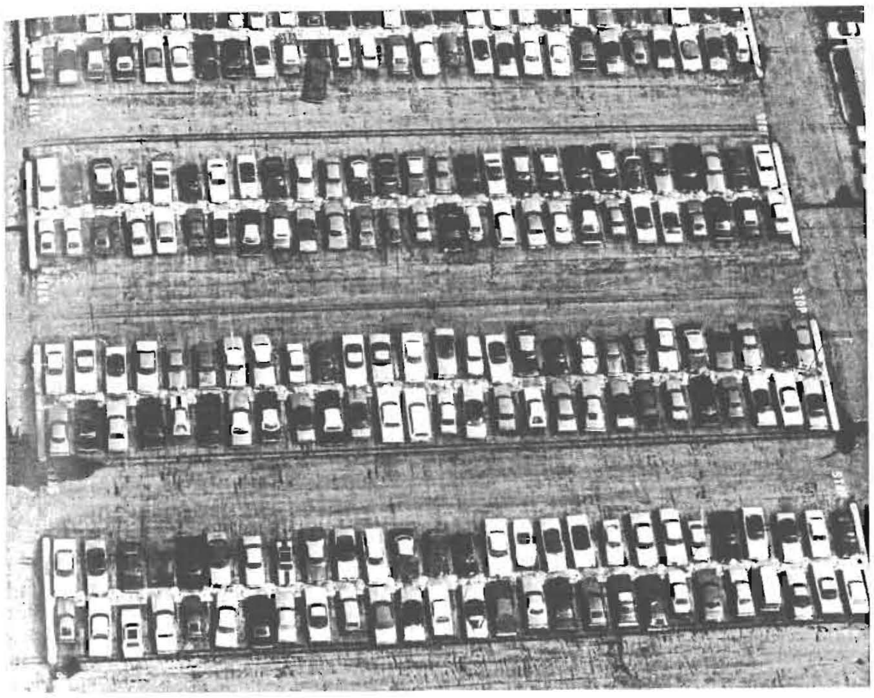
SCPY3

Chap 2



Chapter 24

Band Theory of Solids

24.1 INTRODUCTION

The free electron models of metals that we have presented in the preceding chapter gives us a good deal of insight into several properties of metals. Yet there are many other important properties that these models do not explain. In particular, they do not tell us why, when chemical elements crystallize to become solids, some are good *conductors*, some are *insulators*, and yet others are *semiconductors* with electrical properties that vary greatly with temperature. These differences are not minor, but rather remarkable. The resistivity may vary from $\rho \sim 10^{-8}$ ohm-m for a good conductor to $\rho \sim 10^{22}$ ohm-m for a good insulator.

We can understand the differences between insulators and conductors by extending the free electron model to take into account the interaction of the electrons with the positive ion lattice. In the quantum mechanical free electron (QMFE) model, we assumed that the potential energy inside the solid was uniform. It would be more realistic to assume that it is a periodic (alternating uniformly) function of x, y, z. This is reasonable because of the periodic distribution of the lattice ions in a *crystalline* solid.

When the interaction between the electrons and the lattice ions is considered, we will find some unusual properties possessed by the electrons in the crystal.

1 In the last chapter we saw that the QMFE model gave rise to a series of discrete energy states, about 10^{-15} eV apart, which ranged from $E \approx 0$ to $E = E_F$, the Fermi level. This range of energy levels can be called a *band* of energies, and the energy levels are so close together that they are referred to as quasicontinuous. In the band of the QMFE model there were an infinite number of unoccupied energy states above the Fermi level that could be occupied by excited electrons. When we introduce the potential of the lattice ions, we will see that bands of this type have upper and lower limits of allowable energies. If a band is not filled with electrons, then the electrons may be excited into the empty states and contribute to electrical or thermal conduction. If, however, a band is filled, then there are no states to be occupied and the electrons cannot be excited. For conductors, we will see that the behavior of the electrons in the occupied band of highest energy is almost identical to that predicted by the QMFE model, but not so for insulators and semiconductors. The band theory model will solve the questions concerning the differences between conductors and insulators.

