

QUANTUM THEORY OF ELECTRONS - 1READING ASSIGNMENT

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

SUPPLEMENTARY REFERENCE

"Energy Band Theory", from *Advanced Semiconductor Fundamentals*, R.F. Pierret

LECTURE PROGRAM

1. Bloch Theorem
 - a. Mathematical statement
 - b. 1-dimensional representation
 - c. 3-dimensional representation
 - d. Physical statement
2. Kronig-Penney Model of an Infinite 1-D Crystal
 - a. Form of the periodic potential $V(x)$
 - b. Schrödinger's equation and the form of the wave functions
 - c. Solutions to the differential equations in $u_1(x)$ and $u_2(x)$
 - d. Allowed and forbidden energy bands
3. Crystal Momentum and Effective Mass
 - a. Crystal momentum
 - b. Localized wave function of an electron in the crystal
 - c. Group velocity of the electron wave packet
 - d. Analogue of Newton's law
 - e. Effective mass
4. Reduced Zone Representation: Electrons and Holes
 - a. Expanded representation
 - b. Reduced zone representation
 - c. Physical significance of $k(\epsilon_0)$ and $-k(\epsilon_0)$
 - d. Motion of a free electron in a perfectly periodic lattice
 - e. Effect of deviation from perfect periodicity
 - f. Hole
5. Free-Electron Approximation
 - a. Qualitative description (width of allowed and forbidden bands)
 - b. Mathematical representation
 - c. ϵ versus k
 - d. Effective mass

CHAPTER 8

QUANTUM THEORY OF ELECTRONS IN PERIODIC LATTICES

8.1 INTRODUCTION

The free-electron theory of metals developed in the previous chapter is based upon the notion that the conduction electrons within a metallic substance act like the classical free particles of a gas, subject only to the limitations of the Fermi-Dirac statistics. It is not immediately obvious, from the quantum mechanical point of view, why this should be the case. It is also not clear why some substances have large numbers of free electrons and are thus very good conductors, while others have hardly any and behave as insulators.

The simplest quantum mechanical view of an electron in a crystal is that of a single electron in a perfectly periodic potential which has the periodicity of the lattice. In this *one-electron* model of a solid, the periodic potential may be thought of as arising from the periodic charge distribution associated with the ion cores situated on the lattice sites, *plus* the (constant) average "smeared out" potential contribution due to all the other free electrons belonging to the crystal, so that the interaction of the single electron with all the others is accounted for in an average sense. The solution of Schrödinger's equation for the single electron in this potential then provides a set of "one-electron" states which the single electron may occupy, and in fact, which may be occupied (subject to the limitations of the Pauli principle) by all the electrons of the crystal, since the single electron which is considered initially may be regarded as typical of all electrons of the system. A one-dimensional representation of a periodic crystal potential, such as might be obtained by following a path along one of the $\langle 100 \rangle$ directions of a cubic crystal of lattice constant a is shown in Figure 8.1. So far as the one-electron quantum-mechanical picture is concerned, the crystal periodicity will usually be assumed to extend to infinity in all directions, but, of course, at a surface of any actual crystal the periodicity will be interrupted, the potential function then behaving somewhat as shown at the left-hand edge of Figure 8.1. The lattice spacing will not be *quite* uniform near such a surface, but as a practical matter the lattice periodicity will usually be found to be almost perfect after a few atomic spacings within the crystal.

We shall find that the one-electron wave functions (called Bloch functions) calculated by the above prescription always have certain properties closely related to the lattice periodicity, and that the allowed electronic energies occur in bands of permitted states separated by forbidden energy regions. Within the allowed energy

bands, the dynamical behavior of electrons will be found to be in many ways similar to that of free particles, and the question of whether a crystal is a conductor or an insulator will depend upon whether the electronic states within a given allowed band or set of bands is completely filled or partially empty. In addition, we shall see that the collective behavior of electrons in an *almost* filled band of allowed states is very

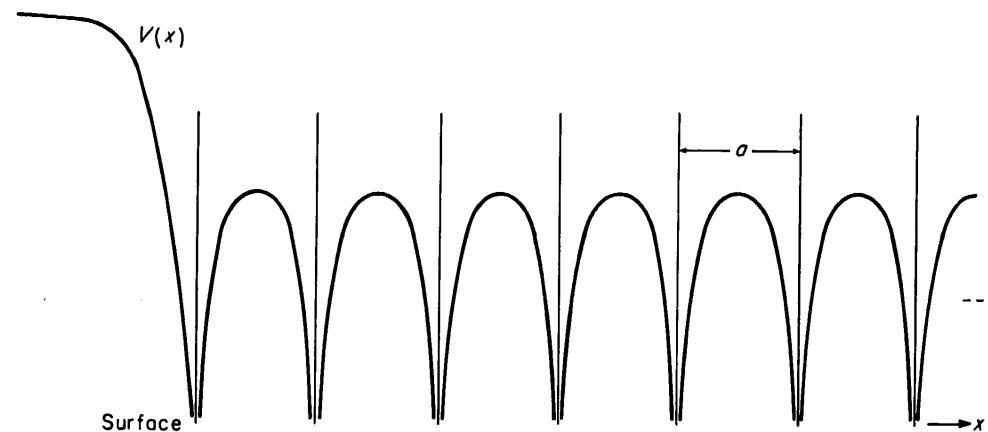


FIGURE 8.1. Schematic representation of the potential within a perfectly periodic crystal lattice. The surface potential barrier is shown at the left.

much like that of a few *positive* charge carriers in an almost empty band! The one-electron picture will thus serve to justify the free-electron model from a fundamental point of view, and also to answer certain rather puzzling questions which the free-electron theory, of itself, could not even attempt to explain. It should be remembered, however, that the one-electron treatment is an approximation in which the *details* of electron-electron interactions are completely overlooked.

8.2

THE BLOCH THEOREM

The Bloch theorem¹ is a mathematical statement regarding the form of the one-electron wave functions for a perfectly periodic potential. Consider a differential equation of the form

$$\frac{d^2\psi}{dx^2} + f(x)\psi(x) = 0 \quad (8.2-1)$$

where $f(x)$ is a given function which is *periodic* with period a so that

$$f(x+a) = f(x). \quad (8.2-2)$$

¹ F. Bloch, *Z. Physik* **52**, 555 (1928); the same result had been known to mathematicians previously as *Floquet's Theorem*. See, for example, E. I. Whittaker and G. N. Watson, *Modern Analysis*, Cambridge University Press (1948), p. 412.

If the potential function $V(x)$ for the one-dimensional Schrödinger equation, of the form (4.8-5), is periodic, then Schrödinger's equation will be a special case of equation (8.2-1). Since (8.2-1) is a linear second-order differential equation, there are (for any given value of the energy ε in 4.8-5) two independent solutions $g(x)$ and $h(x)$ such that

$$\psi(x) = Ag(x) + Bh(x) \quad (8.2-3)$$

represents the most general solution of (8.2-1). Since $f(x+a) = f(x)$, not only $g(x)$ and $h(x)$ but also $g(x+a)$ and $h(x+a)$ satisfy (8.2-1). But any solution of (8.2-1) must be expressable as a linear combination of $g(x)$ and $h(x)$ of the form (8.2-3). In particular, then,

$$\begin{aligned} g(x+a) &= \alpha_1 g(x) + \alpha_2 h(x) \\ h(x+a) &= \beta_1 g(x) + \beta_2 h(x) \end{aligned} \quad (8.2-4)$$

where α_1 , α_2 , β_1 , and β_2 are constants. Then

$$\begin{aligned} \psi(x+a) &= Ag(x+a) + Bh(x+a) \\ &= (\alpha_1 A + \beta_1 B)g(x) + (\alpha_2 A + \beta_2 B)h(x). \end{aligned} \quad (8.2-5)$$

Now $\psi(x+a)$ can always be expressed in the form

$$\psi(x+a) = \lambda\psi(x), \quad (8.2-6)$$

where λ is a constant, provided that the proper value for λ is chosen. Comparing (8.2-5) and (8.2-3) we see that if (8.2-6) is to be satisfied, then

$$(\alpha_1 - \lambda)A + \beta_1 B = 0 \quad (8.2-7)$$

and

$$\alpha_2 A + (\beta_2 - \lambda)B = 0.$$

This system of homogeneous equations in A and B has solutions other than $A = B = 0$ only if

$$\begin{vmatrix} \alpha_1 - \lambda & \beta_1 \\ \alpha_2 & \beta_2 - \lambda \end{vmatrix} = \lambda^2 - (\alpha_1 + \beta_2)\lambda + (\alpha_1\beta_2 - \alpha_2\beta_1) = 0. \quad (8.2-8)$$

The solution of this quadratic equation in λ serves to determine the two possible values of λ for which (8.2-6) is true. Designating them λ_1 and λ_2 , then,

$$\psi(x+a) = \lambda_1 \psi(x) \quad (8.2-9)$$

$$\psi(x+a) = \lambda_2 \psi(x).$$

If we now define k_1 and k_2 such that

$$\lambda_1 = e^{ik_1 a} \quad (8.2-10)$$

$$\lambda_2 = e^{ik_2 a},$$

uation, of the
use of equation
there are (for
 $g(x)$ and $h(x)$)

and define $u_{k_1}(x)$ and $u_{k_2}(x)$ as

$$\begin{aligned} u_{k_1}(x) &= e^{-ik_1 x} \psi(x) \\ u_{k_2}(x) &= e^{-ik_2 x} \psi(x), \end{aligned} \quad (8.2-11)$$

(8.2-3)

then, using (8.2-11), (8.2-9), and (8.2-10) it is clear that

$$\begin{aligned} u_{k_1}(x+a) &= e^{-ik_1(x+a)} \psi(x+a) = e^{-ik_1(x+a)} \lambda_1 \psi(x) \\ &= e^{-ik_1(x+a)} e^{ik_1 a} \psi(x) = e^{-ik_1 x} \psi(x) = u_{k_1}(x). \end{aligned} \quad (8.2-12)$$

(8.2-4)

The function $u_{k_1}(x)$ is thus *periodic* with period a ; the same is found to be true for $u_{k_2}(x)$ in a similar way. According to (8.2-11), then, $\psi(x)$ can then always be written in the form

$$\psi_k(x) = e^{ikx} u_k(x) \quad (8.2-13)$$

(8.2-5)

where $u_k(x)$ is a *periodic* function of period a , and where k represents either k_1 or k_2 as determined above. This is Bloch's theorem; all one-electron wave functions for periodic potentials can be written in this way. In three dimensions Bloch's theorem becomes

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} u_{\mathbf{k}}(\mathbf{r}). \quad (8.2-14)$$

(8.2-6)

The wave functions (8.2-13) and (8.2-14) clearly have the form of plane waves with propagation vector \mathbf{k} modulated by a function whose periodicity is that of the crystal lattice.

(8.2-7)

The Bloch function (8.2-13) can be thought of as the most general way of writing a solution of Schrödinger's equation which leads to the same probability density $\psi^* \psi$ in each unit cell of the crystal. From (8.2-13) it is apparent that in the n th unit cell, due to the periodicity of u_k ,

$$\begin{aligned} \psi(x+na) &= e^{ik(x+na)} u_k(x+na) = e^{ik(x+na)} u_k(x) \\ &= e^{ikna} \psi(x). \end{aligned} \quad (8.2-15)$$

(8.2-8)

Likewise,

$$\psi^*(x+na) = e^{-ikna} \psi^*(x) \quad (8.2-16)$$

(8.2-9)

and

$$\psi^*(x+na) \psi(x+na) = \psi^*(x) \psi(x). \quad (8.2-17)$$

The 3-dimensional wave function (8.2-14) exhibits the same sort of behavior.

In the one-dimensional case, for an *infinite* crystal, one can find a formal solution to Schrödinger's equation for every value of the energy ϵ within certain ranges, and corresponding to each such value of ϵ , by (8.2-8) and (8.2-10), are two values of k . The allowed energy eigenvalues for such a system within these ranges are continuous. If the crystal is of finite extent, however, in which case appropriate physical boundary conditions must be satisfied at the surfaces, then solutions which satisfy Schrödinger's

equation and the boundary conditions can be found only for certain discrete energy eigenvalues. Since for each energy eigenvalue there are only two related values of k , the allowed values of k also form a discrete set. The situation here is very similar to that which was encountered in connection with mechanical vibrations of a periodic lattice and which was discussed in Section 3.3. Of course, when the number of atoms in the crystal becomes very large, the allowed values of ϵ and k , although still discrete, crowd very closely together and in most cases can be regarded as quasi-continuous bands of allowed values. The same qualitative behavior is found for three-dimensional crystals, except that in this case there are more than two allowed values of k for each energy eigenvalue.

For example, if periodic boundary conditions are imposed in a one-dimensional crystal of N atoms (or if the lattice is assumed to have the form of a closed ring of N atoms), then if the wave function ψ is to be single-valued, we must have (using (8.2-15))

$$\psi(x + Na) = e^{ikNa} \psi(x) = \psi(x) \quad (8.2-18)$$

whereby

$$e^{ikNa} = 1$$

or $e^{ika} = (1)^{1/N} = e^{\frac{2\pi in}{N}} \quad (n = 0, 1, 2, \dots, N - 1), \quad (8.2-19)$

recalling the form of the N th roots of unity. Taking the logarithm of both sides of this equation and solving for k , we find that the possible values for k under these boundary conditions are

$$k_n = \frac{2\pi n}{Na} \quad (n = 0, 1, 2, \dots, N). \quad (8.2-20)$$

If N is large, there will be very many allowed values of k as given by (8.2-20), and in this case they may be regarded as forming a quasi-continuous range of values. Of course the N distinct values of k allowed by (8.2-20) do not correspond to the same value of energy. For each constant value of ϵ , there are only the *two* allowed values of k given by (8.2-10).

8.3 THE KRONIG-PENNEY MODEL OF AN INFINITE ONE-DIMENSIONAL CRYSTAL

For the case of the infinite periodic one-dimensional square-well potential shown in Figure 8.2, it is possible to arrive at an exact solution of the Schrödinger equation. This solution was first investigated by Kronig and Penney,² and although it is related to a somewhat idealized periodic potential which is only a crude approximation to that found in the actual crystal, it is nevertheless very useful because it serves to illustrate in a most explicit way many of the important characteristic features of the quan-

² R. de L. Kronig and W. G. Penney, *Proc. Roy. Soc. A* 130, 499 (1931).

ain discrete energ
related values of
e is very similar t
tions of a periodi
e number of atom
ough still discrete
s quasi-continuo
three-dimensiona
alue of k for each

a one-dimensiona
of a closed ring o
must have (using

(8.2-18)

(8.2-19)

n of both sides o
for k under thes

tum behavior of electrons in periodic lattices. The wave functions associated with this model may be calculated in the one-electron approximation by solving Schrödinger's equation

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} (\epsilon - V(x)) \psi(x) = 0 \quad (8.3-1)$$

for a single electron in the periodic potential $V(x)$ as given by Figure 8.2. Since, from

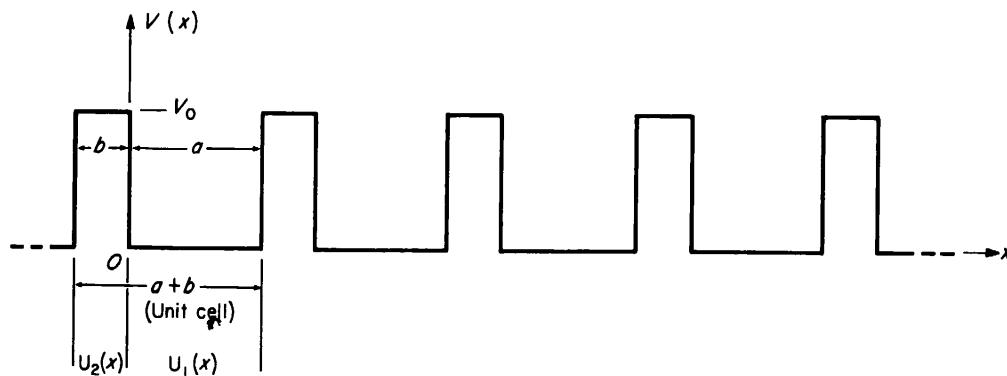


FIGURE 8.2. Ideal periodic square well potential used by Kronig and Penney to illustrate the general characteristics of the quantum behavior of electrons in periodic lattices.

the previous section, the wave functions must have the Bloch form, we may expect that

(8.2-20)

$$\psi(x) = e^{ikx} u(x). \quad (8.3-2)$$

Substituting (8.3-2) into (8.3-1), it is found that the function $u(x)$ must satisfy

$$\frac{d^2u}{dx^2} + 2ik \frac{du}{dx} - \left(k^2 - \alpha^2 + \frac{2mV(x)}{\hbar^2} \right) u(x) = 0 \quad (8.3-3)$$

where

$$\alpha = (2m\epsilon/\hbar^2)^{1/2}. \quad (8.3-4)$$

For the potential of Figure 8.2, then, we shall find that

NITE

$$\frac{d^2u_1}{dx^2} + 2ik \frac{du_1}{dx} - (k^2 - \alpha^2) u_1(x) = 0 \quad (0 < x < a) \quad (8.3-5)$$

and

$$\frac{d^2u_2}{dx^2} + 2ik \frac{du_2}{dx} - (k^2 - \beta^2) u_2(x) = 0 \quad (-b < x < 0), \quad (8.3-6)$$

where $u_1(x)$ represents the value of $u(x)$ in the interval $(0 < x < a)$ and $u_2(x)$ represents the value of $u(x)$ in $(-b < x < 0)$, and where

$$\beta = (2m(\epsilon - V_0)/\hbar^2)^{1/2}. \quad (8.3-7)$$

The differential equations (8.3-5) and (8.3-6) are easily solved by standard procedures

to give

$$u_1(x) = Ae^{i(\alpha-k)x} + Be^{-i(\alpha+k)x} \quad (0 < x < a) \quad (8.3-8)$$

$$u_2(x) = Ce^{i(\beta-k)x} + De^{-i(\beta+k)x} \quad (-b < x < 0), \quad (8.3-9)$$

where A, B, C , and D are arbitrary constants. Note that the quantity β is a pure imaginary one for $0 < \varepsilon < V_0$.

The requirement of continuity for the wave function ψ and its derivative at $x = a$ and $x = -b$ demands that the functions $u(x)$ satisfy these same conditions, since in (8.3-2) e^{ikx} is a well-behaved function. Applying these boundary conditions (and recalling that since $u(x)$ has the periodicity of the lattice, $u_1(a) = u_2(-b)$), we find that

$$A + B = C + D$$

$$i(\alpha - k)A - i(\alpha + k)B = i(\beta - k)C - i(\beta + k)D \quad (8.3-10)$$

$$Ae^{i(\alpha-k)a} + Be^{-i(\alpha+k)a} = Ce^{-i(\beta-k)b} + De^{i(\beta+k)b}$$

$$i(\alpha - k)Ae^{i(\alpha-k)a} - i(\alpha + k)Be^{-i(\alpha+k)a} = i(\beta - k)Ce^{-i(\beta-k)b} - i(\beta + k)De^{i(\beta+k)b}$$

The coefficients A, B, C , and D can thus be determined as the solution of a set of four simultaneous linear homogeneous equations in those quantities. There is no solution other than $A = B = C = D = 0$, of course, unless the determinant of the coefficients vanishes. This requires that

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ \alpha - k & -(\alpha + k) & \beta - k & -(\beta + k) \\ e^{i(\alpha-k)a} & e^{-i(\alpha+k)a} & e^{-i(\beta-k)b} & e^{i(\beta+k)b} \\ (\alpha - k)e^{i(\alpha-k)a} & -(\alpha + k)e^{-i(\alpha+k)a} & (\beta - k)e^{-i(\beta-k)b} & -(\beta + k)e^{i(\beta+k)b} \end{vmatrix} = 0. \quad (8.3-11)$$

Expanding the determinant, one can show after quite a lot of tedious but straightforward algebra that (8.3-11) can be expressed as

$$-\frac{\alpha^2 + \beta^2}{2\alpha\beta} \sin \alpha a \sin \beta b + \cos \alpha a \cos \beta b = \cos k(a + b). \quad (8.3-12)$$

Since in the range ($0 < \varepsilon < V_0$), β as defined by (8.3-7) is imaginary, for these values of energy it is most convenient to express (8.3-12) in a slightly different form. Letting

$$\beta = i\gamma \quad (8.3-13)$$

in this region, and noting that $\cos ix = \cosh x$ and $\sin ix = i \sinh x$, Equation (8.3-12) can be written as

$$\frac{\gamma^2 - \alpha^2}{2\alpha\gamma} \sinh \gamma b \sin \alpha a + \cosh \gamma b \cos \alpha a = \cos k(a + b), \quad (8.3-14)$$

where γ is a real positive quantity in the interval $(0 < \varepsilon < V_0)$, just as β is in the interval $(V_0 < \varepsilon < \infty)$. We may thus use (8.3-12) most conveniently when $(V_0 < \varepsilon < \infty)$ and (8.3-14) when $(0 < \varepsilon < V_0)$.

The wave functions (8.3-2) must, like all wave functions, be well-behaved functions as x approaches $\pm\infty$. Since $u(x)$ is a periodic function whose values are the same in each unit cell, no difficulties arise on its account, provided that the factor e^{ikx} in (8.3-2) is well behaved under these conditions. But e^{ikx} is well behaved at both $+\infty$ and $-\infty$ only if k is real, whereby e^{ikx} is oscillatory. If k were imaginary, e^{ikx} would diverge to infinity either at $+\infty$ or $-\infty$, and the resulting expression for $\psi(x)$ would not behave properly as a wave function. We must therefore accept only functions of the form (8.3-2) with *real* values for k . The above expressions (8.3-12) and (8.3-14) have, on the left-hand side, a function of the form $K_1 \sin \alpha a + K_2 \cos \alpha a$ which must equal $\cos k(a+b)$. If for a given value of energy the function on the left-hand side of these equations should have a value in the range between $+1$ and -1 , then the required value for $\cos k(a+b)$ is obtained with a *real* value for the argument $k(a+b)$. On the other hand, if the value of the function on the left-hand side of (8.3-12) or (8.3-14) should lie outside that range, it would mean that $\cos k(a+b)$ would have to be greater than $+1$ or less than -1 , which would, in turn, require that the argument $k(a+b)$ be a *complex* number with an imaginary part other than zero. Under these circumstances the solutions (8.3-2) would not behave properly at infinity and would not satisfy the physical requirement for wave functions of the system. The energies associated with such values of k would simply be forbidden to the electron.

