n_x \pi x/a} e^{2n_y \pi y/a} e^{2n_z \pi z/a} u(\mathbf{r})] \\ &= e^{i\mathbf{k}' \cdot \mathbf{r}} v(\mathbf{r}), \end{aligned} \quad (8.8-19)$$

where  $v(\mathbf{r})$  is *again* a periodic function having the periodicity of the lattice. The wave function can thus be written in the Bloch form in terms of either  $\mathbf{k}$  or the transformed wave vector  $\mathbf{k}'$ . It should be noted that (8.8-18) may be written equally well as

$$\mathbf{k}' = \mathbf{k} - \mathbf{G} \quad (8.8-20)$$

where  $\mathbf{G}/2\pi$  is a vector connecting two lattice points of the reciprocal lattice.<sup>11</sup> The transformation (8.8-18) or (8.8-20) may always be used to arrive at a reduced zone representation in which the transformed zone occupies the same region of  $\mathbf{k}$ -space as the first zone. The situation is quite similar to that encountered in Section 3.3 in the case of mechanical vibrations of a linear chain of discrete point masses, and discussed in connection with Figure 3.8.

The higher zones may be constructed by an extension of the procedure used to arrive at the form of the first and second zones. In (8.8-15), if we take  $\mu = 0$ ,  $\nu = \pm 2$  or  $\mu = \pm 2$ ,  $\nu = 0$ , the four lines  $k_x = \pm 2\pi/a$ ,  $k_y = \pm 2\pi/a$  are obtained, as shown in Figure 8.13(b). Again, it is clear that Bragg reflection takes place along these lines, and one is initially tempted to assign all the space outside the second zone and within the area enclosed by these four lines to the third Brillouin zone. This is not correct, however, for this area is much larger than that of the first or second zones, and each of the Brillouin zones must encompass the same area in  $\mathbf{k}$ -space. As a matter of fact there are *additional* Bragg reflection lines which run through this area and form boundaries which divide this space into the third, fourth, fifth, and sixth zones. The criterion for which zone is which is simply this; in travelling along a general radial line which does *not* go through any intersection of Bragg reflection lines, starting from the origin, one must pass through the first, second, third, fourth ... etc., zones successively, and each successive Bragg reflection line which is crossed forms the boundary along the radial path between a zone and the next highest neighboring zone. Thus, progressing from the origin along line  $OA$ , one enters the third zone at  $C$  and

<sup>11</sup> As a matter of fact, expression (8.8-20) is really the fundamental definition of this transformation, while (8.8-18) refers specifically to such a transformation in the particular case of the simple cubic lattice.

ta  
he  
nas

leaves it at  $D$ , the next Bragg reflection line intersecting  $OA$ ; beyond  $D$  one enters the fourth zone. Traveling along the line  $OB$ , however, one enters the third zone at  $E$  and leaves it at  $F$ , which lies on the line  $k_x = \pi/a$  which is now the third Bragg reflection line from the origin along path  $OB$ . Beyond  $F$  along line  $OB$  lies a part of the fourth zone. The third zone thus consists of the eight triangular segments shown in Figure 8.13(b). Again, it is clear that these eight segments can be assembled to form a square inside the first zone if they are subjected to appropriate translations according to (8.8-18). The boundaries of the fourth zone are formed, in part, by the lines for which  $\mu = \pm 1$ ,  $v = \pm 2$ , and  $\mu = \pm 2$ ,  $v = \pm 1$  (all possible permutations of plus and minus signs being taken). A family of eight lines (such as  $k_x \pm 2k_x = -5\pi/a$ ) is obtained, leading to boundaries for the fourth and fifth zones as shown in Figure 8.13(c), by the procedures outlined previously. Obviously this method of calculation can be extended *ad infinitum* to find as many zones as desired. In constructing the boundaries of the Brillouin zones in this manner, it is usually advantageous to begin by considering  $\mathbf{G}$  vectors whose magnitudes (as obtained from (8.8-14) or its equivalent in another crystal structure) are as small as possible, then proceeding to the next

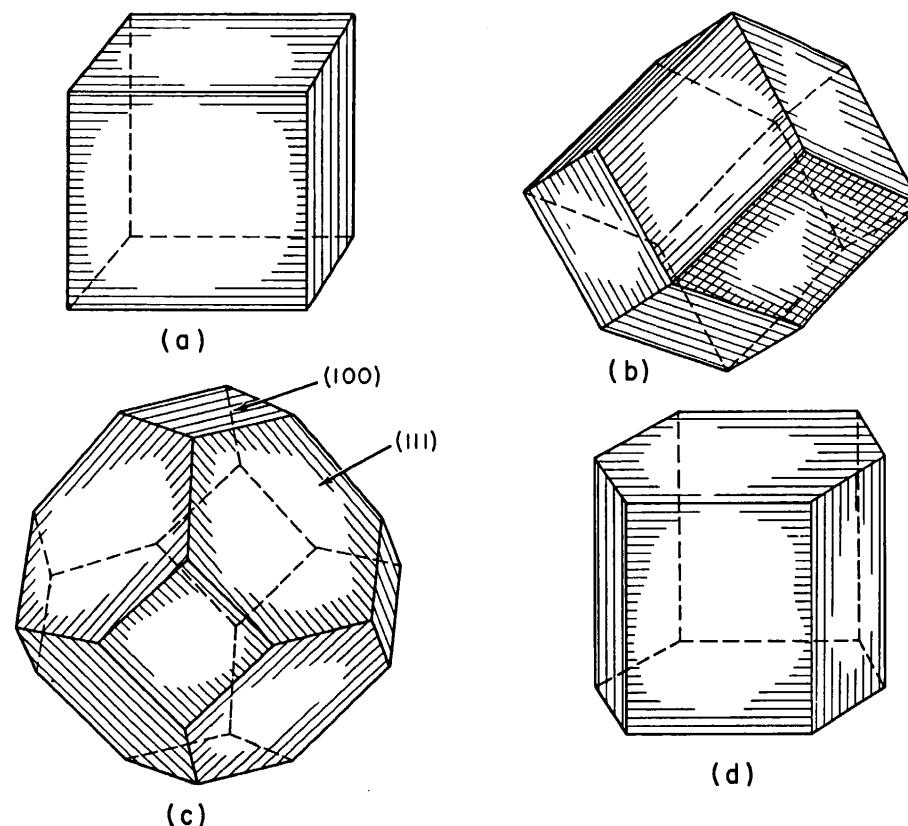


FIGURE 8.14. First Brillouin zone configuration for (a) simple cubic lattice, (b) body-centered cubic lattice, (c) face-centered cubic lattice, and (d) hexagonal lattice.

shortest set of  $\mathbf{G}$  vectors, then to the next shortest, and so on. Since the Bragg reflection condition implies the formation of standing waves at normal incidence, the group velocity and thus the gradient of  $\epsilon$  normal to the boundary of any Brillouin zone must

8.8  
he  
E  
re  
he  
in  
1 a  
ng  
or  
nd  
is  
ire  
on  
he  
gin  
/a-  
ext

## SEC. 8.9

## QUANTUM THEORY

245

*vanish*, that is

$$\left(\frac{\partial\omega}{\partial k}\right)_{\perp \text{ to bdry.}} = \mathbf{n} \cdot \nabla_k \omega = \frac{1}{\hbar} \mathbf{n} \cdot \nabla_k \epsilon = 0 \quad (8.8-21)$$

at a zone boundary, if  $\mathbf{n}$  is a unit vector normal to the boundary. It is possible to verify this general conclusion from the results of the Kronig-Penney potential model, the free-electron approximation and the tight binding approximation. The details are assigned as exercises.

The extension of these results to three-dimensional lattices is wholly straightforward. In this instance, of course, the Brillouin zones are three-dimensional regions of  $\mathbf{k}$ -space bounded by planes along which the Bragg condition is satisfied. The equations of these plane boundaries are obtained from equation (8.8-13) just as for the two-dimensional case. Again, each zone is of equal volume and the separate pieces of higher zones can be assembled to form a replica of the first zone by proper application of the transformation given by (8.8-20). It is obvious that the first Brillouin zone for a simple cubic lattice is a cube of side  $2\pi/a$ . The form of the first Brillouin zone for the face-centered and body-centered cubic lattices and for the hexagonal structure are shown in Figure 8.14. The zone for the b.c.c. lattice is a rhombic dodecahedron, all of whose faces are  $\{110\}$  planes in  $\mathbf{k}$ -space. The zone for the f.c.c. lattice is a tetrakaidecahedron, a polyhedron of 14 faces having the  $\{100\}$  and  $\{111\}$  orientation. The Fermi surface in all cases must have all the symmetry properties of the related Brillouin zone, which are in turn related to the symmetry properties of the crystal lattice.

## 8.9 INSULATORS, SEMICONDUCTORS, AND METALS

In Section 8.5, it was demonstrated that no electric current could arise either from empty bands or from completely filled bands. Any electrical conductivity exhibited by a crystal, due to the motion of free electrons, must therefore arise from the motion of electrons in energy bands which are only *partially* filled. This observation forms the basis for the distinction between insulators, metallic conductors and semiconductors.

In an insulator the number of electrons in the crystal is just sufficient to completely fill a number of energy bands. Above these bands in energy is a series of completely empty bands, but between the full and empty bands is a forbidden energy region so wide that it is virtually impossible at physically realizable temperatures to thermally excite a significant number of electrons across this region from the top of the highest filled band to the bottom of the lowest empty band. All the bands are then either full or empty, and no free-electron current can flow. This situation is illustrated in Figure 8.15(a).

If the energy gap  $\Delta\epsilon$  between the full and empty bands in this type of a crystal is quite small, then there will be appreciable statistical probability that electrons will be excited thermally from states near the top of the filled band across the gap to states

ion  
up  
ust

near the bottom of the empty band. A limited number of free electrons will be available for conduction of electrical currents in the almost empty upper band, and, in addition, the empty electronic states which are left behind near the top of the lower band allow this band to contribute to electric current flow by the mechanism of *hole* conduction.

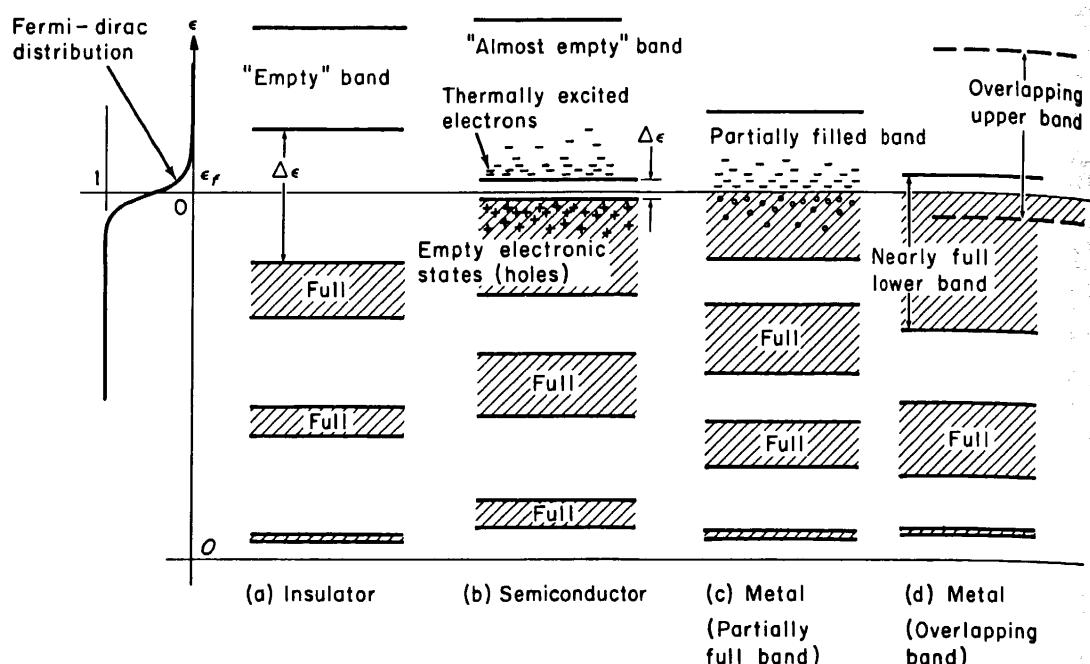


FIGURE 8.15. Energy band diagrams for (a) insulator, (b) semiconductor, (c) metallic conductor (partially filled band) and (d) metallic conductor (overlapping bands).

*conduction.* A material of this sort is called a *semiconductor*; its electrical conductivity is ordinarily much smaller than that of a metal due to the limited concentration of free electrons and holes, and is in addition strongly temperature dependent, rising rapidly with increasing temperature because the probability of thermal excitation rises with temperature. The electrical conductivity of a semiconductor is also, of course, a function of the energy gap  $\Delta\epsilon$ . The distinction between insulators and semiconductors is one of degree only; all semiconductors become ideal insulators as the temperature approaches absolute zero, because then the probability associated with thermal excitation becomes vanishingly small. Likewise, at sufficiently high temperatures (which often cannot be realized experimentally because the crystal melts or vaporizes), all insulators must exhibit semiconductive behavior. The band structure of a semiconductor is shown in Figure 8.15(b).

If the number of electrons in the crystal does not suffice to completely fill the uppermost energy band, but leaves it instead only partially full, then a great number of them can behave as free electrons and can serve as charge or current carriers. This situation is illustrated in Figure 8.15(c). Such a crystal will then exhibit all the characteristic properties of a metallic conductor, for example, high electrical and thermal conductivity and high optical reflectivity.

The determination of the number of electronic states available in a given energy band of a particular substance is not a completely straightforward procedure, and in general involves the use of quantum mechanical calculational techniques which are

C. 8.9

SEC. 8.9

## QUANTUM THEORY

247

beyond the scope of this work. If one were to assume that the tight-binding approximation were always valid, then the energy bands would always arise from the energy levels of the isolated atoms, the wave functions for a given band would always consist of a linear combinations of the wave functions of the corresponding electronic energy state of the isolated atoms, and the degeneracy factor associated with a given energy level in the energy band would be the same as that of the corresponding level in the free atom. Accordingly, there would be a  $1s$  band corresponding to the  $1s$  level, a  $2s$  band, a  $2p$  band, a  $3s$  band, a  $3p$  band, a  $3d$  band, and so on. The number of levels in each band would be  $N$ , according to Equation (8.2-20), and since the degeneracy of each level is the same as that of the corresponding atomic level, the number of electronic quantum states (spin degeneracy included) in each band would be as shown in Table 8.1.

TABLE 8.1.

Band	No. of states
$1s$	$2N$
$2s$	$2N$
$2p$	$6N$
$3s$	$2N$
$3p$	$6N$
$3d$	$10N$

Consider now an alkali metal, such as sodium. Sodium has two  $1s$  electrons, two  $2s$  electrons, six  $2p$  electrons, and a single  $3s$  electron. In a crystal of  $N$  atoms, there are then  $2N$   $1s$  electrons,  $2N$   $2s$  electrons,  $6N$   $2p$  electrons and  $N$   $3s$  electrons. Comparing these figures with those given in Table 8.1, we conclude that the  $1s$ ,  $2s$ , and  $2p$  bands are completely filled, but that the  $3s$  band, which contains  $2N$  states is only half-full. We should therefore expect sodium—along with all the other alkali metals—to be a metallic conductor, as indeed it is. On the other hand, if we adhere strictly to this approach, we must expect magnesium, whose electronic structure is  $1s^2)2s^22p^6)3s^2$ , to be an insulator or semiconductor, since now the  $3s$  band will be filled. The fact that magnesium (and the other alkaline earth elements) are *metals* is due to the fact that the  $3s$  and  $3p$  bands *overlap* in energy, and some of the electrons “spill over” from the  $3s$  band to the  $3p$ , leaving the former not quite full, and rendering the latter partially occupied, as shown in Figure 8.15(d). As a matter of fact, the  $3s$  and  $3p$  wave functions for these bands become strongly intermixed with one another, so that it is not even strictly correct to refer to the bands in that way. The simple tight-binding approach, in effect, is no longer a very good approximation to describe the behavior of  $s$  and  $p$  electrons in the outer valence shell of most metallic substances; this mixing of  $s$  and  $p$  bands for the valence electrons is the rule rather than the exception. In silicon, for example, the  $3s$  and  $3p$  states combine into two nonoverlapping bands each containing  $4N$  states, so that the  $4N$  valence electrons completely occupy the lower band to form the semiconductor band configuration of Figure 8.15(b).

The overlapping of bands occurs despite the fact that there is an energy gap at all points on the boundary of the Brillouin zone which separates the two bands. Using the square zone of Figure 8.12 as an example, it is quite easy to see how this comes about. The Fermi surface in that figure first touches the zone boundary along the

$k_x$ - and  $k_y$ -axes at an energy  $\epsilon = \epsilon_1 + 4\beta$ , but the Brillouin zone and the corresponding energy band are not completely filled until  $\epsilon = \epsilon_1 + 8\beta$ . It may happen that the energy gap between this band and the next, in the direction of the  $k_x$ - or  $k_y$ -axis may be less than  $4\beta$ , the additional energy required to completely fill the lower band. If this is so, then electrons will spill over into the upper band along the  $k_x$ - and  $k_y$ -directions before the corners of the original zone are completely filled, resulting in the overlapping-band situation referred to previously. This state of affairs is illustrated in Figure 8.16. It

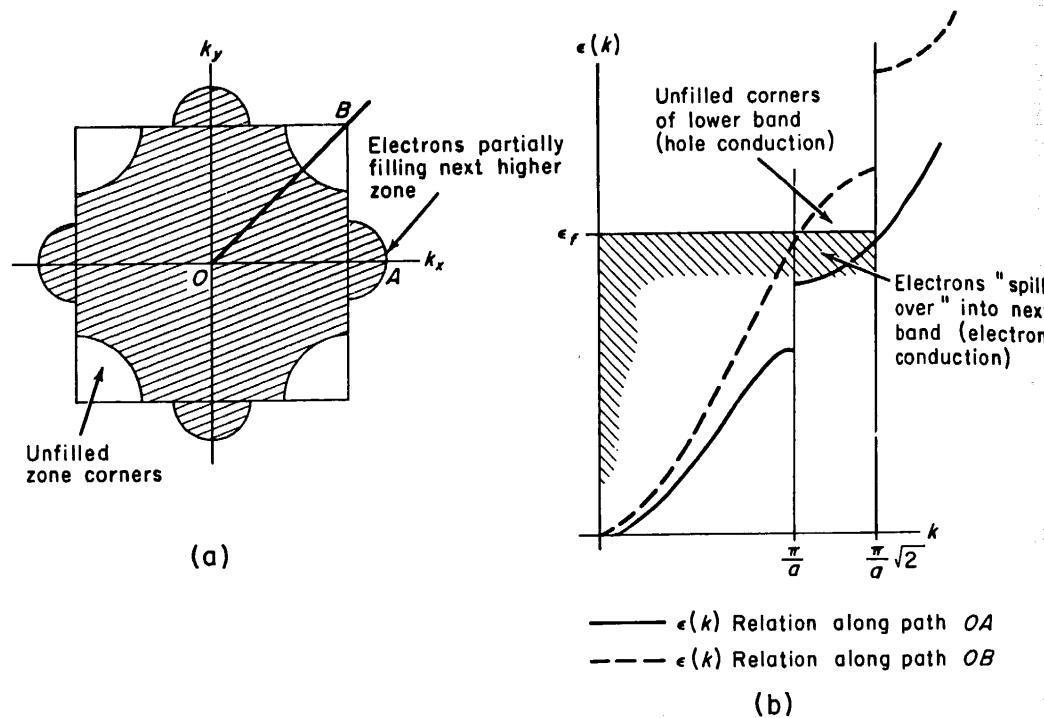


FIGURE 8.16. (a) An instance where the lowest energy states in the higher band (at the centers of the sides of the square zone, on the  $k_x$ - and  $k_y$ -axes) are lower in energy than the highest energy states in the lower band (at the corners of the zone). Under these circumstances electrons start to occupy the higher band before the lower one is completely filled. (b) Schematic representation of the  $\epsilon$  versus  $k$  relation along the direction OA (solid curves) and OB (dashed curves) illustrating in a slightly different way the situation described in (a).

should be emphasized that when this situation occurs, the simple tight-binding assumption of atomic wave functions associated with a single level is no longer valid, and the energy bands must be recalculated assuming a mixture of wave functions associated with both bands. It will also be noted that in such a situation both electron conduction by the electrons which spill over into the higher band and hole conduction by unfilled states in the upper part of the lower band take place simultaneously. The Fermi surface then has two separate branches, one for electrons in the outer zone and one for holes in the original zone. If the hole effective mass is less than the electron effective mass, this situation will lead to a positive Hall coefficient.<sup>12</sup> We are thus led to an

<sup>12</sup> We shall see how this comes about in detail in a later chapter.

C. 8.9  
ding  
ergy  
less  
s so,  
fore  
band  
s. It

understanding of one of the most puzzling discrepancies in the original free electron theory of metals.

We see that the quantum theory enables us to understand in general terms why the electrical properties of metallic conductors and insulators are so vastly different, even though a rather involved wave-mechanical calculation is often required to establish with accuracy the energy surfaces and wave functions associated with the valence electron bands in any particular substance.

## 8.10 THE DENSITY OF STATES FUNCTION AND PHASE CHANGES IN BINARY ALLOYS

In Chapter 5, the density of states function for free particles was derived as Equation (5.2-22). It would appear intuitively obvious that the density of states function for electrons in crystals should be the same as that obtained previously, except possibly that the effective mass rather than the inertial mass might be involved. We shall see shortly that this is indeed true, provided that the constant energy surfaces are spherical. If the constant energy surfaces deviate from the spherical form, however, the density of states function becomes more complex.

To begin with, it is important to note that the allowed values for the crystal momentum  $\hbar\mathbf{k}$ , when periodic boundary conditions are imposed, according to (8.2-20) are the *same* as the allowed values of the inertial momentum of a free particle, as given by (5.2-15). The volume of momentum space per quantum state is thus the same as before, and the task of determining the density of states may be accomplished, as before, by finding the number of quantum states  $dv$  between the constant energy surfaces corresponding to energies  $\epsilon$  and  $\epsilon + d\epsilon$ , computing the volume of momentum space between the two surfaces, and dividing by  $\frac{1}{2}\hbar^3$ , the volume per quantum state. The volume element of momentum space can be written

$$dv_p = p^2 \sin \theta dp d\theta d\phi = \frac{\hbar^3}{8\pi^3} k^2 \sin \theta dk d\theta d\phi, \quad (8.10-1)$$

whereby the number of states within that volume is just

$$\frac{dv_p}{\frac{1}{2}\hbar^3} = \frac{1}{4\pi^3} k^2 \sin \theta dk d\theta d\phi = \frac{1}{4\pi^3} \frac{dk}{d\epsilon} \cdot k^2 \sin \theta d\epsilon d\theta d\phi. \quad (8.10-2)$$

The number of states  $dv = g(\epsilon) d\epsilon$  in the energy range  $d\epsilon$  about  $\epsilon$  is obtained by integrating over the polar angles  $\theta$  and  $\phi$  in  $\mathbf{k}$ -space. Thus,

$$g(\epsilon) = \frac{1}{4\pi^3} \int \int \frac{k^2 \sin \theta d\theta d\phi}{(d\epsilon/dk)} = \frac{1}{4\pi^3} \int_S \frac{dS_k}{(d\epsilon/dk)}, \quad (8.10-3)$$

where  $dS_k$  represents the surface area element  $k^2 \sin \theta d\theta d\phi$  on the constant energy surface within the Brillouin zone. It is clear that if the function  $\epsilon(\mathbf{k})$  is known, the

density of states can in principle be calculated from (8.10-3). Unfortunately, however, the mathematical form of  $\epsilon(\mathbf{k})$  is in most cases of practical interest so complex that the integration can be done only by numerical methods. For the simple cubic lattice in the tight binding approximation, the  $\epsilon(\mathbf{k})$  function for an  $s$ -band is given by (8.7-16), and the corresponding density of states function, as given by (8.10-3), is illustrated in Figure 8.17.

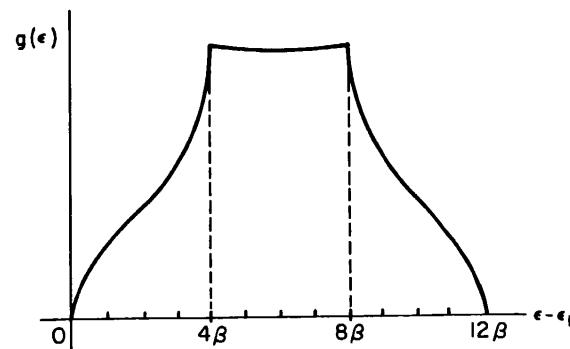


FIGURE 8.17. Density of states curve for the simple cubic lattice in the tight binding approximation, as calculated from (8.10-3) and (8.7-16).