2 The electrons respond to an externally applied electric or magnetic field as if they were endowed with an *effective mass* m^*, which may be greater or smaller than that of the free electron, or even be negative. By this we mean that the electrons can be treated as free in responding to an external electric and magnetic field provided that we assign to them a mass different from the true mass.

3 There are situations in which instead of conduction by electrons it is convenient to attribute the conduction to charge carriers with a positive charge $+e$, called *holes*.

There are several methods (or models) to show the existence of bands and to find the shape of the band. Some work well for certain materials, some with others. Quantitative band calculations must take into account the particular crystal structure, atomic configuration, and type of bonding: These details belong to the realm of solid-state physics research. Our purpose here is simply to show the existence of bands and the general characteristics. This can be achieved with idealized models and by using qualitative arguments.

24.2 BLOCH'S THEOREM

Before we proceed to study the motion of an electron in a periodic potential,[1] we should mention a general property of the wavefunctions in such a periodic potential.

For a free electron with E_p = constant, the space part of the wave function $\psi(x, t)$, called the eigenfunction $\chi(x)$, is written as (see Eq. 20.4)

$$\chi(x) = e^{\pm ikx}$$

If the spacing of the ions in the x direction in a solid is d, then the potential energy of an electron at a point x distance from the origin is equal to the potential energy at a point $x + d$ from the origin. This potential energy is equal in turn to that at point $x + 2d$ from the origin, and so on. Therefore, we can generalize and take any point x in the lattice and state that the potential energy at that point is equal to the potential energy at point $x + d$ or, stated mathematically, $E_p(x) = E_p(x + d)$. This s known as a *periodic potential*. There is a theorem by Bloch which states that for a particle moving in a periodic potential, the eigenfunctions $\chi(x)$ are of the form

$$\chi(x) = u_k(x)\, e^{\pm ikx} \tag{24.1}$$

where

$$u_k(x) = u_k(x + d)$$

These eigenfunctions are plane waves modulated by a function $u_k(x)$, where $u_k(x)$ has he same periodicity as the potential energy. Because the potential energy $E_p(x) = E_p(x + d)$, one expects that the probability of finding a particle at a given x is the same

---

[1] As indicated earlier, in Chapter 20, in quantum mechanics the *potential energy* is often called the *otential*.

as that of finding it at $x + d$. This is guaranteed by the periodicity of u_k and can be seen in the following expression for the probability density.

$$\begin{aligned}\chi^*(x)\,\chi(x) &= u_k^*(x)\,e^{-ikx}\,u_k(x)\,e^{ikx}\\ &= u_k^*(x)\,u_k(x)\end{aligned}$$

Therefore, when

$$u_k(x) = u_k(x + d)$$

then

$$\chi^*(x)\,\chi(x) = \chi^*(x + d)\,\chi(x + d)$$

The specific form of the function $u_k(x)$ will depend on the form of the function $E_p(x)$. We will now consider an idealized, one-dimensional periodic potential.

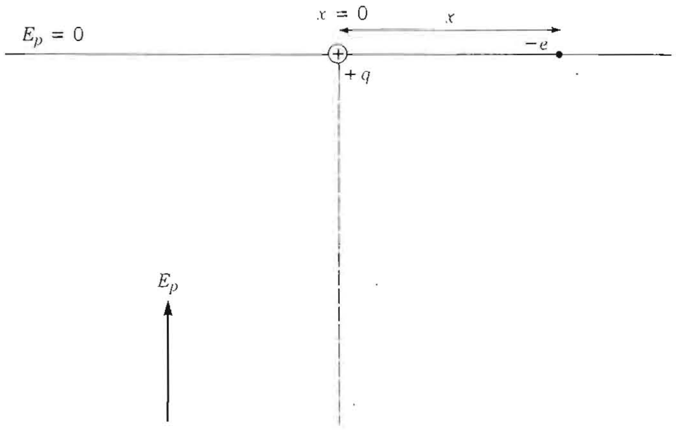
24.3 THE KRONIG-PENNEY MODEL

Let us try to understand what the potential energy of an electron in a crystalline solid may look like. Consider a positively charged ion q and an electron e at a distance x from q as shown in Fig. 24-1. The electric potential energy from the coulomb attraction experienced by the electron is (Eq. 14.9).