It is possible to write the left-hand sides of (8.3-12) and (8.3-14) in the form $K_3 \cos(\alpha a - \delta)$, where $K_3 = (K_1^2 + K_2^2)^{1/2}$ and $\tan \delta = K_1/K_2$. In this way (8.3-12) and (8.3-14) can be written as

$$\left[1 + \frac{(\alpha^2 - \beta^2)^2}{4\alpha^2\beta^2} \sin^2 \beta b \right]^{1/2} \cos(\alpha a - \delta) = \cos k(a+b), \quad (8.3-15)$$

where $\tan \delta = -\frac{\alpha^2 + \beta^2}{2\alpha\beta} \tan \beta b \quad (V_0 < \varepsilon < \infty)$

and $\left[1 + \frac{(\alpha^2 + \gamma^2)^2}{4\alpha^2\gamma^2} \sinh^2 \gamma b \right]^{1/2} \cos(\alpha a - \delta') = \cos k(a+b), \quad (8.3-16)$

where $\tan \delta' = \frac{\alpha^2 + \gamma^2}{2\alpha\gamma} \tanh \gamma b \quad (0 < \varepsilon < V_0)$

From these expressions it is clear that in both cases the left-hand side has the form of a cosine function times a modulating factor whose amplitude is *invariably* greater than unity. The value of this modulating factor is actually a maximum for $\alpha = 0$ (hence for $\varepsilon = 0$) and approaches unity in the limit of large energies, where $\alpha \cong \beta$.

When the left-hand sides of (8.3-15) and (8.3-16) are plotted as a function of energy, in which connection one should note from (8.3-4) and (8.3-7) that

$$\alpha^2 - \beta^2 = \alpha^2 + \gamma^2 = 2mV_0/\hbar^2 = \text{const.}, \quad (8.3-17)$$

the results are as illustrated in Figure 8.3. In this figure the left-hand side of (8.3-15)

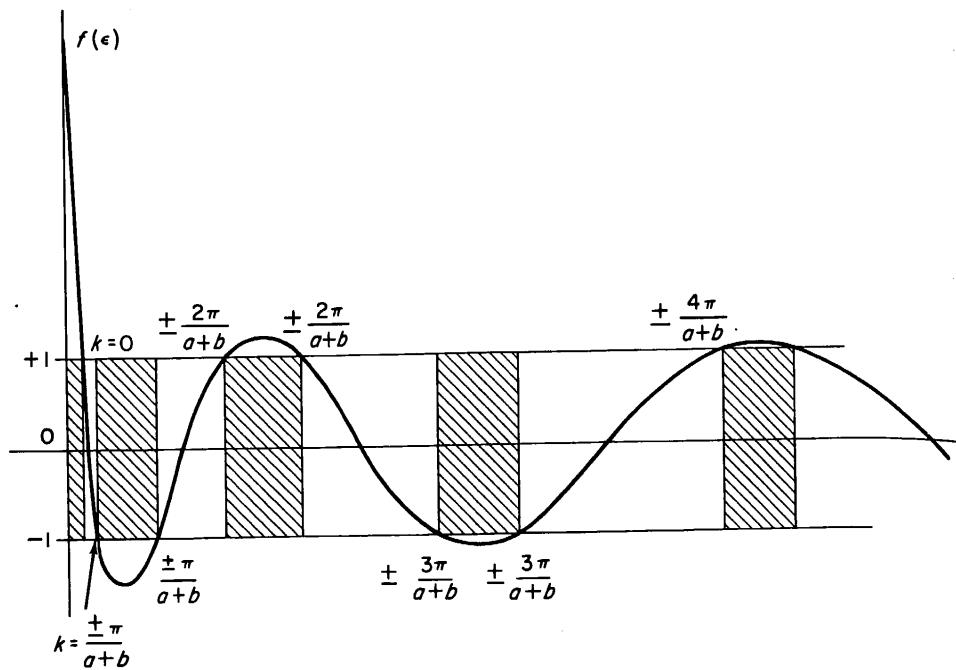


FIGURE 8.3. A plot of the functions on the left-hand side of (8.3-15) and (8.3-16) versus energy. The shaded regions show forbidden energy bands where the value of k is complex, the unshaded regions allowed energy bands corresponding to real values of k .

or (8.3-16) is plotted as a function of energy. When the ordinate of the curve lies between $+1$ and -1 , there exists a real value for k corresponding to physically possible wave functions. Outside these limits, however, k must be complex with a nonzero imaginary part. Such values for k can never lead to physically well-behaved wave functions; the corresponding ranges of energy are forbidden and are shown in Figure 8.3 as shaded regions. There are thus formed alternate regions of allowed energy eigenvalues and forbidden regions. These regions are usually referred to as allowed and forbidden *energy bands*, and the grouping of the permitted energy values into these bands is one of the most important and characteristic features of the behavior of electrons in periodic lattices. It can be shown that energy bands having the same qualitative aspects as those shown in Figure 8.3 are formed no matter what the detailed form of the potential is, so long as it is periodic.

Using (8.3-15), (8.3-16), (8.3-4), and (8.3-7) it is possible to plot a curve showing the energy ϵ as a function of k . The result is illustrated schematically in Figure 8.4. The forbidden energy bands and the relationship of ϵ vs k within the various allowed bands have been assigned in accord with the scheme shown in Figure 8.3. Since $\cos^{-1} k(a + b)$ is not a single valued function this assignment is of necessity somewhat arbitrary. For large energies, it is apparent that the function $\epsilon(k)$ approaches the free-electron relation $\epsilon = \hbar^2 k^2 / 2m$ (shown as a dotted curve in Figure 8.4) quite closely within the allowed bands. Also, for large energies it is found that the allowed bands become very broad and the forbidden regions quite narrow. Nevertheless, the curve $\epsilon(k)$ always has zero slope at the edges of the allowed bands, that is, at $k = \pm n\pi/(a + b)$, where n is an integer. This result can be proved by direct differentiation of (8.3-15) and (8.3-16), but we shall not go through the details of the cal-

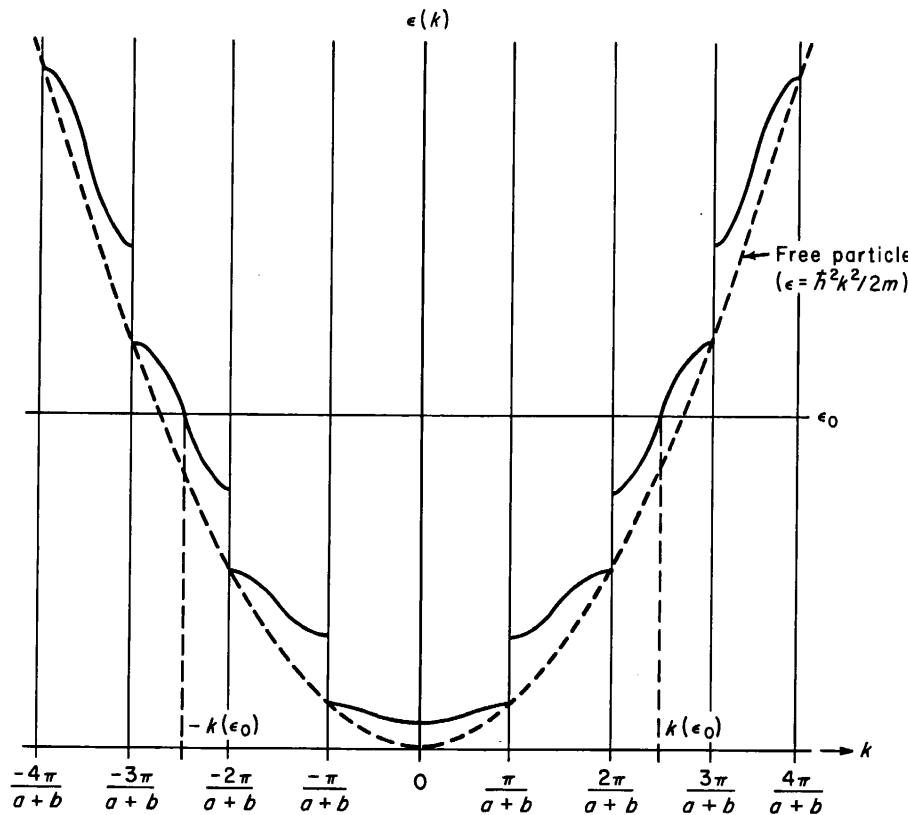


FIGURE 8.4. The energy ϵ plotted as a function of k according to (8.3-15) and (8.3-16).

culation. This feature of the $\epsilon(k)$ curve is quite general, and can be shown to occur even in the three-dimensional case independent of the precise mathematical form of the potential function.

CRYSTAL MOMENTUM AND EFFECTIVE MASS

If the function $u_k(x)$ in (8.2-13) is a constant, then the wave function has the form $e^{\pm ikx}$, corresponding to a perfectly free electron of momentum $p = \pm \hbar k$ whose energy, according to the results of Section 4.9, would be

$$\epsilon = \frac{\hbar^2 k^2}{2m}. \quad (8.4-1)$$

This relation is shown as the dashed curve in Figure 8.4. For large values of ϵ the actual ϵ vs k relation conforms quite closely to this relation. It is clear from (8.2-13) and from the results of the previous section that k is a constant of the motion, that $\hbar k$ will have the dimensions of a momentum, and that as the energy of the electron increases (the particle becoming thereby more nearly "free") the values of k in general approximate those of the free-particle momentum divided by \hbar . It can easily be seen that these conclusions must hold no matter what particular form the periodic potential

takes. It is therefore customary to refer to $\hbar k$ as the *crystal momentum*, and we shall soon see that in many cases the dynamical behavior of the electron in the crystal lattice with respect to crystal momentum is very similar to that of a free particle with respect to the real momentum. To make the distinction between the actual momentum and the crystal momentum perfectly clear, we must note that due to the presence of the lattice potential, the true instantaneous momentum of an electron is not a constant of the motion at all and is not directly calculable by the methods of quantum mechanics, except as an average value, while, for a state of given energy, the crystal momentum $\hbar k$ is a perfectly well-defined constant value, just as is the true momentum of a free particle of given energy.

Let us now consider the motion of an electron in the crystal under the influence of an applied electric field. For convenience in visualizing the motion of an electron in such a situation we shall have to localize the wave function by superposing solutions having different values of k , as in Section (4.9). If this is done, then the group velocity associated with the electron "wave packet" is

$$v_g = \frac{d\omega}{dk} = \frac{1}{\hbar} \frac{d\varepsilon}{dk}, \quad (8.4-2)$$

where, of course, ε and ω are connected by the Planck relation $\varepsilon = \hbar\omega$. Suppose that the electron is acted upon by an external electric field E , acquiring an increase in velocity dv_g over a distance dx in a time dt . Then, using (8.4-2) we see that

$$d\varepsilon = \frac{d\varepsilon}{dk} dk = -eE dx = -eEv_g dt = -\frac{eE}{\hbar} \frac{d\varepsilon}{dk} dt, \quad (8.4-3)$$

whereby

$$dk = -\frac{eE}{\hbar} dt, \quad (8.4-4)$$

or

$$\hbar \frac{dk}{dt} = \frac{dp}{dt} = -eE = F, \quad (8.4-5)$$

where we now use the symbol p to denote the *crystal momentum*. Equation (8.4-5) tells us simply that the time rate of change of crystal momentum equals the force $-eE$. It is thus the analogue of Newton's law, showing that the crystal momentum of the electron in a periodic lattice changes under the influence of an applied field in the same way as does the true momentum of a free electron *in vacuo*.

If one differentiates (8.4-2) with respect to time, the result is

$$\frac{dv_g}{dt} = \frac{1}{\hbar} \frac{d}{dt} \left(\frac{d\varepsilon}{dk} \right) = \frac{1}{\hbar} \frac{d^2\varepsilon}{dk^2} \frac{dk}{dt}, \quad (8.4-6)$$

which, using (8.4-5), can be written as

$$\frac{dv_g}{dt} = \frac{-eE}{\hbar^2} \frac{d^2\varepsilon}{dk^2} = \frac{F}{m^*} \quad (8.4-7)$$

ntum, and we shall
tron in the crystal
a free particle with
actual momentum
to the presence of
is not a constant
of quantum mech-
the crystal momen-
e momentum of a

under the influence
tion of an electron
erposing solutions
the group velocity

(8.4-2)

$\hbar\omega$. Suppose that
ing an increase in
see that

(8.4-3)

(8.4-4)

(8.4-5)

Equation (8.4-5)
als the force $-eE$.
momentum of the
d field in the same

(8.4-6)

(8.4-7)

where the *effective mass* m^* is given by

$$m^* = \frac{\hbar^2}{(d^2\varepsilon/dk^2)}. \quad (8.4-8)$$

Equation (8.4-7) is essentially Newton's force equation. The proportionality factor relating the force eE and the acceleration dv_g/dt may be regarded as the *effective mass* of the electron. The actual gravitational mass of the electron, however, has nothing directly to do with the effective mass; it is the quantity $d^2\varepsilon/dk^2$ which is the important factor. If the electron is really a free electron, then ε and k are related by (8.4-1) and (8.4-8) reduces to $m^* = m$. If the energy is a *parabolic* function of k , having the form

$$\varepsilon = C(k - k_0)^2, \quad (8.4-9)$$

then the effective mass has the *constant* value $m^* = \hbar^2/2C$ and the dynamical behavior of the electron will be the same as that of a free particle with this effective mass. If the energy is not a parabolic function of k , then the effective mass will not be constant with energy, the dynamical behavior will be that of a particle of variable mass, and the situation will be much more involved. In any case, to a first approximation, the entire effect of the periodic crystal potential is to replace the free electron mass with an effective mass.

Fortunately, the relation connecting ε and k is almost always parabolic or nearly parabolic over the range of energies which is accessible to an electron in the crystal. From Figure 8.4 it is apparent that the ε vs k curve is always parabolic in form at the bottom and top of the allowed energy bands. Even when the electrons of the system are in other regions, however, what is important, usually, is that the effective mass be essentially constant over an energy interval of the order of the average amount of energy which a particle may gain or lose in a single scattering event, or between scattering events in a time of the order of the relaxation time. This amount of energy is of the order of kT . But at 300°K, kT amounts to only 0.025 eV, and the total width in energy of a band which might play an important role in transport processes in typical crystalline substances is much greater than this—ordinarily of the order of 1 eV. In single scattering events, or in the interval between successive scattering events, then, an electron will usually be confined to a short segment of the ε vs k curve, which can ordinarily be regarded as approximately parabolic.

It is clear from this that the free-electron picture of a metal is to a very great extent justified, the only important correction being that the free electron mass should be in all cases replaced with an appropriate effective mass related to the form of the ε versus k curve by (8.4-8). We may thus conclude that all the results derived in Chapter 7 on the basis of the free-electron theory are correct, provided that the electron mass is replaced with the proper effective mass. The effective mass can be deduced by comparison of experimental electronic specific heat data with the free-electron result given as Equation (7.7-25); it may also be calculated quantum mechanically starting with self-consistent atomic wave functions for the atoms of the crystal, which are then combined to give a crystal potential appropriate for the substance in question. The values so calculated are usually in reasonable agreement with the experimentally determined ones. For most metals the effective mass ranges between one-half and

twice the free-electron mass, although for many of the transition metals it is much higher, and for certain semiconducting compounds of the III-V type it may be much lower.

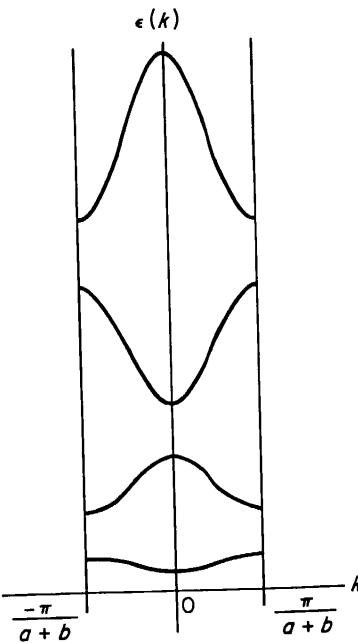
8.5

REDUCED ZONE REPRESENTATION: ELECTRONS AND HOLES

The assignment of ranges of values for k within each allowed energy band, leading to the curve of Figure 8.4, is shown in Figure 8.3. Since the only requirement which must be fulfilled within each band is that $-1 \leq \cos k(a+b) \leq +1$, it is clear that this assignment of values for k for each band is not a unique one. The advantage of the assignment shown in Figure 8.3 is that it leads to the so-called *expanded representation* of Figure 8.4 in which the relationship with the free particle ϵ versus k curve is clearly apparent. We could, however, translate any of the curve segments shown in Figure 8.4 to the right or left parallel to the k -axis at a distance $2n\pi/(a+b)$, where n is an integer, and still satisfy the relations (8.3-15) and (8.3-16), since any such transformation leaves the value of $\cos k(a+b)$ unchanged.

It is sometimes very useful to make a particular transformation of this type, by translating the various segments of the ϵ versus k curve to the right or to the left parallel to the k -axis, through distances which are integral multiples of $2\pi/(a+b)$.

FIGURE 8.5. Schematic representation of the ϵ versus k plot of Figure 8.4 transformed to the reduced-zone representation.



so that they all fit within the interval $-\pi/(a+b) < k < \pi/(a+b)$. The resulting representation of the energy versus propagation constant relationship is called the *reduced zone representation*. The reduced zone representation of the curve of Figure 8.4 is shown in Figure 8.5. It will in either case be noted that the Bloch wave function (8.2-13) satisfy the Bragg reflection condition (2.1-1) at the points $k = n\pi/(a+b)$,

metals it is much
e it may be much

RONS

rgy band, leading
equirement which
it is clear that the
advantage of the
ded representation
k curve is clearly
own in Figure 8.
ere n is an integ
h transformatio

n of this type, b
it or to the le
les of $2\pi/(a + b)$

discussed in connection with mechanical vibrations of the lattice in Section 3.3. At these points the group velocity $\hbar^{-1} de/dk$ is zero, corresponding to a standing wave which represents an electron at rest. The electron may be regarded as undergoing an internal diffraction by the lattice potential under these circumstances. We shall have more to say about this subject in a later section.

In the infinitely large crystal of Section 8.3, the permitted energies form a continuum of values within the allowed bands. If the crystal is of finite extent and contains a total of N atoms, however, we have already seen from (8.2-20) that there will be instead just N distinct allowed eigenstates of the crystal momentum k within each allowed energy band. From Figure 8.4 it is easy to see that for a given energy ϵ_0 the two allowed crystal momentum values are $k(\epsilon_0)$ and $-k(\epsilon_0)$. Physically, the former represents a state in which an electron is moving to the right with positive crystal momentum $k(\epsilon_0)$, while the latter represents a state in which an electron moves to the left with equal and opposite crystal momentum; the energy is clearly the same for either case. The effect of electron spin here, as in isolated atomic systems, is simply to *double* the degeneracy factor associated with all the energy levels of the system, since each quantum state must be regarded as having split into two, one accommodating an electron with "spin up" and the other an electron with "spin down." When the effect of electron spin is included, then, we must conclude that within each allowed energy band of this system there are just $2N$ quantum states.

At a temperature of absolute zero the electrons of the system will occupy these states, one per state, as required by the Pauli exclusion principle, from the lowest state up to a given energy determined by the number of available states, their distribution in energy and the number of electrons in the crystal. This energy is, of course, the Fermi energy of the crystal at zero temperature. In a one-dimensional crystal some of the allowed bands will be entirely filled, some will be entirely empty, and one may be partially filled.³ If an electric field is applied, it is obvious that no current can be contributed from the unfilled bands; it is equally true, however, that no current can arise from the filled bands. This can be understood by noting that the current density arising from a given band will be

$$I = -n_0 e \bar{v} \quad (8.5-1)$$

where \bar{v} is the average velocity, and n_0 is the number of electrons per unit volume belonging to that band. But \bar{v} can be expressed as

$$\bar{v} = \frac{1}{n_0 V} \sum_i v_i, \quad (8.5-2)$$

where the summation is taken over all the velocities associated with individual electrons within the volume V of material. Using (8.5-2), (8.5-1) can then be written as

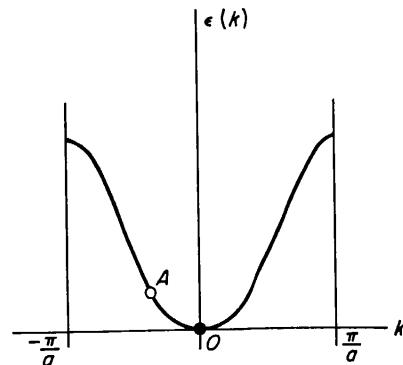
$$I = -\frac{e}{V} \sum_i v_i. \quad (8.5-3)$$

³ The situation in two- and three-dimensional systems is more complex. In those instances different allowed bands may *overlap* in energy, and more than one band may be partially filled. These systems will be discussed in detail later.

This, however, must yield zero when summed over a full band, because due to the symmetry of the curves of Figure 8.4 or Figure 8.5 about the $\epsilon(k)$ axis, for every state of positive velocity $\hbar^{-1} \partial\epsilon/\partial k$ corresponding to a point of positive slope, there is a state corresponding to a negative velocity of equal magnitude (with negative slope) at $k' = -k$. We must conclude that only bands which are *partially* filled can contribute to a current flow. This can be understood on physical grounds by observing that in a partially filled band there are always electrons which can be excited gradually to unoccupied states of higher energy and momentum, whereas in a filled band, on account of the Pauli exclusion principle, this gradual field excitation can never occur, since all the states are already occupied.