Near the bottom of the energy band for this particular example, the form of  $\epsilon(\mathbf{k})$  is given by (8.8-9); we have then for this region

$$\epsilon(\mathbf{k}) \approx \epsilon_1 + \beta a^2 k^2 \quad \text{and} \quad \frac{d\epsilon}{dk} = 2\beta a^2 k. \quad (8.10-4)$$

Since  $\epsilon$  is independent of  $\theta$  and  $\phi$  under these circumstances, (8.10-3) can be integrated over these angles to give

$$g(\epsilon) = \frac{1}{4\pi^3} k^2 \frac{dk}{d\epsilon} \int_0^{2\pi} \int_0^\pi \sin \theta \, d\theta \, d\phi = \frac{k^2}{\pi^2} \frac{dk}{d\epsilon}, \quad (8.10-5)$$

or, using (8.10-4)

$$g(\epsilon) = \frac{\sqrt{\epsilon - \epsilon_1}}{2\pi^2 (\beta a^2)^{3/2}}. \quad (8.10-6)$$

Using (8.7-18) to express  $\beta a^2$  in terms of the effective mass, this can be written as

$$g(\epsilon) = \frac{8\sqrt{2\pi}}{h^3} m^{*3/2} \sqrt{\epsilon - \epsilon_1} \quad (\epsilon - \epsilon_1 \ll 12\beta), \quad (8.10-7)$$

which, provided that the origin of energy is taken at the bottom of the band, is the usual parabolic free-electron density of states expression with the inertial mass replaced by the effective mass  $m^*$ . In a somewhat similar fashion, it can be shown that for energies near the top of the band, where the constant energy surfaces are again spherical in shape, we may write

$$g(\epsilon) = \frac{8\sqrt{2\pi}}{h^3} m^{*3/2} \sqrt{\epsilon_1 + 12\beta - \epsilon} \quad (\epsilon_1 + 12\beta - \epsilon \ll 12\beta). \quad (8.10-8)$$

8.10

## SEC. 8.10

## QUANTUM THEORY

251

ver,  
hat  
tice  
16),  
ted

1 of

0-4)

ited

0-5)

0-6)

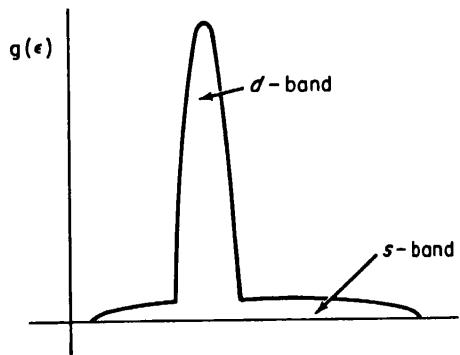
0-7)

the  
nass  
that  
gain

0-8)

This density of states function vanishes at the top of the band, where  $\epsilon = \epsilon_1 + 12\beta$ , and increases parabolically for decreasing energies. It may thus be regarded as a free-particle density of states function for hole conduction in a nearly filled band. The free-electron density of states functions (8.10-7) and (8.10-8) have been obtained with reference to a specific example, but precisely the same expressions may be shown to result for *any* cubic structure, whether or not the tight-binding approximation is applicable. In metals of the transition group, a rather narrow band arising from atomic 3d-states, which will accommodate  $10N$  electrons, overlaps the 4s valence electron band, giving rise to a density of states curve for the valence electrons such as that shown schematically in Figure 8.18. This rather peculiar density of states curve is responsible for many of the characteristic properties of these metals.

FIGURE 8.18. Overlapping s- and d-bands characteristic of the transition metals.



The form of the density of states curve and the concept of the Fermi surface for valence electrons within the Brillouin zone has been used to explain certain phase changes which are associated with binary alloys of copper, silver and gold with divalent or trivalent metals. Pure copper, silver and gold are all monovalent elements which have the face-centered cubic structure. If one begins with pure silver, for example, and adds increasing amounts of cadmium, the f.c.c. structure, which is called the  $\alpha$ -phase, is retained up to a certain concentration of cadmium (called the  $\alpha$ -limit), at which point a phase change occurs, and the crystal structure changes from f.c.c. to b.c.c., the latter being referred to as the  $\beta$ -phase. If the concentration of cadmium is further increased, the  $\beta$ -phase b.c.c. structure is stable over a certain range of concentration, after which another transition point (called the  $\beta$ -limit) is reached and the  $\beta$ -phase is transformed into a  $\gamma$ -phase whose structure is a very complex cubic arrangement with some 52 atoms per unit cell. At still higher cadmium concentrations a  $\gamma$ -limit is reached and a hexagonal close-packed  $\eta$  phase is formed. Alloys of the other noble metals and alloys made with other multivalent additives follow the same general pattern of phase changes, although the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -limits are different in each individual system.

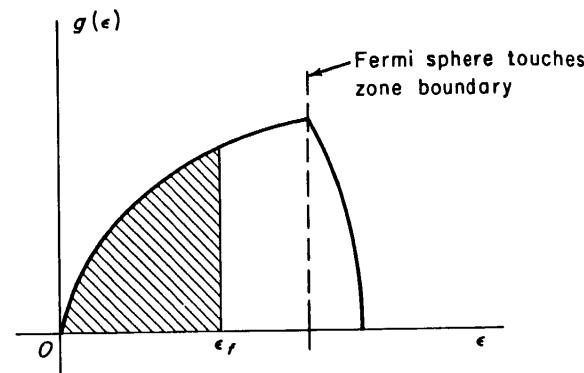
Hume-Rothery<sup>13</sup> pointed out that the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -limits in these alloy systems always occur at certain well-defined valence electron per atom ratios. Thus the  $\beta$ -limit seemed always to be reached at the point where there were just about  $3/2$  valence electrons for each atom in the crystal, and the  $\gamma$ -limit appeared to be reached at a valence electron per atom ratio of about  $21/13$ . This empirical observation was explained on fundamental grounds by Jones.<sup>14</sup> To begin with, one must note that for

<sup>13</sup> W. Hume-Rothery, *The Metallic State*, Clarendon Press, Oxford (1931), p. 328.

<sup>14</sup> H. Jones, *Proc. Roy. Soc.* **144**, 225 (1934); **147**, 396 (1934). See also N. Mott and H. Jones, *op. cit.*, p. 170.

the pure monovalent f.c.c. metal in the  $\alpha$ -phase, the (reduced) Brillouin zone for the valence electrons is as shown in Figure 8.14(c). Since the valence electrons in the noble metals are  $s$ -electrons, we may expect that this Brillouin zone will contain  $2N$  electronic states. For the pure metal the zone is just half full, because the total number of valence electrons is then  $N$ . For simplicity, we shall assume, as Jones did, that the Fermi surface within the zone is spherical at all times.<sup>15</sup> On the basis of this simple assumption, one may calculate the density of states curve for the Brillouin zone quite simply; the result is shown schematically in Figure 8.19. The density of states function

FIGURE 8.19. Picture of the density of states curve for the simple cubic lattice arising from the idea of a spherical Fermi surface expanding within the Brillouin zone, as employed by Jones.



is simply a free-electron parabola, *until the Fermi sphere touches the zone boundary*. From that point on, however, the boundaries of the zone cut off parts of the Fermi sphere, and a given increase in the radius (corresponding to a given increment of energy) encompasses only a relatively small number of states in the corners of the Brillouin zone. The density of states curve thus decreases rapidly, going to zero when the radius of the Fermi sphere equals the radius of the circumscribed sphere. The same general behavior can be noted in the density of states curve of Figure 8.17, which is associated with a simple cubic lattice.

When atoms of a multivalent metal are alloyed with the monovalent  $\alpha$ -phase, the number of free electrons per atom is increased, the Fermi level is raised, and the radius of the Fermi sphere in the Brillouin zone associated with the valence electrons increases. Beyond the point where the Fermi sphere touches the zone boundary, a large increase in electronic energy takes place with each small increment in free-electron concentration, thus with each small increment in concentration of the multivalent additive. At this point it will result in a *lower* total electronic energy if the crystal undergoes a phase change to a lattice whose Brillouin zone will accommodate a *larger* inscribed sphere, and indeed this is what happens. The f.c.c.  $\alpha$ -phase changes to the b.c.c.  $\beta$ -phase, whose Brillouin zone [Figure 8.14(b)], although of the same volume, does in fact have a larger inscribed sphere. Upon further addition of the multivalent alloying element, the radius of the Fermi sphere increases further, finally touching the surface of the  $\beta$ -phase Brillouin zone. Again a phase change takes place, the Brillouin zone of the resulting  $\gamma$ -phase being a complex polyhedron of 36 faces

<sup>15</sup> This assumption is much better here than it would have been in the case of the simple cube whose Fermi surfaces are illustrated in Figure 8.11, because of the fact that the bounding Brillouin zone is more nearly spherical in shape. In any case the exact form of the Fermi surface is not particularly important for energies *larger* than that required to make the Fermi surface barely touch the zone boundary.

8.10  
the  
the  
2N  
ber  
the  
iple  
uite  
tion

## SEC. 8.10

## QUANTUM THEORY

253

which is more nearly spherical<sup>16</sup> in form than the rhombic dodecahedron associated with the  $\beta$ -phase and whose inscribed sphere is therefore still larger.

It is possible, on the basis of these arguments to make quantitative estimates of the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -limits. In the  $\alpha$  phase, the Fermi sphere first reaches the zone boundary along the {111} directions in  $k$ -space. The (111) plane forming part of the Brillouin zone polyhedron has [according to (8.8-13)] the equation

$$k_x + k_y + k_z = 3\pi/a. \quad (8.10-9)$$

At the point of contact, in the center of the (111) face,  $k_x = k_y = k_z = \pi/a$ , whereby the radius of the Fermi sphere at contact,  $k_c$ , must be given by

$$k_c = \sqrt{k_x^2 + k_y^2 + k_z^2} = \pi\sqrt{3}/a. \quad (8.10-10)$$

The volume of the Fermi sphere, under these conditions, is  $4\pi k_c^3/3$ , or  $4\sqrt{3}\pi(\pi/a)^3$ , while it is easily shown that the total volume of the zone is just half the volume of a cube of edge  $4\pi/a$ , or  $32(\pi/a)^3$ . The ratio of the volume of the inscribed sphere to that of the zone is then the ratio of these two numbers, or  $\pi\sqrt{3}/8$ . The average number of valence electrons per atom is just the number per atom when the zone is filled (i.e., 2) times the fraction of the number of states within the zone which are occupied. The number of valence electrons per atom for the  $\alpha$ -limit is thus  $\pi\sqrt{3}/4$  or 1.362. The values for the  $\beta$ - and  $\gamma$ -limits can be worked out using the same general procedures. The values obtained, compared with Hume-Rothery's empirically derived values are shown in Table 8.2.

TABLE 8.2.

## Electron-atom ratio

Phase limit	Hume-Rothery	Jones
$\alpha$	—	1.362
$\beta$	1.5	1.480
$\gamma$	1.615	1.538

Naturally, since the Fermi surfaces cannot be perfectly spherical, the Jones estimates must be only approximate, but they are found nevertheless to agree quite well with experimentally determined phase limits. The phase diagrams of more complex alloy systems can also often be understood in terms of electronic energy surfaces and the density of states function within the Brillouin zone, but the details of calculation and the resulting answers are inevitably more involved.

<sup>16</sup> See F. Seitz, *Modern Theory of Solids*, McGraw-Hill Book Company, Inc., New York (1940), Figure 13, p. 433.

## EXERCISES

1. Show that if in the potential model of Figure 8.2, used in connection with the Kronig-Penney calculation,  $V_0$  is allowed to become infinitely large and  $b$  is allowed to approach zero, such that the product  $P = -\beta^2 ab/2$  remains fixed, then equation (8.3-12) becomes

$$P \frac{\sin \alpha a}{\alpha a} + \cos \alpha a = \cos ka.$$

If a plot of the function on the left-hand side is made, it is seen to behave in a manner similar to (8.3-15) and (8.3-16) as shown in Figure 8.3. The quantity  $P$  may be regarded as the "scattering power" of a single potential spike.

2. Show that in the case discussed in Exercise 1 above, the quantity  $d\varepsilon/dk$  vanishes at the band edges ( $k = n\pi/a$ ).
3. Show that for the one-dimensional free-electron approximation discussed in Section 8.6,  $d\varepsilon/dk$  vanishes at the band edges.
4. Show that for the tight binding approximation in the simple cubic lattice, discussed in Section 8.7,  $\mathbf{n} \cdot \nabla_{\mathbf{k}}\varepsilon$  vanishes at all points on the cubical boundaries of the first Brillouin zone.
5. Show explicitly that the polyhedra shown in Figure 8.14(b) and (c) are the correct forms for the first Brillouin zone of the b.c.c. and f.c.c. lattices, respectively.
6. Find the  $\varepsilon(\mathbf{k})$  relation for an  $s$ -band (spherically symmetric wave functions) in the tight binding approximation for a body-centered cubic crystal. Consider nearest neighbor overlap integrals only. Plot the forms of the constant energy surfaces for several energies within the zone. Show that these surfaces are spherical, as for free electrons, for energies near the bottom of the band. Show that  $\mathbf{n} \cdot \nabla_{\mathbf{k}}\varepsilon$  vanishes on the zone boundaries.
7. Obtain all the results required in Exercise 6 above for an  $s$ -band in a face-centered cubic crystal in the tight binding approximation, considering overlap of nearest neighbor wave functions only.
8. Make a diagram [similar to Figure 8.13(c)] in the  $\mathbf{k}$ -plane, of the first five Brillouin zones of a two-dimensional lattice defined by the basis vectors  $\mathbf{a} = ai_x$ ,  $\mathbf{b} = 3ai_y$ . Note: The formulas (2.5-5) cannot be used to define a two-dimensional reciprocal lattice, because the cross product cannot be defined in a two-dimensional space. Proceed from the fundamental definitions (2.5-1) and (2.5-2).
9. Show explicitly from the results of Section 8.7 that the density of states near the top of the energy band for the simple cubic crystal which was discussed in detail in that section is as given by Equation (8.10-8).
10. Show that the limiting valence electron per atom ratio for the  $\beta$ -phase (b.c.c.) of a binary alloy of the type discussed in Section 8.10 is 1.489. You may make the assumption of the Jones calculation.
11. Show that the wave function (8.7-3) which was used in the tight binding approximation calculation can be expressed as a Bloch function of the form (8.2-14).

## GENERAL REFERENCES

- W. Hume-Rothery, *The Metallic State*, Clarendon Press, Oxford (1931).
- H. Jones, *The Theory of Brillouin Zones and Electronic States in Crystals*, North Holland Publishing Co., Amsterdam (1960).
- N. F. Mott and H. Jones, *Theory of Metals and Alloys*, Dover Publications, Inc., New York (1958).
- S. Raimes, *The Wave Mechanics of Electrons in Metals*, Interscience Publishers, New York (1961).

.8.10

## SEC. 8.10

## QUANTUM THEORY

255

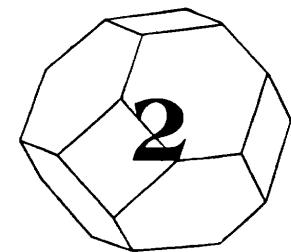
- F. Seitz, *Modern Theory of Solids*, McGraw-Hill Book Co., Inc., New York (1940).  
R. A. Smith, *Wave Mechanics of Crystalline Solids*, John Wiley & Sons, Inc., New York (1961).  
E. Spenke, *Electronic Semiconductors*, McGraw-Hill Book Co., Inc., New York (1958).  
A. H. Wilson, *Theory of Metals*, 2nd Edition, Cambridge University Press, New York (1953).

with an associated intensity

$$\frac{I}{N^2} = \left\{ (N^2 - 1) \sin^2 \left[ \frac{\pi}{2N} (2p + 1) \right] + 1 \right\}^{-1}$$

For the case  $N = 1000$ , determine numerical solutions of the exact relation for the first five secondary maxima and compare these to the approximate solutions. Express the solutions in terms of  $\pi/N$ . [That is,  $(N/\pi)x = (2p + 1) - (\pi/N)\delta_p$ .] Compare the relative refracted intensity ( $I/N^2$ ) to the approximate solutions.

- 1.12. A back-reflection Laue photograph is made using x-rays of wavelength  $\lambda$  directed along a  $\langle 100 \rangle$  direction of a simple cubic sample with lattice constant  $a$ . For  $\lambda = a/5$  and a film width of four times the sample-to-film spacing, determine the pattern on the photograph.



NOV-05-2004 19:30

nazir

## Energy Bands

An electron characterized by its wavefunction,  $\psi(\mathbf{r})$ , and spin orientation,  $\mathbf{s}$ , must satisfy the time-independent Schrödinger equation,

$$\mathbf{H}\psi(\mathbf{r}) = \mathcal{E}\psi(\mathbf{r}) \quad (2.1)$$

where  $\mathcal{E}$  is the total energy of the electron and  $\mathbf{H}$  is the appropriate Hamiltonian operator. The Hamiltonian takes into account all kinetic and potential energy terms, including applied forces and interactions with other particles. If the electron is traveling in a force-free region where it does not interact with other electrons (a free electron), the Hamiltonian contains only a kinetic energy term for the one electron,  $\mathbf{p}^2/2m$ , where the momentum operator,  $\mathbf{p}$ , is

$$\mathbf{p} = \frac{\hbar}{i} \frac{\partial}{\partial \mathbf{r}} = \frac{\hbar}{i} \nabla \quad (2.2)$$

and  $m$  is the free electron mass. In (2.3)  $\hbar = h/2\pi$  where  $h$  is Planck's constant. Under these conditions Schrödinger's equation (2.1) reduces to its free one-electron formulation,

$$-\frac{\hbar^2}{2m} \nabla^2 \psi(\mathbf{r}) = \mathcal{E}\psi(\mathbf{r}) \quad (2.3)$$

which has solutions of the form

$$\psi_{\mathbf{k}}(\mathbf{r}) = A \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (2.4)$$

where  $\mathbf{k}$  is any position-independent vector.

With these solutions we can easily determine the free-electron energy from (2.3) as

$$\mathcal{E}(\mathbf{k}) = \frac{\hbar^2 \mathbf{k}^2}{2m} \quad (2.5)$$

The momentum is determined by operating on (2.4) with (2.2) to give

$$\mathbf{p} = \hbar \mathbf{k} \quad (2.6)$$

Considering the electron as a particle with velocity  $\mathbf{v} = \mathbf{p}/m$ , we arrive at the simple classical expression for the total energy of a free electron,

$$\mathcal{E} = \frac{\mathbf{p}^2}{2m} = \frac{1}{2} m\mathbf{v}^2 \quad (2.7)$$

From (2.4) we can also consider the electron as a plane wave with wavevector  $\mathbf{k}$  and de Broglie wavelength,

$$\lambda = \frac{2\pi}{|\mathbf{k}|} \quad (2.8)$$

The problem we consider in this chapter is how this free-electron description is modified for electrons in a periodic crystal structure. We expect a substantial modification for the following reasons. The atoms in the crystal, consisting of valence electrons, core electrons, and nuclei, produce a potential energy  $U(\mathbf{r})$  with the periodicity of the direct Bravais lattice,

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R}) \quad (2.9)$$

for all direct lattice vectors,  $\mathbf{R}$ . Equation (2.8) tells us that a free-electron wavelength is of the same order of magnitude as the lattice periodicity. Thus we expect electrons to be strongly diffracted by the lattice.

To solve this problem, one would, in principle, have to include in the Hamiltonian of (2.1) terms that take into account interactions among the nuclei, core electrons, and valence electrons. Such a problem would be difficult to formulate, let alone solve. Since, in semiconductors the valence electrons are shared among atoms, a useful approximation is to treat the valence electrons as noninteracting entities that move through the crystal under the influence of an effective potential which includes the combined effects of the nuclei, the core electrons, and other valence electrons. Although not obvious a priori, this *one-electron approximation* provides a good description of semiconductor properties. In this manner we formulate a one-electron Schrödinger equation from (2.1) and (2.9), where  $U(\mathbf{r})$  is taken as an effective one-electron potential: that is, the potential that the nuclei, core electrons, and all the other valence electrons produce for one valence electron. The problem is then a manner of solving (2.1) for the allowed one-electron energy levels.

There are several reasons why this one-electron approximation works as well as it does.

1. Electrons tend to be spatially removed from one another by Coulomb repulsion and by Fermi exclusion when they have the same spin. This reduces the interaction between the one electron and the rest of the valence electrons taken as a whole.
2. The valence electrons tend to cluster around the ion cores (nuclei and core electrons) due to Coulomb attraction. This effectively screens the Coulomb attraction of the ionic cores for the one electron and reduces this interaction.
3. Electrons passing near the ionic cores are accelerated by the Coulomb attraction. Because of this, electrons spend less time in the neighborhood of a core, effectively reducing the Coulomb attraction. (It is this effective *repulsion* that produces the *pseudopotential* discussed in Section 2.5.)

For these reasons we will examine in detail relatively simple one-electron models that illustrate some of the more important properties of electrons in periodic structures. We then discuss the results of more detailed computations on specific crystal structures.

## 2.1 BLOCH ELECTRONS

Let us first examine those properties of electrons in periodic structures which are independent of the specific nature of the potential  $U(\mathbf{r})$ . The term *Bloch electron* is used to refer to an electron that obeys the one-electron Schrödinger equation in a periodic potential. Bloch found that such electrons have wavefunctions in the form of a plane wave multiplied by a function that has the periodicity of the direct lattice. That is,

$$\psi_k(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r})u_k(\mathbf{r}) \quad (2.10)$$

where  $\mathbf{k}$  is a wavevector and

$$u_k(\mathbf{r}) = u_k(\mathbf{r} + \mathbf{R}) \quad (2.11)$$

for all direct lattice vectors  $\mathbf{R}$ . This result is known as *Bloch's theorem* [F. Bloch, *Z. Phys.* 52, 555 (1928)]. From (2.10) we also have

$$\psi_k(\mathbf{r} + \mathbf{R}) = \exp[i\mathbf{k}\cdot(\mathbf{r} + \mathbf{R})]u_k(\mathbf{r} + \mathbf{R}) \quad (2.12)$$

or using (2.11),

$$\psi_k(\mathbf{r} + \mathbf{R}) = \exp(i\mathbf{k}\cdot\mathbf{R})\psi_k(\mathbf{r}) \quad (2.13)$$

for any value of  $\mathbf{k}$  and every  $\mathbf{R}$  in the direct lattice.

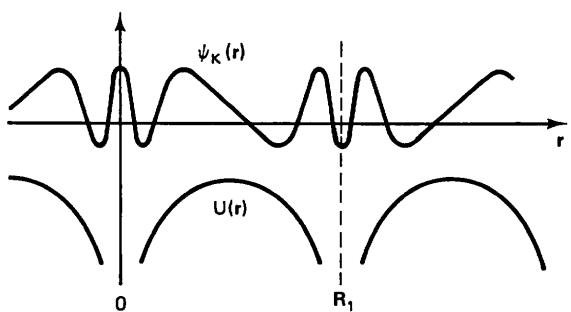


Figure 2.1 Possible Bloch electron wavefunction and periodic potential showing phase difference from one unit cell to the next.  $\mathbf{r}$  and  $\mathbf{R}_1$  are taken to be parallel.

Equation (2.13) is an alternative form of Bloch's theorem. It tells us that the electron wavefunction in any primitive unit cell of the direct lattice differs from that in any other cell only by the factor  $\exp(i\mathbf{k}\cdot\mathbf{R})$ . For real  $\mathbf{k}$  this represents a difference in phase as shown in Fig. 2.1. We can see that this factor is similar to the expression obtained in (1.14), which was

$$\exp(i\mathbf{K}\cdot\mathbf{R}) = 1 \quad (2.14)$$

for all reciprocal lattice vectors  $\mathbf{K}$ . The wavevector,  $\mathbf{k}$ , thus has dimensions of reciprocal length and belongs in reciprocal space with the vectors  $\mathbf{K}$ . Let us assume, for instance, that some electron wavefunction has a wavevector that is equal to a reciprocal lattice vector. From (2.13),

$$\begin{aligned} \psi_{\mathbf{K}}(\mathbf{r} + \mathbf{R}) &= \exp(i\mathbf{K}\cdot\mathbf{R})\psi_{\mathbf{K}}(\mathbf{r}) \\ &= \psi_{\mathbf{K}}(\mathbf{r}) \end{aligned} \quad (2.15)$$

for all  $\mathbf{R}$ . That is, the electron wavefunctions  $\psi_{\mathbf{K}}$  are periodic in  $\mathbf{R}$ .

Let us assume that an electron has a wavevector  $\mathbf{k}$  given by

$$\mathbf{k} = \mathbf{K} + \mathbf{k}' \quad (2.16)$$

where  $\mathbf{k}'$  is some other vector in reciprocal space. From (2.13) and (2.14) we find that

$$\begin{aligned} \psi_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) &= \exp\{i[(\mathbf{K} + \mathbf{k}')\cdot\mathbf{R}]\}\psi_{\mathbf{k}}(\mathbf{r}) \\ &= \exp(i\mathbf{k}'\cdot\mathbf{R})\psi_{\mathbf{k}}(\mathbf{r}) \end{aligned} \quad (2.17)$$

or the wavefunctions  $\psi_{\mathbf{k}}$  obey Bloch's theorem as if they had wavevector  $\mathbf{k}'$ . Thus the wavefunction does not have a unique, wavevector  $\mathbf{k}$ , but a set of wavevectors that differ from each other by the set of reciprocal lattice vectors.