$$E_p(x) = -\frac{1}{4\pi\epsilon_0}\frac{q|e|}{x}$$

The variation of E_p with x is illustrated in Fig. 24-1. Suppose we now place another charge q at a point d away from the first. The potential energy E_p at any point on the x axis will be equal to the algebraic sum of the potential energies due to each individual charge, as illustrated in Fig. 24-2. The dashed lines represent the potential energy due to the individual q's and the solid lines represent the sum of the dashed lines. If we now place a long array of q's separated by a distance d from each other to form a periodic array, the potential energy E_p looks like that shown in Fig. 24-3. The main features of the potential energy in Fig. 24-3 are: (1) it is periodic with a period d, (2) the maxima are halfway between the ions, and (3) the potential energy tends to $-\infty$ as the position of the ions is approached because the electron is bound more strongly to the ion as it comes closer and, because it takes more energy to pull it away, it can be said to lie in a deeper potential energy well the closer it is to the ion.

If one tries to solve the Schrödinger equation for such a potential, one runs into mathematical difficulties that are best solved by a computer. However, we can replace the potential energy of Fig. 24-3 with one that is mathematically simpler to handle while retaining the essential features of the actual one. We replace the potential energy of Fig. 24-3 by one consisting of periodically spaced rectangular wells as shown in

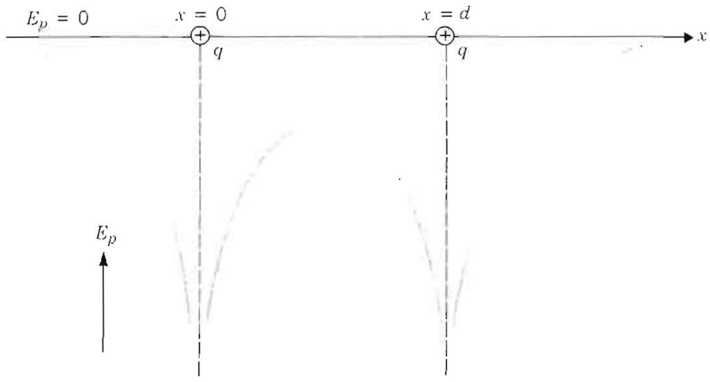


Potential energy E_p associated with the coulomb interaction of an electron with a positive ion $+q$ as a function of the separation x between the electron and the ion.

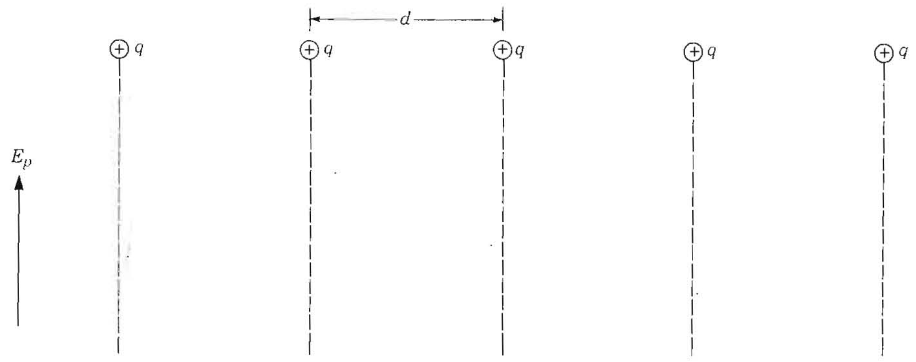
Fig. 24-4. The potential energy is a series of rectangular wells of width c, spaced a distance b apart so that the periodicity $d = b + c$. The energy of the wells is $-E_{p0}$. However, it is convenient to shift the zero of potential energy so that the bottoms of the wells are at potential energy $E_p = 0$ and the tops are at $E_p = E_{p0}$. The potential energy of Fig. 24-4 has the same periodicity as the lattice; the potential energy is lower in the vicinity of the ions and highest between the ions. This potential energy model is known as the Kronig-Penney model.