A single electron at rest in an otherwise unoccupied band will occupy a state of lowest energy at the bottom of that band in the absence of thermal excitation, as shown in Figure 8.6. If an electric field E_0 is now impressed upon the crystal, the

FIGURE 8.6. An electron near the bottom of an allowed band, subject to forces which tend to accelerate it along the $-x$ direction.



electron will experience a force $-eE_0$ and will gradually be excited through states of ever-increasing energy and (negative) momentum, acquiring a higher and higher momentum in the $-x$ direction, according to (8.4-5). According to the predictions of the theory of Sections 8.2–8.4, this process would continue to excite the electron through successively higher energies until it reached the top of the band at $x = \pi/a$, where a is the lattice spacing. Since the points $x = \pi/a$ and $x = -\pi/a$ are equivalent, the electron may then be regarded as “reappearing” at $x = \pi/a$, decreasing in energy along the right-hand portion of the curve as k goes from π/a to zero, arriving finally at the starting point, whereby the whole process is repeated. The motion of the electron under these circumstances would be oscillatory (Zener’s oscillation). It is important to note that there is no mechanism inherent in the theory of Sections 8.2–8.4 which serves to limit this motion; it will continue indefinitely according to the theory outlined there. If the potential were perfectly parabolic instead of having the form shown in Figure 8.6, the electron would acquire unlimited velocity and energy. We must conclude in either case that the motion of a free electron in a *perfectly periodic* potential is *unimpeded* by the lattice, in the sense that there is no scattering process associated with the presence of the lattice itself which stops or randomizes the velocity of an electron.

If there is a *deviation* from the perfect periodicity of the lattice, however, this result no longer holds true. In this case there will always be some probability of a sudden random transition to a state associated with another value of k within the band. On the *average* these random transitions bring the electron back to $k = 0$, and we conclude that there is a *scattering process* associated with the departure of the

use due to the
for every state
slope, there is a
negative slope
can contribute
rving that in a
l gradually to
lled band, on
n never occur,

up a state of
excitation, as
ie crystal. the

lattice potential from perfect periodicity. The deviation from periodicity may be caused by the fact that the lattice may be at a temperature higher than absolute zero, in which case the thermal vibrations which are set up introduce a slight aperiodicity, the instantaneous positions of the atoms no longer coinciding exactly with the periodic lattice of their equilibrium positions. This situation may be discussed in terms of a picture in which the electrons are scattered by phonons. Deviations from periodicity may also be caused by the presence of impurity atoms as well as by lattice vacancies, interstitial atoms, dislocations and other structural defects of the crystal lattice. These effects result in the phenomena of lattice scattering, impurity scattering, and defect scattering which were discussed in the preceding chapter. Discussion of the quantitative aspects of these processes will be postponed to a later chapter.

These scattering processes prevent the Zener oscillation behavior from ever occurring in practice.⁴ What actually happens in all experimentally realizable situations is that an electron starting from O in Figure 8.6 is accelerated as described by the theory of Sections 8.2-8.4 until it arrives at some point A a small distance along the parabolic lower portion of the curve. Its behavior in this range is that of a free electron of mass m^* as given by (8.4-8). At point A it is scattered by one of the mechanisms described previously, arriving *on the average* back at O , whereupon the cycle is repeated. The behavior is then just that of a free particle subject to the scattering processes discussed above and in the last chapter, leading to Ohm's law, the thermal conductivity, the thermoelectric effect, and all the other results of the free-electron theory of metals, the only difference being that the electron mass is replaced by the effective mass throughout. The distance OA along the ϵ versus k curve in Figure 8.6 is exaggerated for the sake of clarity, and in most cases arising in practice would be very much smaller.

For a band containing a relatively small number of electrons, the current obtained when a small voltage is impressed is most simply given by (8.5-3), summed over all the electrons in the band. If the band is nearly full, however, and there are only a few empty states (which in the equilibrium state will be concentrated near the top of the band⁵) the current equation (8.5-3) is best expressed in another way, by writing (8.5-3) in the form

$$I = -\frac{e}{V} \sum_i v_i = -\frac{e}{V} \left[\sum_j v_j - \sum_k v_k \right] = +\frac{e}{V} \sum_k v_k. \quad (8.5-4)$$

Here the sum over i represents the sum over all velocity states occupied by electrons, the sum over j represents the sum over all velocity states in the band and the sum over k represents the sum over all *unoccupied* velocity states. As we have seen before in connection with (8.5-3), the sum over j , taken over all states in the band, must vanish. The remaining sum over the unoccupied states corresponds to a current which could be produced by a corresponding number of *positive* charge carriers! It is possible (and advantageous, as we shall see) to express the current from an almost full band as a

⁴ A quantitative discussion of the Zener oscillation is given by E. Spenke, *Electronic Semiconductors*, McGraw-Hill Book Co., Inc., New York (1958), pp. 229, 257.

⁵ In a nearly empty band, the electrons occupy the lowest available energy states in the absence of thermal excitation. This is also true in a nearly filled band, the result being that the unoccupied states are at the *top* of the band. In an almost completely filled band, then, the empty states "gravitate" upward on a diagram such as that of Figure 8.5, in which the ordinate represents the energy of an electron.

current derived from the motion of a comparatively small number of empty electronic states or *holes*, which behave like positive particles, rather than a very large number of electrons. The velocity associated with a hole is that which an electron would have if it were to occupy the empty energy state, which is ordinarily near the top of the energy band. But since the ϵ versus k relation there is concave downward, $d^2\epsilon/dk^2$ is negative, giving a negative electron effective mass from (8.4-8). A particle with negative effective mass experiences an acceleration in a direction opposite to that of the applied force. A negative particle with negative effective mass would thus be accelerated in the same direction as the applied field, and would thus exhibit the same dynamical behavior as a *positive particle of positive mass*. We may thus regard the situation in a nearly filled band as one involving a relatively small number of positive particles of positive mass, which we shall refer to as holes, whose velocities and momenta are those corresponding to the unoccupied electronic states in the band. We shall see in a later chapter that in certain materials the physical nature and dynamical behavior of holes is very easily visualized in terms of defects in the electronic valence bonds which connect nearest neighbor atoms and provide the cohesive forces which hold the crystal together. We have not discussed in detail the behavior of electrons and holes under the influence of magnetic force, but it can be shown⁶ that they move as would negative or positive particles of effective mass m^* as given by (8.4-8) under the influence of the usual magnetic force $q(v \times B)/c$.

8.6 THE FREE-ELECTRON APPROXIMATION

In the foregoing sections we have seen how a particular periodic potential model (shown in Figure 8.2) led to the formation of allowed energy bands and an ϵ versus k relation of the form shown in Figure 8.4. The form of this relationship was such that the quantity $\hbar k$ could be advantageously regarded as a "crystal momentum," in terms of which the dynamics of electrons in applied force fields could be described quite simply. The resulting picture of electronic behavior was quite similar to the free-electron picture of Chapter 7, except that an effective mass had to be substituted for the free-electron mass, and that the charge carriers in a nearly filled band had to be viewed as positive holes. We have emphasized that these qualitative conclusions are independent of the precise form of the potential function, as long as it is periodic. The potential function of Figure 8.2 was adopted only for illustrative purposes, since it allowed an exact solution of Schrödinger's equation.

In actual crystals the potential function which is used must be somehow related to the actual potential experienced by an electron due to the ion cores and all the other electrons of the crystal. An exact solution of this problem, even in the one-electron approximation, is impossible; it is therefore customary to approach the problem from the viewpoint of either the *free-electron approximation* or the *tight-binding approximation*, whichever seems most appropriate to the particular situation at hand.

In the free-electron approximation, the total energy of the electron is assumed

⁶ R. A. Smith, *Wave Mechanics of Crystalline Solids*, John Wiley and Sons, New York (1955), p. 458 ff.

of empty electronic
ry large number of
electron would have
ear the top of the
vnward, d^2e/dk^2 is
ticle with negative
that of the applied
be accelerated in
e same dynamical
l the situation in a
ositive particles of
and momenta are
1. We shall see in
nodynamical behavior
nic valence bonds
es which hold the
electrons and holes
ey move as would
4-8) under the in-

to be always large compared to the periodic potential energy. Under these conditions the allowed bands will be broad and the forbidden energy regions quite narrow. These circumstances are never perfectly realized in any actual crystal, because the potential always goes to $-\infty$ at the ion cores, but for the outermost electrons in many simple metals, including the alkali metals, the requirements are met fairly well over most of the volume of the crystal.

Adopting for simplicity a one-dimensional crystal potential model, and writing the periodic potential $V(x)$ in the form

$$-\frac{2mV(x)}{\hbar^2} = \gamma f(x), \quad (8.6-1)$$

where $f(x)$ has the periodicity of the lattice, the Schrödinger equation (8.3-1) may be written

$$\frac{d^2\psi}{dx^2} + (k_0^2 + \gamma f(x))\psi(x) = 0. \quad (8.6-2)$$

where k_0 is related to the total energy ϵ by

$$\epsilon = \frac{\hbar^2 k_0^2}{2m}. \quad (8.6-3)$$

Since $f(x)$ is periodic, it can be expressed as a Fourier series of the form

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{-2\pi i n x/a} \quad (8.6-4)$$

with

$$c_n = \frac{1}{a} \int_0^a f(x) e^{2\pi i n x/a} dx, \quad (8.6-5)$$

where a is the lattice constant of the crystal. It is obvious from (8.6-1) that the coefficients c_n are related to the Fourier expansion coefficients of the periodic potential itself by

$$V_n = -\frac{\hbar^2 \gamma}{2m} c_n, \quad (8.6-6)$$

where V_n refer to the Fourier expansion coefficients of $V(x)$.

Since the wave functions must have the Bloch form $e^{ikx} u_k(x)$, and since $u_k(x)$ must be a periodic function of period a , expressible as a Fourier series, we may write

$$u_k(x) = \sum_{n=-\infty}^{\infty} b_n e^{-2\pi i n x/a} \quad (8.6-7)$$

and

$$\psi(x) = e^{ikx} u_k(x) = e^{ikx} \sum_n b_n e^{-2\pi i n x/a}. \quad (8.6-8)$$

For perfectly free electrons $\gamma = 0$ and the solutions become free-particle wave functions

of the form

$$\psi(x) = b_0 e^{ik_0 x}. \quad (8.6-9)$$

In this limit $u_k(x) \rightarrow b_0$ and $k \rightarrow k_0$, so that in the general case, we must expect that as γ approaches zero, all the b_n except b_0 approach zero. We are thus led to write an approximate expression for the wave function of the form

$$\psi(x) = b_0 e^{ikx} + \gamma \left[e^{ikx} \sum_{n \neq 0} b_n e^{-2\pi i n x / a} \right], \quad (8.6-10)$$

which we may expect to be valid for *small* values of γ (that is, for the conditions outlined above under which the free-electron approximation is valid). It is easy to see that this wave function has the Bloch form, with u_k given by

$$u_k(x) = b_0 + \gamma \sum_{n \neq 0} b_n e^{-2\pi i n x / a}. \quad (8.6-11)$$

Substituting (8.6-10) into Schrödinger's equation (8.6-2), one may obtain

$$\begin{aligned} b_0(k_0^2 - k^2)e^{ikx} + \gamma \sum_{n \neq 0} [(k_0^2 - k_n^2)b_n + b_0 c_n] e^{ik_n x} \\ + \gamma^2 \sum_{n \neq 0} \sum_{n' \neq 0} b_{n'} c_n e^{i(k - \frac{2\pi n}{a} - \frac{2\pi n'}{a})x} = 0 \end{aligned} \quad (8.6-12)$$

where

$$k_n = k - \frac{2\pi n}{a}. \quad (8.6-13)$$

Since we are interested in the solution in the limit where $\gamma \rightarrow 0$, we may as a first approximation neglect the γ^2 term in (8.6-12); then, multiplying the equation through by $e^{-ik_m x}$, integrating over the unit cell from $x = 0$ to $x = a$, we find

$$\begin{aligned} b_0(k_0^2 - k^2) \int_0^a e^{2\pi i m x / a} dx + \gamma \sum_{n \neq 0} \left\{ [(k_0^2 - k_n^2)b_n + b_0 c_n] \right. \\ \times \left. \int_0^a e^{2\pi i(m-n)x/a} dx \right\} = 0. \end{aligned} \quad (8.6-14)$$

If $m = 0$, the second integral vanishes for all values of n in the summation, whereby

$$b_0(k_0^2 - k^2)a = 0 \quad \text{or} \quad k = k_0, \quad (8.6-15)$$

while if $m \neq 0$, the first integral vanishes, and the second gives zero except when $n = m$, in which case one obtains

$$\gamma[(k_0^2 - k_m^2)b_m + b_0 c_m]a = 0$$

$$\text{or} \quad b_m = \frac{b_0 c_m}{k_m^2 - k_0^2} = \frac{b_0 c_m}{k_m^2 - k^2}. \quad (8.6-16)$$

To this order of approximation, then, from (8.6-15) and (8.6-3) the relation between ϵ and k is the same as that for a free particle. The wave function is obtained by substituting the values (8.6-16) for b_n into (8.6-10), whence

$$\psi(x) = b_0 e^{ikx} \left[1 - \gamma \sum_{n \neq 0} \frac{c_n}{k^2 - k_n^2} e^{-2\pi i n x/a} \right]. \quad (8.6-17)$$

According to (8.6-15) the first-order correction to the free-particle energy arising from the periodic potential is zero. A second-order energy correction can be obtained, however, by retaining the γ^2 term in (8.6-12). Multiplying (8.6-12) by e^{-ikx} and integrating from $x = 0$ to $x = a$, one may obtain

$$\begin{aligned} b_0(k_0^2 - k^2) + \gamma \sum_{n \neq 0} [(k_0^2 - k_n^2)b_n + b_0 c_n] \int_0^a e^{-2\pi i n x/a} dx \\ + \gamma^2 \sum_{n \neq 0} \sum_{n' \neq 0} b_{n'} c_n \int_0^a e^{-2\pi i (n+n')x/a} dx = 0. \end{aligned} \quad (8.6-18)$$

The first integral above is zero for all allowed values of n , while the second integral is zero unless $n = -n'$. The equation thus reduces to

$$b_0(k_0^2 - k^2)a + \gamma^2 \sum_{n' \neq 0} b_{n'} c_{-n'} a = 0. \quad (8.6-19)$$

If $f(x)$ is given as a Fourier series by (8.6-4), then, substituting $-n$ for n ,

$$f(x) = \sum_{n=-\infty}^{\infty} c_{-n} e^{2\pi i n x/a}, \quad (8.6-20)$$

while, taking the complex conjugate of both sides of (8.6-4) and noting that since $f(x)$ is a real function, $f^*(x) = f(x)$, one may also write

$$f(x) = \sum_{n=-\infty}^{\infty} c_n^* e^{2\pi i n x/a}. \quad (8.6-21)$$

Since the Fourier coefficients associated with the representation of a given function are unique, we must have

$$c_{-n} = c_n^*. \quad (8.6-22)$$

Using the first approximation expression (8.6-16) for the coefficients b_n and the relation (8.6-22), (8.6-19) can be written in the form

$$k_0^2 = k^2 + \gamma^2 \sum_{n \neq 0} \frac{c_n^* c_n}{k^2 - k_n^2} = k^2 + \gamma^2 \sum_{n \neq 0} \frac{c_n^* c_n}{k^2 - \left(k - \frac{2\pi n}{a} \right)^2}. \quad (8.6-23)$$

Using (8.6-3) and (8.6-6) to express this result in terms of ϵ and V_n , it is easy to show

that

$$\epsilon = \frac{\hbar^2 k^2}{2m} + \sum_{n \neq 0} \frac{|V_n|^2}{\left(\frac{\hbar^2 k^2}{2m}\right) - \frac{\hbar^2}{2m} \left(k - \frac{2\pi n}{a}\right)^2}. \quad (8.6-24)$$

These solutions are satisfactory only so long as k^2 does *not* approach one of the values k_n^2 . If $k^2 \cong k_n^2$ for some value of n , one of the quantities $k^2 - k_n^2$ in the denominators of the summation terms in (8.6-17) will approach zero, and that term will remain quite large despite the fact that γ is very small. In this case the assumed expression (8.6-10) is not a good approximation for the wave function, and another form must be used. If $k^2 \cong k_n^2$, then

$$k = \pm k_n = \pm k \mp \frac{2\pi n}{a}. \quad (8.6-25)$$

For the upper sign, this equation gives $n = 0$, which is excluded from (8.6-17) in any case and hence is of no particular interest; for the lower sign one obtains

$$k = \frac{n\pi}{a} \quad (8.6-26)$$

which, since n may take on both positive and negative values, refers to all the band edge points. We shall have to use a separate treatment for values of k in the neighborhood of these points.

Suppose now that $k \cong n\pi/a$, whence $k_n = k - (2\pi n/a) \cong -n\pi/a$. Then $k_n^2 \cong k^2$ and b_n will be very large, so that in (8.6-17) the quantity γb_n will no longer be small even though γ is small. In this case we may write the wave function, approximately, as

$$\psi(x) = b_0 e^{ikx} + \gamma b_n e^{ikx} e^{-2\pi n x/a} = b_0 e^{ikx} + \gamma b_n e^{ik_n x}, \quad (8.6-27)$$

regarding the other terms in the sum of (8.6-17) as negligible compared to the n th. Since $k_n = -k$ at the band edge, this wave function (8.6-27) is a superposition of a wave propagating along the positive x -axis and one propagating in the opposite direction; it has thus the character of a standing wave. This physical situation is in accord with the view that the electron wave function undergoes Bragg reflection at $k = \pm n\pi/a$. If the wave function (8.6-27) is substituted into Schrödinger's equation (8.6-2), using (8.6-4), one obtains

$$\begin{aligned} & b_0(k_0^2 - k^2)e^{ikx} + \gamma b_n(k_0^2 - k_n^2)e^{i(k-\frac{2\pi n}{a})x} + \gamma b_0 \sum_{n' \neq 0} c_{n'} e^{i(k-\frac{2\pi n'}{a})x} \\ & + \gamma^2 b_n \sum_{n' \neq 0} c_{n'} e^{i(k-\frac{2\pi n}{a}-\frac{2\pi n'}{a})x} = 0. \end{aligned} \quad (8.6-28)$$

Multiplying (8.6-28) through by e^{-ikx} and integrating from $x = 0$ to $x = a$, it is easily seen that the second and third terms contribute nothing and that a contribution from the fourth term is obtained only when $n' = -n$; the final result of this operation is

then

$$\alpha(k_0^2 - k^2)b_0 + \gamma^2 c_n^* b_n = 0. \quad (8.6-29)$$

(8.6-24)

Likewise, multiplying (8.6-28) through by $e^{-ik_n x} = e^{-i(k - \frac{2\pi n}{a})x}$ and integrating as above it can be shown that

$$c_n b_0 + (k_0^2 - k_n^2)b_n = 0. \quad (8.6-30)$$

The two above equations may be regarded as a set of homogeneous equations determining b_0 and b_n . As such, solutions other than $b_0 = b_n = 0$ exist only if the determinant of the system vanishes, that is, only if

$$\begin{vmatrix} k_0^2 - k^2 & \gamma^2 c_n^* \\ c_n & k_0^2 - k_n^2 \end{vmatrix} = (k_0^2 - k^2)(k_0^2 - k_n^2) - \gamma^2 c_n^* c_n = 0. \quad (8.6-31)$$

This equation is quadratic in k_0^2 , and may be solved in the usual way to give

$$k_0^2 = \frac{1}{2}[(k^2 + k_n^2) \pm \sqrt{(k^2 - k_n^2)^2 + 4\gamma^2 c_n^* c_n}], \quad (8.6-32)$$

(8.6-26) which, using (8.6-3), (8.6-16) and (8.6-13), can be expressed as

$$\varepsilon(k) = \frac{\hbar^2}{4m} \left[k^2 + \left(k - \frac{2\pi n}{a} \right)^2 \pm \sqrt{\left(k^2 - \left(k - \frac{2\pi n}{a} \right)^2 \right)^2 + \left(\frac{4m|V_n|}{\hbar^2} \right)^2} \right]. \quad (8.6-33)$$

At the band edge, $k = k_n = n\pi/a$, and at these points (8.6-33) reduces to

$$\varepsilon = \varepsilon_n \pm |V_n| \quad (8.6-34)$$

where

$$\varepsilon_n = \frac{\hbar^2}{2m} \left(\frac{n\pi}{a} \right)^2 \quad (8.6-35)$$

represents the free-particle energy associated with the band edge points.

From the results of Section 8.3, we know that at the band edge points $k = \pm n\pi/a$, internal Bragg reflection takes place, and is accompanied by a discontinuity or "energy gap" in the ε versus k curve.⁷ Accordingly, Equation (8.6-34) should be interpreted as implying an energy gap, or forbidden band, of width $2|V_n|$, where V_n is the n th Fourier coefficient in the Fourier series expansion of the periodic lattice potential. For values of k greater in magnitude than $n\pi/a$ we should expect $\varepsilon(k)$ to be greater than the value given by (8.6-34), and to approach the free electron value predicted by (8.6-15) where k^2 differs from k_n^2 by an amount sufficient to render γb_n small enough so that the treatment leading to Equations (8.6-16) and (8.6-24) is legitimate. Likewise, for values of k smaller in magnitude than $n\pi/a$, $\varepsilon(k)$ must be expected to be less than

⁷ The fact that the electron group velocity is indeed zero at these points, as one would expect for the standing waves which are always associated with Bragg reflection can be proved from (8.6-33); the details of this proof are assigned as an exercise.

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 ϵ and k such as that illustrated in Figure 8.7.