As indicated in Fig. 2.2, we can define a wavevector uniquely by reducing it with the appropriate reciprocal lattice vector to the first Brillouin zone. The prescription for this reduction is as follows. We choose the value

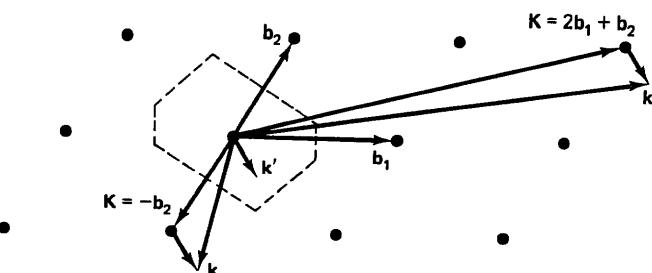


Figure 2.2 A two-dimensional reciprocal lattice indicating that any wavevector  $\mathbf{k}$  in a higher Brillouin zone can be reduced to a value  $\mathbf{k}'$  in the first Brillouin zone by choosing the appropriate reciprocal lattice vector  $\mathbf{K}$ .

of  $\mathbf{K}$  that will make the point  $\mathbf{k}'$  lie as close to the origin as is possible. Since the value of  $\mathbf{K}$  can be selected in increments of  $\mathbf{b}_i$ , the primitive vectors, the point  $\mathbf{k}'$  can be made to lie closer to the origin than to any other lattice point in reciprocal space. This, of course, is the first Brillouin zone. Therefore, we have shown that any wavevector in higher Brillouin zones in reciprocal space is equivalent to one in the first Brillouin zone. It is for this reason that the first Brillouin zone is to be preferred over other primitive unit cells in the reciprocal lattice.

Let us now consider the number of allowed values of  $\mathbf{k}$  in the first Brillouin zone. This can be determined by introducing boundary conditions at the outer surfaces of the crystal. If there are a total of  $N$  primitive unit cells of volume  $\Omega$  in a crystal of volume  $V$  so that  $N\Omega = V$ , the  $N$  unit cells can be divided into  $N_i$  unit cells in the directions of the  $\mathbf{a}_i$  primitive vectors,  $i = 1, 2, 3$ , so that

$$N = N_1 N_2 N_3 \quad (2.18)$$

Thus the boundary of the crystal in the  $\mathbf{a}_i$  direction is at  $N_i \mathbf{a}_i$ . To avoid standing electron waves we impose a cyclic or Born-von Kármán [M. Born and T. von Kármán, *Z. Phys.* 13, 297 (1912)] condition at these boundaries,

$$\psi(\mathbf{r}) = \psi(\mathbf{r} + N_i \mathbf{a}_i), \quad i = 1, 2, 3 \quad (2.19)$$

According to Bloch's theorem (2.13), we have

$$\psi_{\mathbf{k}}(\mathbf{r} + N_i \mathbf{a}_i) = \exp(iN_i \mathbf{k} \cdot \mathbf{a}_i) \psi_{\mathbf{k}}(\mathbf{r}) \quad (2.20)$$

or

$$\exp(iN_i \mathbf{k} \cdot \mathbf{a}_i) = 1 \quad (2.21)$$

Since the  $\mathbf{a}_i$  are real, the  $\mathbf{k}$  must also be real to satisfy (2.21). If we define the wavevectors  $\mathbf{k}$  in terms of the primitive vectors for the reciprocal lattice  $\mathbf{b}_i$ , we can write

$$\mathbf{k} = k_1 \mathbf{b}_1 + k_2 \mathbf{b}_2 + k_3 \mathbf{b}_3 \quad (2.22)$$

where the components  $k_i$  are to be determined. By inserting (1.16) for the  $\mathbf{b}_i$  into (2.22) and (2.22) into (2.21), we obtain

$$\exp(i2\pi N_i k_i) = 1 \quad (2.23)$$

or

$$N_i k_i = m_i, \quad i = 1, 2, 3 \quad (2.24)$$

where the  $m_i$  takes on all integer values. Thus the allowed values of  $\mathbf{k}$  in reciprocal space are

$$\mathbf{k} = \frac{m_1}{N_1} \mathbf{b}_1 + \frac{m_2}{N_2} \mathbf{b}_2 + \frac{m_3}{N_3} \mathbf{b}_3 \quad (2.25)$$

Let us recall from (1.17) that the reciprocal lattice vector which defines the reciprocal lattice primitive unit cell is

$$\mathbf{K} = \mathbf{b}_1 + \mathbf{b}_2 + \mathbf{b}_3 \quad (2.26)$$

and from (1.20) that the volume of this cell is

$$\Omega_K = \mathbf{b}_1 \cdot (\mathbf{b}_2 \times \mathbf{b}_3) = \frac{(2\pi)^3}{\Omega} \quad (2.27)$$

In a similar manner, the volume of reciprocal space occupied by an allowed value of  $\mathbf{k}$  is defined by

$$\mathbf{k} = \frac{\mathbf{b}_1}{N_1} + \frac{\mathbf{b}_2}{N_2} + \frac{\mathbf{b}_3}{N_3} \quad (2.28)$$

and given as

$$\Omega_k = \frac{\mathbf{b}_1}{N_1} \cdot \left( \frac{\mathbf{b}_2}{N_2} \times \frac{\mathbf{b}_3}{N_3} \right) = \frac{(2\pi)^3}{N\Omega} = \frac{(2\pi)^3}{V} \quad (2.29)$$

Therefore,

$$\Omega_K = N\Omega_k \quad (2.30)$$

and since the volume of a primitive cell is independent of how it is chosen, there are  $N$  allowed values of  $\mathbf{k}$  in the first Brillouin zone. These results are summarized in Fig. 2.3 for a two-dimensional reciprocal lattice.

Since the number of unit cells  $N$  in a crystal of volume  $V$  is equal to or has the same order of magnitude as the number of atoms ( $10^{22}$  to  $10^{23}$  per cubic centimeter), the number of allowed values of  $\mathbf{k}$  in the first Brillouin zone is quite large and  $\Omega_k$  is very small. Because of this it is sometimes convenient to treat reciprocal space and the first Brillouin zone as a continuum for  $\mathbf{k}$  values. However, when  $\mathbf{k}$  is used to index the energy levels in each energy band, it is treated as discrete.

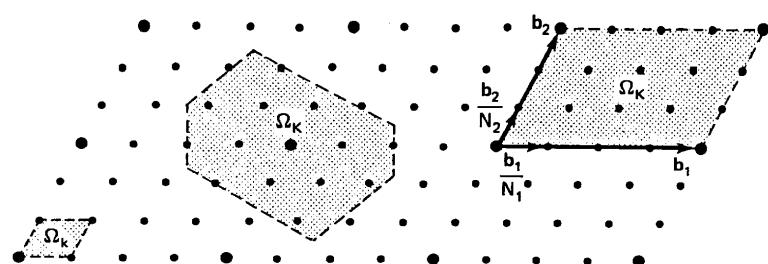


Figure 2.3 Allowed values of  $\mathbf{k}$  (small points) in a two-dimensional reciprocal lattice. The figure is drawn with  $N_1 = 4$ ,  $N_2 = 3$ . Both  $\Omega_K$  equal  $12 \Omega_k$ .

It is interesting to compare the wavevector  $\mathbf{k}$  for Bloch electrons to the wavevector for free electrons. From (2.6) we know that for free electrons  $\mathbf{k}$  is proportional to the electron momentum,

$$\mathbf{p} = \hbar\mathbf{k} \quad (2.31)$$

For Bloch electrons, however, this is not the case. To determine the relationship between  $\mathbf{p}$  and  $\mathbf{k}$  for Bloch electrons, we operate on (2.10) with (2.2) to give

$$\mathbf{p}\psi_k(\mathbf{r}) = \hbar\mathbf{k}\psi_k(\mathbf{r}) + \exp(i\mathbf{k}\cdot\mathbf{r}) \frac{\hbar}{i} \frac{\partial}{\partial\mathbf{r}} u_k(\mathbf{r})$$

which for a periodic potential, is not a constant times the wavefunction. Thus  $\hbar\mathbf{k}$  is not the momentum of a Bloch electron. It is, nevertheless, useful and convenient to define a *crystal momentum* for Bloch electrons as

$$\mathbf{P} = \hbar\mathbf{k} \quad (2.32)$$

We will find in Section 2.8 that this crystal momentum,  $\mathbf{P}$ , behaves as a momentum only for externally applied forces. The “real” momentum,  $\mathbf{p}$ , must take into account the response of the Bloch electrons to externally applied forces *and* the internal periodic potential of the crystal.

## 2.2 EMPTY LATTICE MODEL

We have seen that a substantial amount of information can be obtained about the wavevector of a Bloch electron without reference to the specific nature of the effective one-electron potential energy,  $U(\mathbf{r})$ . We can also obtain the general form and degeneracy (states at the same energy) of Bloch electron energy bands by solving the one-electron Schrödinger equation (2.3) for  $U(\mathbf{r}) = 0$ . The solutions, of course, are identical to those obtained for free electrons (2.4) and consist of plane waves with wavevectors,  $\mathbf{k}$ , which are con-

tinuous throughout the reciprocal lattice. For this reason, the resulting energy values (2.5), given by

$$\mathcal{E}(k) = \frac{\hbar^2}{2m} k^2 \quad (2.33)$$

are referred to as *free-electron* or *empty lattice* energy bands. The relationship between  $\mathcal{E}$  and  $k$  in (2.33) is referred to as a *parabolic* energy band.

To determine the general form of the energy bands for Bloch electrons, we wish to reduce the free-electron wavevector,  $\mathbf{k}$ , to the first Brillouin zone. As we saw in Fig. 2.2 this can be done for any wavevector (including a free-electron wavevector) by a suitable choice of reciprocal lattice vector,

$$\mathbf{k} = \mathbf{k}' + \mathbf{K} \quad (2.34)$$

The wavefunction (2.4) is then

$$\psi_{\mathbf{k}}(\mathbf{r}) = A \exp [i(\mathbf{k}' + \mathbf{K}) \cdot \mathbf{r}] \quad (2.35)$$

which is a Bloch wavefunction (2.10) with

$$u_{\mathbf{k}}(\mathbf{r}) = A \exp (i\mathbf{K} \cdot \mathbf{r}) \quad (2.36)$$

From (2.34) and (2.33) the empty lattice energy bands are

$$\mathcal{E}(\mathbf{k}) = \frac{\hbar^2}{2m} (\mathbf{k}' + \mathbf{K})^2 \quad (2.37)$$

where  $\mathbf{K}$  is given by (1.17) as

$$\mathbf{K} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3 \quad (2.38)$$

and  $h, k, l$  are integers. Thus it can be seen that  $\mathbf{K}$  serves as an index for the different energy bands. Since there are an infinite number of  $h, k, l$ , there are an infinite number of energy bands.

We can demonstrate these free-electron energy bands most simply by looking at values of  $\mathbf{k}$  in the direction of  $\mathbf{K}$  without reference to a specific crystal structure. This gives the one-dimensional energy bands shown in Fig. 2.4. The lowest-lying band in the first Brillouin zone is obtained by setting  $\mathbf{K} = 0$  in (2.37) and letting  $\mathbf{k}$  range from  $-\frac{1}{2}\mathbf{K}$  to  $+\frac{1}{2}\mathbf{K}$ . Continuing this procedure, an infinite number of energy bands are generated in the first Brillouin zone, each indexed with a separate reciprocal lattice vector. These free-electron energy bands, however, are of most interest in three dimensions, where they illustrate the form and degeneracy of electron states for a specific lattice.

As an example, we consider the general form and degeneracies of the empty lattice energy bands for the face-centered cubic lattice. The reciprocal space for this lattice is shown in Fig. 2.5. Values of wavevector,  $\mathbf{k}$ , for some

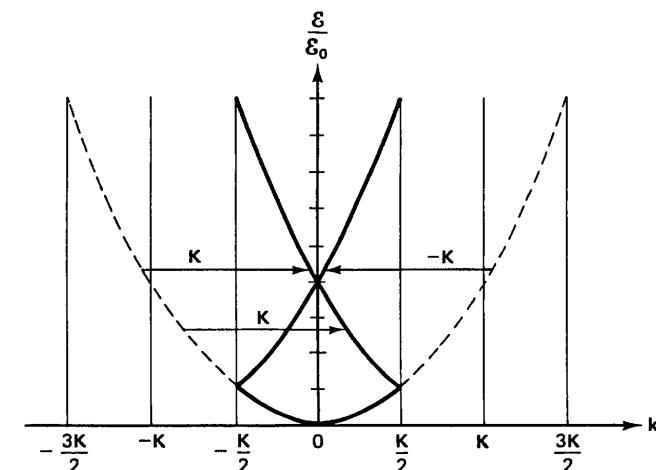


Figure 2.4 Empty lattice energy bands for  $\mathbf{k}$  in the direction of  $\mathbf{K}$ . The dashed line shows the parabolic free-electron description.  $\mathcal{E}_0 = (\hbar^2/2m)(\frac{1}{2}\mathbf{K})^2$  is the energy at  $\pm \frac{1}{2}\mathbf{K}$ , the boundaries of the first Brillouin zone.

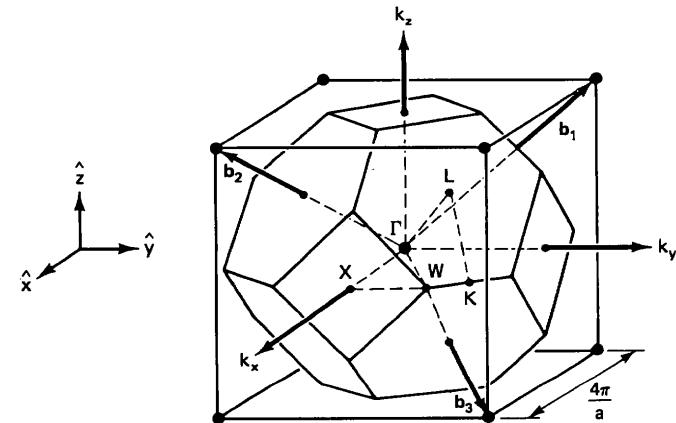


Figure 2.5 Reciprocal lattice points, primitive vectors, first Brillouin zone, and several high symmetry points for the face-centered cubic direct lattice. This is reciprocal space for crystals with the diamond, sphalerite, and sodium chloride structures.

of the high-symmetry points in the Brillouin zone are:

$$\Gamma\text{-point: } \mathbf{k} = 0$$

$$X\text{-points: } \mathbf{k} = \pm \frac{2\pi}{a} \hat{x}, \quad \pm \frac{2\pi}{a} \hat{y}, \quad \pm \frac{2\pi}{a} \hat{z}$$

$$L\text{-points: } \mathbf{k} = \pm \frac{\pi}{a} (\hat{x} + \hat{y} + \hat{z}), \quad \pm \frac{\pi}{a} (-\hat{x} + \hat{y} + \hat{z}), \\ \pm \frac{\pi}{a} (\hat{x} - \hat{y} + \hat{z}), \quad \pm \frac{\pi}{a} (\hat{x} + \hat{y} - \hat{z}) \quad (2.39)$$

$$K\text{-points: } \mathbf{k} = \pm \frac{3\pi}{2a} (\hat{x} + \hat{y}), \text{ etc.}$$

$$W\text{-points: } \mathbf{k} = \pm \frac{\pi}{a} (2\hat{x} + \hat{y}), \text{ etc.}$$

In the discussion to follow it is convenient to express the reciprocal lattice vectors (points) in terms of their components along orthogonal unit vectors. Using (1.23) in (2.38), we obtain

$$\mathbf{K} = \frac{2\pi}{a} [(-h + k + l)\hat{x} + (h - k + l)\hat{y} + (h + k - l)\hat{z}] \quad (2.40)$$

Let us first consider the lowest-lying energy band given by (2.37). This is given by  $h = k = l = 0$  or  $\mathbf{K} = 0$ . The minimum of the band is  $\mathcal{E} = 0$  at  $\mathbf{k} = 0$  or at the  $\Gamma$ -point. The energy of this band increases parabolically with increasing  $\mathbf{k}$  in the Brillouin zone until it reaches the values,  $\mathcal{E} = (\hbar^2/2m)(2\pi/a)^2$  at the  $X$ -points,  $\mathcal{E} = (\hbar^2/2m)(\sqrt{3}\pi/a)^2$  at the  $L$ -points, and so on in different directions. From (2.35) there is only one wavefunction that corresponds to this range of energies up to the zone boundaries, so this lowest-lying energy band is nondegenerate (except for the twofold spin degeneracy).

We see from (2.37) that there are an infinite number of energy values for  $\mathbf{k} = 0$ , corresponding to the infinite number of reciprocal lattice points. The next energy value above zero at the  $\Gamma$ -point occurs for  $h = \pm 1$ , or  $k = \pm 1$ , or  $l = \pm 1$  in (2.38). From (2.40) these reciprocal lattice vectors are given by

$$\mathbf{K} = \frac{2\pi}{a} (-\hat{x} + \hat{y} + \hat{z}) \quad (2.41)$$

for  $h = \pm 1$ , and so on. Since the magnitude of all these  $\mathbf{K}$  vectors is the same,  $(2\pi/a)\sqrt{3}$ , the energy level is

$$\mathcal{E}(0) = \frac{\hbar^2}{2m} \left( \frac{2\sqrt{3}\pi}{a} \right)^2 \quad (2.42)$$

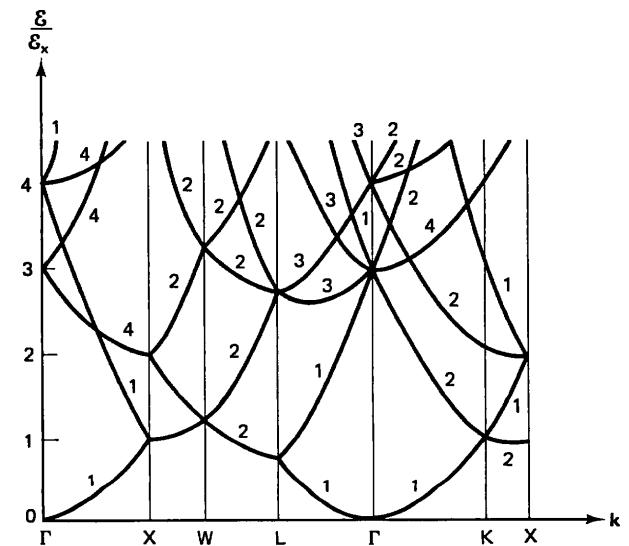
From (2.35) there are eight wavefunctions that have this energy at  $\mathbf{k} = 0$ , so this point is eightfold degenerate. The next highest level above (2.42) occurs for  $h = k = l = 1$ ,  $l = 0$  and so on, or for

$$\mathbf{K} = \frac{2\pi}{a} (2\hat{z}) \quad (2.43)$$

and so on. The energy of this point is sixfold degenerate and has the value

$$\mathcal{E}(0) = \frac{\hbar^2}{2m} \left( \frac{4\pi}{a} \right)^2 \quad (2.44)$$

By generating a sufficient number of energy levels at the  $\Gamma$ -point and following them out to the desired high-symmetry points on the zone boundary with (2.37), we obtain the empty lattice energy bands for the face-centered cubic lattice shown in Fig. 2.6. As can be seen, the free-electron energy bands are fairly complicated, with multiple degeneracy at high-symmetry points, and the electrons can range through all energy values. When we apply a periodic potential to the one-electron Schrödinger equation, however, some of the degeneracy at the zone boundaries will be removed and the electrons will be constrained to certain energy values. These effects are demonstrated in the nearly free electron model.



**Figure 2.6** Empty lattice energy bands for the face-centered cubic direct lattice. The numbers indicate the degeneracy of each band.  $\mathcal{E}_x = (\hbar^2/2m)(2\pi/a)^2$  is the energy at the  $X$ -points. [After F. Herman in *An Atomistic Approach to the Nature and Properties of Materials*, ed. J. A. Pask (New York: Wiley, 1967).]

### 2.3 NEARLY FREE ELECTRON MODEL

When the kinetic energy of the electrons is large compared to the periodic energy of the lattice, the behavior of the electrons can be approximated by nearly free electron wavefunctions. Although often considered a pedagogical exercise, with the advent of pseudopotential theory [J. C. Phillips and L. Kleinman, *Phys. Rev.* **116**, 287 (1959)] this method has become important in the calculation of semiconductor energy bands. In the nearly free electron model, gaps occur in allowed electron energy values for the following reasons. Traveling electron waves reflected from adjacent atoms interfere constructively to produce standing waves. Some of these standing waves pile up charge at the atomic sites, where their energy is lowered, while other standing waves pile up charge between atomic sites, where their energy is increased over the free-electron values. This shift in energy between the standing-wave states produces an *energy gap*.

Let us consider a Bloch electron with a wavefunction given by (2.10) and (2.11). From (1.9) the periodic part of the wavefunction can be expanded in a Fourier series with reciprocal lattice vector,  $\mathbf{K}$ , as index:

$$u_k(\mathbf{r}) = \sum_{\mathbf{K}} A_{\mathbf{K}} \exp(i\mathbf{K} \cdot \mathbf{r}) \quad (2.45)$$

where

$$A_{\mathbf{K}} = \frac{1}{\Omega} \int_{\Omega} u_k(\mathbf{r}) \exp(-i\mathbf{K} \cdot \mathbf{r}) d\mathbf{r} \quad (2.46)$$

The wavefunction then has the form

$$\psi_k(\mathbf{r}) = \sum_{\mathbf{K}} A_{\mathbf{K}} \exp[i(\mathbf{k} + \mathbf{K}) \cdot \mathbf{r}] \quad (2.47)$$

The one-electron Schrödinger equation (2.1) is then

$$\left[ -\frac{\hbar^2}{2m} \nabla^2 + U(\mathbf{r}) - \varepsilon \right] \sum_{\mathbf{K}} A_{\mathbf{K}} \exp[i(\mathbf{k} + \mathbf{K}) \cdot \mathbf{r}] = 0 \quad (2.48)$$

or

$$\sum_{\mathbf{K}} \left[ \frac{\hbar^2}{2m} (\mathbf{k} + \mathbf{K})^2 + U(\mathbf{r}) - \varepsilon \right] A_{\mathbf{K}} \exp[i(\mathbf{k} + \mathbf{K}) \cdot \mathbf{r}] = 0 \quad (2.49)$$

Multiplying on the left by  $(1/\Omega) \exp[-i(\mathbf{k} + \mathbf{K}') \cdot \mathbf{r}]$ , where  $\mathbf{K}'$  is another reciprocal lattice vector and integrating over a primitive unit cell, (2.49) becomes

$$\frac{1}{\Omega} \sum_{\mathbf{K}} A_{\mathbf{K}} \int_{\Omega} \left[ \frac{\hbar^2}{2m} (\mathbf{k} + \mathbf{K})^2 + U(\mathbf{r}) - \varepsilon \right] \exp[i(\mathbf{k} - \mathbf{K}') \cdot \mathbf{r}] d\mathbf{r} = 0 \quad (2.50)$$

But

$$\frac{1}{\Omega} \int_{\Omega} \exp[i(\mathbf{k} - \mathbf{K}') \cdot \mathbf{r}] d\mathbf{r} = \delta_{\mathbf{k}, \mathbf{K}'} \quad (2.51)$$

where the Kronecker delta is

$$\delta_{\mathbf{k}, \mathbf{K}'} = \begin{cases} 1 & \text{for } \mathbf{k} = \mathbf{K}' \\ 0 & \text{for } \mathbf{k} \neq \mathbf{K}' \end{cases} \quad (2.52)$$

We now have

$$\begin{aligned} \sum_{\mathbf{K}} A_{\mathbf{K}} \left[ \frac{\hbar^2}{2m} (\mathbf{k} + \mathbf{K})^2 - \varepsilon \right] \delta_{\mathbf{k}, \mathbf{K}'} \\ = -\frac{1}{\Omega} \sum_{\mathbf{K}} A_{\mathbf{K}} \int_{\Omega} U(\mathbf{r}) \exp[i(\mathbf{k} - \mathbf{K}') \cdot \mathbf{r}] d\mathbf{r} \end{aligned} \quad (2.53)$$

or

$$\begin{aligned} A_{\mathbf{K}'} \left[ \varepsilon - \frac{\hbar^2}{2m} (\mathbf{k} + \mathbf{K}')^2 \right] \\ = \frac{1}{\Omega} \sum_{\mathbf{K}} A_{\mathbf{K}} \int_{\Omega} U(\mathbf{r}) \exp[i(\mathbf{k} - \mathbf{K}') \cdot \mathbf{r}] d\mathbf{r} \end{aligned} \quad (2.54)$$

Since from (2.9)  $U(\mathbf{r})$  is periodic in  $\mathbf{R}$ , we can expand it in a Fourier series with reciprocal lattice vector,  $\mathbf{K}''$ , as index:

$$U(\mathbf{r}) = \sum_{\mathbf{K}''} B_{\mathbf{K}''} \exp(i\mathbf{K}'' \cdot \mathbf{r}) \quad (2.55)$$

where

$$B_{\mathbf{K}''} = \frac{1}{\Omega} \int_{\Omega} U(\mathbf{r}) \exp(-i\mathbf{K}'' \cdot \mathbf{r}) d\mathbf{r} \quad (2.56)$$

Using (2.55) in (2.54), we obtain

$$\begin{aligned} A_{\mathbf{K}'} \left[ \varepsilon - \frac{\hbar^2}{2m} (\mathbf{k} + \mathbf{K}')^2 \right] \\ = \frac{1}{\Omega} \sum_{\mathbf{K}} \sum_{\mathbf{K}''} A_{\mathbf{K}} B_{\mathbf{K}''} \int_{\Omega} \exp[i(\mathbf{k} - \mathbf{K}' + \mathbf{K}'') \cdot \mathbf{r}] d\mathbf{r} \end{aligned} \quad (2.57)$$

But just as in (2.51),

$$\frac{1}{\Omega} \int_{\Omega} \exp[i(\mathbf{k} - \mathbf{K}' + \mathbf{K}'') \cdot \mathbf{r}] d\mathbf{r} = \delta_{\mathbf{k}, \mathbf{K}' - \mathbf{K}''} \quad (2.58)$$

where

$$\delta_{K,K'-K''} = \begin{cases} 1 & \text{for } K = K' - K'' \\ 0 & \text{for } K \neq K' - K'' \end{cases} \quad (2.59)$$

Thus (2.57) becomes

$$\begin{aligned} A_{K'} \left[ \mathcal{E} - \frac{\hbar^2}{2m} (k + K')^2 \right] &= \sum_{K''} A_K B_{K''} \delta_{K,K'-K''} \\ &= \sum_{K''} A_{K'-K''} B_{K''} \end{aligned} \quad (2.60)$$

which, with a change of notation, gives us

$$A_K \left[ \mathcal{E} - \frac{\hbar^2}{2m} (k + K)^2 \right] = \sum_{K'} A_{K-K'} B_{K'} \quad (2.61)$$

Equation (2.61) is an exact expression relating the Fourier coefficients  $A_K$  for the expansion of the Bloch function  $u(r)$  to the Fourier coefficients  $B_{K'}$  for the expansion of the effective one-electron lattice potential energy  $U(r)$ . Formally, we could now solve this problem by taking self-consistent expressions for  $U(r)$  and  $u(r)$  and determining the coefficients  $A_K$  and  $B_{K'}$  from (2.46), (2.56), and (2.61) in an iterative fashion. We can obtain approximate analytical solutions, however, by making certain assumptions regarding the relative sizes of the various coefficients. The simplest assumption to make is that only the terms for  $K$  and  $K'$  equal to zero are important. This assumption just leads to free-electron solutions.