To find the behavior of the electrons in such a periodic potential, we have to find the eigenfunction that we can associate with them by solving the Schrödinger equation. We will not show all the mathematical details, but instead we will outline the method in order to understand the origin of the final result.

Because E_p is either 0 or E_{p0}, we solve for χ separately in both regions I and II. We then impose the conditions of continuity for χ and $d\chi/dx$ discussed in Chapter 20, while meeting the periodicity requirements. We will consider the case where $E < E_{p0}$.



Potential energy E_p when an electron interacts with two ions (solid lines). The dashed lines correspond to the potential energy associated with the interaction of the electron with the individual ions. The solid lines represent the sum of the dashed lines.



Potential energy of an electron in a one-dimensional array of periodically spaced ions. The periodicity of the ions is d.

Region I

In region I $E_p = 0$ and the Schrödinger equation is written as (Eq. 20.18)

$$-\frac{\hbar^2}{2m}\frac{d^2\chi_I}{dx^2} = E\,\chi_I$$

where χ_I is the eigenfunction in region I. Rearranging terms, we get

$$\frac{d^2\chi_I}{dx^2} + \gamma^2\chi_I = 0 \quad \text{where } \gamma = \sqrt{\frac{2mE}{\hbar^2}}$$

From Bloch's theorem, $\chi_I = u_I(x)\,e^{ikx}$ (Eq. 24.1). If we substitute this χ in Eq. 24.2, we get a differential equation for u_I,

$$\frac{d^2u_I}{dx^2} + 2ik\frac{du_I}{dx} + (\gamma^2 - k^2)\,u_I = 0$$

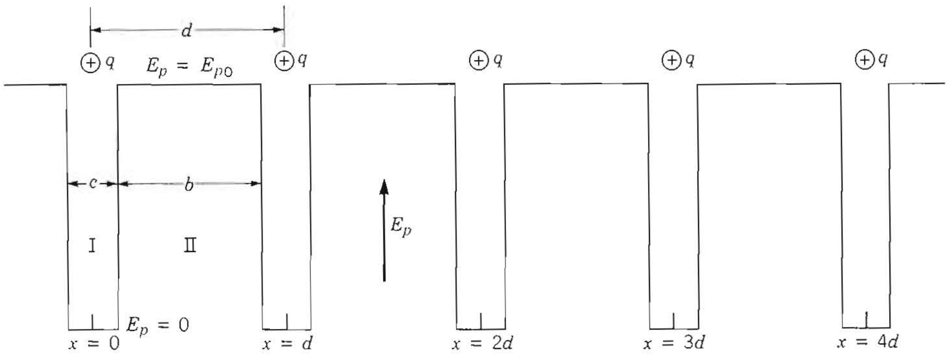


FIGURE 24-4 Periodic array of rectangular potential energy wells of depth E_{p0}, width c, separated by a distance b. The periodicity of the wells is $d = b + c$. Such an array of potential energy wells can be used to approximate the potential energy of Fig. 24-3.

The solution of this equation can be found by the methods that we used before in Chapter 20 and is

$$u_{\mathrm{I}}(x) = A\,e^{i(\gamma-k)x} + B\,e^{-i(\gamma+k)x}$$

where A and B are arbitrary constants.

Region II

In region II $E_p = E_{p0}$ and the Schrödinger equation is written as

$$-\frac{\hbar^2}{2m}\frac{d^2\chi_{\mathrm{II}}}{dx^2} + E_{p0}\,\chi_{\mathrm{II}} = E\,\chi_{\mathrm{II}}$$

where χ_{II} is the eigenfunction in region II. Rearranging terms, we get

$$\frac{d^2\chi_{\mathrm{II}}}{dx^2} - \xi^2\chi_{\mathrm{II}} = 0 \quad \text{where } \xi = \sqrt{\frac{2m(E_{p0}-E)}{\hbar^2}}$$

If we substitute $\chi_{\mathrm{II}} = u_{\mathrm{II}}(x)\,e^{ikx}$, we will get a differential equation for u_{II} that can be solved by the same method yielding for u_{II}

$$u_{\mathrm{II}} = C\,e^{(\xi-ik)x} + D\,e^{-(\xi+ik)x}$$

where C and D are arbitrary constants.