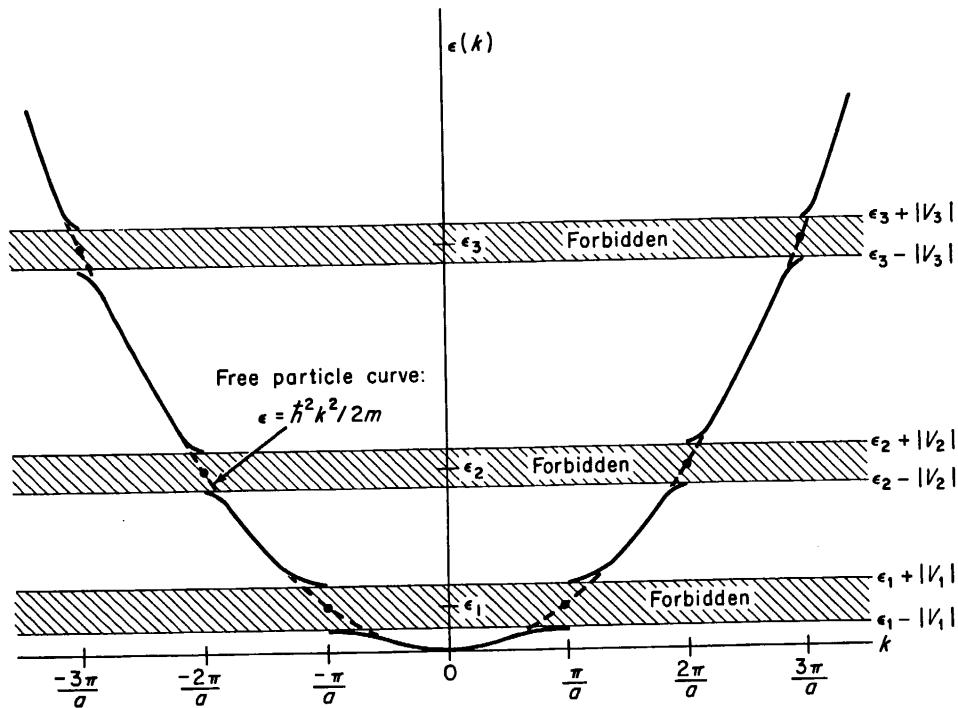


FIGURE 8.7. Schematic representation of the ϵ versus k relation for the free electron approximation as obtained from (8.6-24) and (8.6-33).

It is apparent from this figure that the ϵ 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 π/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

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

ously
sfied
The
ween
dges
sh as

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 ε_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).

CHAPTER 3

Energy Band Theory

The results and concepts of band theory essential for performing device analyses are routinely presented in introductory texts. Extrapolating from the discrete energy states available to electrons in isolated atoms, it is typically argued that the interaction between atoms leads to the formation of energy bands, ranges of allowed electron energies, when the atoms are brought into close proximity in forming a crystal. The highest energy band containing electrons at temperatures above absolute zero is identified as the conduction band; the next-lower-lying band, separated from the conduction band by an energy gap on the order of an electron-volt in semiconductors, and mostly filled with electrons at temperatures of interest, is identified as the valence band. The carriers involved in charge transport or current flow are associated with filled states in the conduction band and empty states (holes) in the valence band, respectively.

With the quantum mechanical foundation established in the preceding chapter, we are able to present a straightforward development of the energy band model and a more sophisticated treatment of related concepts. Specifically, we will show that energy bands arise naturally when one considers the allowed energy states of an electron moving in a periodic potential—the type of potential present in crystalline lattices. The “essential” energy-band-related concepts found in introductory texts, such as the effective mass, will be expanded and examined in greater detail. Additional concepts encountered in advanced device analyses will also be presented and explained. The overall goal is to establish a working knowledge of the energy band description of electrons in crystals.

The chapter begins with a simplified formulation of the electron-in-a-crystal problem, and the introduction of a powerful mathematical theorem that is of use in dealing with periodic potentials. A one-dimensional analysis is then performed that leads to the direct prediction of energy bands. The one-dimensional result is used as a basis for introducing and discussing energy-band-related terms and concepts. The development is next generalized to three dimensions, with special emphasis being placed on the interpretation of commonly encountered informational plots and constructs.

3.1 PRELIMINARY CONSIDERATIONS

3.1.1 Simplifying Assumptions

Electrons moving inside a semiconductor crystal may be likened to particles in a three-dimensional box with a very complicated interior. In a real crystal at operational temperatures there will be lattice defects (missing atoms, impurity atoms, etc.), and the semiconductor atoms will be vibrating about their respective lattice points. To simplify the problem it will be our assumption that lattice defects and atom core vibrations lead to a second-order perturbation—i.e., we begin by considering the lattice structure to be perfect and the atoms to be fixed in position. Moreover, in our initial considerations we treat a one-dimensional analog of the actual crystal. This procedure yields the essential features of the electronic behavior while greatly simplifying the mathematics.

The potential energy function, $U(x)$, associated with the crystalline lattice is of course required before one can initiate the quantum mechanical analysis. The general form of the function can be established by considering the one-dimensional lattice shown in Fig. 3.1(a). Atomic cores (atomic nuclei plus the tightly bound core electrons) with a net charge $+Z'q$ and separated by a lattice constant a are envisioned to extend from $x = 0$ to $x = (N - 1)a$, where N is the total number of atoms in the crystal. If the atomic core-electron interaction is assumed to be purely coulombic, the attractive force between the $x = 0$ atomic core and an electron situated at an arbitrary point x would give rise to the potential energy versus x dependence pictured in Fig. 3.1(b). Adding the attractive force associated with the $x = a$ atomic core yields the potential energy dependence shown in Fig. 3.1(c). Ultimately, accounting for the electron interaction with all atomic cores, one obtains the periodic crystalline potential sketched in Fig. 3.1(d). This result, we should point out, neglects any non-core electron-electron interaction which may occur in the crystal. However, it is reasonable to assume that the non-core electron-electron interaction approximately averages out to zero, and that the allowed electron states within the crystal can be determined to first order by considering a single electron of constant energy E moving in a periodic potential well of the form pictured in Fig. 3.1(d).

3.1.2 The Bloch Theorem

The Bloch theorem is of great utility in quantum mechanical analyses involving periodic potentials. The theorem basically relates the value of the wavefunction within any “unit cell” of a periodic potential to an equivalent point in any other unit cell, thereby allowing one to concentrate on a single repetitive unit when seeking a solution to Schrödinger’s equation. For a one-dimensional system the statement of the Bloch theorem is as follows:

IF $U(x)$ is periodic such that $U(x + a) = U(x)$

THEN

$$\psi(x + a) = e^{ika}\psi(x) \quad (3.1a)$$

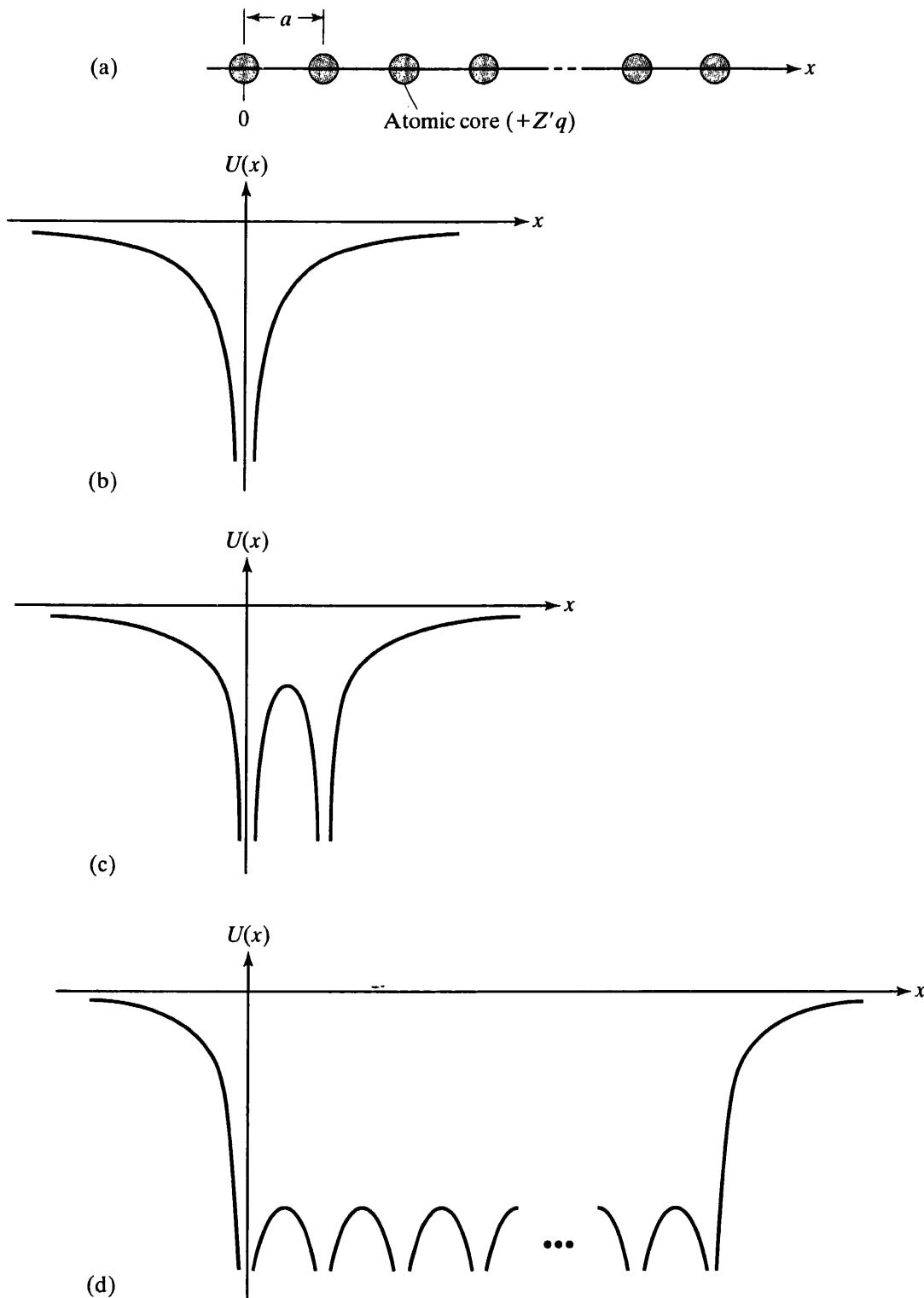


Figure 3.1 (a) One-dimensional crystalline lattice. (b-d) Potential energy of an electron inside the lattice considering (b) only the atomic core at $x = 0$, (c) the atomic cores at both $x = 0$ and $x = a$, and (d) the entire lattice chain.

54 CHAPTER 3 ENERGY BAND THEORY

or, equivalently,

$$\psi(x) = e^{ikx}u(x) \quad (3.1b)$$

where the unit cell wavefunction $u(x)$ has the same periodicity as the potential; i.e., $u(x + a) = u(x)$.

Similarly, for a three-dimensional system characterized by a translational symmetry vector \mathbf{a} and a periodic potential where $U(\mathbf{r} + \mathbf{a}) = U(\mathbf{r})$, the Bloch theorem states

$$\psi(\mathbf{r} + \mathbf{a}) = e^{i\mathbf{k}\cdot\mathbf{a}}\psi(\mathbf{r}) \quad (3.2a)$$

or

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{r}) \quad (3.2b)$$

where $u(\mathbf{r} + \mathbf{a}) = u(\mathbf{r})$.

Examining the one-dimensional statement of the theorem, please note that since

$$\psi(x + a) = e^{ik(x+a)}u(x + a) = e^{ika}e^{ikx}u(x) = e^{ika}\psi(x) \quad (3.3)$$

the alternative forms of the theorem are indeed equivalent. Also note that $\psi(x)$ itself is not periodic from unit cell to unit cell as one might expect intuitively. Rather, $\psi(x)$ has the form of a plane wave, $\exp(ikx)$, modulated by a function that reflects the periodicity of the crystalline lattice and the associated periodic potential.

The boundary conditions imposed at the end points of the periodic potential (or at the surfaces of the crystal) totally determine the permitted values of the Bloch function k in any given problem. Nevertheless, certain general statements can be made concerning the allowed values of k .

- (1) It can be shown that, for a one-dimensional system, two and only two distinct values of k exist for each and every allowed value of E .
- (2) For a given E , values of k differing by a multiple of $2\pi/a$ give rise to one and the same wavefunction solution. Therefore, a complete set of distinct k -values will always be obtained if the allowed k -values (assumed to be real) are limited to a $2\pi/a$ range. It is common practice to employ the Δk range $-\pi/a \leq k \leq \pi/a$.
- (3) If the periodic potential (or crystal) is assumed to be infinite in extent, running from $x = -\infty$ to $x = +\infty$, then there are no further restrictions imposed on k other than k must be real—i.e., k can assume a continuum of values. k must be real if the crystal is taken to be infinite because the unit cell function $u(x)$ is well behaved for all values of x , while $\exp(ikx)$, and therefore $\psi(x)$, will blow up at either $-\infty$ or $+\infty$ if k contains an imaginary component.
- (4) In dealing with crystals of finite extent, information about the boundary conditions to be imposed at the crystal surfaces may be lacking. To circumvent this

problem while still properly accounting for the finite extent of the crystal, it is commonplace to utilize what are referred to as *periodic boundary conditions*. The use of periodic boundary conditions is equivalent either to considering the ends of the crystal to be one and the same point, or to envisioning the lattice (Fig. 3.1(a)) to be in the form of a closed N -atom ring. For an N -atom ring with interatomic spacing a , one must have

$$\psi(x) = \psi(x + Na) = e^{ikNa}\psi(x) \quad (3.4)$$

which in turn requires

$$e^{ikNa} = 1 \quad (3.5)$$

or

$$k = \frac{2\pi n}{Na} \quad \dots n = 0, \pm 1, \pm 2, \dots \pm N/2 \quad (3.6)$$

Thus, for a finite crystal, k can only assume a set of discrete values. Note that k has been limited to $-\pi/a \leq k \leq \pi/a$ in accordance with the discussion in (2) above, and the total number of distinct k -values is equal to N . Practically speaking, the large number of atoms N in a typical crystal will cause the Eq. (3.6) k -values to be very closely spaced, thereby yielding a quasi-continuum of allowed k -values.

3.2 APPROXIMATE ONE-DIMENSIONAL ANALYSIS

3.2.1 Kronig–Penney Model

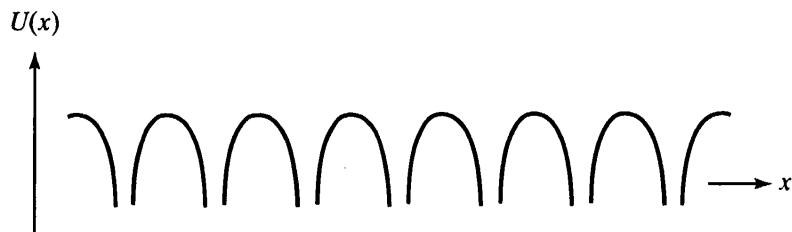
Even with the simplifications presented in Section 3.1, the solution of Schrödinger's equation for an electron in a crystal remains quite formidable. Specifically, solution difficulties can be traced to the shape of the periodic potential. We therefore propose a further simplification, an idealization of the periodic potential as shown in Fig. 3.2. This idealization of the actual crystal potential is referred to as the Kronig–Penney model. Note that the modeled crystal is assumed to be infinite in extent.

The Kronig–Penney analysis must be considered a “classic”—required knowledge for anyone with a serious interest in devices. The value of the admittedly crude model stems from the fact that the associated analysis illustrates energy band concepts in a straightforward manner, with a minimum of math, and in a quasi-closed form. General features of the quantum mechanical solution can be applied directly to real crystals.

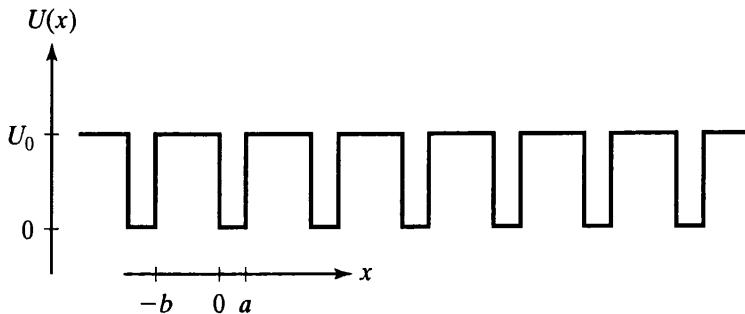
3.2.2 Mathematical Solution

The Kronig–Penney analysis closely parallels the finite potential well problem addressed in Subsection 2.3.3. We consider a particle, an electron, of mass m and fixed energy E subject to the periodic potential of Fig. 3.2(b). As in the finite potential well

56 CHAPTER 3 ENERGY BAND THEORY



(a)



(b)

Figure 3.2 Kronig–Penney idealization of the potential energy associated with a one-dimensional crystalline lattice. (a) One-dimensional periodic potential. (b) Kronig–Penney model.

problem, we expect mathematically distinct solutions for the energy ranges $0 < E < U_0$ and $E > U_0$. Here, however, the two energy ranges will be handled simultaneously. The Bloch theorem of course relates the solution in an arbitrarily chosen unit cell of length $a + b$ to any other part of the crystal. For convenience let us choose $x = -b$ and $x = a$ as the unit cell boundaries, with the subscripts b and a identifying the wavefunctions and solution constants in the regions $-b < x < 0$ and $0 < x < a$, respectively. Schrödinger's equation, the equation to be solved in the two spatial regions, then assumes the form

$$\frac{d^2\psi_a}{dx^2} + \alpha^2\psi_a = 0 \quad 0 < x < a \quad (3.7)$$

$$\alpha = \sqrt{2mE/\hbar^2} \quad (3.8)$$

and

$$\frac{d^2\psi_b}{dx^2} + \beta^2\psi_b = 0 \quad -b < x < 0 \quad (3.9)$$

$$\beta = \begin{cases} i\beta_-; & \beta_- = \sqrt{2m(U_0 - E)/\hbar^2} & 0 < E < U_0 \\ \beta_+; & \beta_+ = \sqrt{2m(E - U_0)/\hbar^2} & E > U_0 \end{cases} \quad (3.10a)$$

$$(3.10b)$$

Written in the most convenient form, the general solutions to Eqs. (3.7) and (3.9) are

$$\psi_a(x) = A_a \sin \alpha x + B_a \cos \alpha x \quad (3.11a)$$

$$\psi_b(x) = A_b \sin \beta x + B_b \cos \beta x \quad (3.11b)$$

(We will subsequently replace the $\sin \beta x$ and $\cos \beta x$ with their hyperbolic equivalents when $\beta = i\beta_-$ is purely imaginary.) Now, the wavefunction and its derivative must be continuous at $x = 0$. Likewise, the wavefunction and its derivative evaluated at the cell boundaries must obey the periodicity requirements imposed by the Bloch theorem [Eq. (3.1a)]. These requirements translate into four boundary conditions:

$$\psi_a(0) = \psi_b(0) \quad (3.12a)$$

$$\left. \frac{d\psi_a}{dx} \right|_0 = \left. \frac{d\psi_b}{dx} \right|_0 \quad \left. \begin{array}{l} \text{Continuity} \\ \text{requirements} \end{array} \right\} \quad (3.12b)$$

$$\psi_a(a) = e^{ik(a+b)} \psi_b(-b) \quad (3.12c)$$

$$\left. \frac{d\psi_a}{dx} \right|_a = e^{ik(a+b)} \left. \frac{d\psi_b}{dx} \right|_{-b} \quad \left. \begin{array}{l} \text{Periodicity} \\ \text{requirements} \end{array} \right\} \quad (3.12d)$$

The Eq. (3.12) boundary conditions give rise to a set of four simultaneous equations:

$$B_a = B_b \quad (3.13a)$$

$$\alpha A_a = \beta A_b \quad (3.13b)$$

$$A_a \sin \alpha a + B_a \cos \alpha a = e^{ik(a+b)} [-A_b \sin \beta b + B_b \cos \beta b] \quad (3.13c)$$

$$\alpha A_a \cos \alpha a - \alpha B_a \sin \alpha a = e^{ik(a+b)} [\beta A_b \cos \beta b + \beta B_b \sin \beta b] \quad (3.13d)$$

Equations. (3.13a) and (3.13b) can be used to readily eliminate A_b and B_b in Eqs. (3.13c) and (3.13d), yielding

$$A_a [\sin \alpha a + (\alpha/\beta) e^{ik(a+b)} \sin \beta b] + B_a [\cos \alpha a - e^{ik(a+b)} \cos \beta b] = 0 \quad (3.14a)$$

$$A_a [\alpha \cos \alpha a - \alpha e^{ik(a+b)} \cos \beta b] + B_a [-\alpha \sin \alpha a - \beta e^{ik(a+b)} \sin \beta b] = 0 \quad (3.14b)$$

Paralleling the finite potential well problem, we could next proceed to eliminate B_a between the two remaining equations. It is expedient in the present situation,

however, to make use of a well-known mathematical result—namely, that a set of n homogeneous equations linear in n unknowns has a non-trivial solution (a solution where the unknowns are non-zero) only when the determinant formed from the coefficients of the unknowns is equal to zero. Thus the first bracketed expression in Eq. (3.14a) times the second bracketed expression in Eq. (3.14b) minus the first bracketed expression in Eq. (3.14b) times the second bracketed expression in Eq. (3.14a) must be equal to zero.[†] Performing the required cross-multiplication and simplifying the result as much as possible, one obtains

$$-\frac{\alpha^2 + \beta^2}{2\alpha\beta} \sin\alpha a \sin\beta b + \cos\alpha a \cos\beta b = \cos k(a + b) \quad (3.15)$$

Finally, reintroducing $\beta = i\beta_-$ for $0 < E < U_0$ and $\beta = \beta_+$ for $E > U_0$, noting $\sin(ix) = isinhx$ and $\cos(ix) = coshx$, and defining

$$\alpha_0 \equiv \sqrt{2mU_0/\hbar^2} \quad (3.16)$$

$$\xi \equiv E/U_0 \quad (3.17)$$

such that $\alpha = \alpha_0\sqrt{\xi}$, $\beta_- = \alpha_0\sqrt{1 - \xi}$ and $\beta_+ = \alpha_0\sqrt{\xi - 1}$, we arrive at the result

$$\begin{aligned} \frac{1 - 2\xi}{2\sqrt{\xi(1 - \xi)}} \sin\alpha_0 a \sqrt{\xi} \sinh\alpha_0 b \sqrt{1 - \xi} + \cos\alpha_0 a \sqrt{\xi} \cosh\alpha_0 b \sqrt{1 - \xi} \\ = \cos k(a + b) \quad \dots 0 < E < U_0 \end{aligned} \quad (3.18a)$$

$$\begin{aligned} \frac{1 - 2\xi}{2\sqrt{\xi(\xi - 1)}} \sin\alpha_0 a \sqrt{\xi} \sin\alpha_0 b \sqrt{\xi - 1} + \cos\alpha_0 a \sqrt{\xi} \cos\alpha_0 b \sqrt{\xi - 1} \\ = \cos k(a + b) \quad \dots E > U_0 \end{aligned} \quad (3.18b)$$

Other than system constants, the left-hand sides of Eqs. (3.18a) and (3.18b) depend only on the energy E , while the right-hand sides depend only on k . Consequently, Eqs. (3.18) specify the allowed values of E corresponding to a given k .