The next simplest assumption is that only the terms for  $K$  and  $K'$  equal to zero and  $K$  equal to  $K'$  are important. Under this assumption we obtain two equations for the coefficients from (2.61):

$$\begin{aligned} A_0 \left[ \mathcal{E} - \frac{\hbar^2}{2m} k^2 \right] &= A_0 B_0 + A_{-K} B_K & \text{for } K = 0 \\ A_K \left[ \mathcal{E} - \frac{\hbar^2}{2m} (k + K)^2 \right] &= A_0 B_K + A_K B_0 & \text{for } K = K' \end{aligned} \quad (2.62)$$

For simplicity we have ignored the degenerate  $-K'$  term in each of these equations. If you wish to include them, simply replace every  $B_K$  in the rest of the analysis by  $2B_K$ . Since from (2.46) and (2.56)  $A_{-K}B_K = A_KB_{-K}$ , we can put this coupled set of equations in the matrix form

$$\begin{bmatrix} \mathcal{E} - \mathcal{E}_0 & -B_{-K} \\ -B_K & \mathcal{E} - \mathcal{E}_K \end{bmatrix} \begin{bmatrix} A_0 \\ A_K \end{bmatrix} = 0 \quad (2.63)$$

where

$$\mathcal{E}_0 = \frac{\hbar^2 k^2}{2m} + B_0 \quad \text{and} \quad \mathcal{E}_K = \frac{\hbar^2}{2m} (k + K)^2 + B_0 \quad (2.64)$$

Note that  $\mathcal{E}_0$  and  $\mathcal{E}_K$  have the form of free-electron energies in a constant potential energy  $B_0$  and correspond to the first two terms in the wavefunction (2.47),

$$\psi_k(r) = A_0 \exp(i\mathbf{k}\cdot\mathbf{r}) + A_K \exp[i(\mathbf{k} + \mathbf{K})\cdot\mathbf{r}] \quad (2.65)$$

To keep the coefficients of the wavefunction expansion,  $A_0$  and  $A_K$ , from vanishing identically, the determinant of the matrix in (2.63) must be zero. Since  $B_{-K}B_K = B_K^*B_K = B_K^2$  for crystals with inversion symmetry, we obtain a quadratic equation for the energy of the electrons,

$$\mathcal{E}^2 - (\mathcal{E}_0 + \mathcal{E}_K)\mathcal{E} + (\mathcal{E}_0\mathcal{E}_K - B_K^2) = 0 \quad (2.66)$$

which has the solutions

$$\mathcal{E}^\pm = \frac{1}{2}(\mathcal{E}_0 + \mathcal{E}_K) \pm \frac{1}{2}[(\mathcal{E}_0 - \mathcal{E}_K)^2 + 4B_K^2]^{1/2} \quad (2.67)$$

By comparing (2.67) with (2.64) and (2.65), we see that the states with energy  $\mathcal{E}_0$  and  $\mathcal{E}_K$  are combined into two states  $\psi^+$  and  $\psi^-$  with energies  $\mathcal{E}^+$  and  $\mathcal{E}^-$  by the energy perturbation  $B_K$ .

We can examine these nearly free electron bands most simply for  $\mathbf{k}$  in the direction of  $\mathbf{K}$ . The lowest-lying band is obtained for  $\mathbf{K} = 0$  in (2.67) and looking at  $\mathcal{E}^-$  for  $\mathbf{k}$  near zero. This gives us

$$\mathcal{E}^- = \frac{\hbar^2 \mathbf{k}^2}{2m} \quad (2.68)$$

so the  $\mathcal{E}^-$  versus  $\mathbf{k}$  is parabolic near  $\mathbf{k} = 0$ . At the zone boundary,  $\mathbf{k} = \frac{1}{2}\mathbf{K}$ , and for  $\mathbf{K} = -\mathbf{K}$ , we have, from (2.67),

$$\mathcal{E}^\pm = \mathcal{E}_0 \pm |B_K| \quad (2.69)$$

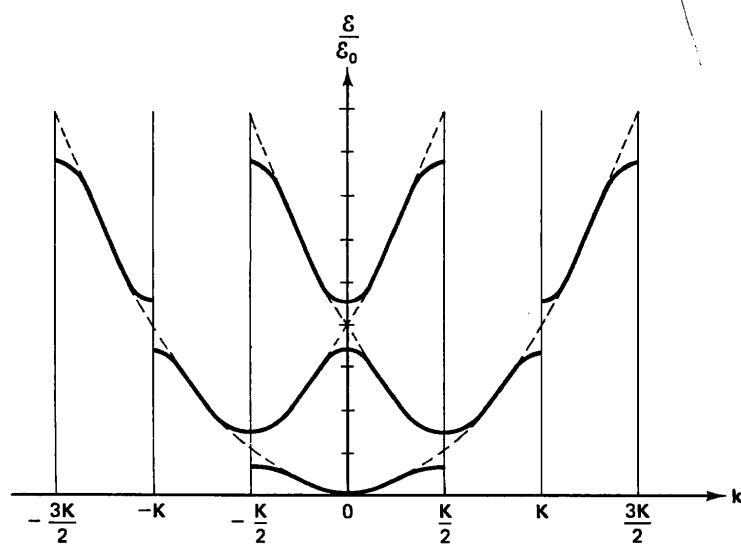
Thus we see that at the zone boundary the energy of the  $\psi^-$  state is lower than the free-electron energy by  $|B_K|$ , and the energy of the  $\psi^+$  state is higher by  $|B_K|$ . It is evident, then, that the periodic lattice potential has created an *energy gap* of magnitude  $2|B_K|$  between these two states. In Fig. 2.7 the higher-energy bands are shown inside the first Brillouin zone and also outside. The bands outside are referred to as an extended zone scheme. They tend to show more clearly how the nearly free electron bands are perturbed from the free-electron parabola with discontinuities at the zone boundaries. Note that because of this perturbation,  $K$  is no longer a band index.

It is interesting to examine the wavefunctions that correspond to these energy bands. From (2.65) the wavefunctions at the zone boundary are

$$\psi(r) = A_0 \exp(i\frac{1}{2}\mathbf{k}\cdot\mathbf{r}) + A_K \exp(-i\frac{1}{2}\mathbf{k}\cdot\mathbf{r}) \quad (2.70)$$

The ratio of the coefficients can be obtained from (2.62) and (2.64) as

$$\frac{A_K}{A_0} = \frac{B_K}{\mathcal{E} - \mathcal{E}_K} \quad (2.71)$$



**Figure 2.7** Nearly free electron energy bands for  $k$  in the direction of  $K$ . The dashed line shows the parabolic free-electron  $\mathcal{E}$  versus  $k$  dependence. The higher-lying bands outside the first Brillouin zone is referred to as an extended zone scheme;  $\mathcal{E}_0 = (\hbar^2/2m)(\frac{1}{2}K)^2 + B_0$ .

or using (2.67),

$$\frac{A_K}{A_0} = \frac{2B_K}{(\mathcal{E}_0 - \mathcal{E}_K) \pm [(\mathcal{E}_0 - \mathcal{E}_K)^2 + 4B_K^2]^{1/2}} \quad (2.72)$$

At the zone boundary  $\mathcal{E}_0 = \mathcal{E}_K$ , and

$$\frac{A_K}{A_0} = \mp \frac{B_K}{|B_K|} \quad (2.73)$$

Let us assume that  $B_K$  is negative. This corresponds to a periodic potential that is negative in the neighborhood of each atom and thus tends to attract electrons. Then

$$\begin{aligned} \psi^-(r) &= A_0[\exp(i\frac{1}{2}\mathbf{K}\cdot\mathbf{r}) + \exp(-i\frac{1}{2}\mathbf{K}\cdot\mathbf{r})] \\ &= 2A_0 \cos(\frac{1}{2}\mathbf{K}\cdot\mathbf{r}) \end{aligned} \quad (2.74)$$

and

$$\psi^+(r) = 2A_0 i \sin(\frac{1}{2}\mathbf{K}\cdot\mathbf{r}) \quad (2.75)$$

Therefore, we find that the periodic potential has converted the traveling electron plane waves of (2.65) into the standing waves of (2.74) and (2.75) at the zone boundaries. This corresponds to Bragg reflection of the electrons.

If we determine the electron charge density for these two standing

waves, as indicated in Fig. 2.8, we find that the charge for  $\psi^-$  is concentrated near the sites or atoms in the direct lattice, while the charge for  $\psi^+$  is concentrated between atoms. We conclude, then, that the energy gap is caused by the potential well ( $B_K$  negative) at the atoms, which attracts the charge of  $\psi^-$  (lowers its energy) and repulses the charge of  $\psi^+$  (raises its energy). Because of this position dependence, the  $\psi^-$  state is said to be "s-like," in analogy to atomic s-levels, which do not vanish at the ion, and the  $\psi^+$  state is referred to as "p-like," since its charge vanishes at the atoms.

With the nearly free electron model we have developed the concept of allowed and "forbidden" energy bands. In a relatively simple manner, two other very important properties of electrons in a periodic potential can also be obtained from this model. Since  $\mathcal{E}_0$  is about equal to  $\mathcal{E}_K$  near the zone boundary, we can approximate the term in square brackets in (2.67) by

$$(1 + x)^{1/2} \approx 1 + \frac{1}{2}x$$

or

$$\mathcal{E}^\pm = \frac{1}{2}(\mathcal{E}_0 + \mathcal{E}_K) \pm |B_K| \left[ 1 + \frac{1}{2} \left( \frac{\mathcal{E}_0 - \mathcal{E}_K}{2|B_K|} \right)^2 \right] \quad (2.76)$$

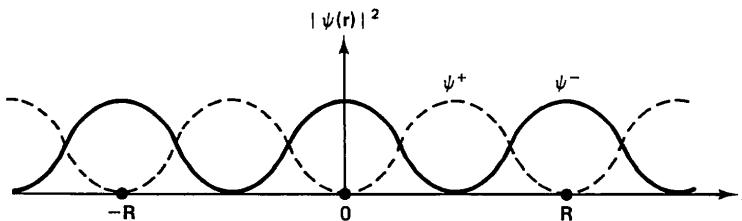
Using (2.64) and  $\mathbf{k} = \pm \frac{1}{2}\mathbf{K}$  near a boundary, (2.76) can be put in the form

$$\mathcal{E}^\pm \approx B_0 \pm |B_K| + \frac{\hbar^2 \mathbf{k}^2}{2m^*} \quad (2.77)$$

where

$$m^* = \frac{m}{1 \pm \hbar^2 \mathbf{K}^2 / m |B_K|} \quad (2.78)$$

Equations (2.77) and (2.78) show us that the electron energies near the zone boundaries have an approximately parabolic dependence on wave-vector in a manner similar to free electrons. The Bloch electrons, however, behave as though they have mass different from the free-electron mass. This behavior is characterized by an *effective mass*,  $m^*$ , as given in (2.78). Equa-



**Figure 2.8** Distribution of electron charge density in direct space for  $r$  in the direction of  $R$ , the direct lattice vectors, for values of  $k$  at the zone boundaries in reciprocal space.

tion (2.78) also shows us that the effective mass for  $\psi^-$  states near the top of the lowest-lying band in Fig. 2.7 is *negative*, while the effective mass for  $\psi^+$  states near the bottom of the next-highest band is *positive*.

The concept of an electron in a periodic potential energy having an effective mass different from its free mass should not be surprising. We found in Section 2.1 that the wavevector,  $\mathbf{k}$ , is not proportional to the real electron momentum, which takes into account the response of the electron to external forces and the internal forces due to the periodic potential. It reflects only the response of the electron to external forces. The response of the electron to the periodic potential energy of the crystal is accounted for with the effective mass,  $m^*$ , which can be greater or smaller than the free mass,  $m$ . A *negative* effective mass for an electron simply means that in the process of increasing its  $\mathbf{k}$  vector under the influence of an external force, the momentum transfer from the crystal to the electron is larger than and in the opposite direction to the applied external force. We consider the effects of external forces in more detail in Chapter 5.

In the discussion above we took into account only the periodic potential due to the direct Bravais lattice or, more accurately, a crystal structure with one atom located at each site of the Bravais lattice. An interesting effect is obtained in a crystal with a basis of more than one atom per unit cell. Let us assume that we can approximate the periodic potential in (2.55) by taking the sum of potentials coming from each atom in the unit cell and each unit cell in the crystal structure. That is, let

$$U(\mathbf{r}) = \sum_{R,d} U_d(\mathbf{r} - \mathbf{R} - \mathbf{d}) \quad (2.79)$$

where  $U_d$  represents the atomic potentials with the index  $d$  allowing for different atoms in the unit cell. If we substitute (2.79) into (2.56), we obtain for the Fourier coefficients of  $U(r)$ ,

$$\begin{aligned} B_K &= \frac{1}{\Omega} \int_{\Omega} \sum_{R,d} U_d(\mathbf{r} - \mathbf{R} - \mathbf{d}) \exp(-i\mathbf{K}\cdot\mathbf{r}) d\mathbf{r} \\ &= \frac{1}{\Omega} \sum_d \int_V U_d(\mathbf{r} - \mathbf{d}) \exp(-i\mathbf{K}\cdot\mathbf{r}) d\mathbf{r} \end{aligned} \quad (2.80)$$

where  $V$  is the volume of the whole crystal. Making the substitution  $\mathbf{r} - \mathbf{d} = \mathbf{r}'$ , we have

$$B_K = \sum_d \exp(-i\mathbf{K}\cdot\mathbf{d}) \frac{N}{\Omega} \int_{\Omega} U_d(\mathbf{r}') \exp(-i\mathbf{K}\cdot\mathbf{r}') d\mathbf{r}' \quad (2.81)$$

$$= \sum_d B_{Kd} \exp(-i\mathbf{K}\cdot\mathbf{d}) \quad (2.82)$$

where  $B_{Kd}$  are the Fourier coefficients of  $U_d$ .

The exponential in (2.82) has a simple physical interpretation: It shows

the relative phase of the traveling electron waves reflected from the various atoms in the unit cell. Depending on the basis and the direction of the electron wave, these reflections can add constructively or destructively, changing the values of  $B_K$  or possibly reducing it to zero. From (2.69),  $B_K$ , of course determines the gap of forbidden energies between allowed energy bands. Thus the addition of a basis to the nearly free electron model can significantly alter the energy gap.

## 2.4 TIGHTLY BOUND ELECTRON MODEL

In the nearly free electron model it was assumed that the kinetic energy of the electrons was large compared to the periodic potential energy due to the lattice. Under these conditions the wavefunctions were found to be plane waves with a perturbation from the lattice, and the allowed energy bands were large compared to the forbidden energy regions. In the tightly bound electron model this situation is reversed. The periodic potential energy due to the lattice is assumed to be large compared to the kinetic energy of the electrons, so that the electrons are largely bound to the atomic cores. In this situation we expect the electron wavefunctions to be more like atomic orbitals than plane waves. That is, we expect the wavefunction overlap between adjacent atoms in the lattice to be sufficiently small that the band structure will be closely related to the wavefunctions and discrete energies of electrons in isolated atoms.

In this manner we consider  $N$  isolated atoms with electron wavefunctions  $\psi_a$  and discrete energy levels  $\mathcal{E}_a$  located at the lattice sites  $\mathbf{R}$ . The one-electron Schrödinger equation for these atoms is

$$\mathbf{H}_a \psi_a(\mathbf{r} - \mathbf{R}) = \mathcal{E}_a \psi_a(\mathbf{r} - \mathbf{R}) \quad (2.83)$$

where

$$\mathbf{H}_a = -\frac{\hbar^2}{2m} \nabla^2 + U_a(\mathbf{r} - \mathbf{R}) \quad (2.84)$$

The Bloch electrons satisfy (2.1):

$$\mathbf{H} \psi_k(\mathbf{r}) = \mathcal{E} \psi_k(\mathbf{r}) \quad (2.85)$$

where

$$\mathbf{H} = -\frac{\hbar^2}{2m} \nabla^2 + U(\mathbf{r}) \quad (2.86)$$

and  $U(\mathbf{r})$  satisfies (2.9). We construct a wavefunction for the Bloch electron from a linear combination of atomic orbitals (L.C.A.O.).

$$\psi_k(\mathbf{r}) = \sum_R \exp(i\mathbf{k}\cdot\mathbf{R}) \psi_a(\mathbf{r} - \mathbf{R}) \quad (2.87)$$

which must satisfy Bloch's theorem (2.13). From (2.87) we have

$$\begin{aligned}\psi_k(\mathbf{r} + \mathbf{R}') &= \sum_{\mathbf{R}} \exp(i\mathbf{k} \cdot \mathbf{R}) \psi_a(\mathbf{r} + \mathbf{R}' - \mathbf{R}) \\ &= \exp(i\mathbf{k} \cdot \mathbf{R}') \sum_{\mathbf{R}} \exp[i\mathbf{k} \cdot (\mathbf{R} - \mathbf{R}')] \psi_a[\mathbf{r} - (\mathbf{R} - \mathbf{R}')] \\ &= \exp(i\mathbf{k} \cdot \mathbf{R}') \psi_k(\mathbf{r})\end{aligned}\quad (2.88)$$

which is Bloch's theorem.

The Bloch electron Hamiltonian (2.86) can be put in the form

$$\begin{aligned}\mathbf{H} &= -\frac{\hbar^2}{2m} \nabla^2 + U_a(\mathbf{r} - \mathbf{R}) + \Delta U(\mathbf{r} - \mathbf{R}) \\ &= \mathbf{H}_a + \Delta U(\mathbf{r} - \mathbf{R})\end{aligned}\quad (2.89)$$

where

$$\Delta U(\mathbf{r} - \mathbf{R}) \equiv U(\mathbf{r}) - U_a(\mathbf{r} - \mathbf{R}) \quad (2.90)$$

indicates the extent to which the periodic potential energy of the crystal deviates from the isolated atomic potential (Fig. 2.9). If the electrons are tightly bound, this deviation is small. Operating on (2.87) with (2.84), we find that

$$\begin{aligned}\mathbf{H}_a \psi_k(\mathbf{r}) &= \left[ -\frac{\hbar^2}{2m} \nabla^2 + U_a(\mathbf{r} - \mathbf{R}) \right] \sum_{\mathbf{R}} \exp(i\mathbf{k} \cdot \mathbf{R}) \psi_a(\mathbf{r} - \mathbf{R}) \\ &= \sum_{\mathbf{R}} \exp(i\mathbf{k} \cdot \mathbf{R}) \mathbf{H}_a \psi_a(\mathbf{r} - \mathbf{R}) \\ &= \mathcal{E}_a \psi_k(\mathbf{r})\end{aligned}\quad (2.91)$$

Using (2.89) in (2.85), we have

$$\mathcal{E} \psi_k(\mathbf{r}) = [\mathbf{H}_a + \Delta U(\mathbf{r} - \mathbf{R})] \psi_k(\mathbf{r}) \quad (2.92)$$

$$(\mathcal{E} - \mathcal{E}_a) \psi_k(\mathbf{r}) = \Delta U(\mathbf{r} - \mathbf{R}) \psi_k(\mathbf{r}) \quad (2.93)$$

Multiplying on the left by  $\psi_k^*(\mathbf{r})$  and integrating over the crystal volume yields

$$(\mathcal{E} - \mathcal{E}_a) \int_V \psi_k^*(\mathbf{r}) \psi_k(\mathbf{r}) d\mathbf{r} = \int_V \psi_k^*(\mathbf{r}) \Delta U(\mathbf{r} - \mathbf{R}) \psi_k(\mathbf{r}) d\mathbf{r} \quad (2.94)$$

Since the wavefunctions are normalized over the volume of the unit cell  $\Omega$ , the integral on the left-hand side of (2.94) is just equal to the number of atoms in the crystal,

$$\mathcal{E} - \mathcal{E}_a = \frac{1}{N} \int_V \psi_k^*(\mathbf{r}) \Delta U(\mathbf{r} - \mathbf{R}) \psi_k(\mathbf{r}) d\mathbf{r} \quad (2.95)$$

We now substitute the atomic wavefunctions of (2.87) for the Bloch

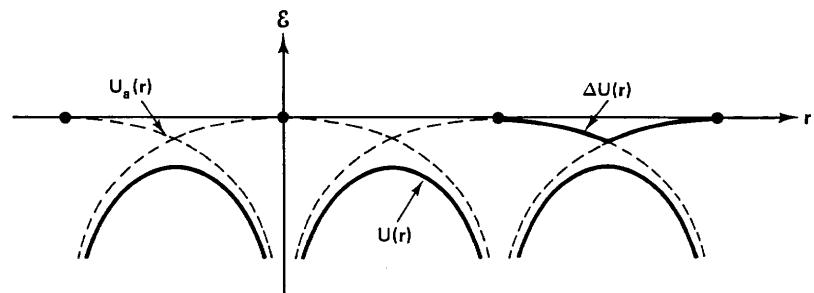


Figure 2.9 The periodic lattice potential energy, atomic potential energies, and their difference for  $\mathbf{r}$  in the direction of  $\mathbf{R}$ .

wavefunctions in (2.95), with the result

$$\mathcal{E} - \mathcal{E}_a = \frac{1}{N} \sum_{\mathbf{R}} \sum_{\mathbf{R}'} \exp[i\mathbf{k} \cdot (\mathbf{R} - \mathbf{R}')] \int_V \psi_a^*(\mathbf{r} - \mathbf{R}') \Delta U(\mathbf{r} - \mathbf{R}) \psi_a(\mathbf{r} - \mathbf{R}) d\mathbf{r} \quad (2.96)$$

Because of the periodicity of the lattice, we can set the origin at  $\mathbf{R}$  for each term in the summation over  $\mathbf{R}$  (set  $\mathbf{R} = 0$ ). It is then easy to see that all  $N$  terms in the summation over  $\mathbf{R}$  are identical and (2.96) becomes

$$\mathcal{E} - \mathcal{E}_a = \sum_{\mathbf{R}'} \exp(-i\mathbf{k} \cdot \mathbf{R}') \int_V \psi_a^*(\mathbf{r} - \mathbf{R}') \Delta U(\mathbf{r}) \psi_a(\mathbf{r}) d\mathbf{r} \quad (2.97)$$

If we take out the term for  $\mathbf{R}' = 0$ , we have

$$\mathcal{E} = \mathcal{E}_a - \alpha - \sum_{\mathbf{R}'} \beta_{\mathbf{R}'} \exp(i\mathbf{k} \cdot \mathbf{R}') \quad \text{for } \mathbf{R}' \neq 0 \quad (2.98)$$

where

$$\alpha \equiv - \int_V \psi_a^*(\mathbf{r}) \Delta U(\mathbf{r}) \psi_a(\mathbf{r}) d\mathbf{r} \quad (2.99)$$

$$\beta_{\mathbf{R}'} \equiv - \int_V \psi_a^*(\mathbf{r} - \mathbf{R}') \Delta U(\mathbf{r}) \psi_a(\mathbf{r}) d\mathbf{r} \quad (2.100)$$

and  $\mathbf{R}'$  is now a vector joining an atom at the origin to all other atoms in the crystal. Since the integrals in (2.99) and (2.100) are difficult to calculate, we will express our results in terms of  $\alpha$  and  $\beta_{\mathbf{R}'}$ .