The next step is to impose the requirements of continuity and periodicity, between the regions I and II. It is seen in Fig. 24-4 that regions I and II join at $x = c/2$, therefore, following the discussion in Section 20.2e, we recall that both the eigenfunctions and their first derivatives must be continuous across a boundary; mathematically these criteria are

$$\chi_{\mathrm{I}}\left(\frac{c}{2}\right) = \chi_{\mathrm{II}}\left(\frac{c}{2}\right)$$

$$\frac{d\chi_{\mathrm{I}}}{dx}\left(\frac{c}{2}\right) = \frac{d\chi_{\mathrm{II}}}{dx}\left(\frac{c}{2}\right)$$

In addition, the periodicity requirements must be satisfied. This can be done by choosing points separated by the period of the lattice d, such as $x = -c/2$ and $x = b+c/2$. When we substitute Bloch functions for the χ functions, that is, $\chi = u(x)\,e^{+ikx}$, the periodicity requirement on the function $u(x)$ yields

$$u_{\mathrm{I}}\left(-\frac{c}{2}\right) = u_{\mathrm{II}}\left(b+\frac{c}{2}\right)$$

$$\frac{du_{\mathrm{I}}}{dx}\left(-\frac{c}{2}\right) = \frac{du_{\mathrm{II}}}{dx}\left(b+\frac{c}{2}\right)$$

These four conditions on the eigenfunctions lead to four linear algebraic equations for the constant A, B, C, D. In solving these equations, it is found that a solution

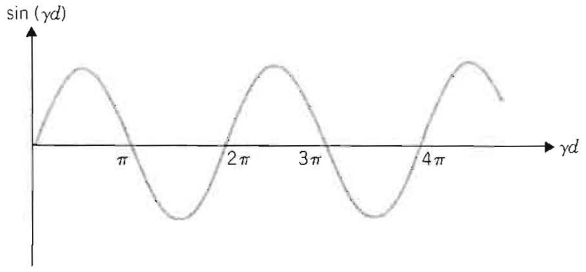


FIGURE 24-5

exists only if[2]

$$P\frac{\sin \gamma d}{\gamma d} + \cos \gamma d = \cos kd \qquad (24.3)$$

where

$$P = \frac{mE_{p0}bd}{\hbar^2} \quad \text{and, from Eq. 24.2,} \quad \gamma = \sqrt{\frac{2mE}{\hbar^2}}$$

(Note: This P is a new term with no relation to momentum). If this condition is not satisfied, the boundary conditions on χ cannot be satisfied, and the corresponding χ's are not acceptable solutions. In arriving at Eq. 24.3 we used one of the forms of Bloch functions, namely, $\chi = u(x)\,e^{+ikx}$. The same result will be obtained if we use the other form, that is, $\chi(x) = u(x)\,e^{-ikx}$.

24.3a. Allowed and Forbidden Energy Bands

The main result of the solution of the Schrödinger equation for the periodic potential of Fig. 24-4 is that the only acceptable χ's are those for which Eq. 24.3 holds.

Let us try to understand the significance of Eq. 24.3. Remember that

$$\gamma = \sqrt{\frac{2mE}{\hbar^2}}$$

Therefore γ is a measure of the total energy, E, whereas k is the wave vector $k = 2\pi/\lambda$. From de Broglie's hypothesis, Eq. 19.1, $\lambda = h/p$ and therefore $p = h\hbar$. Therefore k is a measure of the momentum of the particle. Equation 24.3 relates the total energy E of the electron to its momentum p. (The corresponding expression for the free particle is $E = p^2/2m = k^2\hbar^2/2m$.) Finding a direct analytical expression $E(k)$ is not possible because Eq. 24.3 is a transcendental equation that cannot be solved analytically. We can, however, solve it numerically: We choose a γ (an E), insert it