3.2.3 Energy Bands and Brillouin Zones

We are at long last in a position to confirm that energy bands arise naturally when one considers the allowed energy states of an electron moving in a periodic potential.

[†]If so inclined, we could have applied the determinant rule to the original set of four equations and to the four simultaneous equations encountered in the finite potential well problem. The end result is the same.

Because the crystal under analysis was assumed to be infinite in extent, the k in Eqs. (3.18) can assume a continuum of values and must be real. (This follows from observation #3 presented in Subsection 3.1.2.) The $\cos k(a + b)$ can therefore take on any value between -1 and $+1$. E -values which cause the left-hand side of Eq. (3.18a) or (3.18b), call this $f(\xi)$, to lie in the range $-1 \leq f(\xi) \leq 1$ are then the allowed system energies.

For a given set of system constants, the allowed values of E can be determined by graphical or numerical methods. To illustrate the graphical procedure and the general nature of the results, we have plotted $f(\xi)$ versus ξ in Fig. 3.3 for the specific case where $\alpha_0 a = \alpha_0 b = \pi$. From Fig. 3.3, $f(\xi)$ is seen to be an oscillatory-type function that alternately drops below -1 and rises above $+1$. This same behavior is observed for any set of system constants. Thus, as we have anticipated, there are extended ranges of allowed system energies (the shaded regions in Fig. 3.3). The ranges of allowed energies are called *energy bands*; the excluded energy ranges, *forbidden gaps* or *band gaps*. Relative to the crystal potential, the energy bands formed inside a crystal with $\alpha_0 a = \alpha_0 b = \pi$ would be roughly as envisioned in Fig. 3.4.

If the allowed values of energy are plotted as a function of k , one obtains the E - k diagram shown in Fig. 3.5. In constructing this plot the allowed values of k were limited to the $2\pi/(a + b)$ range between $-\pi/(a + b)$ and $+\pi/(a + b)$. As noted previously in Subsection 3.1.2, confining the allowed k -values to a $2\pi/(\text{cell length})$ range properly

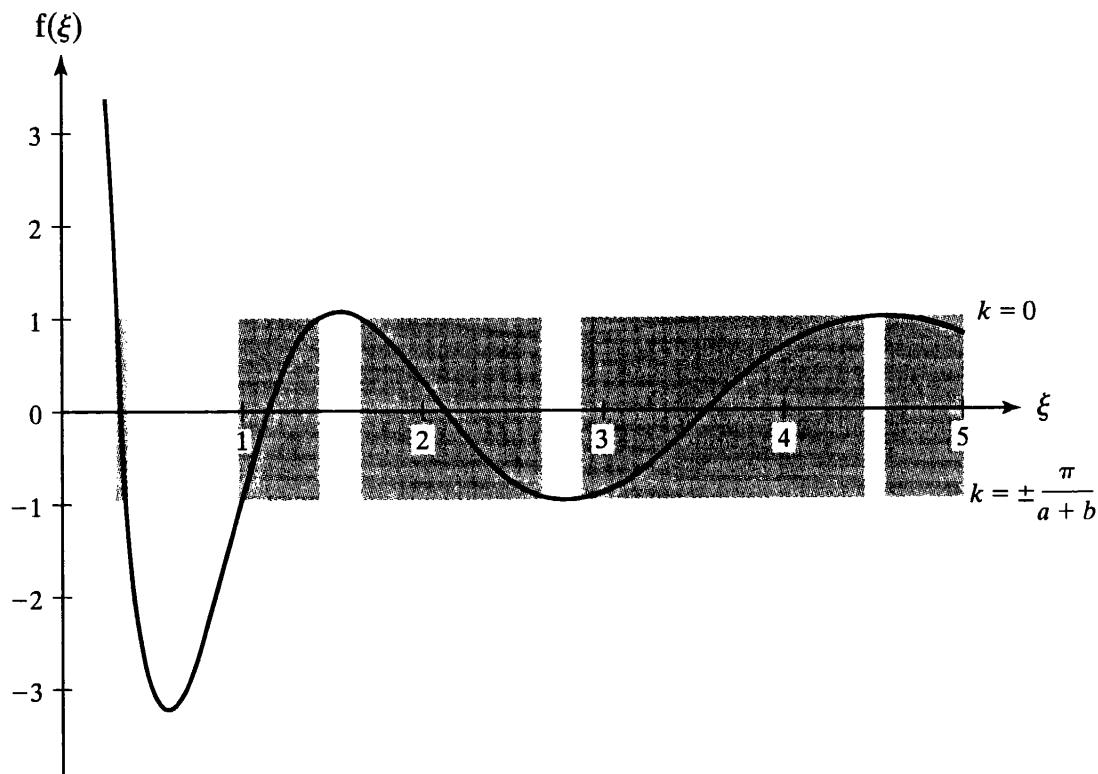


Figure 3.3 Graphical determination of allowed electron energies. The left-hand side of the Eqs. (3.18) Kronig-Penney model solution is plotted as a function of $\xi = E/U_0$. The shaded regions where $-1 \leq f(\xi) \leq 1$ identify the allowed energy states ($\alpha_0 a = \alpha_0 b = \pi$).

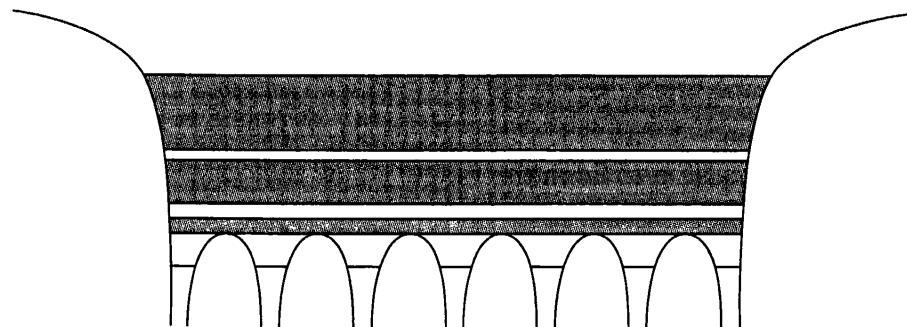


Figure 3.4 Visualization of the energy bands in a crystal.

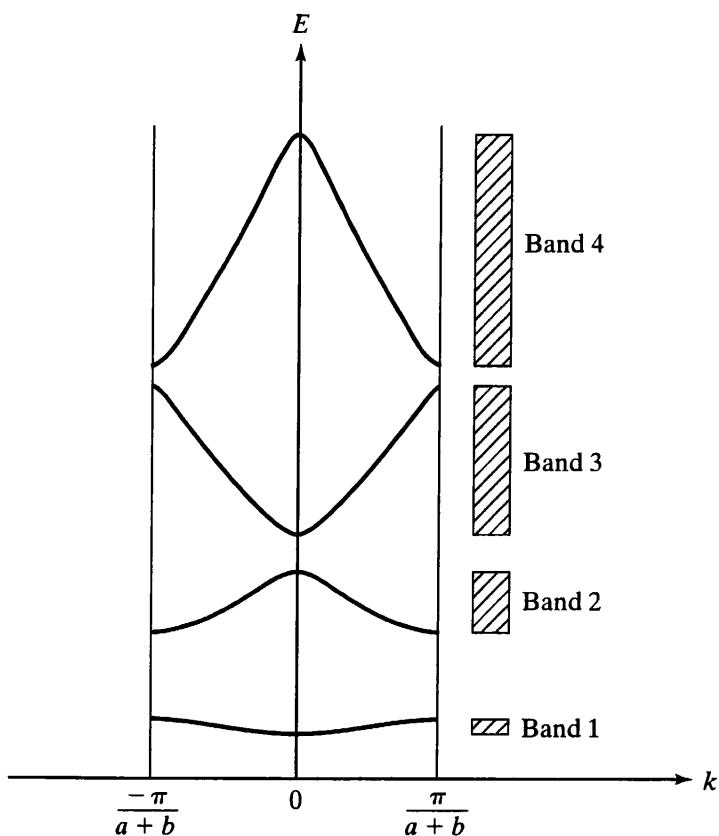


Figure 3.5 Reduced-zone representation of allowed E - k states in a one-dimensional crystal (Kronig–Penney model with $\alpha_0 a = \alpha_0 b = \pi$).

accounts for all distinct k -values—a complete set of distinct solutions lies within the cited range. Increasing or decreasing k in Eqs. (3.18) by a multiple of $2\pi/(a + b)$ has no effect on the allowed system energy and thus an E - k solution lying outside the chosen range simply duplicates one of the E - k solutions inside the chosen range. Also note in Fig. 3.5 that, consistent with observation #1 presented in the Bloch theorem discussion, there are two and only two k -values associated with each allowed energy.

E - k diagrams are very important in the characterization of materials and we will have a great deal to say about their interpretation and use. At this point, however, we

would merely like to point out that the energy band slope, dE/dk , is zero at the k -zone boundaries; i.e., $dE/dk = 0$ at $k = 0$ and $k = \pm\pi/(a + b)$ in Fig. 3.5. This is a feature common to all E - k plots, even those characterizing real materials.

Whereas Fig. 3.5 exemplifies the preferred and most compact way of presenting actual E - k information, valuable insight into the electron-in-a-crystal solution can be gained if the E - k results are examined from a somewhat different viewpoint. Specifically, instead of restricting k to the values between $\pm\pi/(a + b)$, one could alternatively associate increasing values of allowed E deduced from Eqs. (3.18) with monotonically increasing values of $|k|$. This procedure yields the E - k diagram shown in Fig. 3.6. As indicated in Fig. 3.6, the same result could have been achieved by starting with Fig. 3.5 and translating half-segments of the various bands along the k -axis by a multiple of $2\pi/(a + b)$. k -value solutions differing by $2\pi/(a + b)$ are of course physically indistinct, and therefore Figs. 3.5 and 3.6 are totally equivalent. When presented in the Fig. 3.6 format, however, the relationship between the periodic potential and the free-particle E - k solutions becomes obvious. The periodic potential introduces a perturbation

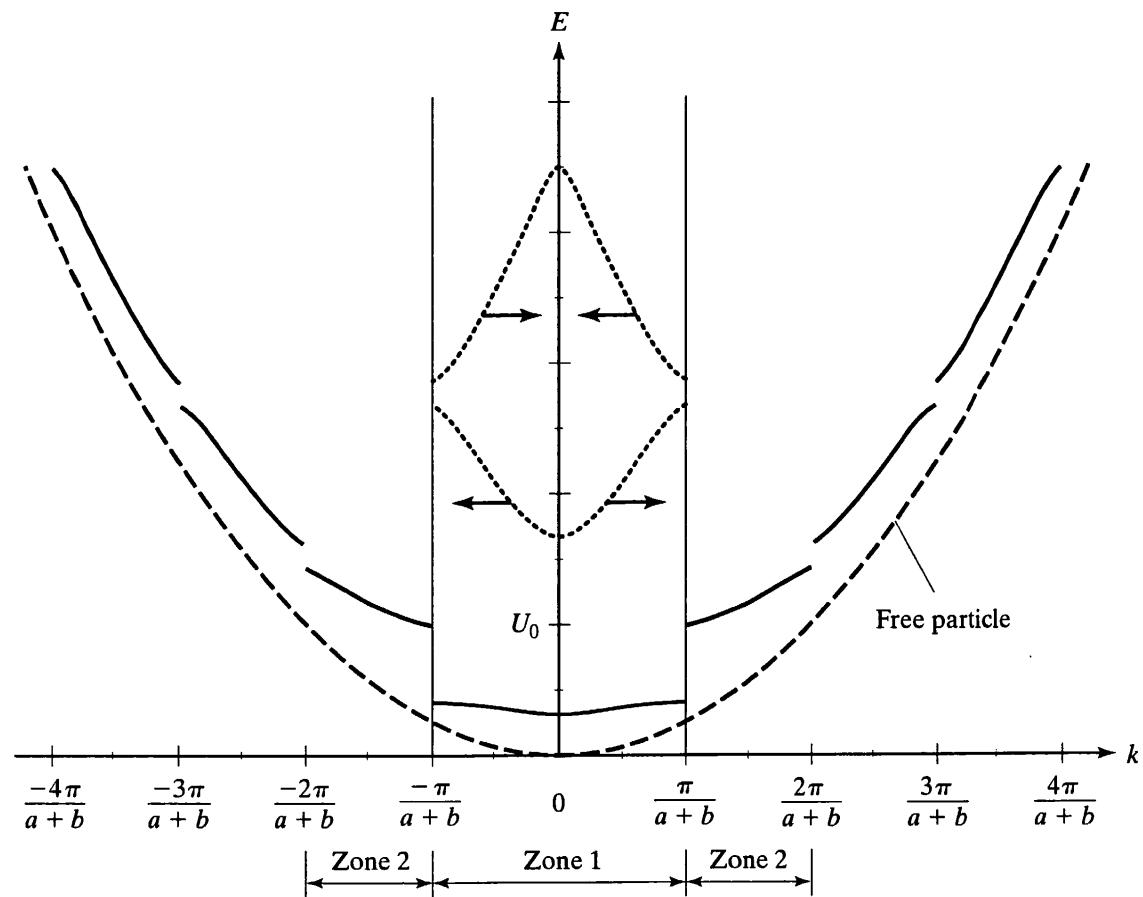


Figure 3.6 Extended-zone representation of allowed E - k states in a one-dimensional crystal (Kronig-Penney model with $\alpha_0a = \alpha_0b = \pi$). Shown for comparison purposes are the free-particle E - k solution (dashed line) and selected bands from the reduced-zone representation (dotted lines). Arrows on the reduced-zone band segments indicate the directions in which these band segments are to be translated to achieve coincidence with the extended-zone representation. Brillouin zones 1 and 2 are also labeled on the diagram.

that segments and distorts the free-particle solution. The modification is greatest at the lower energies, with the two solutions essentially merging at the higher energies. This seems reasonable from an intuitive standpoint, since the greater the electron energy, the smaller the relative importance of the periodic potential within the crystal.

Figure 3.6 also allows us to introduce relevant band-theory terminology. Those k -values associated with a given energy band are said to form a *Brillouin zone*. Brillouin zones are numbered consecutively beginning with the lowest energy band. The first Brillouin zone in Fig. 3.6 runs from $-\pi/(a + b)$ to $\pi/(a + b)$, the second from $\pm\pi/(a + b)$ to $\pm2\pi/(a + b)$, etc. For obvious reasons, an E - k diagram of the type presented in Fig. 3.6 is called an *extended-zone representation*. When the bands are all folded back into the first Brillouin zone as in Fig. 3.5, the diagram is called a *reduced-zone representation*.

To conclude this subsection, a comment is in order concerning the interpretation of the Bloch parameter k , the k found in E - k diagrams. For a free particle, k is of course the wavenumber and $\hbar k = \langle p \rangle$ is the particle momentum. Given the similarity between the free-particle solution and the extended-zone representation of the periodic potential solution, it is not surprising that k in the latter case is also referred to as the wavenumber and $\hbar k$ as the *crystal momentum*. However, as the addition of the word "crystal" implies, $\hbar k$ is not the actual momentum of an electron in a crystal, but rather a momentum-related constant of the motion which incorporates the crystal interaction. One might have suspected the $\hbar k$ in crystal plots to be different from the actual momentum since $\pm2\pi/(\text{cell length})$ can be added to the crystal momentum without modifying the periodic potential solution. \hbar times the k appearing in the reduced-zone representation, it should be mentioned, is often called the reduced crystal momentum, or simply the *reduced momentum*.

Although accepting the difference between $\hbar k$ (crystal) and the actual momentum, one might still wonder how $\pm2\pi/(\text{cell length})$ can be added to k (crystal) without modifying the solution. How, in particular, can the periodic potential solution with an "adjustable" k approach the free-particle solution with a fixed k in the limit where $E \gg U_0$? In this regard it must be remembered that the wavefunction solution for an electron in a crystal is the product of two terms, $\exp(ikx)$ and $u(x)$, where $u(x)$ is also a function of k . Increasing or decreasing k by a multiple of $2\pi/(\text{cell length})$ modifies both $\exp(ikx)$ and $u(x)$ in such a way that the product of the two terms is left unchanged. It is the product of the two terms, not just $\exp(ik_{\text{crystal}}x)$, that approaches the free-particle solution in the $E \gg U_0$ limit.

3.2.4 Particle Motion and Effective Mass

It was noted earlier that our ultimate goal is to model the *action* of electrons in crystals. The energy band solution we have achieved tells us about the allowed energy and reduced momentum states of an electron inside a crystal, but it is intrinsically devoid of action information. By assuming that the electron has a given energy E , we are automatically precluded from determining anything about the time evolution of the particle's position. Likewise, having specified k with absolute precision, the best we can do is compute the probability of finding the particle inside the various regions of the crystal.

The cited inability to deduce information about the position and motion of the particle when the energy and momentum are precisely specified is fundamental to the formulation of quantum mechanics. The fundamental property to which we refer is usually stated in terms of the *Heisenberg uncertainty principle*. The uncertainty principle observes that there is a limitation to the precision with which one can simultaneously determine conjugate dynamical variables. Specifically, for the E - t and p_x - x variable pairs, the precision is limited to

$$\Delta E \Delta t \geq \hbar \quad (3.19a)$$

$$\Delta p_x \Delta x \geq \hbar \quad (3.19b)$$

where the Δ in Eqs. (3.19) is to be read “the uncertainty in.” Clearly, if the E of a particle is specified with absolute precision, the uncertainty in t is infinite; one is precluded from determining anything about the time evolution of the particle’s position. Consequently, a *superposition* of fixed- E wavefunction solutions must be used to describe a particle if it is experimentally or conceptually confined to a given segment of a crystal at a given instant in time. In other words, to address the question of particle motion inside the crystal one must work with “wavepackets.”

The wavepacket is the quantum mechanical analog of a classical particle localized to a given region of space. The wavepacket, literally a packet of waves, consists of a linear combination of constant- E wavefunction solutions closely grouped about a peak or center energy. The wavefunctions are assumed to be combined in such a way that the probability of finding the represented particle in a given region of space is unity at some specified time. Completely analogous to the Fourier series expansion of an electrical voltage pulse, the smaller the width of the wavepacket, the more constant- E solutions (analogous to Fourier components) of appreciable magnitude needed to accurately represent the wavepacket.

Reaction of the wavepacket to external forces and its spatial evolution with time provide the sought-after equation of motion for an electron in a crystal. Corresponding to the center of mass of a classical particle moving with a velocity v , one can speak of the wavepacket’s center moving with a group velocity $v_g = dx/dt$. For a packet of traveling waves with center frequency ω and center wavenumber k , classical wave theory yields the dispersion relationship

$$v_g = \frac{d\omega}{dk} \quad (3.20)$$

As is most readily evident from a comparison of free-particle and traveling-wave expressions given in Subsection 2.3.1, E/\hbar in the quantum mechanical formulation replaces ω in the classical formulation. The wavepacket group velocity is therefore concluded to be

$$v_g = \frac{1}{\hbar} \frac{dE}{dk} \quad (3.21)$$

where E and k are interpreted as the center values of energy and crystal momentum, respectively.

We are now in a position to consider what happens when an “external” force F acts on the wavepacket. F could be any force other than the crystalline force associated with the periodic potential. The crystalline force is already accounted for in the wavefunction solution. The envisioned force might arise, for example, from dopant ions within the crystal or could be due to an externally impressed electric field. The force F acting over a short distance dx will do work on the wavepacket, thereby causing the wavepacket energy to increase by

$$dE = Fdx = Fv_g dt \quad (3.22)$$

We can therefore assert

$$F = \frac{1}{v_g} \frac{dE}{dt} = \frac{1}{v_g} \frac{dE}{dk} \frac{dk}{dt} \quad (3.23)$$

or, making use of the group-velocity relationship,

$$F = \frac{d(\hbar k)}{dt} \quad (3.24)$$

Next, differentiating the group-velocity relationship with respect to time, we find

$$\frac{dv_g}{dt} = \frac{1}{\hbar} \frac{d}{dt} \left(\frac{dE}{dk} \right) = \frac{1}{\hbar^2} \frac{d^2E}{dk^2} \frac{d(\hbar k)}{dt} \quad (3.25)$$

which when solved for $d(\hbar k)/dt$ and substituted into Eq. (3.24) yields

$$F = m^* \frac{dv_g}{dt}$$

(3.26)

$$m^* = \frac{1}{\frac{1}{\hbar^2} \frac{d^2E}{dk^2}}$$

(3.27)

The foregoing is a very significant result; its importance cannot be overemphasized. Equation (3.26) is identical to Newton’s second law of motion except that the actual particle mass is replaced by an *effective mass* m^* . This implies that the motion of electrons in a crystal can be visualized and described in a quasi-classical manner. In most instances the electron can be thought of as a “billiard ball,” and the electronic motion

3.2 APPROXIMATE ONE-DIMENSIONAL ANALYSIS

can be modeled using Newtonian mechanics, provided that one accounts for the effect of crystalline forces and quantum mechanical properties through the use of the effective-mass factor. Practically speaking, because of the cited simplification, device analyses can often be completed with minimal direct use of the quantum mechanical formalism. The effective-mass relationship itself, Eq. (3.27), also underscores the practical importance of the E - k diagrams discussed previously. Having established the crystal band structure or E - k relationship, one can determine the effective mass exhibited by the carriers in a given material.