In (2.98)  $\alpha$  is referred to as the Coulomb energy. It determines the shift in the atomic core levels,  $\mathcal{E}_a$ , caused by the interactions among atoms.  $\beta_{\mathbf{R}'}$  is referred to as the exchange energy and it determines the extent of broadening of the atomic levels into energy bands (Fig. 2.10). Since  $\Delta U(\mathbf{r})$  is negative and becomes larger with decreasing atomic spacing, both  $\alpha$  and  $\beta_{\mathbf{R}'}$  are positive and increase in magnitude with decreasing atomic spacing. Thus

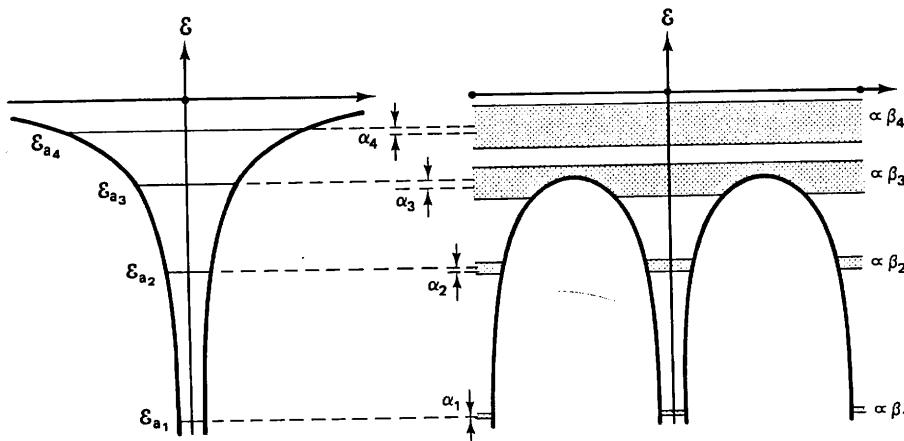


Figure 2.10 Diagram showing how energy bands are formed from the terms in (2.98). Each band is constructed from one atomic level.

these terms contain the overlap of atomic orbitals and account for the difference in potential energy between isolated atoms and a crystal of interacting atoms. The series in (2.98) allows for a summation of the overlap between an atom at the origin and all other atoms in the crystal. From Fig. 2.9, however, it can be seen that most of  $\Delta U(\mathbf{r})$  can be accounted for by summing over nearest-neighbor atoms only.

As an example, let us consider the tight-binding energy bands for a face-centered cubic direct lattice with a basis of one atom. For spherically-symmetric atomic wavefunctions,  $\beta_{R'} = \beta$  is the same for all nearest neighbors. Figure 2.11 shows that there are 12 nearest neighbors in this structure located at

$$\mathbf{R} = \frac{a}{2} (\pm \hat{x} \pm \hat{y}), \quad \frac{a}{2} (\pm \hat{y} \pm \hat{z}), \quad \frac{a}{2} (\pm \hat{x} \pm \hat{z}) \quad (2.101)$$

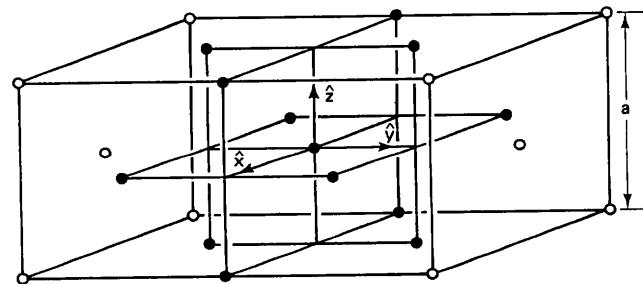


Figure 2.11 The 12 nearest-neighbor atoms (solid spheres) in face-centered cubic structure.

For  $\mathbf{k}$  in the form

$$\mathbf{k} = k_x \hat{x} + k_y \hat{y} + k_z \hat{z} \quad (2.102)$$

(2.98) becomes

$$\mathcal{E} = \mathcal{E}_a - \alpha - 4\beta [\cos \frac{1}{2}k_x a \cos \frac{1}{2}k_y a + \cos \frac{1}{2}k_y a \cos \frac{1}{2}k_z a + \cos \frac{1}{2}k_z a \cos \frac{1}{2}k_x a] \quad (2.103)$$

Equation (2.103) shows that the energy of a Bloch electron, derived from one atomic state, consists of some constant value and an expression that varies with wavefunction between well-defined limits. Thus we find that for every electron state in the free atom, there exists a band of energies in the crystal.

Figure 2.12 shows this tight-binding energy band for several high-symmetry directions in the first Brillouin zone. This figure can be compared directly to the lowest-lying empty lattice energy band in Fig. 2.6. It can be noted that the width of the band depends on  $\beta$  and varies for different directions of wavevector in a manner similar to the lowest-lying empty lattice band. This similarity shows the strong dependence of energy band structure on the crystal lattice. The number of nondegenerate electronic states in the band is equal to  $N$ , the number of atoms in the crystal, and each state can be occupied by two electrons of opposite spin. To obtain the complete energy band diagram in the tight-binding approximation, it is necessary to add additional atomic states to this picture, as indicated in Figs. 2.10 and 2.13.

Figure 2.13 illustrates conceptually how the complete energy band diagram for a crystal can be obtained from discrete atomic levels. We take  $N$

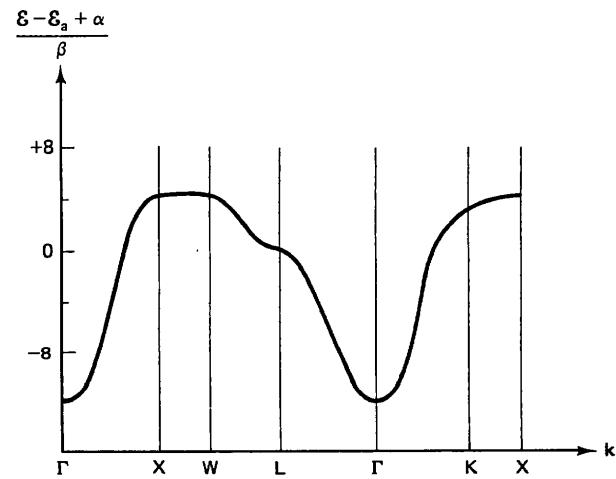
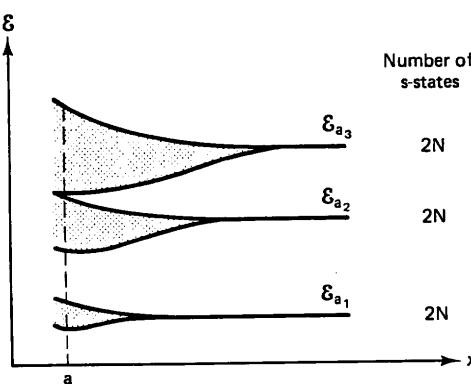


Figure 2.12 Tight-binding energy band, corresponding to one atomic state, for a face-centered cubic lattice with a basis of one atom.



**Figure 2.13** Diagram showing how energy bands in tightly bound electron model are formed from discrete atomic energy levels as the distance between atoms,  $x$ , is decreased to the lattice constant,  $a$ .

atoms and locate them on the sites of the appropriate Bravais lattice, except that the distance between atoms is so large that they do not interact. Each atomic level is then  $N$ -fold degenerate (excluding spin and angular momentum degeneracy). As we bring the atoms closer together, the higher-energy atomic states begin to interact, because of their large orbitals, and the  $N$  degenerate states split into a band of  $N$  discrete states. As the distance between atoms decreases further, lower-lying atomic levels begin to split into bands, and so on. In this manner, band states can be classified according to their atomic origin. That is, bands can be referred to as *s*-bands, *p*-bands, and so on. For higher-lying bands that overlap, however, this classification is not as obvious.

To take into account a basis in the tightly bound electron model, (2.87) for the Bloch wavefunctions can be changed to

$$\psi_k(\mathbf{r}) = \sum_{\mathbf{R}} \exp(i\mathbf{k} \cdot \mathbf{R}) \sum_d \psi_{ad}(\mathbf{r} - \mathbf{R} - \mathbf{d}) \quad (2.104)$$

If the molecular wavefunctions are known, however, the Bloch wavefunctions could be formed from a linear combination of *molecular* orbitals.

## 2.5 OTHER BAND MODELS

We have used the general models discussed up to now to demonstrate the existence of energy bands and to gain some qualitative understanding of their properties. These simple analytical models, however, are not sufficiently accurate to obtain results that can be compared with experiment. The nearly free electron model, for example, tends to overemphasize the plane wave aspects of the wavefunction, while the tightly bound electron model overemphasizes the atomic core aspects. Also, we were not able to obtain an analytical expression for the periodic potential.

To overcome some of these problems, other methods for calculating energy bands have been devised. Even with these more sophisticated models it is necessary to base some of the parameters, such as energy gap or effective mass, on experimental results. In addition, the calculation of semiconductor energy bands is sufficiently complex that numerical methods must be used to obtain reliable results. We will discuss some of these methods and look at the resulting energy bands for specific semiconductors.

### 2.5.1 Orthogonalized Plane Wave Method

The orthogonalized plane wave or OPW method [C. Herring, *Phys. Rev.* 57, 1169 (1940)] resolves some of the problems associated with the nearly free electron (plane wave) approximation. Near the atomic core, we expect the Bloch wavefunctions to differ substantially from plane waves, and many plane waves with large  $\mathbf{k}$  values are required in the series expansion of (2.47) to approximate its behavior. Thus (2.54) or (2.61) must be solved for a large number of coefficients to obtain accurate results. The OPW method helps to resolve this problem by constructing a wavefunction that behaves like a plane wave between atoms and approximates more closely an atomic wavefunction near the atomic core. The wavefunction consists of an atomic orbital part which is made orthogonal to a plane wave part. This wavefunction is used instead of (2.47), and fewer coefficients have to be considered in (2.54) or (2.61) to obtain reliable results.

### 2.5.2 Pseudopotential Method

Another technique used to improve the nearly free electron model is the pseudopotential method [J. C. Phillips and L. Kleinman, *Phys. Rev.* 116, 287 (1959)]. The concept of a pseudopotential arises in the use of OPW wavefunctions instead of the plane waves in (2.47). When these wavefunctions are operated on with the one-electron Hamiltonian, the orthogonalized atomic orbital part produces a term on the right-hand side of (2.54) in addition to that involving the real potential  $U(\mathbf{r})$ . The combination of the two terms defines an effective or pseudopotential for the valence electrons. An examination of these two terms individually shows that the term due to the real potential is negative, while the term due to the orthogonalization of the wavefunction is positive. This cancellation produces a net pseudopotential for the electrons which is weaker than the real potential. Physically, the reason for this is that the electron wavefunctions in the atomic core oscillate rapidly with a high kinetic energy. This kinetic energy acts as a repulsive potential in the attractive potential well of the core and tends to cancel part of the negative potential energy. In this manner, the behavior of the valence electrons is less sensitive to the form chosen for  $U(\mathbf{r})$ , such as the number

of terms in (2.55). If the pseudopotential is regarded as an adjustable parameter to be determined from experimental data, this method can produce reliable results for specific semiconductor energy bands.

### 2.5.3 Cellular Methods

Some of the other techniques used to calculate energy bands can be classified as **cellular methods** [E. P. Wigner and F. Seitz, *Phys. Rev.* 43, 804 (1933)]. The basic idea behind these methods is that due to the Bloch equation (2.13), it is only necessary to solve Schrödinger's equation for the wavefunction within a primitive unit cell (Wigner–Seitz unit cell) of the direct lattice. The wavefunction for any other cell can then be obtained from (2.13). The problem is that not every solution to Schrödinger's equation within a primitive cell is valid for the entire crystal. That is, the wavefunctions and their gradients must be continuous and approximate plane waves at the boundaries of the cell. These conditions can be met by the use of a periodic lattice potential which has the topology of a pan used for baking rolls. This so-called “muffin-tin potential” consists of an attractive atomic potential within a sphere of arbitrary radius (usually less than half the nearest-neighbor distance) around each lattice site and zero everywhere else.

Two methods often used for calculating energy bands with a muffin-tin potential are the augmented plane wave or APW method [J. C. Slater, *Phys. Rev.* 51, 846 (1937)] and the KKR method [J. Korringa, *Physica* 13, 392 (1947); W. Kohn and N. Rostoker, *Phys. Rev.* 94, 1411 (1954)]. In the APW method, a solution to Schrödinger's equation is obtained by first matching plane wave solutions outside the core radius to atomic solutions within the core. A wavefunction is then constructed from these augmented plane waves. Unfortunately, these augmented plane waves have a discontinuous gradient at the boundary between the core and flat potential regions, so that substantial calculation is required to obtain reliable results. The KKR method is essentially a Green's function approach to the solution of Schrödinger's equation with a muffin-tin potential. It has certain similarities to the APW method and the results from the two are in substantial agreement when the same potential is used in each. Both, however, are first-principles calculations which should be viewed with caution.

### 2.5.4 Spin-Orbit Coupling

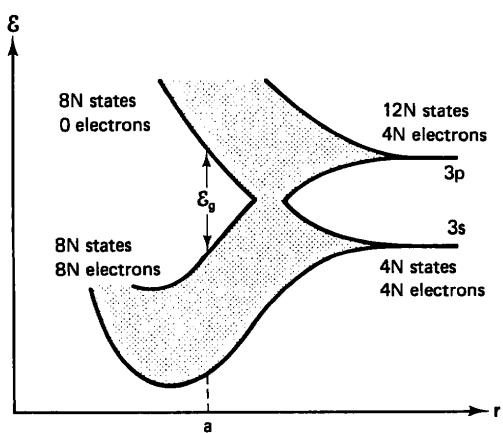
In the discussion of energy band calculations we have ignored any effects due to the electron spin operator,  $\mathbf{s}$ . We have simply assumed that two electrons of opposite spin can occupy any state defined by a Bloch wavefunction. In free atoms we know, however, that the electrons moving through the electric field of the ion core experience a potential that depends on the scalar product of the spin magnetic moment with the vector product

of its velocity and the electric field. This *spin-orbit coupling* tends to remove the degeneracy of states with the same wavefunction and opposite spin. We should then expect similar effects for electrons moving in the periodic potential of a lattice. The effects of spin-orbit coupling on energy bands can be determined by using a Hamiltonian that includes this potential and solving the one-electron Schrödinger equation with a perturbation technique known as the  $\mathbf{k}\cdot\mathbf{p}$  method [E. O. Kane, *J. Phys. Chem. Solids* 1, 83 (1956)]. The results indicate that even strong spin-orbit coupling will not remove the spin degeneracy for crystals with inversion symmetry, provided that a band is not degenerate with one of different wavefunction. Even in crystals that lack inversion symmetry, the spin-orbit splitting is small. For any crystal, however, if several bands with different wavefunction are degenerate, the spin-orbit splitting can be substantial.

## 2.6 VALENCE BANDS AND BONDS

Before we consider the details of semiconductor energy bands, it is worthwhile to pursue the relationship between energy bands and valence bonds. We learned in Section 2.1 that the first Brillouin zone for a direct Bravais lattice has  $N$  allowed values of  $\mathbf{k}$ , where  $N$  is the number of primitive unit cells in the direct lattice. Each allowed  $\mathbf{k}$  value corresponds to a wavefunction or electron state which can have two electrons of opposite spin. Including spin, we then have  $2N$  electron states per energy band or two electron states per unit cell per energy band for a Bravais lattice. If we now add a basis (atoms) to the lattice to obtain a crystal structure, we can count the number of electrons per unit cell and determine the occupancy of the energy bands.

As an example, let us consider the occupancy of energy bands in silicon. Silicon has the diamond crystal structure, which is a face-centered cubic Bravais lattice with a basis of two atoms per primitive unit cell. The atomic structure is  $(1s)^2(2s)^2(2p)^6(3s)^2(3p)^2$ . Since the first two shells are completely filled and tightly bound to the nucleus, we will consider only the third shell of valence electrons. Here the two 3s states (including spin states) are completely filled with two electrons, and the six 3p states have only two electrons. If we use a tight-binding approach to form the crystal from the atoms, we will need  $2N$  silicon atoms. As shown in Fig. 2.14, the resulting  $12N$  3p-states and the  $4N$  3s-states interact and cross over when the atoms are brought together to the appropriate lattice constant. The result for 0 K is  $8N$  lower-lying states occupied by  $8N$  electrons separated by an energy gap from  $8N$  higher-lying states with zero electrons. Since there are  $2N$  states per energy band, the crystal has three 3p bands and one 3s band (all spin degenerate) in both the low- and high-energy regions. The four bands appear to be only one in an  $E$  versus  $r$  diagram, as in Fig. 2.14.

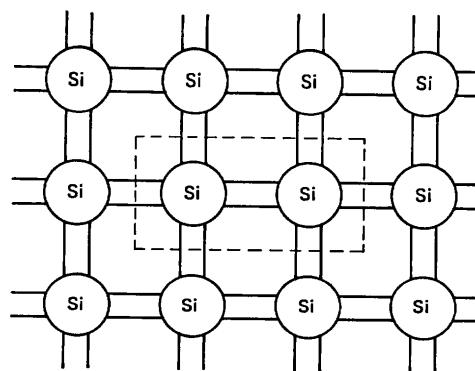


**Figure 2.14** The formation of energy bands in silicon from the two-highest occupied atomic levels. The  $8N$  available electrons completely fill the  $8N$  states in the lowest bands (valence bands), leaving the  $8N$  states in the highest bands (conduction bands) completely empty.

In molecules such as  $\text{CH}_4$  this combination of  $s$  and  $p$  states is known to give a tetrahedral bonding arrangement, called *covalent bonding*, where two electrons of opposite spin are shared between atoms. For the silicon crystal, however, there are  $8N$  electrons in the four lower-lying bands, or eight electrons per unit cell. This means that every silicon atom in the crystal is surrounded by eight outer electrons. From the crystal structure of silicon (Fig. 1.10), we see that every silicon atom has four nearest neighbors, so that each silicon atom must be covalently bonded to its nearest neighbors, with the  $8N$  electrons in the lowest-lying energy bands. For this reason these occupied bands are referred to as *valence bands*. The decrease in the energy of these bands over the atomic states, as shown in Fig. 2.14, reflects the binding or cohesive energy. In a similar manner the four higher-lying bands correspond to an antibonding molecular state and are referred to as *conduction bands*. It should be pointed out, however, that this analogy with covalently bonded molecules is not precise. The wavefunction overlap, which produces the energy bands in the crystal, demonstrates that the *valence electrons are not localized in the bonds*. There is, nevertheless, a high concentration of electron charge in the regions between atoms, which is equivalent to a covalent electron pair.

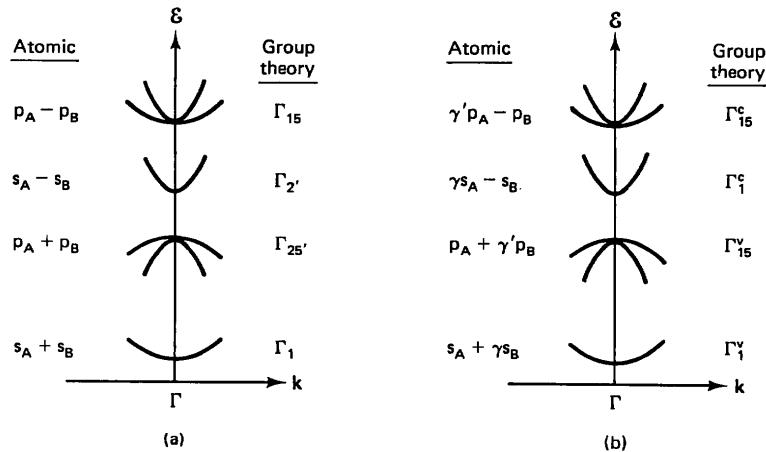
This covalent bonding arrangement can conveniently be illustrated by the two-dimensional schematic shown in Fig. 2.15. Each solid line between atoms represents one of the paired electrons in the covalent bond. Adding up the number of electrons in the Wigner–Seitz unit cell, we find a total of eight. These are the eight electrons that completely fill the four valence bands

**Figure 2.15** Two-dimensional schematic of covalent bonding arrangement in silicon. The dashed lines show the Wigner–Seitz unit cell.



in Fig. 2.14. The other column IV semiconductors can be represented in a similar fashion.

To illustrate the origin of the various bands, we can look at the  $E$  versus  $k$  curve near the  $\Gamma$ -point as in Fig. 2.16(a). The lowest-lying valence band, labeled  $\Gamma_1$ , is twofold (spin) degenerate and is derived from the bonding  $s$  orbitals for the two atoms in the unit cell. The other three valence bands ( $\Gamma_{25'}$ ) are derived from  $p$  orbitals and are sixfold degenerate at  $\Gamma$ . When spin orbit effects are included, these bands are split into two fourfold degenerate bands, labeled  $\Gamma_{25'}^{3/2}$ , and one lower-lying twofold degenerate band, labeled  $\Gamma_{25'}^{1/2}$ , at  $\Gamma$ . These three valence bands are also referred to as the heavy mass,



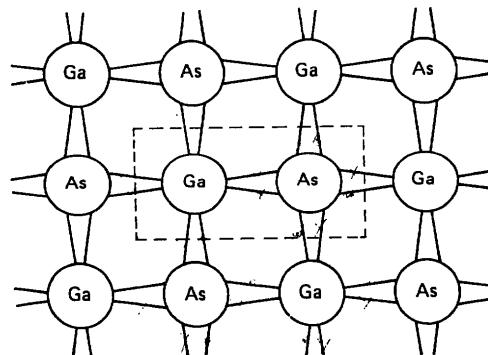
**Figure 2.16** Tight-binding prescription near  $\Gamma$  for the conduction and valence bands of crystals with (a) the diamond structure and (b) the sphalerite structure. The eight bands are labeled with both atomic and group theory notation. Spin-orbit coupling effects are neglected. A and B refer to the two atoms in the unit cell.

light mass, and split-off bands, in order of decreasing energy. As indicated in Fig. 2.16, the four conduction bands are derived from antibonding *s* and *p* orbitals, with properties that reflect these atomic states.

It is interesting to see what happens to the bonding arrangement when the column IV elements in the diamond structure are replaced by two different elements in the sphalerite structure. As an example, let us replace the two Ge atoms in a unit cell by one Ga (column III) and one As (column V) atom. These atoms have the same core as Ge, with an outer atomic configuration of  $(4s)^2(4p)$ ,  $(4s)^2(4p)^2$ , and  $(4s)^2(4p)^3$  for Ga, Ge, and As, respectively. In the sphalerite structure each As atom is surrounded by four Ga atoms, and vice versa. In the unit cell the Ga atom can contribute three electrons to the bonds, and the As atom can contribute five electrons, for a total of eight. This is the same number of electrons a unit cell of Ge has, so we can expect the same kind of bonding arrangement. If these electrons are distributed uniformly in the unit cell, however, the +5 charge of the As core is not quite neutralized by its surrounding average electron charge of -4, and the +3 charge of the Ga core is in a similar situation. We thus have a residual positive charge in the region of the As core, which must be neutralized. As indicated in Fig. 2.17, this is easily realized by having a greater electronic charge around the As atom than around a Ga atom. The unit cell, however, still has eight electrons.

If we proceed to the II–VI compounds with the sphalerite structure, this accumulation of electronic charge around the nonmetallic ions becomes even more pronounced and the bonding arrangement looks less and less covalent. The result of this transfer of electrons from the metallic to the nonmetallic ion is that the bonding becomes more ionic in nature. Thus compound semiconductors have mixed covalent and ionic bonding.

The degree of ionic bonding can be estimated by the difference in electronegativity between the atoms forming the compound: that is, the greater the electronegativity difference, the larger the ionic component of bonding. Electronegativity describes the ability of an atom to attract electrons. Elec-



**Figure 2.17** Two-dimensional schematic of bonding arrangement in GaAs. The tilt in the lines, which indicate the bonding electrons, is meant to show more charge around the As atoms than around the Ga atoms. The unit cell still has eight electrons.

**TABLE 2.1** Electronegativity Values for the Elements in Tetrahedral Environments<sup>a</sup>

Li	Be	B	C	N	O	F
1.0	1.5	2.0	2.5	3.0	3.5	4.0
Na	Mg	Al	Si	P	S	Cl
0.72 (0.9)	0.95 (1.2)	1.18 (1.5)	1.41 (1.8)	1.64 (2.1)	1.87 (2.5)	2.1 (3.0)
Cu	Zn	Ga	Ge	As	Se	Br
0.79 (1.9)	0.91 (1.6)	1.13 (1.6)	1.35 (1.8)	1.57 (2.0)	1.79 (2.4)	2.01 (2.8)
Ag	Cd	In	Sn	Sb	Te	I
0.57 (1.9)	0.83 (1.7)	0.99 (1.7)	1.15 (1.8)	1.31 (1.9)	1.47 (2.1)	1.63 (2.5)
Au	Hg	Tl	Pb	Bi		
0.64 (2.4)	0.79 (1.9)	0.94 (1.8)	1.09 (1.8)	1.24 (1.9)		

<sup>a</sup> The values in parentheses are from L. Pauling, *The Nature of the Chemical Bond* (Ithaca, N.Y.: Cornell University, 1960).