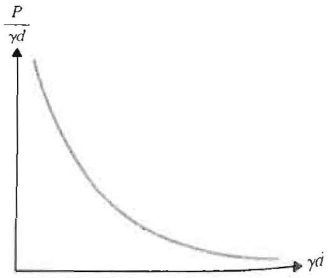


FIGURE 24-6

---

[2] In arriving at Eq. 24.3, an additional simplification is made, namely E_{p0} is assumed to approach infinity and b to tend to zero while their product $E_{p0}b$ remains constant.

into the equation, and solve for k. This is not difficult; in fact, it is a simple exercise in computer programming. When we do this we will find a rather interesting result. There will be ranges of γ for which k will be a real number. Those ranges of γ will be separated by other ranges for which k is imaginary. The momentum of a particle cannot be imaginary, and the conclusion is as follows: A particle in this periodic potential cannot have values of γ for which k is imaginary; therefore, the corresponding values for E for these γ's are not allowed.

We do not have to go through the tedious task of selecting a γ and substituting it into the Eq. 24.3 to show this. We can get the result by qualitatively plotting the left side of Eq. 24.3, calling it $f(\gamma d)$.

$$f(\gamma d) = P\frac{\sin \gamma d}{\gamma d} + \cos \gamma d$$

For example, let $P = 5/2\pi$. The sine function is periodic, as in Fig. 24-5, and $P/\gamma d$ behaves as in Fig. 24-6. When we multiply these two functions to get the first term of $f(\gamma d)$, an oscillating function similar to $\sin \gamma d$ results, but the amplitude will decrease with increasing γd. Some values are shown in Table 24-1.

TABLE 24-1

γd	π	$\frac{3}{2}\pi$	2π	$\frac{5}{2}\pi$	3π	$\frac{7}{2}\pi$	4π
$P\frac{\sin \gamma d}{\gamma d}$	0	$-\frac{5}{3}$	0	1	0	$-\frac{5}{7}$	0

Between 0 and π we have to be careful, particularly close to $\gamma d = 0$, because when $\gamma d = 0$, $P \sin \gamma d/\gamma d = P\,0/0$, which is undetermined. We can, however, use the L'Hospital rule on limits that

$$\lim_{x \to 0} \frac{\sin x}{x} = \lim_{x \to 0} \frac{\dfrac{d}{dx}\sin x}{\dfrac{d}{dx}x} = \lim_{x \to 0} \frac{\cos x}{1} = 1$$

Thus, in the limit as $\gamma d \to 0$, $\sin \gamma d/\gamma d \to 1$. As γd begins to increase from 0, both $\sin \gamma d$ and γd increase; however, the ratio decreases. If we express γd in radians, we can easily show with a calculator that $\sin \gamma d/\gamma d$ is a decreasing function of γd, which becomes 0 when $\gamma d = \pi$. We can put all these facts together to obtain the plot of the first terms of $f(\gamma d)$; this is shown in Fig. 24-7. Note that the larger P, the greater the slopes will be, because the zero positions are fixed. To get the entire function $f(\gamma d)$, we must add to Fig. 24-7 the term $\cos \gamma d$, Fig. 24-8. Between 0 and π both functions decrease; therefore, $f(\gamma d)$ decreases and becomes -1 when $\gamma d = \pi$. After π, the first term continues decreasing, while the second begins to increase. Because $\cos \gamma d$ changes slowly