Mathematically, the effective mass is inversely proportional to the curvature of an E versus k plot. It is therefore possible to deduce certain general facts about the effective mass from an E - k diagram simply by inspection. Consider, for example, the two band segments pictured in Fig. 3.7. In the vicinity of the respective energy minima, the curvature of segment (b) is greater than the curvature of segment (a). With $(d^2E/dk^2)_b > (d^2E/dk^2)_a$, one concludes that $m_a^* > m_b^*$. This example illustrates that the relative size of the carrier m^* 's in different bands can be readily deduced by inspection.

Consider next the band segment of the Kronig–Penney type reproduced in Fig. 3.8(a). Using graphical techniques, one finds the first and second derivatives of the

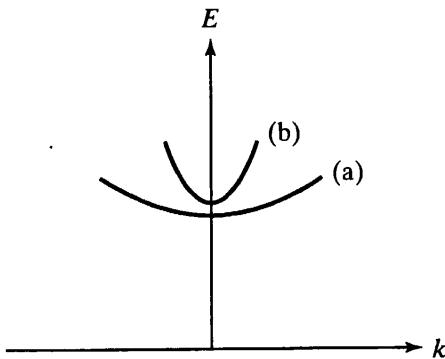


Figure 3.7 Hypothetical band segments used to illustrate how the relative magnitudes of the effective masses can be deduced from curvature arguments. By inspection, $m_a^* > m_b^*$ near $k = 0$.

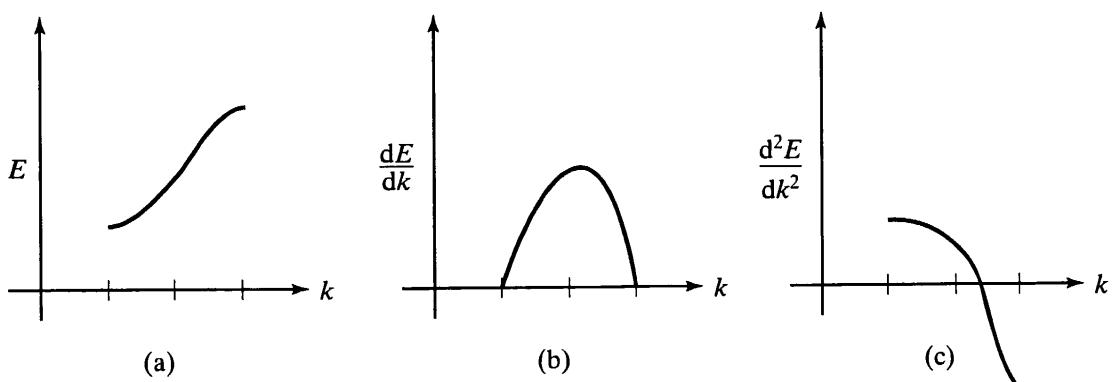


Figure 3.8 Deducing the sign of the effective-mass factor. (a) Sample band segment. (b) Graphically deduced first derivative and (c) second derivative of energy with respect to wavenumber.

band segment are roughly as sketched in Fig. 3.8(b and c), respectively. From Fig. 3.8(c) and Eq. (3.27), one in turn concludes that $m^* > 0$ near the band-energy minimum and $m^* < 0$ near the band-energy maximum. Since the shape of the Fig. 3.8(a) band segment is fairly typical of the band contours encountered in real materials, the preceding result is quite universal:

m^* is positive near the bottoms of all bands.
 m^* is negative near the tops of all bands.

A negative effective mass simply means that, in response to an applied force, the electron will accelerate in a direction opposite to that expected from purely classical considerations.

In general, the effective mass of an electron is a function of the electron energy E . This is clearly evident from Fig. 3.8(c). However, near the top or bottom band edge—the region of the band normally populated by carriers in a semiconductor[†]—the E - k relationship is typically parabolic; i.e., as visualized in Fig. 3.9,

$$E - E_{\text{edge}} \simeq (\text{constant})(k - k_{\text{edge}})^2 \quad (3.28)$$

and therefore

$$\frac{d^2E}{dk^2} \simeq \text{constant} \quad \dots E \text{ near } E_{\text{edge}} \quad (3.29)$$

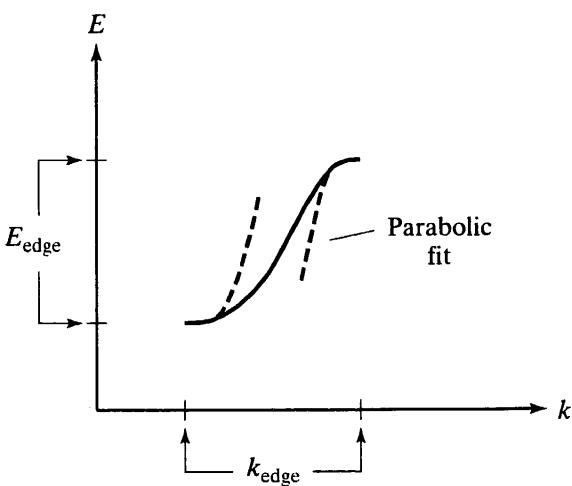


Figure 3.9 Parabolic nature of the bands near energy extrema points.

[†]For confirmation of this assertion, see Subsection 4.3.1

Thus, *carriers in a crystal with energies near the top or bottom of an energy band typically exhibit a CONSTANT (energy-independent) effective mass*. The usefulness of the effective-mass approximation would be questionable, and the correlation with classical behavior decidedly more complicated, if the effective carrier masses were not constant near the band edges.

3.2.5 Carriers and Current

The reader is no doubt aware that two types of carriers are present in semiconductors; namely, conduction-band electrons and valence-band holes. In this subsection we approach the origin and identification of these carriers from a band-theory standpoint. The band-theory approach provides additional insight, allowing one to answer certain questions that are difficult to address on an elementary level.

As a basis for discussion we consider a large one-dimensional crystal maintained at room temperature. The band structure of the crystal is assumed to be generally characterized by the Kronig–Penney model solution of Fig. 3.5. Because the crystal in the Kronig–Penney analysis was taken to be infinite in extent, there are an infinite number of k -states associated with each band in the cited solution. Restricting the number of atoms to some large but finite number, N , would have little effect on the energy-band structure. The number of distinct k -values in each band, however, would then be limited to N and spaced at $2\pi/N(a + b)$ intervals in accordance with observation #4 presented in Subsection 3.1.2. We assume this to be the case. For the sake of discussion, we further assume that each atom contributes two electrons to the crystal as a whole (there are two non-core electrons per atom), giving a grand total of $2N$ electrons to be distributed among the allowed energy states. At temperatures approaching zero Kelvin, the electrons would assume the lowest possible energy configuration: the available $2N$ electrons would totally fill the two lowest energy bands which contain N allowed states each. At room temperature, however, a sufficient amount of thermal energy is available to excite a limited number of electrons from the top of the second band into the bottom of the third band. We therefore conclude the electronic configuration within our crystal will be roughly as pictured in Fig. 3.10.

If a voltage is impressed across the crystal, a current will flow through the crystal and into the external circuit. Let us examine the contributions to the observed current from the various bands. The fourth band is of course totally devoid of electrons. Without “carriers” to transport charge there can be no current: *totally empty bands do not contribute to the charge-transport process*. Going to the opposite extreme, consider next the first band, where all available states are occupied by electrons. The individual electrons in this band can be viewed as moving about with velocities $v(E) = (1/\hbar)(dE/dk)$. However, because of the band symmetry and the filling of all available states, for every electron with a given $|v|$ traveling in the $+x$ direction, there will be another electron with precisely the same $|v|$ traveling in the $-x$ direction. Please note that this situation cannot be changed by the applied force: the absence of empty states precludes a modification of the electron velocity distribution within the band. Consequently, the first band, like the fourth band, does not contribute to the observed current: *totally filled bands do not contribute to the charge-transport process*.

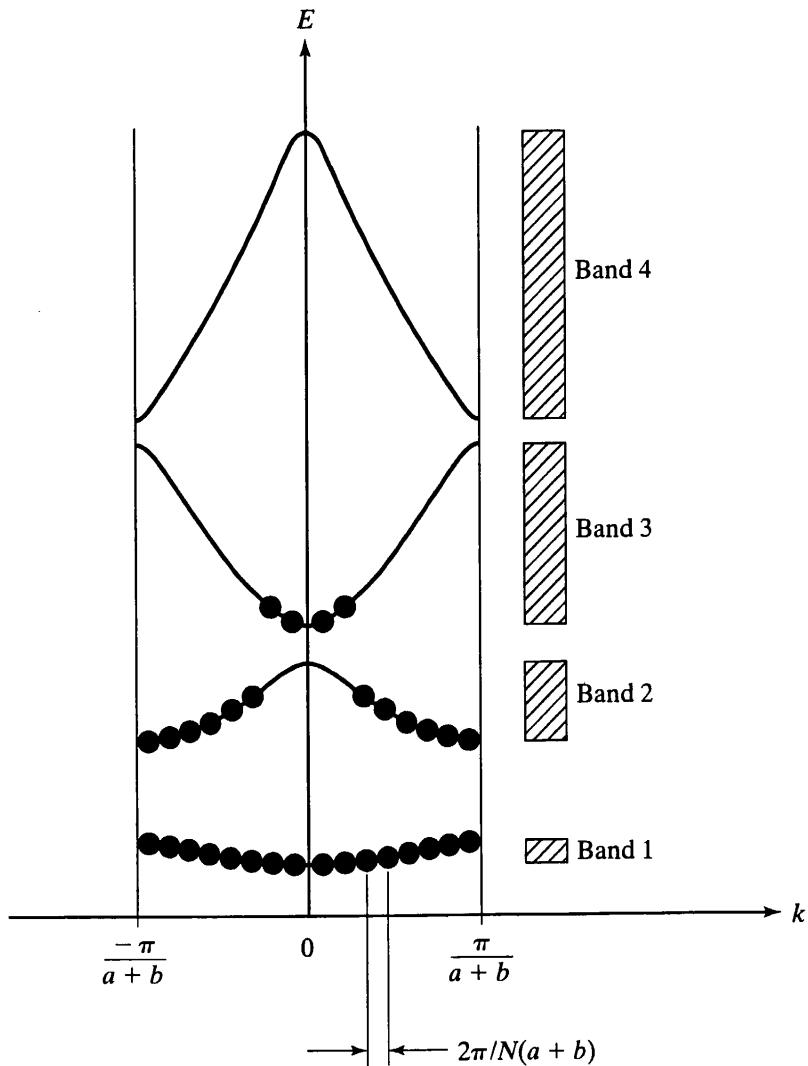


Figure 3.10 Filled and empty electronic states in the envisioned one-dimensional N -atom crystal maintained at room temperature. Each atom is assumed to contribute two electrons to the electronic configuration, and equilibrium conditions prevail. (Electrons are represented by filled circles.)

It follows from the foregoing discussion that only partially filled bands can give rise to a net transport of charge within the crystal. Under equilibrium conditions (Fig. 3.10), the filled-state distribution in the partially filled bands is symmetric about the band center and no current flows. Under the influence of an applied field, however, the filled-state distribution becomes skewed as envisioned in Fig. 3.11(a), and a current contribution is to be expected from both the second and third bands.

For the nearly empty third band, the contribution to the overall current will be

$$I_3 = -\frac{q}{L} \sum_{i(\text{filled})} v_i \quad (3.30)$$

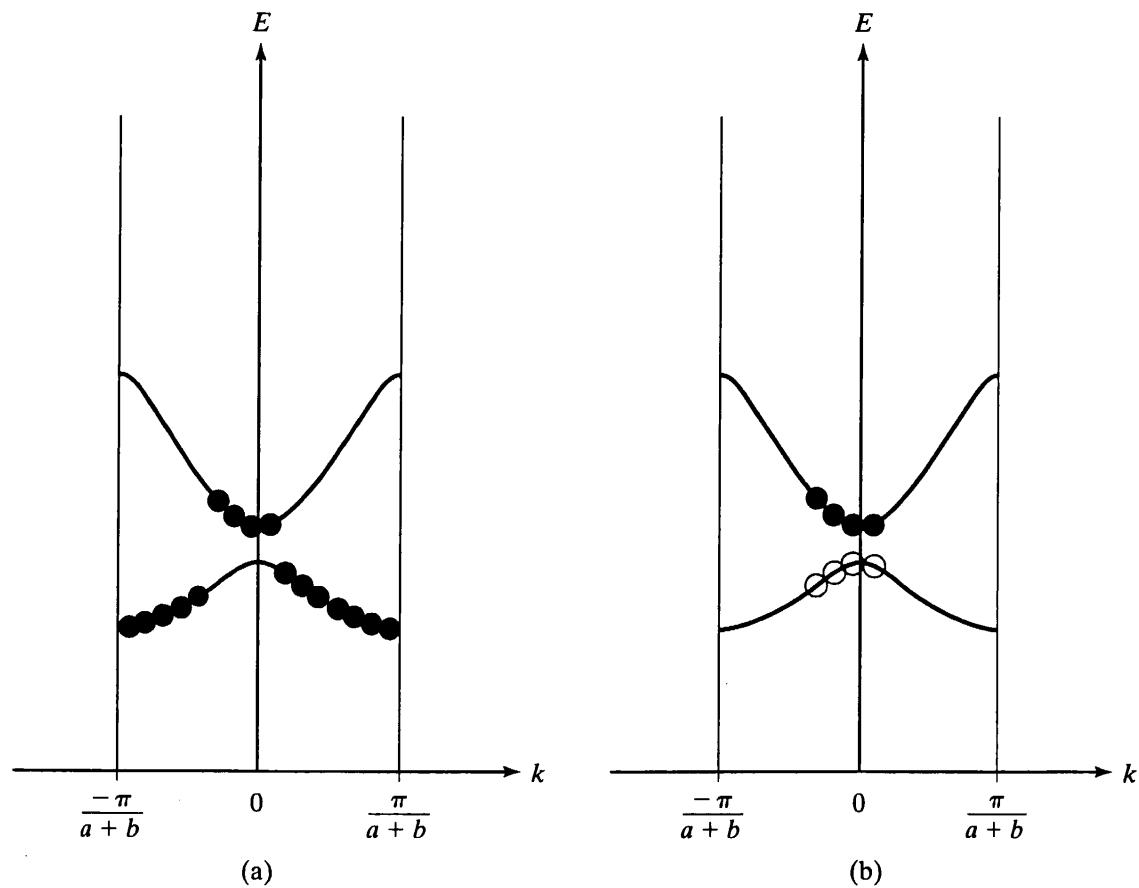


Figure 3.11 (a) The skewed filled-state distribution under steady-state conditions subsequent to the application of an external force. (b) Introduction of the hole. Alternative description of the electronic configuration in the lower energy band.

where I_3 is the current, L is the length of the one-dimensional crystal, and the summation is understood to be over all filled states. The Eq. (3.30) result is clearly analogous to the current attributed to conduction-band electrons in real crystals.

For the nearly filled second band, one could likewise write

$$I_2 = -\frac{q}{L} \sum_{i(\text{filled})} v_i \quad (3.31)$$

The summation in this case is rather cumbersome since it extends over a very large number of states. To simplify the result, we note that (as established previously) the summation of v_i over all states in a band is identically zero:

$$\sum_{i(\text{all})} v_i = 0 \quad (3.32)$$

It is therefore permissible to write

$$I_2 = -\frac{q}{L} \sum_{i(\text{filled})} v_i + \frac{q}{L} \sum_{i(\text{all})} v_i = \frac{q}{L} \sum_{i(\text{empty})} v_i \quad (3.33)$$

where v_i in the last summation is understood to be the velocity $v(E) = (1/\hbar)(dE/dk)$ associated with the empty states. Interestingly, the form of the last result is what one would expect if a *positively charged entity* were placed in the empty electronic states and the remainder of the states in the band were considered to be *unoccupied* by the positively charged entity. The suggested conceptual revision is pictured in Fig. 3.11(b). Pursuing this idea, one finds the overall motion of the electrons in the nearly filled band can likewise be described by considering just the empty electronic states—PROVIDED THAT the effective mass associated with the empty states is taken to be the negative of the m^* deduced from Eq. (3.27). We know from the effective-mass discussion, however, that the m^* deduced from Eq. (3.27) is itself negative near the tops of energy bands. Thus, conceptually and mathematically, we can model the action of the electrons in a nearly filled band in terms of a positively charged entity with *positive* effective mass occupying empty electronic states. The cited entity is called a *hole* and the arguments we have presented apply to the valence-band holes in real materials.

3.3 EXTRAPOLATION OF CONCEPTS TO THREE DIMENSIONS

Basic energy-band concepts were introduced in the previous section using a simplified one-dimensional model of a crystalline lattice. In this section we examine the modifications required in extending these basic concepts to three dimensions and real crystals. No attempt will be made to present a detailed derivation of the 3-D results. Rather, emphasis will be placed on identifying the special features introduced by the 3-D nature of real materials and on interpreting oft-encountered plots containing band-structure information. The specific topics to be addressed are Brillouin zones, $E-k$ diagrams, constant-energy surfaces, effective mass, and the band gap energy.

3.3.1 Brillouin Zones

The band structure of a one-dimensional lattice was described in terms of a one-dimensional or scalar k . 1-D Brillouin zones were in turn simply lengths or ranges of k associated with a given energy band. In progressing to the band-structure description of three-dimensional space lattices, the Bloch wavenumber becomes a vector and Brillouin zones become volumes. Specifically, a Brillouin zone (3-D) is the volume in k -space enclosing the set of k -values associated with a given energy band.

The first Brillouin zone for materials crystallizing in the diamond and zincblende lattices (Si, GaAs, etc.) is shown in Fig. 3.12. Geometrically, the zone is an octahedron which has been truncated by $\{100\}$ planes $2\pi/a$ from the zone center, a being the cubic lattice constant. The markings in the figure are group-theory symbols for high-symmetry points. (Don't panic. Familiarity with group theory is not required.) The most widely employed of the group-theory symbols are

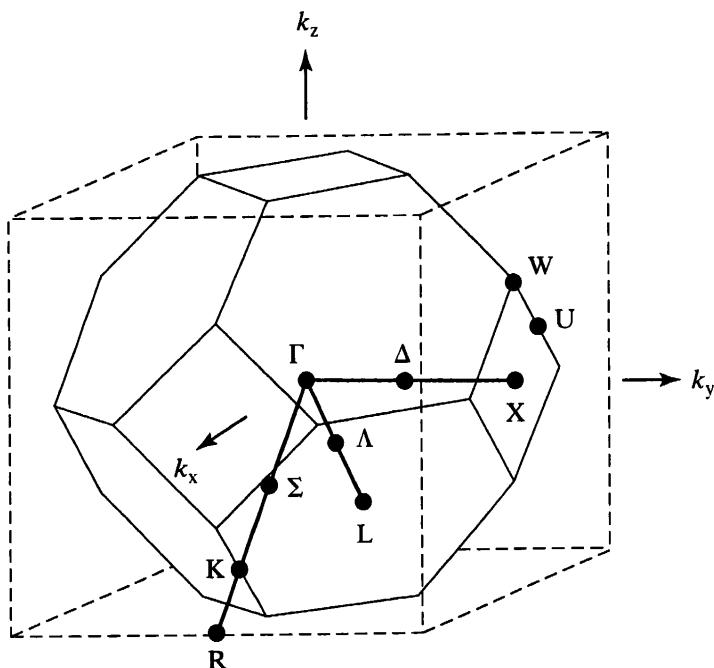


Figure 3.12 First Brillouin zone for materials crystallizing in the diamond and zincblende lattices. (After Blakemore.^[1] Reprinted with permission.)

Γ ... identifies the zone center ($k = 0$)
 X ... denotes the zone end along a $\langle 100 \rangle$ direction, and
 L ... denotes the zone end along a $\langle 111 \rangle$ direction.

Also note from Fig. 3.12 that the maximum magnitude of k varies with direction. In particular, $\Gamma \rightarrow L$, the length from the zone center to the zone boundary along a $\langle 111 \rangle$ direction, is $\sqrt{3}/2 \approx 0.87$ times $\Gamma \rightarrow X$, the distance from the zone center to the zone boundary along a $\langle 100 \rangle$ direction. This should serve to explain the different widths of the $E-k\langle 100 \rangle$ and $E-k\langle 111 \rangle$ diagrams to be considered next.

3.3.2 E-k Diagrams

The presentation of E - k information characterizing 3-D crystals poses a fundamental problem. Since three dimensions are required to represent the k -vector, real-material E - k plots are intrinsically four-dimensional. A plot totally characterizing the band structure of a 3-D lattice is obviously impossible to construct. Constructing reduced-dimension plots where one or more of the variables is held constant is a possible solution, but the random production of such plots could become quite laborious. Fortunately, in semiconductor work only those portions of the bands normally occupied by carriers—the near vicinity of the conduction-band minima and the valence-band maxima—are ro utinely of interest. In the case of the diamond and zincblende

72 CHAPTER 3 ENERGY BAND THEORY

lattices, the extrema points invariably occur at the zone center or lie along the high-symmetry $\langle 100 \rangle$ and $\langle 111 \rangle$ directions. Consequently, the information of greatest interest can be derived from plots of allowed energy versus the magnitude of k along these high-symmetry directions.