**TABLE 2.2** Ionic Component of Bonding in Percent for Selected Semiconductors

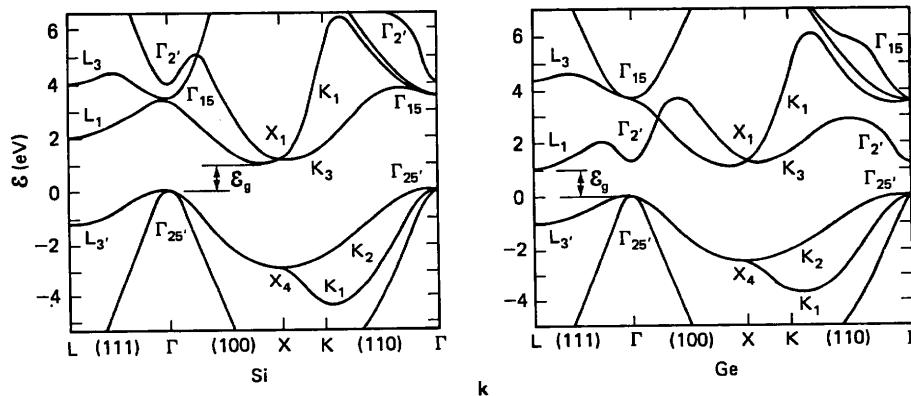
Si	0	GaAs	32	CdO	79	ZnO	62
Ge	0	GaSb	26	CdS	69	ZnS	62
SiC	18	InP	44	CdSe	70	ZnSe	63
		InAs	35	CdTe	67	ZnTe	61
		InSb	32				

Source: J. C. Phillips, *Phys. Rev. Lett.* 22, 705 (1969).

tronegativity values for the elements in tetrahedrally coordinated environments [J. C. Phillips, *Bands and Bonds in Semiconductors* (New York: Academic Press, 1973)] are shown in Table 2.1. In general, the farther the elements are removed from one another in the periodic table, the greater their electronegativity difference is and their ionic component of bonding in compound formation. Table 2.2 shows the extent of ionic bonding for a number of semiconductors, calculated on this basis.

## 2.7 ENERGY BAND STRUCTURES

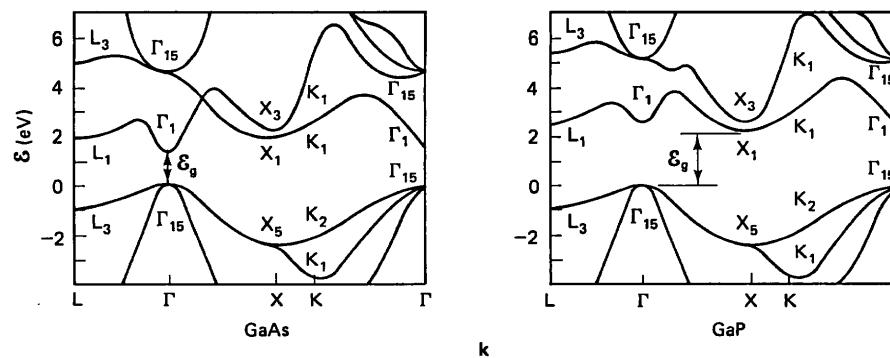
Figure 2.18 shows the results of an empirical pseudopotential calculation for the energy bands of the elemental semiconductors Si and Ge. The results of this calculation are in good agreement with experiments, except that spin-orbit effects are not included. These materials have the diamond crystal



**Figure 2.18** Empirical pseudopotential results for the energy bands of the elemental semiconductors Si and Ge. [After M. L. Cohen and T. K. Bergstresser, *Phys. Rev.* **141**, 789 (1966).]

structure. As can be seen, the  $\Gamma_{25'}$  valence bands for these two materials are quite similar, with the maximum at the  $\Gamma$ -point. The  $\Gamma_1$ -band is off-scale on the lower end of these figures. Although the valence bands are similar, there are significant differences in the conduction band structure. For many applications the most important energy band parameter is the energy gap,  $\epsilon_g$ , defined as the minimum separation between the conduction and valence bands. For Si the energy gap (1.17 eV at 0 K) is between the valence band maximum at  $\Gamma$  and the conduction band minimum at about 0.8X in the (100) direction. This energy gap is referred to as an *indirect bandgap*, since a change in  $\mathbf{k}$  value is required for an electron to make a transition from this conduction band minimum to the valence band maximum. Also, we note that there are six {100} directions in reciprocal space, so that the minimum at 0.8X in Si is actually *six equivalent minima*. For Ge the minimum in the conduction band is at  $L$ , on the outer surface of the first Brillouin zone. Thus it also has an indirect band gap of value 0.74 eV at 0 K. We can see from Fig. 2.5 that there are eight  $L$ -points (or eight {111} directions) in the first Brillouin zone. Since these points are shared with the next Brillouin zone, we find that Ge has *four equivalent minima* at  $L$  in the first Brillouin zone.

Most of the III-V compound semiconductors have the sphalerite crystal structure, so we expect the form of the energy bands to be similar to those for the diamond structure. The energy bands for GaAs and GaP, shown in Fig. 2.19, are fairly typical of this class of materials. As can be seen, the valence band structures are quite similar to Si and Ge. The minimum conduction band energy for GaP is at the  $X$ -point, so that GaP has an indirect bandgap with three equivalent conduction band minima. GaAs, however, has a minimum in the conduction band at  $\Gamma$ . Since no change in  $\mathbf{k}$  value is

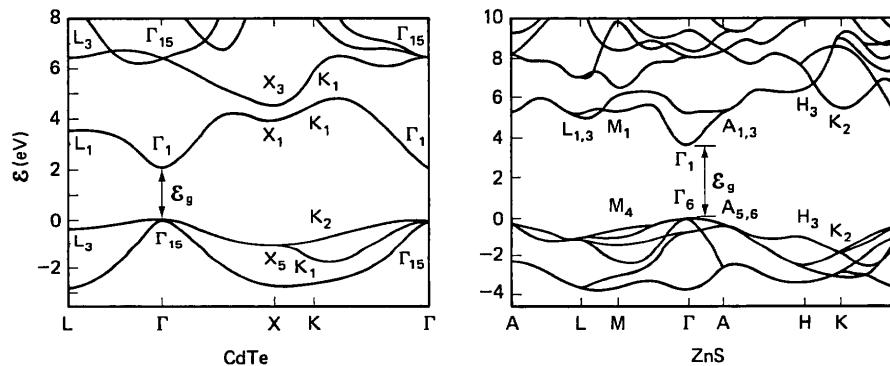


**Figure 2.19** Energy bands for the III-V compound semiconductors GaAs and GaP. [After M. L. Cohen and T. K. Bergstresser, *Phys. Rev.* **141**, 789 (1966).]

required for an electron to go from this conduction band to the maximum in the valence band, GaAs has a *direct bandgap*. Notice, however, that the minima at  $L_1$  and  $X_1$  are not far above  $\Gamma_1$ . This has important implication which will be discussed later.

The II-VI compound semiconductors crystallize into sphalerite or wurtzite structures. Figure 2.20 shows typical band structures for these materials. Sphalerite CdTe has a band structure similar to the III-V compound with a direct bandgap. Wurtzite ZnS is also shown. The band structure in this case is more complex. The bandgap, however, is direct, which is typical of II-VI compounds in either the sphalerite or wurtzite crystal structure.

The II-IV-V<sub>2</sub> compounds crystallize in the chalcopyrite structure (Fig. 1.13), which is very similar to the sphalerite structure. Since the primitive unit cell is four times as large as for sphalerite, the Brillouin zone shown in Fig. 1.17 is four times smaller. These ternary compounds have simple an-



**Figure 2.20** Energy bands for sphalerite CdTe and wurtzite ZnS. [After T. K. Bergstresser and M. L. Cohen, *Phys. Rev.* **164**, 1069 (1967).]

alogs in the III-V binary compounds, and their energy band structure can be obtained by mapping the analog III-V bands into the chalcopyrite Brillouin zone. Because of the 4-to-1 size ratio,  $\Gamma$ ,  $X$ , and  $W$  for sphalerite map into  $\Gamma$  for chalcopyrite;  $L$  and  $\Sigma$  map into  $N$ ; and  $X$  and  $L$  map into  $T$ . Results for  $ZnGeAs_2$ , which has GaAs as its analog, and  $ZnGeP_2$ , with GaP as its analog, are shown in Fig. 2.21.  $ZnGeAs_2$  has a direct energy gap at  $\Gamma$ , with conduction band minima  $\Gamma_1$  derived from  $\Gamma_1$  in GaAs (Fig. 2.19);  $\Gamma_3$  and  $T_1 + T_2$  from  $X_1$ ;  $\Gamma_2$  and  $T_5$  from  $X_3$ ; and  $N_1$  from  $L_1$ . The three degenerate  $\Gamma_{15}$  valence bands in GaAs form the one nondegenerate  $\Gamma_4$  and the two degenerate  $\Gamma_5$  valence bands in  $ZnGeAs_2$ . In  $ZnGeP_2$  the  $X_1$  minima, which cause GaP to have an indirect bandgap, map into  $\Gamma_3$  and  $T_1 + T_2$  minima. If  $\Gamma_3$  occurs at the lowest energy,  $ZnGeP_2$  would have a direct bandgap while its III-V analog, GaP, is indirect. Experiments indicate that this could be the situation for several II-IV-V<sub>2</sub> compounds with indirect gap III-V analogs.

The IV-VI compounds have the NaCl structure or an orthorhombic structure very close to that of NaCl. Figure 2.22 shows the energy bands for PbTe, which is fairly typical of these compounds. The maximum in the valence bands is at  $L_6^+$ , with the minimum in the conduction bands at  $L_6^-$ . Thus PbTe has a direct energy gap at  $L$  which is small, 0.18 eV, and both bands have four equivalent extrema. Small, direct energy gaps are typical of these materials. However, the ordering of the bands is not the same in all IV-VI compounds. For example, the minimum in the conduction band of SnTe is  $L_6^+$  and the maximum valence band is  $L_6^-$ , the reverse of

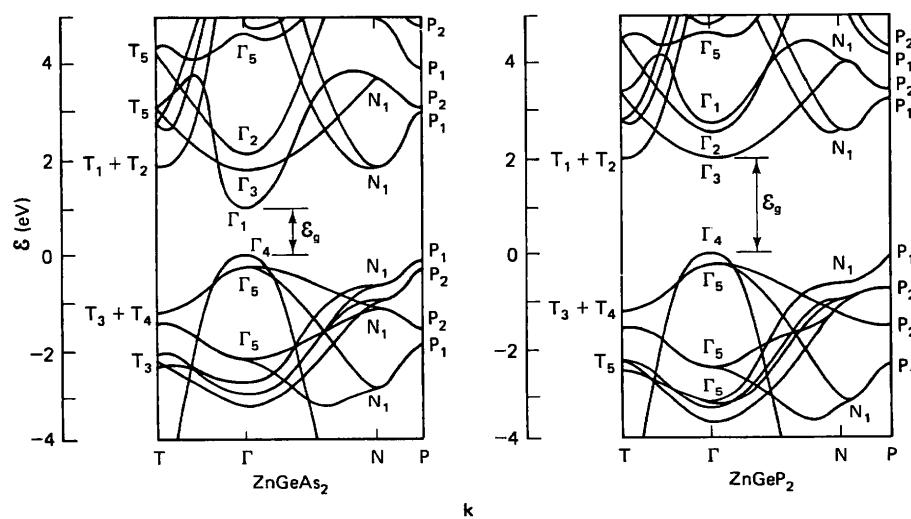


Figure 2.21 Energy bands for chalcopyrite  $ZnGeAs_2$  and  $ZnGeP_2$ . [After A. Shileika, *Surf. Sci.* 37, 730 (1973).]

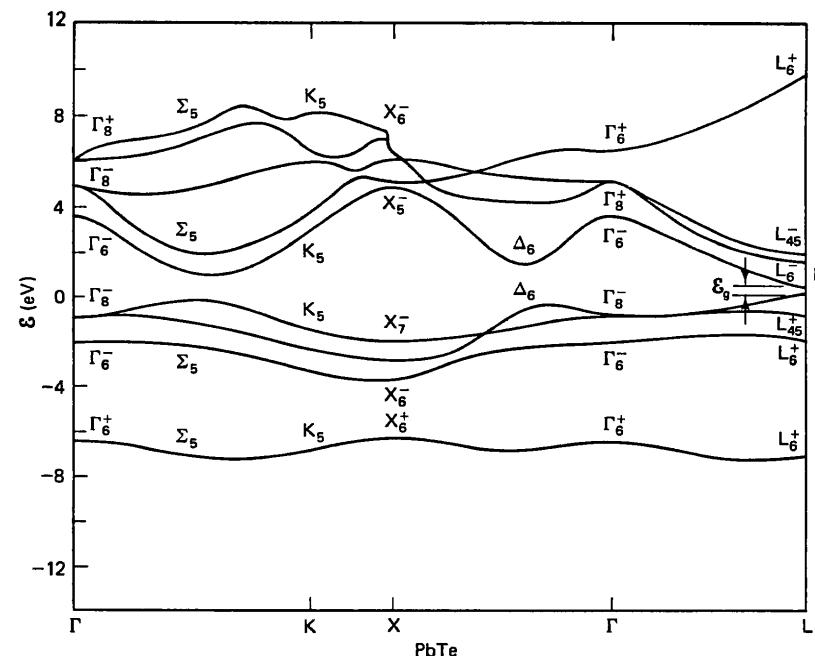


Figure 2.22 Energy bands for PbTe, which has the sodium chloride structure. [After Y. W. Tung and M. L. Cohen, *Phys. Rev.* 180, 823 (1969).]

PbTe [J. O. Dimmock et al., *Phys. Rev. Lett.* 16, 1193 (1966)]. For this reason, the energy gap of the alloy PbSnTe can approach zero at some intermediate composition [I. Melngailis, *J. Phys. Paris* 29, C4-84 (1968)].

In the energy band calculations above, spin-orbit effects are not included. When these effects are taken into account, some of the degeneracy of the highest-lying valence bands is removed, resulting in a lower split-off band. The energy separation between this split-off band and the degenerate light- and heavy-hole bands is referred to as the spin-orbit energy,  $\Delta$ , shown

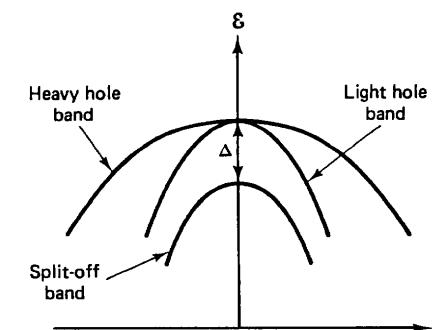


Figure 2.23 Heavy-hole, light-hole, and split-off valence bands.

TABLE 2.3 Spin-Orbit Splitting Energies

	$\Delta$ (eV)		$\Delta$ (eV)
C	0.006	InP	0.11
Si	0.044	InAs	0.38
Ge	0.29	InSb	0.82
$\alpha$ -Sn	0.80	ZnO	-0.005
AlN	0.012	ZnS	0.07
AlP	0.060	ZnSe	0.43
AlAs	0.29	ZnTe	0.93
AlSb	0.75	CdS	0.066
GaN	0.011	CdSe	0.42
GaP	0.127	CdTe	0.92
GaAs	0.34	HgS	0.13
GaSb	0.80	HgSe	0.48
InN	0.08	HgTe	0.99

in Fig. 2.23. Values of this parameter for a number of semiconductors are listed in Table 2.3.

## 2.8 ELECTRON DYNAMICS

There is an intimate relationship between band structure and the transport properties of electrons. By *transport properties* we mean the response of the electrons to applied forces. The proper approach to the problem of electronic motion in a periodic lattice would be to solve the time-dependent Schrödinger equation. A simpler solution, however, can be obtained, by first constructing a wave packet from plane wave solutions to the time-dependent Schrödinger equation. Then since it is well known that this wave packet behaves as a classical particle (correspondence principle), we can treat the problem semiclassically.

Let us then construct a wave packet (particle) from solutions to the time-dependent equation of the form

$$\psi(\mathbf{r}, t) = \psi(\mathbf{r}) \exp(-i\omega t) \quad (2.105)$$

where  $\psi(\mathbf{r})$  is given in (2.10). The group velocity of the wave packet (or the average velocity of the particle) is

$$\mathbf{v} = \nabla_{\mathbf{k}}\omega = \frac{\partial\omega}{\partial\mathbf{k}} \quad (2.106)$$

If we use the quantum scalar operator

$$\mathcal{E} = -\frac{\hbar}{i} \frac{\partial}{\partial t} \quad (2.107)$$

on the wave packet formed from (2.105), we obtain

$$\mathcal{E} = \hbar\omega \quad (2.108)$$

which is just Planck's relationship. With this, (2.106) becomes

$$\mathbf{v} = \frac{1}{\hbar} \nabla_{\mathbf{k}}\mathcal{E} = \frac{1}{\hbar} \frac{\partial\mathcal{E}}{\partial\mathbf{k}} \quad (2.109)$$

It should be pointed out that (2.109) is a rather surprising result. It tells us that an electron in a state specified by wavevector  $\mathbf{k}$  in a given energy band has a velocity determined by the slope of the  $\mathcal{E}$  versus  $\mathbf{k}$  curves. An examination of the  $\mathcal{E}$  versus  $\mathbf{k}$  diagrams in Section 2.7 shows that for most allowed values of  $\mathbf{k}$ , the electrons will have a finite velocity. Thus an electron moving in a perfect periodic potential with no applied forces has a constant velocity and is *not scattered* by the atoms of the crystal. In other words, a perfect crystal offers no resistance to the motion of an electron. Only crystal imperfections serve to scatter electrons. We examine these scattering processes in Chapter 6.

### 2.8.1 Effective Mass Concept

In the meantime, let us apply an external force  $\mathbf{F}$  to the wave packet. This force changes the energy of the electron according to

$$\mathbf{v} \cdot \mathbf{F} = \frac{d\mathcal{E}}{dt} = \frac{1}{\hbar} \nabla_{\mathbf{k}}\mathcal{E} \cdot \hbar \frac{d\mathbf{k}}{dt} \quad (2.110)$$

Then from (2.109),

$$\hbar \frac{d\mathbf{k}}{dt} = \mathbf{F} \quad (2.111)$$

This is an important relationship. It tells us that  $\hbar\mathbf{k}$  behaves as a momentum for external forces applied to a Bloch electron. It is for this reason that we defined a crystal momentum in (2.32). Let us now take the time derivative of the velocity in (2.109), to get

$$\frac{d\mathbf{v}}{dt} = \frac{1}{\hbar} \frac{d}{dt} \nabla_{\mathbf{k}}\mathcal{E} = \frac{1}{\hbar} \nabla_{\mathbf{k}} \left[ \nabla_{\mathbf{k}}\mathcal{E} \cdot \frac{d\mathbf{k}}{dt} \right] \quad (2.112)$$

Using (2.111), equation (2.112) becomes

$$\mathbf{a} = \frac{1}{\hbar^2} \nabla_{\mathbf{k}} [\nabla_{\mathbf{k}}\mathcal{E} \cdot \mathbf{F}] \quad (2.113)$$

where  $\mathbf{a}$  is the electron acceleration. This equation tells us that  $\mathbf{F}$  can produce a change in  $\mathbf{v}$  in directions other than the direction of  $\mathbf{F}$ . Comparing this to the Newtonian force equation, we see that the closest thing to a mass for

the electron is an inverse tensor which depends on the curvature of the  $\mathcal{E}$  versus  $\mathbf{k}$  diagrams. This can be determined by resolving (2.113) into components along three arbitrary axes. Then

$$a_i = \frac{1}{\hbar^2} \sum_j \frac{\partial^2 \mathcal{E}}{\partial k_i \partial k_j} F_j \quad \text{for } i, j = 1, 2, 3 \quad (2.114)$$

and we obtain an inverse effective mass tensor with components

$$\frac{1}{m_{ij}^*} = \frac{1}{\hbar^2} \frac{\partial^2 \mathcal{E}}{\partial k_i \partial k_j} \quad (2.115)$$

If we examine the curvature of some of the energy bands in Section 2.7, we see that at band minima the effective mass is positive and at band maxima the effective mass is negative. This is the same as the result that we found for the nearly free electron energy bands in Section 2.3. Another conclusion that results from (2.115) is that the effective mass becomes infinite at some point within a band. This merely indicates the point in the band where the externally applied force stops accelerating and begins decelerating the electron. From (2.115) we can also see why the  $\Gamma_{25'}$  and  $\Gamma_{15}$  valence bands in Figs. 2.18 and 2.19, respectively, are referred to as light and heavy bands.

Generally, the effective mass tensor can be diagonalized by a suitable choice of axes. If  $\mathbf{k}$  is given by (2.22), a general expression for a band extrema at  $\Gamma$  or  $\mathbf{k} = 0$  would be

$$\mathcal{E} = \mathcal{E}_0 \pm \sum_i A_i \mathbf{k}_i^2 \quad \text{for } i = 1, 2, 3 \quad (2.116)$$

where the + and - are for conduction and valence bands, respectively. Using (2.115) this becomes

$$\mathcal{E} = \mathcal{E}_0 \pm \frac{\hbar^2}{2} \mathbf{k} \cdot \mathbf{M} \cdot \mathbf{k} \quad (2.117)$$

where  $\mathbf{M}$  is the inverse effective mass tensor. For parabolic bands of spherical symmetry (2.117) reduces to

$$\mathcal{E} = \mathcal{E}_0 \pm \frac{\hbar^2 \mathbf{k}^2}{2m^*} \quad (2.118)$$

the equation for a sphere in three dimensions.

For cylindrical symmetry, equal-energy surfaces are ellipsoidal with the form

$$\mathcal{E} = \mathcal{E}_0 \pm \frac{\hbar^2}{2} \left( \frac{\mathbf{k}_1^2}{m_1^*} + \frac{\mathbf{k}_2^2 + \mathbf{k}_3^2}{m_2^*} \right) \quad (2.119)$$

For a band extrema at  $\mathbf{k} = \mathbf{k}_0$ , the equal-energy surfaces would have the form

$$\mathcal{E} = \mathcal{E}_0 \pm \frac{\hbar^2}{2} \sum_i \frac{(\mathbf{k}_i - \mathbf{k}_0)^2}{m_i^*} \quad (2.120)$$

and so on. Cubic crystals typically have extrema at  $\mathbf{k} = 0$  with spherical symmetry and extrema at  $\mathbf{k} = \mathbf{k}_0$  with cylindrical symmetry. Hexagonal crystals have extrema at  $\mathbf{k} = 0$  with cylindrical symmetry.

### 2.8.2 Effective Mass Limitations

Effective mass is one of the more important concepts in the analysis of semiconductors and semiconductor devices, so it is desirable to examine the conditions under which it can be used. Basically, it is obtained by neglecting terms in the one-electron Schrödinger equation which can produce interband transitions. In Chapter 3, for example, where impurity energy levels are derived in the effective mass approximation, we tacitly assume that the impurity Coulomb potential in the Hamiltonian of (3.10) is not large enough to produce band-to-band transitions by Zener (field) tunneling. In Chapter 7, however, where optical absorption is developed, we include a photon vector potential in the Hamiltonian of (7.8) to provide for band-to-band transitions. The resulting equations for absorption depend on the free-electron mass.

Specifically, the effective mass approximation fails when the potentials in the one-electron Hamiltonian change so rapidly in time or in space that interband transitions occur in the semiconductor. For variations of potential with time, this sets an upper bound on the radial frequency of

$$\omega \simeq \frac{\mathcal{E}_g}{\hbar} \quad (2.121)$$

where  $\mathcal{E}_g$  is the minimum energy gap of the semiconductor. For variations of potential with space, this sets an upper bound on the electric field of

$$E \simeq \frac{\mathcal{E}_g}{qa} \quad (2.122)$$

where  $a$  is the atomic spacing. When frequencies or fields of the magnitudes indicated by (2.121) and (2.122) are involved, the effective mass approximation is clearly not valid. There are, however, other limitations on the effective mass approximation that may be more restrictive than these.