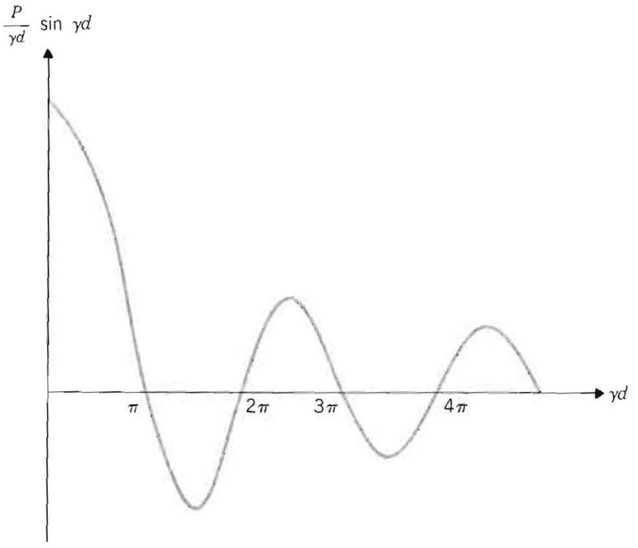


FIGURE 24-7 Plot of the first term of the left side of Eq. 24.3 as a function of γd. The plot, obtained by multiplying the sine function of Fig. 24-5 by the function of Fig. 24-6, is an oscillating function of γd with decreasing amplitude.

near the maximum and minimum, $f(\gamma d)$ continues to decrease below -1. Somewhere between π and $3/2\ \pi$ the trend reverses itself and $f(\gamma d)$ begins to increase, reaching the value of $+1$ at 2π. After 2π the first term of $f(\gamma d)$ continues increasing while the second decreases. Again, just as before, the first term increases at a faster rate than the rate of decrease of the second term, and as a result $f(\gamma d)$ continues (for a while) to increase past $+1$. Somewhere between 2π and $5/2\ \pi$ the trend will reverse.

These arguments are reflected in a plot of $f(\gamma d)$ versus γd, Fig. 24-9. The most important fact to note is that there are ranges of γd (shaded regions) for which the values of $f(\gamma d)$ vary between $+1$ and -1. These ranges of γd are separated by others for which $f(\gamma d)$ is either greater than $+1$ or less than -1. The width of the shaded region, the ranges of γd for which $f(\gamma d)$ varies between $+1$ and -1, increases as γd increases. The condition that had to be satisfied for the solutions to the Schrödinger equation to be acceptable was, Eq. 24.3,

$$f(\gamma d) = \cos kd$$

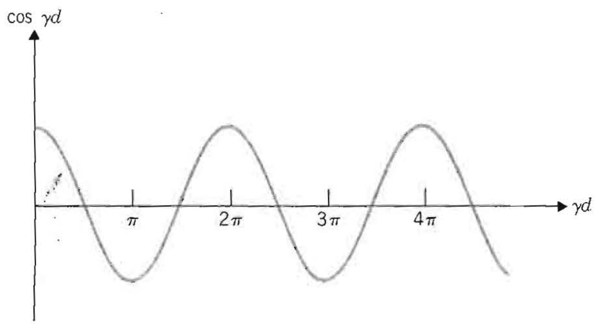


FIGURE 24-8

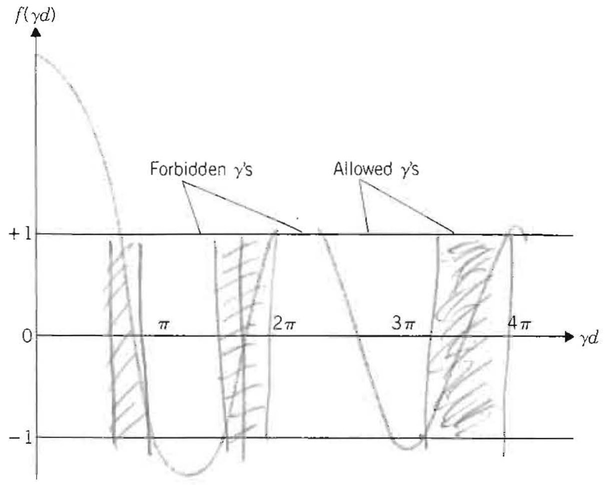


FIGURE 24-9 Plot of the function $f(\gamma d)$, that is, the left side of Eq. 24.3 versus γd (oscillating solid line). The shaded areas represent the values of γd for which $f(\gamma d)$ varies between 1 and -1. These values of γd correspond to allowed energy values. They are separated by ranges of γd for which $f(\gamma d)$ is either greater than 1 of less than -1. These values of γd correspond to forbidden energy values.