Figure 3.13 displays $\langle 100 \rangle/\langle 111 \rangle E-k$ diagrams characterizing the band structures in Ge (3.13a), Si (3.13b), and GaAs (3.13c and d). Before examining these figures it should be explained that the plots are two-direction composite diagrams. Because of crystal symmetry, the $-k$ portions of the $\langle 100 \rangle$ and $\langle 111 \rangle$ diagrams are just the mirror images of the corresponding $+k$ portions of the diagrams: no new information is conveyed by including the negative portions of the diagrams. It is therefore standard practice to delete the negative portions of the diagrams, turn the $\langle 111 \rangle$ diagrams so that the $+k$ direction faces to the left, and abut the two diagrams at $k = 0$. The respective positioning of L, denoting the zone boundary along a $\langle 111 \rangle$ direction, and X, denoting the zone boundary along a $\langle 100 \rangle$ direction, at the left- and right-hand ends of the k -axis corroborates the above observation. Likewise, the left-hand portions ($\Gamma \rightarrow L$) of the diagrams are shorter than the right-hand portions ($\Gamma \rightarrow X$) as expected from Brillouin-zone considerations. Also note that the energy scale in these diagrams is referenced to the energy at the top of the valence band. E_v is the maximum attainable valence-band energy, E_c the minimum attainable conduction-band energy, and $E_G = E_c - E_v$ the band gap energy.

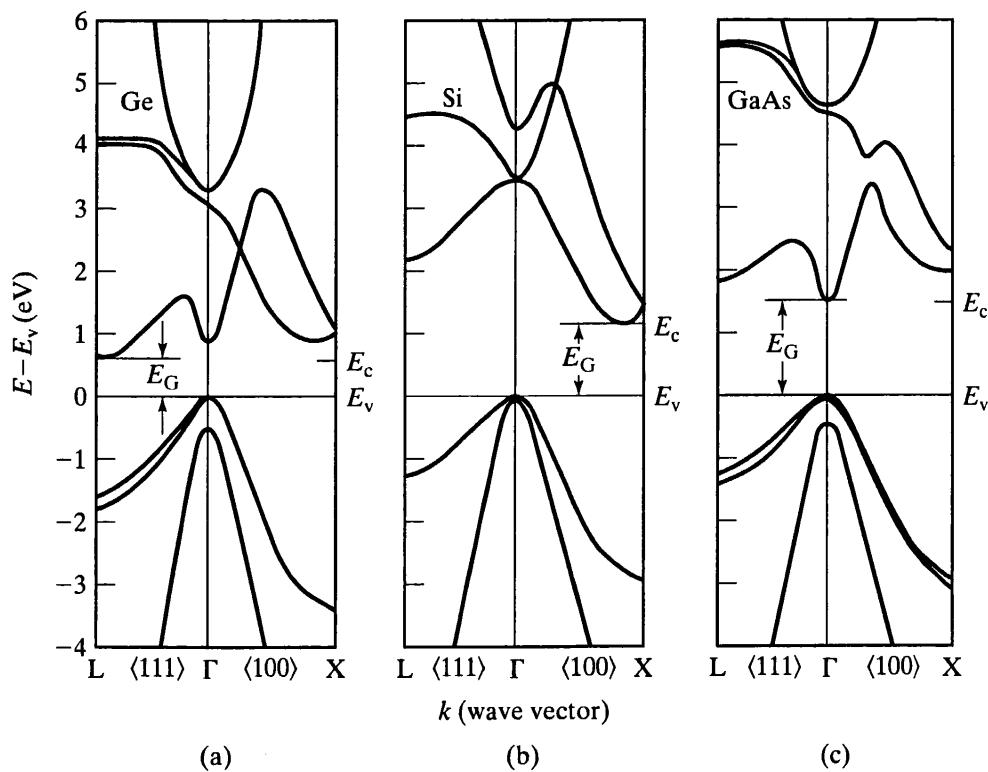
Let us now examine the diagrams for factual information. One observes the following:

VALENCE BAND

- (1) In all cases the valence-band maximum occurs at the zone center, at $k = 0$.
- (2) The valence band in each of the materials is actually composed of three subbands. Two of the bands are degenerate (have the same allowed energy) at $k = 0$, while the third band maximizes at a slightly reduced energy. (In Si the upper two bands are indistinguishable on the gross energy scale used in constructing Fig. 3.13(b). Likewise, the maximum of the third band is a barely discernible 0.044 eV below E_v at $k = 0$.) Consistent with the effective-mass/energy-band-curvature discussions presented in Subsection 3.2.4, the $k = 0$ degenerate band with the smaller curvature about $k = 0$ is called the *heavy-hole* band, and the $k = 0$ degenerate band with the larger curvature is called the *light-hole* band. The subband maximizing at a slightly reduced energy is the *split-off* band.
- (3) Near $k = 0$ the shape and therefore curvature of the subbands is essentially orientation independent. The significance of this observation will be explained later.

CONDUCTION BAND

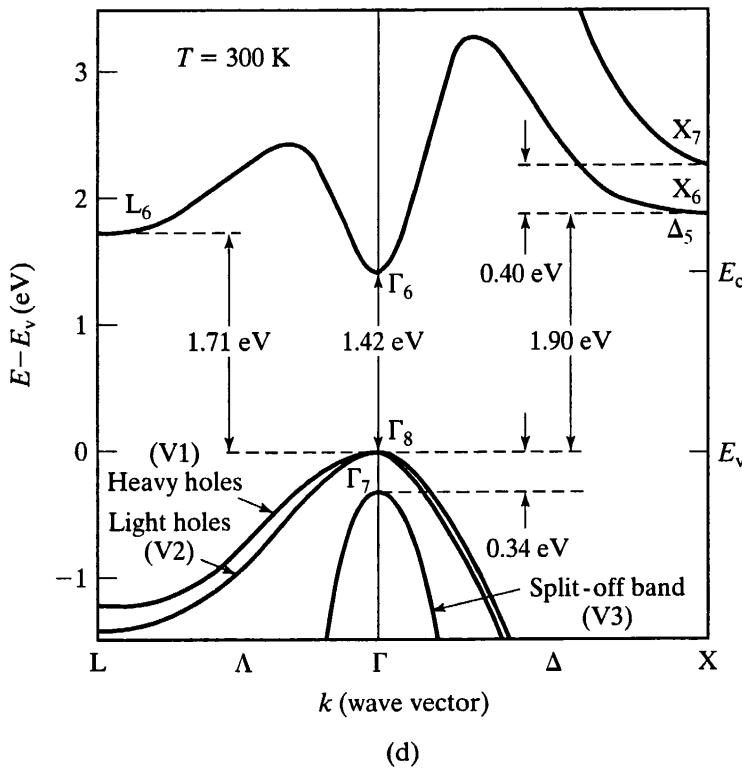
- (1) The gross features of the Ge, Si, and GaAs conduction-band structures are again somewhat similar. The conduction band in each case is composed of a number of subbands. The various subbands in turn exhibit localized and absolute minima at the zone center or along one of the high-symmetry directions. However—and this is very significant—the position of the overall conduction-band minimum, the “valley” where the electrons tend to congregate, varies from material to material.



(a)

(b)

(c)



(d)

Figure 3.13 $\langle 100 \rangle / \langle 111 \rangle$ E - k diagrams characterizing the conduction and valence bands of (a) Ge, (b) Si, and (c, d) GaAs. [(a-c) after Sze^[2]; (d) from Blakemore.^[1] Reprinted with permission.]

- (2) In Ge the conduction-band minimum occurs right at the zone boundary along the pictured $\langle 111 \rangle$ direction. Actually, there are *eight equivalent conduction-band minima* since there are eight equivalent $\langle 111 \rangle$ directions. Other minima in the conduction-band structure occurring at higher energies are seldom populated and may be ignored under most conditions.
- (3) The Si conduction-band minimum occurs at $k \approx 0.8(2\pi/a)$ from the zone center along the pictured $\langle 100 \rangle$ direction. The six-fold symmetry of $\langle 100 \rangle$ directions gives rise of course to *six equivalent conduction-band minima* within the Brillouin zone. Other minima in the Si conduction-band structure occur at considerably higher energies and are typically ignored.
- (4) Of the materials considered, GaAs is unique in that the conduction-band minimum occurs at the zone center directly over the valence-band maximum. Moreover, the L-valley at the zone boundary along $\langle 111 \rangle$ directions lies only 0.29 eV above the conduction-band minimum. Even under equilibrium conditions the L-valley contains a non-negligible electron population at elevated temperatures. Electron transfer from the Γ -valley to the L-valley provides the phenomenological basis for the transferred-electron devices (the Gunn-effect diode, etc.) and must be taken into account whenever a large electric field is impressed across the material.

Having discussed the properties of the conduction-band and valence-band structures separately, we should point out that the relative positioning of the band extrema points in k -space is in itself an important material property. When the conduction-band minimum and the valence-band maximum occur at the same value of k the material is said to be *direct*. Conversely, when the conduction-band minimum and the valence-band maximum occur at different values of k the material is said to be *indirect*. Electronic transitions between the two bands in a direct material can take place with little or no change in crystal momentum. On the other hand, conservation of momentum during an interband transition is a major concern in indirect materials. Of the three semiconductors considered, GaAs is an example of a direct material, while Ge and Si are indirect materials. The direct or indirect nature of a semiconductor has a very significant effect on the properties exhibited by the material, particularly the optical properties. The direct nature of GaAs, for example, makes it ideally suited for use in semiconductor lasers and infrared light-emitting diodes.

3.3.3 Constant-Energy Surfaces

E - k diagrams with the wavevector restricted to specific k -space directions provide one way of conveying relevant information about the band structures of 3-D crystals. An alternative approach is to construct a 3-D k -space plot of all the allowed k -values associated with a given energy E . For semiconductors, E is chosen to lie within the energy ranges normally populated by carriers ($E \lesssim E_v$ and $E \gtrsim E_c$). The allowed k -values form a surface or surfaces in k -space. The geometrical shapes, being associated with a given energy, are called *constant-energy surfaces*. Figure 3.14(a–c) displays the constant-energy surfaces characterizing the conduction-band structures near E_c in Ge, Si, and GaAs, respectively.

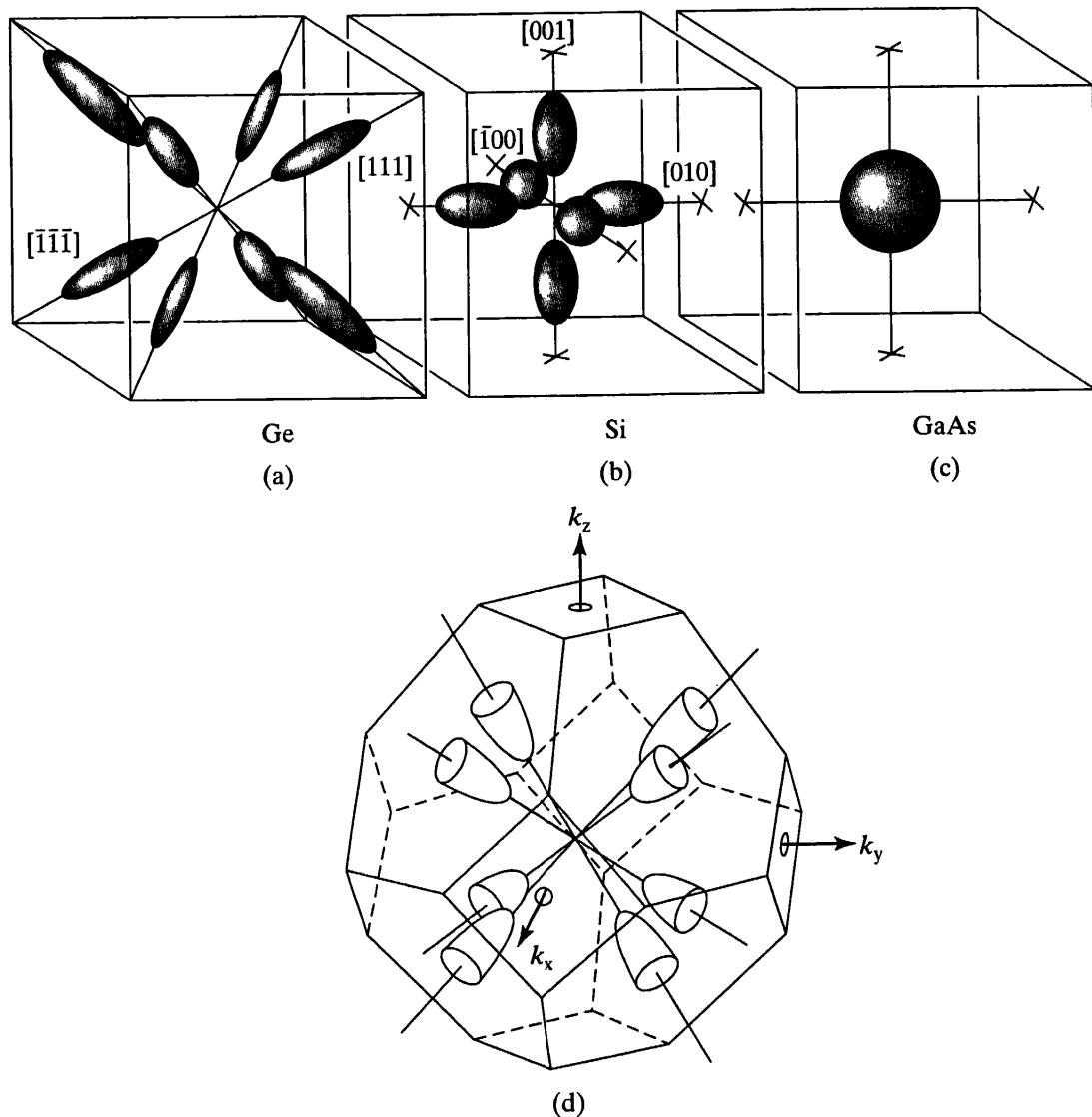


Figure 3.14 Constant-energy surfaces characterizing the conduction-band structure in (a, d) Ge, (b) Si, and (c) GaAs. (d) Shows the truncation of the Ge surfaces at the Brillouin-zone boundaries. [(a-c) after Sze^[2] and Ziman^[3]; (d) from McKelvey.^[4] Reprinted with permission; the latter from Robert E. Krieger Publishing Co., Malabar, FL.]

As pointed out in the E - k diagram discussion, a Ge conduction-band minimum ($E = E_c$) occurs along each of the eight equivalent $\langle 111 \rangle$ directions; a Si conduction-band minimum, along each of the six equivalent $\langle 100 \rangle$ directions. This explains the numbers and positions of the surfaces in the Ge and Si plots. (As clarified in Fig. 3.14(d), however, only one-half of the Ge surfaces actually lie within the first Brillouin zone—the Ge conduction-band minima occur right at the zone boundary.) The GaAs conduction-band minimum is of course positioned at the zone center, giving rise to a single constant-energy surface.

Although the preceding discussion covers most plot features, the geometrical shapes of the constant-energy surfaces still need to be explained. Working toward this end we recall that the E - k relationship characterizing one-dimensional crystals is approximately

parabolic in the vicinity of the band extrema points. The three-dimensional case is completely analogous: for energies only slightly removed from E_c one can write in general

$$E - E_c \approx Ak_1^2 + Bk_2^2 + Ck_3^2 \quad (3.34)$$

where A , B , and C are constants and k_1 , k_2 , and k_3 are k -space coordinates measured from the center of a band minimum along principal axes. (Using the [111] Ge minimum of Fig. 3.14(a) as an example, the k_1 - k_2 - k_3 coordinate system would be centered at the [111] L-point and one of the coordinate axes, say the k_1 -axis, would be directed along the k_x - k_y - k_z [111] direction.) For cubic crystals such as Ge, Si, and GaAs, at least two of the constants in Eq. (3.34) must be equal to satisfy symmetry requirements. Thus, for energies near the conduction-band minima in these materials, the allowed E - k relationships are

$$E - E_c \approx A(k_1^2 + k_2^2 + k_3^2) \quad (A = B = C) \quad (3.35)$$

and (writing down only one of the three possible variants)

$$E - E_c \approx Ak_1^2 + B(k_2^2 + k_3^2) \quad (B = C) \quad (3.36)$$

With E constant, Eq. (3.35) is readily recognized as the equation for a sphere centered at the band minimum. Eq. (3.36) with E held constant, on the other hand, is the mathematical expression for an ellipsoid of revolution, with k_1 being the axis of revolution. The GaAs conduction-band structure exhibits approximately spherical constant-energy surfaces described by Eq. (3.35)[†]; the Ge and Si constant-energy surfaces are all ellipsoids of revolution described by Eq. (3.36).

Examining the valence-band structure, one finds that the three subbands in Ge, Si, and GaAs are each characterized to first order by a plot identical to Fig. 3.14(c). In other words, the constant-energy surfaces about the $k = 0$ valence-band maxima are approximately spherical and are described to first order by Eq. (3.35) with $E - E_c \rightarrow E_v - E$. This is consistent with the earlier E - k diagram observation concerning the orientation independence of these subbands.[‡]

[†]For a more precise description of the GaAs conduction-band structure the reader is referred to Blakemore,^[1] pp. R157-R160.

[‡]The interaction between the heavy- and light-hole subbands gives rise to an E - k perturbation which must be taken into account in more exacting computations. Including the perturbation one finds^[5]

$$E_v - E = Ak^2 \pm [B^2k^4 + C^2(k_x^2k_y^2 + k_x^2k_z^2 + k_y^2k_z^2)]^{1/2}$$

with $k^2 = k_x^2 + k_y^2 + k_z^2$. Where the (\pm) appears in the above equation, the $(+)$ is used in treating the light-hole band and the $(-)$ in treating the heavy-hole band. To be precise, therefore, the constant-energy surfaces about the $k = 0$ valence-band maximum are actually somewhat distorted spheres. A visualization of the Si heavy-hole band distortion can be found in Ziman,^[3] p. 119.

3.3 EXTRAPOLATION OF CONCEPTS TO THREE DIMENSIONS

Having investigated the construction and interpretation of constant-energy plots, one might wonder about the general utility of the plots. First and foremost, the constant-energy plots are very helpful visual and conceptual aids. From these plots one can ascertain by inspection the positions and multiplicity of band extrema points. As we will see, the shapes of the surfaces also provide information about the carrier effective masses. References to the plots are often encountered in advanced device analyses, particularly those involving orientation-dependent phenomena. Herein we will make specific use of the Fig. 3.14 plots during the effective mass discussion and in the subsequent density-of-states derivation (Subsection 4.1.2).

3.3.4 Effective Mass

General Considerations

In the one-dimensional analysis the electron motion resulting from an impressed external force was found to obey a modified form of Newton's second law, $dv/dt = F/m^*$. The scalar parameter, $m^* = \hbar^2/(d^2E/dk^2)$, was identified as the electron effective mass. In three-dimensional crystals the electron acceleration arising from an applied force is analogously given by

$$\frac{dv}{dt} = \frac{1}{m^*} \cdot \mathbf{F} \quad (3.37)$$

where

$$\frac{1}{m^*} = \begin{pmatrix} m_{xx}^{-1} & m_{xy}^{-1} & m_{xz}^{-1} \\ m_{yx}^{-1} & m_{yy}^{-1} & m_{yz}^{-1} \\ m_{zx}^{-1} & m_{zy}^{-1} & m_{zz}^{-1} \end{pmatrix} \quad (3.38)$$

is the inverse effective mass tensor with components

$$\frac{1}{m_{ij}} = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k_i \partial k_j} \quad \dots i, j = x, y, z \quad (3.39)$$

An interesting consequence of the 3-D equation of motion is that the acceleration of a given electron and the applied force will not be colinear in direction as a general rule. For example, given a force pointing in the $+x$ direction, one obtains

$$\frac{dv}{dt} = m_{xx}^{-1} F_x \mathbf{a}_x + m_{yx}^{-1} F_x \mathbf{a}_y + m_{zx}^{-1} F_x \mathbf{a}_z \quad (3.40)$$

with \mathbf{a}_x , \mathbf{a}_y , and \mathbf{a}_z being unit vectors directed along the x , y , and z axes, respectively. Fortunately, the crystal and therefore the k -space coordinate system can always be rotated so as to align the k -space axes parallel to the principal axis system centered at a

78 CHAPTER 3 ENERGY BAND THEORY

band extrema point. Since the E - k relationship is parabolic about the band extrema point, all $1/m_{ij}$ where $i \neq j$ will then vanish, thereby eliminating off-diagonal terms in the effective-mass tensor. In general, therefore, a maximum of three effective-mass components are needed to specify the motion of carriers energetically confined to the vicinity of extrema points. Moreover, the equation of motion in the rotated coordinate system drastically simplifies to $dv_i/dt = F_i/m_{ii}$.

Ge, Si, and GaAs

An even further simplification results when one considers cubic crystals such as Ge, Si, and GaAs. For GaAs, the k_x - k_y - k_z coordinate system is a principal-axes system and the conduction-band structure is characterized to first order by the “spherical” E - k relationship

$$E - E_c = A(k_x^2 + k_y^2 + k_z^2) \quad (3.41)$$

Thus, not only do the $1/m_{ij}$ components with $i \neq j$ vanish, but

$$m_{xx}^{-1} = m_{yy}^{-1} = m_{zz}^{-1} = \frac{2A}{\hbar^2} \quad (3.42)$$

Defining $m_{ii} = m_e^*$, we can therefore write

$$E - E_c = \frac{\hbar^2}{2m_e^*}(k_x^2 + k_y^2 + k_z^2) \quad \dots \text{GaAs} \quad (3.43)$$

and

$$\frac{dv}{dt} = \frac{\mathbf{F}}{m_e^*} \quad (3.44)$$

For the conduction-band electrons in GaAs, the effective mass tensor reduces to a simple scalar, giving rise to an orientation-independent equation of motion like that of a classical particle. Obviously, spherical energy bands are the simplest type of band structure, necessitating a single effective-mass value for carrier characterization.