For many applications we would like the charge carriers to be small enough (in some sense) compared to the smallest device dimension that they are localized within the device structure, and fast enough compared to the oscillation period that they can respond to the applied field. That is, the

application may require that the charge carriers behave more like particles than like waves. (The effective mass approximation is valid only for a wave packet.) The limitation is simply that particles are constructed from a superposition of states indexed by  $k$ , within a range  $\Delta k$ , that are thermally distributed at some temperature,  $T$ . The range of energy  $\Delta \mathcal{E}$  within the wave packet corresponding to  $\Delta k$  is then

$$\Delta \mathcal{E} = \frac{\hbar^2 (\Delta k)^2}{2m^*} = kT \quad (2.123)$$

With (2.123) and the Heisenberg uncertainty relations,

$$\Delta x \Delta p \geq \hbar, \quad \Delta t \Delta \mathcal{E} \geq \hbar \quad (2.124)$$

we can estimate how localized in time and space a particle can be with an energy range of  $kT$ . The results are

$$\Delta t \geq \frac{\hbar}{kT} \quad (2.125)$$

and

$$\Delta x \geq \frac{\hbar}{(2m^* kT)^{1/2}} \quad (2.126)$$

For a semiconductor at 300 K with an energy gap of 1 eV, (2.125) gives a bandwidth of  $6.3 \times 10^{12}$  Hz, whereas (2.121) gives  $2.4 \times 10^{14}$  Hz, more than an order of magnitude larger. However, even  $6.3 \times 10^{12}$  Hz is much higher than the anticipated operating frequency of any current semiconductor device, so there should be no problem with the effective mass approximation in this aspect. At 300 K (2.126) gives

$$\Delta x \geq 76 \left( \frac{m}{m^*} \right)^{1/2} \text{ Å} \quad (2.127)$$

as the minimum size of a wave packet with an energy spread of  $kT$ . This value can be interpreted as the smallest dimension a device can have and still be analyzed using the effective mass approximation. For GaAs with an effective mass of  $0.067m$ , (2.127) gives a minimum device dimension of 294 Å. Since this is much larger than the dimensions of many quantum well devices, the effective mass approximation would not be expected to be a very reliable tool in evaluating these devices.

## PROBLEMS

- 2.1. Plot the two lowest-lying empty lattice energy bands for the simple hexagonal lattice with ideal  $c/a$  ratio from  $A$  to  $L$  to  $M$  to  $\Gamma$  to  $A$  to  $H$  to  $K$  to  $\Gamma$ . Indicate the degeneracy of each band.

- 2.2. Use the nearly free electron approximation to obtain an expression in terms of  $a$ ,  $b$ , and  $U_0$  for the forbidden energy gaps of the periodic square-well potential shown in Fig. P2.2.

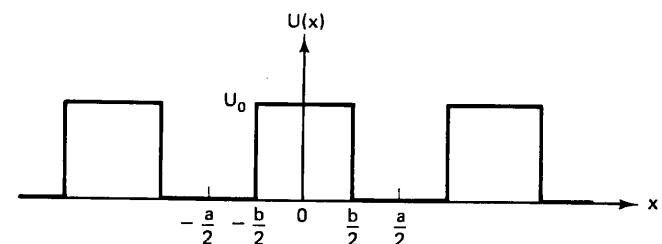


Figure P2.2

- 2.3. (a) Use the tight-binding approximation to obtain an expression for the energy bands of a simple cubic direct lattice with a lattice constant of  $a$ . Use nearest neighbors only.  
(b) Determine the electron velocity,  $v$ , and the effective mass tensor,  $m_{ij}$ , for these bands.
- 2.4. Obtain an expression similar to (2.103) for a tight-binding energy band in the simple hexagonal lattice and plot as in Fig. 2.12 from  $L$  to  $M$  to  $\Gamma$ .
- 2.5. (a) Assuming no collisions, obtain an expression for the time it takes an electron to completely traverse the first Brillouin zone in a face-centered cubic crystal in the direction of  $L$  under an applied electric field  $E$ .  
(b) If the average time between collisions is  $10^{-12}$  s, what fraction of the zone can an electron travel in an electric field of 10 V/cm? Use a lattice constant of 5 Å.
- 2.6. Consider a hypothetical solid made up of  $N$  hydrogen atoms at positions  $R_j$  on a simple cubic lattice of spacing  $a$ .  
(a) Write the expression for a one-electron wave function  $\psi_k$  which is a linear combination of  $1s$  atomic orbitals  $\phi_{1s}$ .  
(b) Show that if the  $\psi_k$  is properly constructed, it satisfies the Bloch theorem.  
(c) Using the tight-binding method, obtain  $\mathcal{E}(k) = \mathcal{E}_0 - \mathcal{E}_0' - 2\mathcal{E}_{\pm} \cos ka$ .
- 2.7. Calculate the four lowest empty lattice free-electron energy bands for a simple square lattice in the  $\Delta(\Gamma - X)$  and  $\Sigma(\Gamma - M)$  directions and the two lowest bands in the  $Z(X - M)$  directions, and plot in terms of energy in units of  $\hbar^2/m a^2$  and wave vector in units of  $2\pi/a$ .
- 2.8. The  $sp^3$  bond hybridization is very important in the bonding of diamond, Si, and Ge. These bonds can be described by linear combinations of the  $s$ ,  $p_x$ ,  $p_y$ , and  $p_z$  wavefunctions of the valence electrons. The appropriate linear combinations for diamond are

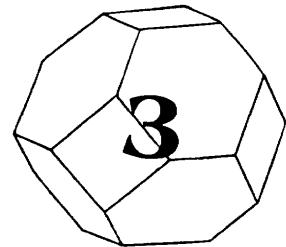
$$\psi_1 = \frac{1}{2} (\phi_s + \phi_{px} + \phi_{py} + \phi_{pz}), \quad \text{where } \phi_s = \frac{1}{(4\pi)^{1/2}}$$

$$\psi_2 = \frac{1}{2} (\phi_s - \phi_{px} - \phi_{py} + \phi_{pz}), \quad \text{where } \phi_{px} = \left( \frac{3\pi}{4} \right)^{1/2} \sin \theta \cos \phi$$

$$\psi_3 = \frac{1}{2} (\phi_s + \phi_{px} - \phi_{py} - \phi_{pz}), \quad \text{where } \phi_{py} = \left(\frac{3\pi}{4}\right)^{1/2} \sin \theta \sin \phi$$

$$\psi_4 = \frac{1}{2} (\phi_s - \phi_{px} + \phi_{py} - \phi_{pz}), \quad \text{where } \phi_{pz} = \left(\frac{3\pi}{4}\right)^{1/2} \cos \theta$$

- (a) Considering only the angular parts of the wave functions, find the directions of the maximum charge densities  $\{q |\psi_1|^2, q |\psi_2|^2, q |\psi_3|^2, \text{ and } q |\psi_4|^2\}$  of these bonds.
- (b) What are the angles between the directions of maximum charge concentration for  $\psi_1$  and  $\psi_2$ ,  $\psi_3$  and  $\psi_4$ , and  $\psi_1$  and  $\psi_3$ ?
- 2.9. Calculate and plot the constant-energy contours in the  $k_x k_y$ -plane of the Brillouin zone (reduced zone scheme) of the simple cubic lattice in
- (a) The tight-binding approximation
  - (b) The nearly free electron approximation
  - (c) The free-electron approximation
- 2.10. Use degenerate perturbation theory to calculate the splitting and form for small  $k$  of the second and third energy bands at the center of the zone of a one-dimensional crystal in the nearly free electron model. What is the effective mass of an electron at the bottom of the third energy band? What is the effective mass of an electron at the top of the second energy band?
- 2.11. Calculate (estimate) and plot the electron velocity for  $\mathbf{k}$  from  $\Gamma$ -X along  $\Delta$  for GaAs. The effective mass at the bottom of the  $\Gamma$  minimum is  $0.0665m_0$ , and the effective mass at the bottom of the X minima can be assumed to be spherical and equal to  $0.4m_0$ .



# Crystal Imperfections

Up to now we have considered only the properties of perfect crystals. We assumed that the atoms of the crystal were arranged at the sites of a mathematically precise Bravais lattice. This resulted in a perfectly periodic lattice potential energy for the electrons and we were able to establish relationships between the energy and wavevector of an electron on this basis. We also found, incidentally, that electrons can have a constant average velocity in a perfect periodic potential without the application of external forces. This, of course, is contrary to experimental observation. Because of this and other important experimental results, we realize that the perfect crystal is an overly simplified picture of a more complicated situation. Thus, to take into account many of the more important properties of semiconductors, we must examine possible crystal imperfections. We will base our analysis of these imperfections on the assumption that they can be, in general, regarded as perturbations in a perfect crystal.

## 3.1 ELECTRONS AND HOLES

Let us first consider the situation where one of the bonds between atoms in a perfect crystal is broken. From a chemical point of view, sufficient energy has been given to an electron to move it from a bonding state to an antibonding state. In the energy band picture, an electron is excited from somewhere near the top of a completely filled valence band to somewhere near the bottom of a completely empty conduction band. After the excitation,

" BAND STRUCTURE HANDOUT "

## Appendix 2.3

### Band Structure of Column IV Elements Calculated by the LCAO Method

We give here a first idea of the band structure of column IV elements, calculated by the tight binding method, for energies close to the band gap, and thus obtain a theory of the covalent bond in semiconductors. This calculation formally holds for the sequence diamond, silicon, germanium, and grey tin. (See G. Leman, *Annales de Physique*, Paris, 7 1962 p. 505.)

The crystal is face-centered cubic with two atoms per cell, one at the origin, the other translated in the direction [1 1 1] of the quarter  $\mathbf{b}$  of the principal diagonal of the cube (see Figs. 2.1 and 2.2). Each atom is at the center of a regular tetrahedron. The primitive cell, shown in Fig. 2.1, is rhombohedral with basis vectors:

$$\begin{aligned} \mathbf{a}_1 &= (\mathbf{a}/2)(\mathbf{i} + \mathbf{j}), \\ \mathbf{a}_2 &= (\mathbf{a}/2)(\mathbf{i} + \mathbf{k}), \\ \mathbf{a}_3 &= (\mathbf{a}/2)(\mathbf{j} + \mathbf{k}). \end{aligned} \quad (2.133)$$

The electron configuration of column IV atoms is  $ns^2np^2$  with the atomic Hamiltonian

$$\mathcal{H}_{at} = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}), \quad \text{or} \quad \mathcal{H}_{at} = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r} - \mathbf{b}), \quad (2.134)$$

and eigenstates  $\phi_s, \phi_p$  of energies  $E_s, E_p$ . We consider

$$\begin{aligned} \phi_0 &= (\phi_s + \phi_x + \phi_y + \phi_z)/2, \\ \phi_1 &= (\phi_s - \phi_x - \phi_y + \phi_z)/2, \\ \phi_2 &= (\phi_s - \phi_x + \phi_y - \phi_z)/2, \\ \phi_3 &= (\phi_s + \phi_x - \phi_y - \phi_z)/2. \end{aligned} \quad (2.135)$$

These functions, called hybrid orbitals  $sp_3$ , are not atomic eigenfunctions but form a basis for the tensor product  $s \otimes p$ . They have the essential property of "pointing," respectively, in the directions [1 1 1], [-1,-1,1],

## Band Structure of Column IV Elements

55

$[-1,1,1]$ ,  $[1,-1,-1]$ , i.e., toward the vertices of a tetrahedron in the observed direction of the covalent bond. The matrix elements of  $\mathcal{H}_{\text{at}}$  in this basis are

$$\begin{aligned} \langle \phi_i | \mathcal{H}_{\text{at}} | \phi_i \rangle &= \frac{E_s + 3E_p}{4}, \\ \langle \phi_i | \mathcal{H}_{\text{at}} | \phi_l \rangle &= \frac{E_s - E_p}{4}, \end{aligned} \quad (2.136)$$

where  $l$  denotes an orbital other than  $i$  localized at the same atom.

Similarly we consider hybrid orbitals pointing in the opposite directions:

$$\begin{aligned} \phi'_0 &= (\phi_s - \phi_x - \phi_y - \phi_z)/2 \text{ pointing toward } [-1, -1, -1], \\ \phi'_1 &= (\phi_s + \phi_x + \phi_y - \phi_z)/2 \text{ pointing toward } [+1, +1, -1], \\ \phi'_2 &= (\phi_s + \phi_x - \phi_y + \phi_z)/2 \text{ pointing toward } [+1, -1, +1], \\ \phi'_3 &= (\phi_s - \phi_x + \phi_y + \phi_z)/2 \text{ pointing toward } [-1, +1, +1]. \end{aligned} \quad (2.137)$$

We seek a Bloch function solution for the crystal of the form

$$\psi_{n,k} = C \sum_j \exp(i \mathbf{k} \cdot \mathbf{R}_j) \sum_{i=0,1,2,3} [A_i \phi_i(\mathbf{r} - \mathbf{R}_j) + A'_i \phi'_i(\mathbf{r} - \mathbf{R}_j - \mathbf{b})], \quad (2.138)$$

where the index  $j$  denotes the particular site in the lattice, the index  $i$  one of the four orbitals Eq. (2.135), and  $C$  is a normalization coefficient.

The crystal Hamiltonian is

$$\mathcal{H} = -\frac{\hbar^2}{2m} \Delta + \sum_j V(\mathbf{r} - \mathbf{R}_j) + V(\mathbf{r} - \mathbf{R}_j - \mathbf{b}), \quad (2.139)$$

whence the eigenstate problem:

$$\mathcal{H} \psi_{n,k}(\mathbf{r}) = E_{n,k} \psi_{n,k}(\mathbf{r}). \quad (2.140)$$

Consider the functions  $\phi_i$  and  $\phi'_i$  centered, respectively, at the origin and at  $(1/4, 1/4, 1/4)$ , and let us take their product with Eq. (2.140). We get

$$\langle \phi_i | \mathcal{H} | n, k \rangle = E_{n,k} \langle \phi_i | n, k \rangle, \quad (2.141)$$

$$\langle \phi'_i | \mathcal{H} | n, k \rangle = E_{n,k} \langle \phi'_i | n, k \rangle. \quad (2.142)$$

We can rewrite Eq. (2.139) in the form

$$\mathcal{H} = \mathcal{H}_{\text{at}} + \sum_{j \neq 0} V(\mathbf{r} - \mathbf{R}_j) + \sum_j V(\mathbf{r} - \mathbf{R}_j - \mathbf{b}), \quad (2.143)$$

where  $\mathcal{H}_{\text{at}}$  is the electron Hamiltonian for the atom at the origin. One can thus see that the only contributions on the left in Eqs. (2.141) and (2.142) come either from the atomic terms or from the interaction term coupling  $\phi_i$  or  $\phi'_i$  to the only orbital from the nearest neighbor which points toward it (neglecting terms of the form  $\langle \phi_i | V_{i \neq j} | \phi_i \rangle$  and interactions between

## 56 Appendix 2.3

orbitals centered on two neighboring atoms but not pointing to each other). The interaction or "transfer" integrals are all equal and negative since the potential is attractive and the functions  $\phi$  and  $\phi'$  have the same sign in the region where they overlap. We set

$$\langle \phi | V | \phi' \rangle = -\lambda. \quad (2.144)$$

Let us consider the site at the origin:  $\phi_0$  points toward the site  $b = (a/4)(1,1,1)$  and is thus only coupled to the orbital  $\phi'_0$  centered at  $b$ , with the coupling coefficient  $\exp ik \cdot a_0 = 1$ .  $\phi_1$ , being directed along  $[-1,-1,1]$  is coupled to the orbital centered at the site  $(a/4)(-1,-1,1) = b - a_1$ , or  $\phi'_1[r - (-a_1)]$  with the coefficient  $\exp ik \cdot R_j = \exp(-ik \cdot a_1)$ , etc. Finally Eqs. (2.141) and (2.142) can be written

$$\left[ \frac{E_s + 3E_p}{4} - E_{n,k} \right] A_i + \frac{E_s - E_p}{4} \sum_{l \neq i} A_l - \lambda \exp(-i \cdot k \cdot a_i) A'_i = 0, \quad (2.145)$$

$$-\lambda \exp(i \cdot k \cdot a_i) A_i + \left[ \frac{E_s + 3E_p}{4} - E_{n,k} \right] A'_i + \frac{E_s - E_p}{4} \sum_{l \neq i} A'_l = 0. \quad (2.146)$$

This system of homogeneous equations gives the eight coefficients  $A_i$  and  $A'_i$  which specify the wave function Eq. (2.138), provided the determinant of the coefficients vanishes. One thus obtains the following secular equation, with

$$x = E_{n,k} - E_p, \quad \delta = \frac{E_p - E_s}{4}, \quad \text{and} \quad \alpha_n = \exp(-i \cdot k \cdot a_n), \quad (2.147)$$

$$\begin{vmatrix} x + \delta & \delta & \delta & \delta & \lambda \alpha_0 & 0 & 0 & 0 \\ \delta & x + \delta & \delta & \delta & 0 & \lambda \alpha_1 & 0 & 0 \\ \delta & \delta & x + \delta & \delta & 0 & 0 & \lambda \alpha_2 & 0 \\ \delta & \delta & \delta & x + \delta & 0 & 0 & 0 & \lambda \alpha_3 \\ \lambda \alpha_0^* & 0 & 0 & 0 & x + \delta & \delta & \delta & \delta \\ 0 & \lambda \alpha_1^* & 0 & 0 & \delta & x + \delta & \delta & \delta \\ 0 & 0 & \lambda \alpha_2^* & 0 & \delta & \delta & x + \delta & \delta \\ 0 & 0 & 0 & \lambda \alpha_3^* & \delta & \delta & \delta & x + \delta \end{vmatrix} = 0. \quad (2.148)$$

Setting  $\Phi = |\frac{1}{4} \sum_n \alpha_n|$ , a careful calculation shows that Eq. (2.148) can be written as

$$(x^2 - \lambda^2)^2(x^2 + 4\delta x - \lambda^2 + 4\delta\lambda\Phi),$$

$$(x^2 + 4\delta x - \lambda^2 - 4\delta\lambda\Phi) = 0, \quad (2.149)$$

with

$$\Phi(\mathbf{k}) = \frac{1}{2} \left[ 1 + \cos \frac{k_x a}{2} \cos \frac{k_y a}{2} + \cos \frac{k_x a}{2} \times \cos \frac{k_z a}{2} + \cos \frac{k_y a}{2} \cos \frac{k_z a}{2} \right]^{1/2} \quad (2.150)$$

Equation (2.149) gives  $x(\Phi)$ , i.e., the dispersion relation  $E_n(\mathbf{k})$ .

## Resulting Band Structure

We obtain four flat bands (for which  $E_n$  does not depend on  $\mathbf{k}$ ) which correspond to the doubly degenerate solutions  $x_1 = \lambda$  and to the similarly doubly degenerate solutions  $x_2 = -\lambda$ .

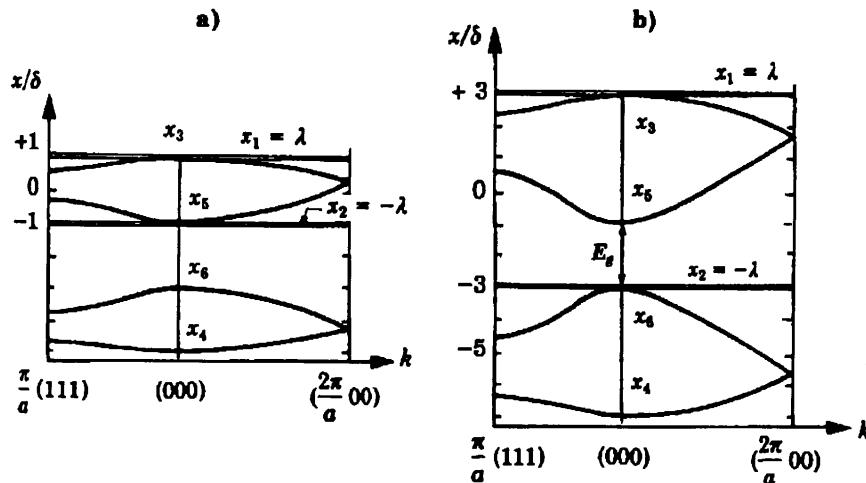
The broad bands are associated with the other solutions of Eq. (2.149), i.e.,

$$\begin{aligned} x_3 &= -2\delta + \sqrt{4\delta^2 + \lambda^2 + 4\delta\lambda\Phi}, \\ x_4 &= -2\delta - \sqrt{4\delta^2 + \lambda^2 + 4\delta\lambda\Phi}, \\ x_5 &= -2\delta + \sqrt{4\delta^2 + \lambda^2 - 4\delta\lambda\Phi}, \\ x_6 &= -2\delta - \sqrt{4\delta^2 + \lambda^2 - 4\delta\lambda\Phi}. \end{aligned} \quad (2.151)$$

The band structure has a different shape for  $\lambda > 2\delta$  or  $\lambda < 2\delta$ . It is shown in Fig. 2.20(a) for two directions of the vector  $\mathbf{k}$  in the zone and in the special case  $\lambda = \delta$ . There are  $8N$  electrons to be placed in these levels since the primitive cell contains two atoms each possessing four valence electrons. Taking account of the spin degeneracy of each atomic level we see that at zero temperature the bands  $x_4$ ,  $x_6$ , and  $x_2$  are filled. We are thus dealing here with a metal, as the band  $x_5$  is empty and very near in energy to the filled levels.

By contrast if  $\lambda > 2\delta$  we have a semiconductor. This is shown in Fig. 2.20(b) for the special case  $\lambda = 3\delta$ . The  $8N$  electrons fill the  $x_4$ ,  $x_6$ , and  $x_2$  bands, and the material is an insulator at zero temperature. The width of the band gap is

$$E_g = x_{5,k=0} - x_2 = 2\lambda - 4\delta. \quad (2.152)$$



**Fig. 2.20.** Band structure of an element of column IV of the Periodic Table, calculated by the LCAO method. The transfer integral is  $-\lambda$ , and the distance between the atomic  $s$  and  $p$  levels is  $4\delta$ . The bands  $x_4$ ,  $x_5$ , and  $x_6$  are full. (a)  $\lambda = \delta$ ; the material is a conductor; (b)  $\lambda = 36$ ; the material is an insulator.

### Application: Binding Energy of Semiconductors of Column IV

This very simplified picture of band structure already gives an approximate understanding of several properties of these solids, for example the relation between binding energy and band gap width. The following values are measured for the binding energy  $E_C$  and the width of the band gap  $E_g$ , expressed in eV (P. Manca, Journal of Physics and Chemistry of Solids 20, 268, 1961).

	C	Si	Ge	Sn
$E_C$	14.7	7.55	6.52	5.5
$E_g$	5.2	1.12	0.66	$\sim 0$

These values are plotted in Fig. 2.21. The experimental points fall approximately on a line with the equation

$$E_C(\text{eV}) = 1.85 E_g + 5.36. \quad (2.153)$$

In our model the binding energy is the difference between the energy  $2E_s + 2E_p$  of an atom ( $2 s$  and  $2 p$  electrons) and the energy per atom of the crystal. This energy  $E_X$  is, taking account of Eqs. (2.147) and (2.151),

$$E_X = 4 E_p + 2 x_2 + (\text{zone average of } x_6 + x_4). \quad (2.154)$$

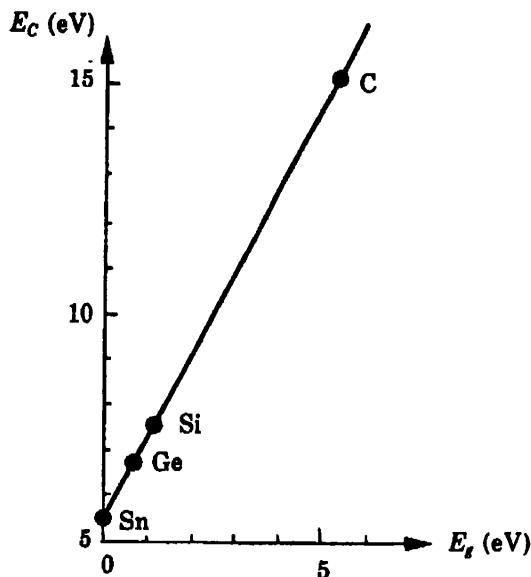


Fig. 2.21. The relation between cohesive energy and band gap width obtained by the LCAO method for the sequence of crystals of elements of column IV of the Periodic Table.

As a first approximation for the average we can take the value  $(x_{6,k=0} + x_{4,k=0})$ . We then find

$$E_X = 4 E_p - 4 \delta - 4 \lambda, \quad (2.155)$$

and using Eqs. (2.152) and (2.155) the binding energy can be written as

$$\begin{aligned} E_C &= 2(E_s + E_p) - E_X, \\ E_C &= 2 E_g + 4 \delta. \end{aligned} \quad (2.156)$$

This relation should be compared with the experimental relation Eq. (2.153). The energy  $4\delta = E_p - E_s$  is the excitation energy for an atomic electron from level  $s$  to level  $p$ . This quantity, provided by atomic physics, is almost constant over column IV and is about 5 eV. The agreement with the experiment is very good. This description of the binding energy neglects the effect of the repulsive term that must exist if the crystal structure is not to collapse. If the repulsive interactions vary very rapidly with distance then the above description holds; this is for example true of the hard-sphere repulsive potential.

We can understand also, at least qualitatively, the influence of the temperature or pressure on  $E_g$ . When the temperature decreases or the pressure increases, the mean distance between atoms decreases, raising the value of the integral  $\lambda$ , and simultaneously the band gap.

## 60 Appendix 2.3

However, the above calculation does not allow us to get more than the order of magnitude of these variations. The fundamental reason is that by limiting ourselves to combinations of a small number of atomic orbitals and only considering nearest neighbors we are working in a Hilbert space of too small a dimension. Nonetheless, this method gives an excellent approximation for the deep levels in solids.