The characterization of conduction-band electrons in Ge and Si is only slightly more involved. For any of the ellipsoidal energy surfaces encountered in these materials one can always set up a k_1 - k_2 - k_3 principal-axis system where k_1 lies along the axis of revolution (see Fig. 3.15). The constant-energy surfaces are then described [repeating Eq. (3.36)] by

$$E - E_c = Ak_1^2 + B(k_2^2 + k_3^2) \quad (3.45)$$

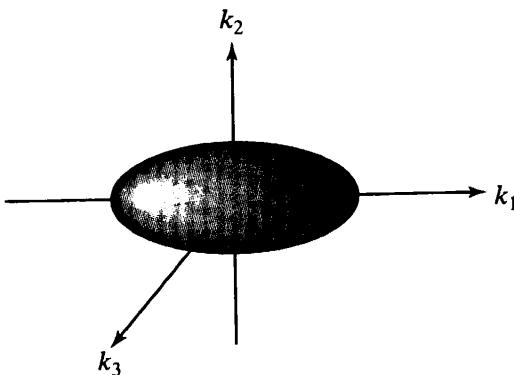


Figure 3.15 Principal-axis system for the ellipsoidal constant-energy surfaces in Ge and Si.

In the principal axis system the $i \neq j$ ($i, j = 1, 2, 3$) components of the inverse effective-mass tensor are again zero and

$$m_{11}^{-1} = \frac{2A}{\hbar^2} \quad (3.46a)$$

$$m_{22}^{-1} = m_{33}^{-1} = \frac{2B}{\hbar^2} \quad (3.46b)$$

Because m_{11} is associated with the k -space direction lying along the axis of revolution, it is called the *longitudinal* effective mass and is usually given the symbol m_ℓ^* . Similarly, $m_{22} = m_{33}$, being associated with a direction perpendicular to the axis of revolution, is called the *transverse* effective mass and is given the symbol m_t^* . In terms of the newly introduced symbols we can write

$$E - E_c = \frac{\hbar^2}{2m_\ell^*}k_1^2 + \frac{\hbar^2}{2m_t^*}(k_2^2 + k_3^2) \quad \dots \text{Ge, Si} \quad (3.47)$$

Now, Eq. (3.47) models any of the ellipsoidal energy surfaces in Ge and Si. For a given material, however, all the ellipsoids of revolution have precisely the same shape. Consequently, the two effective-mass parameters, m_ℓ^* and m_t^* , totally characterize the conduction-band electrons in Ge and Si.

The relative sizes of m_ℓ^* and m_t^* , it should be noted, can be deduced by inspection from the Si and Ge constant-energy plots. By comparing Eq. (3.47) with the general expression for an ellipsoid of revolution, one finds

$$\frac{m_\ell^*}{m_t^*} = \left(\frac{\text{Length of the ellipsoid along the axis of revolution}}{\text{Maximum width of the ellipsoid perpendicular to the axis of revolution}} \right)^2 \quad (3.48)$$

Examining Fig. 3.14(a) and (b) we therefore conclude $m_\ell^* > m_t^*$ for both Ge and Si. The greater elongation of the Ge ellipsoids further indicates that the m_ℓ^*/m_t^* ratio is greater for Ge than for Si.

Finally, let us turn to the characterization of holes in Ge, Si, and GaAs. As established previously, the valence-band structure in these materials is approximately spherical and composed of three subbands. Thus, the holes in a given subband can be characterized by a single effective-mass parameter, but three effective masses are technically required to characterize the entire hole population. (The split-off band, being depressed in energy, is only sparsely populated and is often ignored.) The subband parameters are m_{hh}^* , the heavy hole effective mass; m_{lh}^* , the light hole effective mass; and m_{so}^* , the effective mass of holes in the split-off band.

Measurement

All of the effective masses introduced in the preceding discussion (m_e^* , m_ℓ^* , m_t^* , m_{hh}^* , m_{lh}^* , and m_{so}^*) are directly measurable material parameters. The parametric values have been obtained in a relatively straightforward manner from cyclotron resonance experiments. The near-extrema point band structure, multiplicity and orientation of band minima, etc. were, in fact, all originally confirmed by cyclotron resonance data.

In the basic cyclotron resonance experiment, the test material is situated in a microwave resonance cavity and cooled to liquid helium temperatures ($T \approx 4$ K). A static magnetic field B and an rf electric field \mathcal{E} oriented normal to the B -field are applied across the sample; the Q -factor of the resonant cavity, reflecting the absorption of rf \mathcal{E} -field energy, is monitored as a function of B -field strength.

The force exerted by the B -field causes the carriers in the sample to move in an orbit-like path about the direction of the B -field (see Fig. 3.16). The frequency of the orbit, called the cyclotron frequency ω_c , is directly proportional to the B -field strength and inversely dependent on the effective mass (or masses) characterizing the carrier orbit. When the B -field strength is adjusted such that ω_c equals the ω of the rf electric field, the carriers absorb energy from the electric field and a resonance, a peak in the Q -factor, is observed. From the B -field strength, the B -field orientation, and the ω at resonance, one can deduce the effective mass or the effective mass combination corresponding

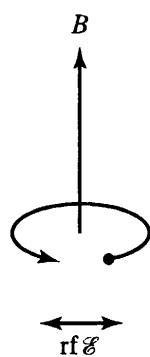


Figure 3.16 Carrier orbit and applied field orientations in the cyclotron resonance experiment.

3.3 EXTRAPOLATION OF CONCEPTS TO THREE DIMENSIONS

Table 3.1 Electron and Hole Effective Masses in Ge,^[6] Si,^[7] and GaAs^[1] at 4 K. (All values referenced to the free electron rest mass m_0 .)

Effective Mass	Ge	Si	GaAs
m_e^*/m_0	1.588	0.9163	—
m_t^*/m_0	0.08152	0.1905	—
m_e^*/m_0	—	—	0.067 [†]
m_{hh}^*/m_0	0.347	0.537	0.51
m_{th}^*/m_0	0.0429	0.153	0.082
m_{so}^*/m_0	0.077	0.234	0.154

[†] Band edge effective mass. The E - k relationship about the GaAs conduction-band minimum becomes non-parabolic and m_e^* increases at energies only slightly removed from E_c .

to a given experimental configuration. Repeating the experiment for different B -field orientations allows one to separate out the effective mass factors (deduce both m_e^* and m_t^* , for example) and to ascertain the orientational dependencies. The experiment is performed at low temperatures to maximize the number of orbits completed by the carriers between scattering events. Orbit disruption due to carrier scattering increases with temperature and tends to broaden and eventually eliminate resonance peaks. (For more information about the cyclotron resonance experiment, the reader is referred to ref. [5] for experimental details and to ref. [4] for an extended explanatory discussion.)

The 4 K effective mass values deduced from cyclotron resonance experiments for Ge, Si, and GaAs are listed in Table 3.1. The entries in this table were derived from definitive review-type works by Paige^[6] on Ge, by Barber^[7] on Si, and by Blakemore^[1] on GaAs. Note that, as previously inferred from the Fig. 3.14 constant-energy plots, $m_e^* > m_t^*$ (Ge, Si) and m_e^*/m_t^* (Ge) > m_e^*/m_t^* (Si).

In practical computations one would often like to know the temperature dependence of the effective mass parameters and, in particular, the parametric values at room temperature. Unfortunately, cyclotron resonance experiments cannot be performed at room temperature and limited data is available concerning the temperature dependence of these parameters. A theoretical extrapolation of the effective mass values to and above room temperature has nevertheless been performed by Barber^[7] for Si and by Blakemore^[1] for GaAs. As a matter of expediency, the effective masses are often implicitly assumed to be temperature independent and the 4 K values simply employed for all operational temperatures. In most cases the errors thereby introduced appear to be relatively minor.

3.3.5 Band Gap Energy

The band gap energy, $E_G = E_c - E_v$, is perhaps the most important parameter in semiconductor physics. At room temperature (roughly $T = 300$ K), $E_G(\text{Ge}) \approx 0.66$ eV,

82 CHAPTER 3 ENERGY BAND THEORY

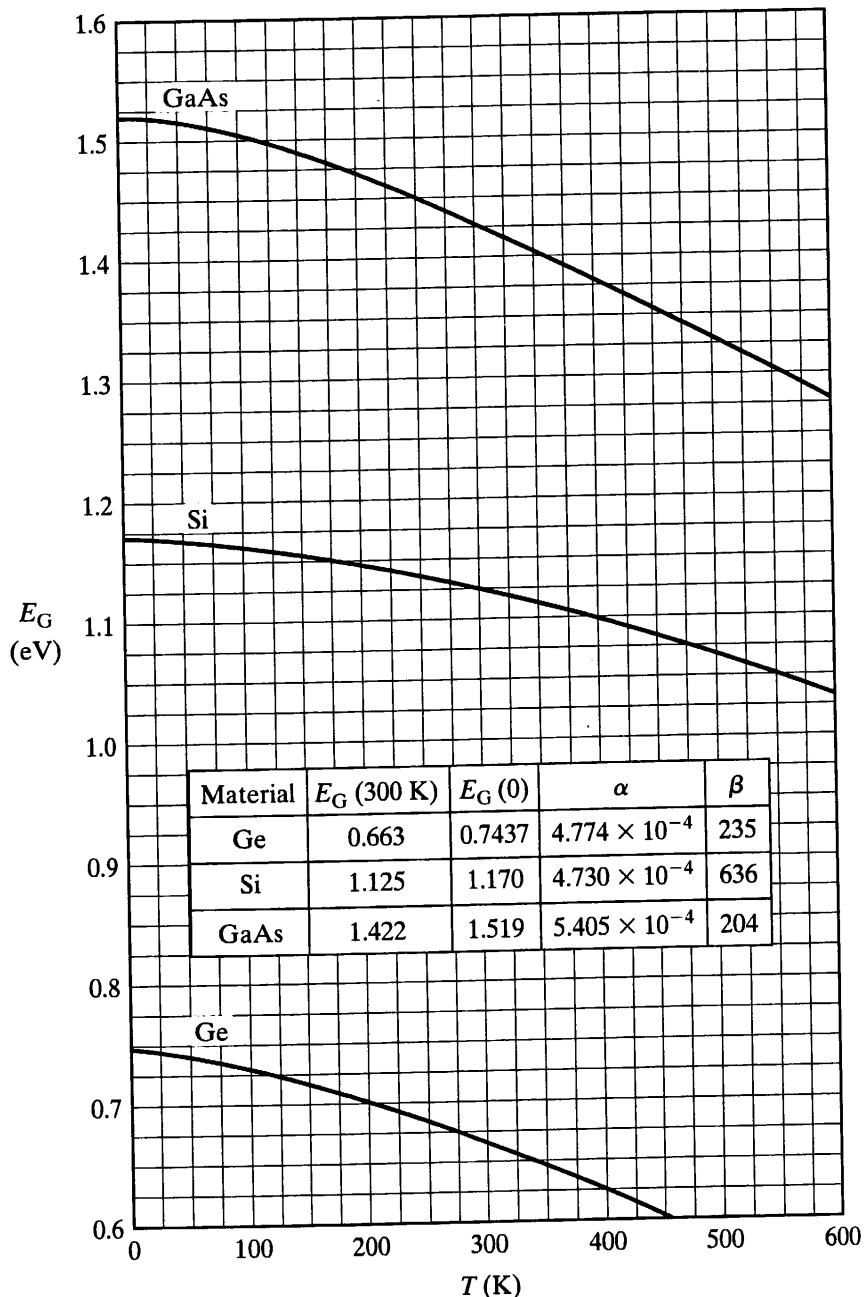


Figure 3.17 The band gap energy in Ge, Si, and GaAs as a function of temperature. The insert gives the 300 K values and the Eq. (3.49) fit parameters. (After Sze.^[2] Reprinted with permission.)

$E_G(\text{Si}) \approx 1.12 \text{ eV}$, and $E_G(\text{GaAs}) \approx 1.42 \text{ eV}$. With decreasing temperature a contraction of the crystal lattice usually leads to a strengthening of the interatomic bonds and an associated increase in the band gap energy. This is true for the vast majority of semiconductors including Ge, Si, and GaAs. To a very good approximation, the cited variation of band gap energy with temperature can be modeled by the “universal” empirical relationship

$$E_G(T) = E_G(0) - \frac{\alpha T^2}{(T + \beta)} \quad (3.49)$$

where α and β are constants chosen to obtain the best fit to experimental data and $E_G(0)$ is the limiting value of the band gap at zero Kelvin. The band gap versus temperature dependencies deduced from Eq. (3.49) for Ge, Si, and GaAs are plotted in Fig. 3.17; the fit parameters are specified in the figure insert. A more complete tabulation of semiconductor band gaps and other pertinent E - k information can be found in Appendix A of Wolfe et al.^[8]

REFERENCES

- [1] J. S. Blakemore, "Semiconducting and Other Major Properties of Gallium Arsenide," *J. Appl. Phys.*, 53, R123 (Oct., 1982).
- [2] S. M. Sze, *Physics of Semiconductor Devices*, 2nd edition, John Wiley & Sons, Inc., New York, 1981.
- [3] J. M. Ziman, *Electrons and Phonons, The Theory of Transport Phenomena in Solids*, Oxford University Press, London, 1960.
- [4] J. P. McKelvey, *Solid State and Semiconductor Physics*, Harper and Row, New York, 1966.
- [5] G. Dresselhaus, A. F. Kip, and C. Kittel, "Cyclotron Resonance of Electrons and Holes in Silicon and Germanium," *Phys. Rev.*, 98, 368 (1955).
- [6] E. G. S. Paige, *The Electrical Conductivity of Germanium (Progress in Semiconductors*, Vol. 8, edited by A. F. Gibson and R. E. Burgess), John Wiley & Sons, Inc., New York, 1964.
- [7] H. D. Barber, "Effective Mass and Intrinsic Concentration in Silicon," *Solid-State Electronics*, 10, 1039 (1967).
- [8] C. M. Wolfe, N. Holonyak Jr., and G. E. Stillman, *Physical Properties of Semiconductors*, Prentice Hall, Englewood Cliffs, NJ, 1989.

PROBLEMS

- 3.1** Answer the following questions as concisely as possible.
- (a) State the Bloch theorem for a one-dimensional system.
 - (b) The current associated with the motion of electrons in a totally filled energy band (a band in which all allowed states are occupied) is always identically zero. Briefly explain why.
 - (c) Define in words what is meant by a "Brillouin zone."
 - (d) Because of crystal symmetry one would expect the one-dimensional E versus k plots characterizing cubic crystals to be symmetrical about the Γ point. Why aren't the E versus k plots in Fig. 3.13 symmetrical about the Γ point?
- 3.2** The Kronig-Penney model solution presented in Subsection 3.2.2 is somewhat unconventional. To indicate why a nonstandard solution approach was presented (and, more generally, to coerce

84 CHAPTER 3 ENERGY BAND THEORY

the reader into examining the Kronig-Penney model solution), let us review the usual textbook approach. (You may wish to consult reference [4], pp. 213–214, to check your answers.)

(a) From the Bloch theorem we know

$$\psi(x) = e^{ikx}u(x)$$

where $u(x)$ is the unit cell wavefunction. Substitute the above expression for $\psi(x)$ into Schrödinger's equation and obtain the simplest possible differential equation for $u_a(x)$ [$u(x)$ for $0 < x < a$] and $u_b(x)$ [$u(x)$ for $-b < x < 0$]. Introduce α and β as respectively defined by Eqs. (3.8) and (3.10).

- (b) Record the general solutions for $u_a(x)$ and $u_b(x)$.
 - (c) Indicate the boundary conditions on the unit cell wavefunctions.
 - (d) Apply the part (c) boundary conditions to obtain a set of four simultaneous equations analogous to Eqs. (3.13).
 - (e) Show that Eq. (3.15) again results when one seeks a nontrivial solution to the part (d) set of equations. (NOTE: This part of the problem involves a considerable amount of straightforward but very tedious algebra.)
- 3.3 The comparison between the free-particle solution and the extended-zone representation of the Kronig-Penney model solution (Fig. 3.6) can be found in a number of texts. In the majority of texts the free-particle solution is pictured to be coincident with the lower energy band of the Kronig-Penney solution at all zone boundaries. Verify that Fig. 3.6 was constructed properly; i.e., verify that the dashed line, the free-particle solution, was drawn through the correct points in the figure.
- 3.4 A certain hypothetical material with cubic crystal symmetry is characterized by the E - k plot sketched in Fig. P3.4.
- (a) Which set of holes, band A holes or band B holes, will exhibit the greater [100]-direction (m_{xx}) effective mass? Explain.
 - (b) Sketch the expected form of the valence-band constant-energy surfaces for the represented cubic material. Assume that the E - k relationship is parabolic (i.e., an ellipsoid of revolution)

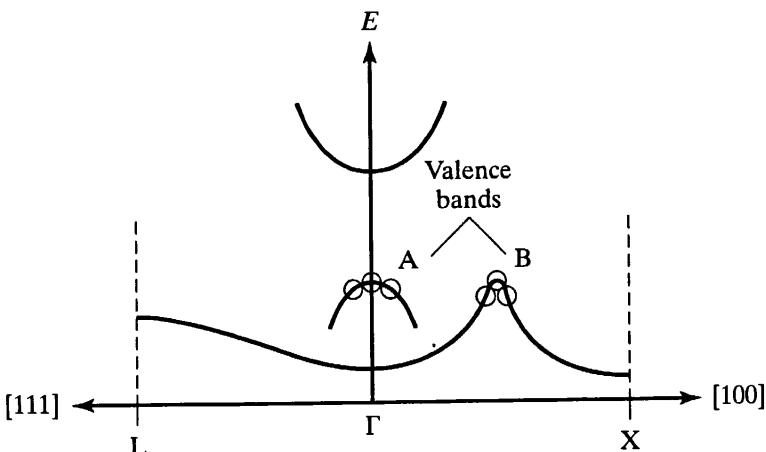


Figure P3.4

in the vicinity of the B maxima. Include the constant energy surfaces associated with both the A and B bands.

- 3.5** The E - k relationship characterizing an electron confined to a *two-dimensional* surface layer is of the form

$$E - E_c = \frac{\hbar^2 k_x^2}{2m_1} + \frac{\hbar^2 k_y^2}{2m_2} \quad \dots m_1 \neq m_2$$

An electric field is applied in the x - y plane at a 45 degree angle to the x -axis. Taking the electron to be initially at rest, determine its direction of motion in the x - y plane.

- 3.6** Consider the detailed E - k diagram for GaAs reproduced in text Fig. 3.13(d).
- (a) How does one deduce the 300 K value of E_G from the diagram. Is the value of E_G (300 K) deduced from the diagram consistent with that quoted in the Fig. 3.17 insert?
 - (b) How does one deduce that GaAs is a direct semiconductor from the given E - k diagram?
 - (c) As best as can be determined from Fig. 3.13(d), is the pictured E - k diagram consistent with the effective masses for GaAs quoted in Table 3.1? Explain.
- 3.7** (a) The E - k relationship about the GaAs conduction-band minimum becomes non-parabolic at energies only slightly removed from E_c and is more accurately described by

$$E - E_c = ak^2 - bk^4 \quad (a > 0, b > 0)$$

What effect will the cited fact have on the effective mass of electrons in the GaAs conduction band? Substantiate your conclusion. (Is your answer here in agreement with the Table 3.1 footnote?)

- (b) Electrons in GaAs can transfer from the Γ minimum to the L minima at sufficiently high electric fields. If electrons were to transfer from the Γ minimum to the L minimum shown in Fig. 3.13(d), would their effective mass increase or decrease? Explain. (The constant-energy surfaces about the L minima are actually ellipsoidal, but for simplicity assume the surfaces to be spherical in answering this question.)
- 3.8** Like GaAs, GaP crystallizes in the zincblende lattice and the valence band maxima occur at the Γ -point in the first Brillouin zone. Unlike GaAs, the conduction band minima in GaP occur at the X-points in the Brillouin zone.
- (a) Where are the X-points located in k -space?
 - (b) Is GaP a direct or indirect material? Explain.
 - (c) Given that the constant energy surfaces at the X-points are ellipsoidal with $m_\ell^*/m_0 = 1.12$ and $m_t^*/m_0 = 0.22$, what is the ratio of the longitudinal length to the maximum transverse width of the surfaces?
 - (d) Picturing only that portion of the constant energy surfaces within the first Brillouin zone, construct a constant-energy surface diagram characterizing the conduction-band structure in GaP.

86 CHAPTER 3 ENERGY BAND THEORY

- 3.9 (a)** Derive Eq. (3.48).

HINT: Show that Eq. (3.47) can be manipulated into the form

$$\frac{k_1^2}{\alpha^2} + \frac{k_2^2 + k_3^2}{\beta^2} = 1$$

where

$$\alpha \equiv \sqrt{(2m_t^*/\hbar^2)(E - E_c)}$$

$$\beta \equiv \sqrt{(2m_t^*/\hbar^2)(E - E_c)}$$

Confirm (quote your reference) that the above k_1 - k_2 - k_3 expression is the defining equation for a prolate spheroid—the ellipsoidal surface formed by the rotation of an ellipse about its major axis. 2α and 2β are, respectively, the lengths of the major and minor axes of the rotated ellipse.

- (b) Are the Fig. 3.14 ellipsoidal surfaces for Ge and Si in general agreement with the m_t^*/m_t^* ratios deduced from Table 3.1? Explain.
- 3.10** In Table 3 on p. 1318 of R. Pässler, Solid State Electronics, 39, 1311 (1996), there is a listing of what the author considers to be the most accurate silicon E_G versus temperature data and the corresponding computed E_G values from “superior” empirical fits to the experimental data. Using Eq. (3.49) and the parameters listed in the text Fig. 3.17 insert, compute E_G to 5 places at 50 K intervals from 50 K to 500 K. Compare your computed E_G values with the tabulated values in the cited reference.