**Remark:** for tin (the allotropic form called "grey tin") the gap is actually negative. This means that there is an overlap in energy between the top of the valence bands and the bottom of the conduction band. In this situation the highest energy electrons in the valence band populate the lowest energy states of the conduction band. There can thus exist a metallic-type conductivity due simultaneously to conduction-band electrons and to "holes" or empty places in the valence band, with equal numbers (see Chap. 3). We say that we are dealing with a "semimetal."

## $k \cdot p$ Perturbation

In the absence of the spin-orbit interaction, the Schrödinger equation reads

$$\left[ -\frac{\hbar^2}{2m} \nabla^2 + V(r) \right] \Phi(r) = \varepsilon \Phi(r) \quad (11.30)$$

where  $V(r)$  represents the potential energy of the electron, which among other things, consists primarily of the periodic lattice potential. Since at present our aim is to work out some theoretical expressions and to correlate the parameters involved in these expressions with experiments, it is not necessary for us to know the explicit form of  $V(r)$ . In a periodic lattice, the one electron wave function can always be written as the well-known Bloch function

$$\Phi_k(r) = u_k(r) \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (11.31)$$

where  $u_k(r)$  is periodic from lattice cell to cell. Substituting Eq. (11.31) into Eq. (11.30), we have

$$\left[ -\frac{\hbar^2}{2m} \nabla^2 - \frac{\hbar^2}{m} i\mathbf{k} \cdot \nabla + V(r) + \frac{\hbar^2 k^2}{2m} \right] u_k(r) = \varepsilon u_k(r) \quad (11.32)$$

Let  $u_s$  be the electronic wave function for the conduction-band electrons and  $u_x$ ,  $u_y$ , and  $u_z$  be those for the valence-band electrons at  $k[000]$ . By definition, these wave functions should satisfy the following relations

$$\left[ -\frac{\hbar^2}{2m} \nabla^2 + V(r) \right] u_s = \varepsilon_c u_s \quad (11.33)$$

$$\left[ -\frac{\hbar^2}{2m} \nabla^2 + V(r) \right] u_{x,y,z} = \varepsilon_v u_{x,y,z} \quad (11.34)$$

where  $\varepsilon_c$  and  $\varepsilon_v$ , respectively, are the energies of the conduction- and valence-band edges at  $k[000]$ .

For electrons near  $k[000]$ , the cell wave function  $u_k(r)$  can be expressed in terms of  $u_s$ ,  $u_x$ ,  $u_y$ ,  $u_z$  as

$$u_k(r) = A u_s + B u_x + C u_y + D u_z \quad (11.35)$$

The operational properties of the  $s$  and  $p$  wave functions  $u_s$  and  $u_{x,y,z}$  have been discussed in Sec. 8.4. Note that the  $s$  wave function is spherically symmetrical while the  $p$  wave function is odd in  $x$ ,  $y$ , or  $z$ . Substituting Eq. (11.35) into Eq. (11.32) and using Eqs. (11.33) and (11.34), we find

$$A(\varepsilon_c - \varepsilon' + h_1)u_s + B(\varepsilon_s - \varepsilon' + h_1)u_x + C(\varepsilon_s - \varepsilon' + h_1)u_y + D(\varepsilon_s - \varepsilon' + h_1)u_z = 0 \quad (11.36)$$

where  $\varepsilon' = \varepsilon - \frac{\hbar^2 k^2}{2m}$  and

$$h_1 = \frac{\hbar^2}{mi} \left( k_x \frac{\partial}{\partial x} + k_y \frac{\partial}{\partial y} + k_z \frac{\partial}{\partial z} \right)$$

It is interesting to observe that the function  $(\partial/\partial x)u_s$  is an odd function in  $x$ ; therefore the matrix element  $\int u_j^* (\partial/\partial x)u_s dx$  is nonzero only when  $j = \dot{x}$ . The same is true for  $\int u_s^* (\partial/\partial x)u_j dx$ . Since the function  $(\partial/\partial x)u_s$  is odd in both  $x$  and  $y$ , the integral

$$\iint_{-\infty}^{\infty} u_j^* \frac{\partial}{\partial x} u_s dx dy$$

is always zero,  $j = s, x, y$ , or  $z$ . In summary, the only nonvanishing matrix elements are the ones defined by the following expression

$$P = -\frac{\hbar^2}{m} \int_{-\infty}^{\infty} u_s \frac{\partial}{\partial j} u_s dq = \frac{\hbar^2}{m} \int_{-\infty}^{\infty} u_s \frac{\partial}{\partial j} u_j dq \quad (11.37)^{**}$$

where  $j = x, y$ , or  $z$ , and  $dq = dx dy dz$ .

Multiplying Eq. (11.36) by  $u_s$  and integrating over  $dq$ , we obtain a homogeneous equation in  $A, B, C$ , and  $D$ . We repeat the process by multiplying Eq. (11.36) with  $u_x, u_y$ , and  $u_z$  separately, and note that for nontrivial solutions, the determinant of the coefficients for  $A, B, C$ , and  $D$  must be zero. Thus, we obtain:

$$\begin{vmatrix} \varepsilon_c - \varepsilon' & -ik_x P & -ik_y P & -ik_z P \\ ik_x P & \varepsilon_v - \varepsilon' & 0 & 0 \\ ik_y P & 0 & \varepsilon_v - \varepsilon' & 0 \\ ik_z P & 0 & 0 & \varepsilon_r - \varepsilon' \end{vmatrix} = 0 \quad (11.38)$$

The four roots of Eq. (11.38) are

$$\begin{aligned} \varepsilon_{1,2} &= \frac{\varepsilon_c + \varepsilon_v}{2} \pm \sqrt{\left(\frac{\varepsilon_c - \varepsilon_v}{2}\right)^2 + k^2 P^2 + \frac{\hbar^2 k^2}{2m}} \\ \varepsilon_{3,4} &= \varepsilon_v + \frac{\hbar^2 k^2}{2m} \end{aligned} \quad (11.39)$$

In Eq. (11.39),  $\varepsilon_1$  taking the positive sign represents the energy of conduction-band electrons while  $\varepsilon_2$  taking the negative sign,  $\varepsilon_3$ , and  $\varepsilon_4$  are the energies of electrons in the three valence bands. From Eq. (11.39), the effective masses can be calculated and the values thus obtained can be compared with those found from cyclotron resonance experiments. Such a comparison indicates that further refinements of the theory must be made. First, the spin-orbit interaction must be added and, second, a higher order perturbation calculation is needed for the valence bands.

The spin-orbit interaction can be incorporated into the formulation

<sup>\*\*</sup> The wave functions  $u_x, u_y$ , and  $u_z$ , being similar in nature to  $\psi_A, \psi_B$ , and  $\psi_C$  of Eq. (8.26), are real and so is  $u_s$ ; therefore the complex conjugate sign is not needed here.

## Appendix 2.4

### The $\mathbf{k} \cdot \mathbf{p}$ Method

The  $\mathbf{k} \cdot \mathbf{p}$  method is a semi-empirical method which uses quantities found from experiment in the theoretical calculation of the band structure. We start from the equation

$$\left[ \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}) \right] \psi_{n,\mathbf{k}}(\mathbf{r}) = E_{n,\mathbf{k}} \psi_{n,\mathbf{k}}(\mathbf{r}). \quad (2.157)$$

Replacing  $\psi_{n,\mathbf{k}}$  by  $u_{n,\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r})$  we note that

$$\begin{aligned} \mathbf{p} \psi_{n,\mathbf{k}} &= \exp(i\mathbf{k} \cdot \mathbf{r}) (\mathbf{p} + \hbar \mathbf{k}) u_{n,\mathbf{k}}(\mathbf{r}), \\ p^2 \psi_{n,\mathbf{k}} &= \exp(i\mathbf{k} \cdot \mathbf{r}) (\mathbf{p} + \hbar \mathbf{k})^2 u_{n,\mathbf{k}}(\mathbf{r}), \end{aligned} \quad (2.158)$$

and thus

$$\left[ \frac{(\mathbf{p} + \hbar \mathbf{k})^2}{2m} + V(\mathbf{r}) \right] u_{n,\mathbf{k}}(\mathbf{r}) = E_{n,\mathbf{k}} u_{n,\mathbf{k}}(\mathbf{r}). \quad (2.159)$$

The periodic part of the Bloch function obeys an equation resembling the original equation apart from the vector  $\hbar \mathbf{k}$ . We can rewrite Eq. (2.159) in the form

$$\left[ \frac{\mathbf{p}^2}{2m} + \frac{\hbar \mathbf{k} \cdot \mathbf{p}}{m} + \frac{\hbar^2 k^2}{2m} + V(\mathbf{r}) \right] u_{n,\mathbf{k}}(\mathbf{r}) = E_{n,\mathbf{k}} u_{n,\mathbf{k}}(\mathbf{r}). \quad (2.160)$$

For a free electron in a box  $V(\mathbf{r}) = 0$  an obvious solution is  $u = \text{constant}$  or  $E_k = \hbar^2 k^2 / 2m$  and  $\psi = \Omega^{-1/2} \exp i\mathbf{k} \cdot \mathbf{r}$ . Equation (2.160) takes a particularly simple form for  $\mathbf{k} = 0$ :

$$\left[ \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}) \right] u_{n,0}(\mathbf{r}) = E_{n,0} u_{n,0}(\mathbf{r}). \quad (2.161)$$

We note that when the atoms are very far apart, the  $E_{n,0}$  are the atomic levels and the  $u_{n,0}(\mathbf{r})$  atomic eigenfunctions. We note also that the equation giving  $u_{n,0}(\mathbf{r})$  has the symmetries of the crystal potential  $V(\mathbf{r})$ .

## 62 Appendix 2.4

The  $\mathbf{k} \cdot \mathbf{p}$  method assumes that we know the values  $E_{n,0}$  either from theory or experiment. We then consider small values of  $\mathbf{k}$  close to  $\mathbf{k} = 0$  and treat the operator  $(\hbar/m)\mathbf{k} \cdot \mathbf{p}$  as a perturbation in the Hamiltonian.

Although the  $\mathbf{k} \cdot \mathbf{p}$  method is more general, we shall assume for simplicity that the crystal has a center of symmetry (in the diamond structure this center is halfway between two atoms) and we confine ourselves to the study of a given non-degenerate level  $n$ .

We first use the symmetry of Eq. (2.161). The Hamiltonian is invariant under inversion, the symmetry operation sending  $\mathbf{r}$  to  $-\mathbf{r}$ . Then if  $u_{n,0}(\mathbf{r})$  is an eigenfunction of energy  $E_{n,0}$ ,  $u_{n,0}(-\mathbf{r})$  is also an eigenfunction of the same energy, hence also  $[u_{n,0}(\mathbf{r}) + u_{n,0}(-\mathbf{r})]$  and  $[u_{n,0}(\mathbf{r}) - u_{n,0}(-\mathbf{r})]$ . The eigenfunctions can thus be classified into even and odd functions.

In perturbation theory one starts by considering the first-order diagonal matrix elements of the perturbation Hamiltonian  $\mathcal{H}_p$ . If we consider a non-degenerate level  $|n, 0\rangle$  the first-order term  $\langle n, 0 | \mathbf{k} \cdot \mathbf{p} | n, 0 \rangle$  vanishes:

$$\int u_{n,0}^*(\mathbf{r}) \frac{\partial}{\partial \mathbf{r}} u_{n,0}(\mathbf{r}) d^3\mathbf{r} = 0 \quad (2.162)$$

because  $u_{n,0}$  is either odd or even. We note that  $u_{n,0}(\mathbf{r})$  is an eigenfunction of Eq. (2.160) with the eigenvalue  $E_{n,0} + \hbar^2 k^2 / 2m$ . There remains only the second-order correction. To this order the energies are given by

$$E_{n,\alpha} = E_{n,0} + \frac{\hbar^2 k^2}{2m} + \hbar^2 \sum_{n' \neq n} \frac{\langle n', 0 | \mathbf{k} \cdot \mathbf{p} / m | n, 0 \rangle \langle n, 0 | \mathbf{k} \cdot \mathbf{p} / m | n', 0 \rangle}{E_{n,0} - E_{n',0}}, \quad (2.163)$$

and thus

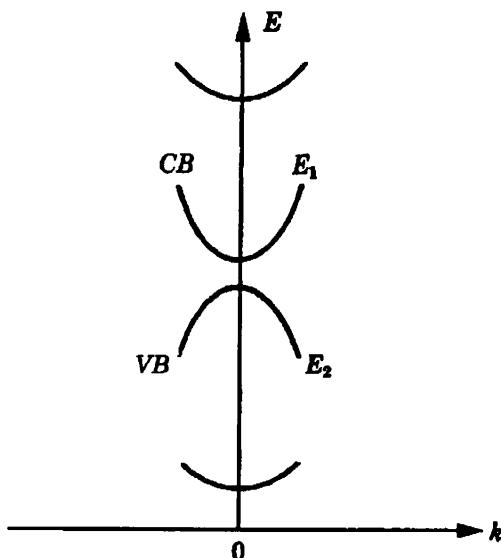
$$E_{n,\mathbf{k}} = E_{n,0} + \sum_{\alpha, \beta} \frac{\hbar^2}{2m} \left( \frac{m}{m^*} \right)_{\alpha\beta} k_\alpha k_\beta, \quad (2.164)$$

with

$$\left( \frac{m}{m^*} \right)_{\alpha\beta} = \delta_{\alpha\beta} + \frac{2}{m} \sum_{n'} \frac{\langle n', 0 | p_\alpha | n, 0 \rangle \langle n, 0 | p_\beta | n', 0 \rangle}{E_{n,0} - E_{n',0}}. \quad (2.165)$$

This so-called  $\mathbf{k} \cdot \mathbf{p}$  method allows us to find the effective masses directly either from the energy spectrum at  $\mathbf{k} = 0$  (the values of  $E_{n,0}$ ) and from the parameters  $|\langle n', 0 | p_\alpha | n, 0 \rangle|^2$ . This is the most useful theoretical procedure for predicting and analyzing details of the band structure of semiconductors near the band extrema, and thus in the region of interest.

We note that if we know the  $u_{n,0}(\mathbf{r})$  we can calculate the parameters directly. Generally these matrix elements are deduced from experiment. The energy differences which appear in the denominator are most usually found from optical absorption or reflectivity.



**Fig. 2.22.** The curvatures at  $k = 0$  of the bands  $E_1$  and  $E_2$  are principally determined by their mutual interaction: the other bands are clearly further away, and the energy denominators make their interactions negligible (cf. Eq. (2.166)). In consequence the curvatures, and thus the effective masses of the bands  $E_1$  (conduction) and  $E_2$  (light holes) are practically opposite.

The order of magnitude of  $m^*$  is given by

$$\frac{m}{m^*} \sim 1 + 2 < \frac{p_x^2}{m} > \frac{1}{E_g}. \quad (2.166)$$

Here  $p_x^2/m$  is of the order of the ionization energy of an atom, for example 5 eV, thus for  $E_g = 0.5$  eV,  $m/m^*$  will be of order 20.

If we are interested in two bands 1 and 2 close in energy with their extrema at  $\mathbf{k} = 0$ , we can confine ourselves to including just their interaction in the expression (2.165), the corresponding term being much bigger than all the others (Fig. 2.22).

*Example:* for a cubic crystal the constant energy surfaces are spheres around  $\mathbf{k} = 0$ , and we have

$$\begin{aligned} \frac{1}{m_1^*} &\simeq \frac{1}{m} + \frac{2}{m^2} \frac{|\langle 1|p_x|2 \rangle|^2}{E_1 - E_2}, \\ \frac{1}{m_2^*} &\simeq \frac{1}{m} + \frac{2}{m^2} \frac{|\langle 1|p_x|2 \rangle|^2}{E_2 - E_1}, \end{aligned} \quad (2.167)$$

or

## 64 Appendix 2.4

$$\frac{1}{m_1^*} = -\frac{1}{m_2^*} + \frac{2}{m}. \quad (2.168)$$

For germanium at  $\mathbf{k} = 0$  the "light hole" band (see Chap. 3) has an effective mass  $m_2^* = -0.042m$ , while the conduction band has an effective mass  $m_1^*$  (at  $k = 0$ ) of  $0.036m$ , in good agreement with the preceding expression: we sometimes say that in the region of small  $\mathbf{k}$ , these two bands are "mirror images" of each other, since  $m_1^* \simeq -m_2^*$ .

For the perturbation theory to provide a good approximation the matrix elements of the perturbation  $(\hbar/m)\mathbf{k} \cdot \mathbf{p}$  between the functions  $u_{n,0}$  and  $u_{n',0}$  must be such that

$$\left| \frac{\langle n, 0 | \hbar \mathbf{k} \cdot \mathbf{p}/m | n', 0 \rangle}{E_n - E_{n'}} \right| < 1, \quad (2.169)$$

or, for two close bands, from Eqs. (2.167) and (2.169):

$$\begin{aligned} \frac{1}{m_1^*} &\sim \frac{2}{m^2} \frac{| \langle 1 | p_x | 2 \rangle |^2}{E_1 - E_2}, \\ \frac{\hbar k_x}{m} &< \frac{E_1 - E_2}{| \langle 1 | p_x | 2 \rangle |}. \end{aligned}$$

Combining these two relations we get

$$\frac{\hbar^2 k^2}{2m_1^*} < E_1 - E_2. \quad (2.170)$$

The quantity  $\hbar^2 k^2 / 2m_1^*$  is the kinetic energy in the  $E_1$  band. For perturbation theory to give a good approximation, we require the energy in the band to be small compared with the band gap.

*Example:*  $E_1 - E_2 = E_g = 1$  eV. For typical values of the electron energy in the band, of order the thermal energy  $kT = 25$  meV at room temperature, the approximation easily holds.

## 2.5 The True Band Structure

41

$$\int d^3r \psi_k^{c*}(\mathbf{r}) \Phi_{\mathbf{k}} = 0$$

or

$$b_c = - \int d^3r \psi_k^{c*}(\mathbf{r}) \exp(i \mathbf{k} \cdot \mathbf{r}). \quad (2.101)$$

The  $\Phi_{\mathbf{k}}$  functions are known precisely if we know the  $\psi_{\mathbf{k}}^c$  accurately. This is true if we use the tight binding method for calculating the deep Bloch states, as the overlap of atomic functions of very negative energies is small. The tight binding approximation is excellent in this case.

The  $\Phi_{\mathbf{k}}$  functions have, by construction, the properties we seek: They are orthogonal to the deep states. They have a very localized part which oscillates rapidly, like the wave functions of the atomic core, and between atoms they appear like plane waves. The method then consists of seeking a solution to the Schrödinger equation for the crystal of the form

$$\psi = \sum_{\mathbf{k}} c_{\mathbf{k}} \Phi_{\mathbf{k}}. \quad (2.102)$$

Again, the  $\Phi_{\mathbf{k}}$  are only coupled by the periodic potential to functions  $\Phi_{\mathbf{k}+\mathbf{G}}$  so that this method is analogous to the plane wave method described above. But we obtain excellent results using expansions of  $\psi$  in a few tens of orthogonalized plane waves. This is at present the most powerful method of calculating the band structure.

Appendix 2.4 describes another semiempirical method of determining the band structure, the so-called  $\mathbf{k} \cdot \mathbf{p}$  method.

## 2.5 The True Band Structure

Before describing real band structure for silicon we note that Appendix 2.3 shows, using the tight binding method, that crystals of the face-centered cubic form like silicon have triply degenerate bands at  $\mathbf{k} = 0$  (levels  $x_2$  and  $x_6$ ). This occurs at the top of the valence band. We also know from the  $\mathbf{k} \cdot \mathbf{p}$  method (Appendix 2.4) that for  $\mathbf{k} = 0$  the states resemble atomic states, and are thus possibly degenerate. In fact, we have up to now neglected the role of spin in the electron Hamiltonian by taking account of its influence only via the Pauli principle. It can be shown that in its motion in the electric potential, the spin sees a magnetic field which results in a Hamiltonian of the form

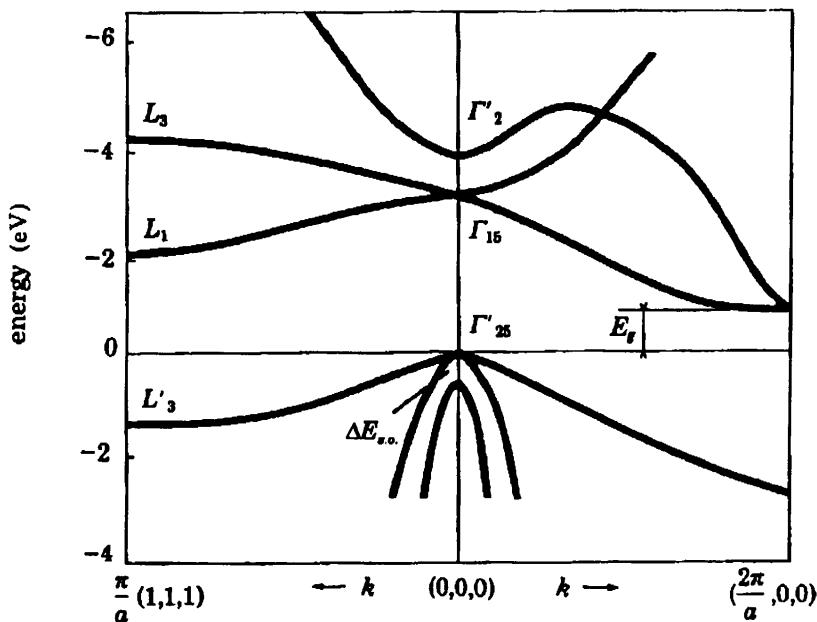
$$\mathcal{H}_{s.o.} = \frac{1}{4 m^2 c^2} \boldsymbol{\sigma} \times \text{grad } V(\mathbf{r}) \cdot \mathbf{p}, \quad (2.103)$$

where  $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$  are the Pauli matrices.

This so-called "spin-orbit" interaction has the effect of partially lifting the degeneracy mentioned above. This is seen in Fig. 2.12 which shows the

## 42 2. Quantum States of a Perfect Semiconductor

band structure of silicon around the band gap of width  $E_g$ . We note that the maximum of the valence band is doubly degenerate (the point  $\Gamma^{25'}$ ). The spin-orbit interaction has split one of the valence bands by an amount  $\Delta E_{s.o.}$ .



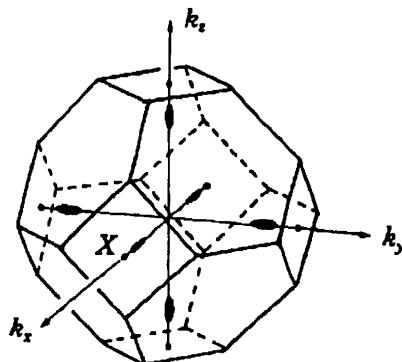
**Fig. 2.12.** Band structure of silicon: the letters  $L, \Gamma$  denote particular points of high symmetry in the zone.  $\Gamma$  is the center of the zone.  $L$  is the point at the edge of the zone in the direction (111). The distance  $L_1 L'_3$  is experimentally determined, while the absolute energies of the  $L'_3$  and  $L_1$  levels are found by a calculation by the method of orthogonalized plane waves (D. Brust, J.C. Phillips, and F. Bassani, Physical Review Letters 9, 94, 1962).

Some of the energy values given in the figure are found from experiment and others from calculations based on the method of orthogonalized plane waves. The form of the constant energy surfaces near the top of the valence band is complex (to describe it we need a  $k \cdot p$  theory for a degenerate level). We assume that it consists of two spheres, one called "heavy holes" and the other "light holes" (the concept of a hole will be introduced in Chap. 3).

We note that the minimum of the conduction band is not at  $k = 0$ , in contrast to the maximum of the valence band. In such a case the gap is called indirect. The value of  $E_g$  is 1.12 eV for silicon. The mass of the heavy holes is  $m_{hh} = 0.49m$  and the mass of the light holes is  $m_{lh} = 0.16m$ .

## 2.6 Experimental Study of Band Structure

43



**Fig. 2.13.** Constant energy surfaces near the bottom of the conduction band of Si.

The minimum of the conduction band is in the direction [100], and by symmetry on the equivalent directions  $\langle 100 \rangle$ . There are thus six minima of the conduction band around  $|k| \sim 0.8 \times (2\pi/a)$ .

By symmetry each ellipsoid of constant energy in the conduction band must have two equal axes. These are prolate ellipses as shown in Fig. 2.13. The dispersion relation near a minimum of the conduction band has the form

$$E(\mathbf{k}) = \frac{\hbar^2}{2} \left( \frac{k_x^2 + k_y^2}{m_T} + \frac{\Delta k_z^2}{m_L} \right) \quad (2.104)$$

for the ellipsoids [001] and [00\bar{1}]. Here we have set  $\Delta k_z = k_z - k_{z0}$ , with  $k_{z0} = (0, 0, 1.6\pi/a)$ . There appears a longitudinal effective mass  $m_L = 0.98m$  and a transverse mass  $m_T = 0.19m$ .

This shows the complexity of the real situation. However to understand many properties it is often enough to consider a band structure with a direct gap as shown below, using appropriate effective masses. We will call this representation "standard band structure" in the remainder of this book (Fig. 2.14).