Because cos kd takes values ranging from $+1$ to -1, this means that this condition can be satisfied only by those values of γ for which $f(\gamma d)$ lies within those limits. The values of γ for which $f(\gamma d)$ is outside these limits correspond to γ's for which the boundary conditions cannot be satisfied and, therefore, these γ's (and the corresponding E's) are not physically acceptable. We conclude that *the electron may possess energies within certain bands of energy but not outside of them: There are allowed and forbidden bands of energy available to electrons moving in a periodic lattice.*

Another conclusion to be drawn from Fig. 24-9 is that the width of the allowed energy bands increases with increasing γ (increasing energy E). The physical reason for this will become clear when we look at an alternative way of showing how the bands come about in Section 24.4.

Dispersion Relation

For a free particle the relation between the energy E and the momentum p is $E = p^2/2m$. From de Broglie's relation $p = h/\lambda$ (Eq. 19.1), and because $\lambda = 2\pi/k$ (Eq. 11.12), it follows that

$$p = \frac{h}{\lambda} = \frac{hk}{2\pi} = \hbar k$$

Substitution of this relation for p into the energy relation shows that the energy of the particle can be expressed in terms of the wave vector as

$$E = \frac{\hbar^2 k^2}{2m} \tag{24.4}$$

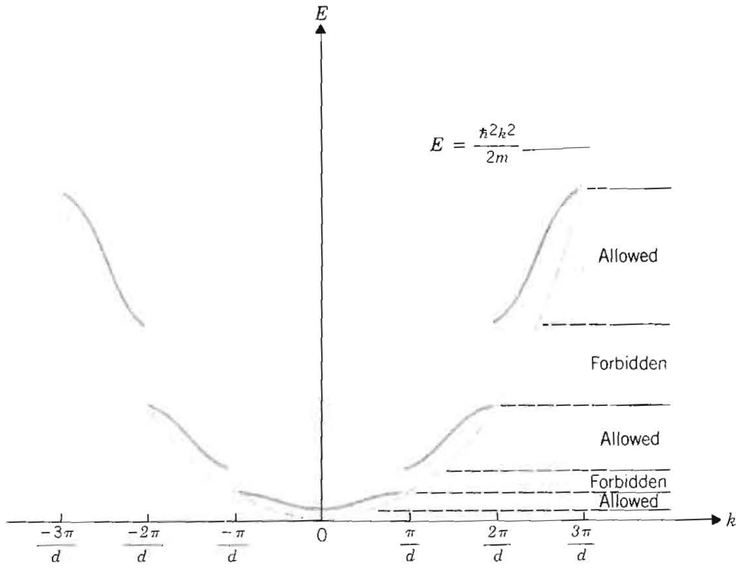


FIGURE 24-10 The solid lines show the dependence of the energy E on the wave vector k (and therefore, the momentum p) for an electron moving in the periodic array of rectangular potential wells of Fig. 24-4. The dashed line represents the relation between the energy and the mementum for the free electron case. The values of E for which Eq. 24.3 yields a real number for k are projected to the right of the figure as allowed energy bands (shaded bands). These are separated by ranges of E for which Eq. 24.3 yields an imaginary solution for k and correspond to forbidden energy values.

The relation between the energy of a particle and its wave vector is often referred to as the *dispersion relation*. For a free particle, this relation is parabolic; that is, $E \propto k^2$, Eq. 24.4. This dependence is illustrated by the dashed lines of Fig. 24-10.

When the particle is not free, the dispersion relation is usually more complicated. Thus, as we have seen in Section 24.3a, for an electron moving in a one-dimensional array of potential wells the dispersion relation is given by Eq. 24.3

$$P\frac{\sin \gamma d}{\gamma d} + \cos \gamma d = \cos kd \qquad (24.3)$$

where

$$\gamma = \sqrt{\frac{2mE}{\hbar^2}}$$

We saw that Eq. 24.3 is transcendental and must be solved numerically. We pick a value of E, substitute it in Eq. 24.3, and obtain the value of k for which the relation holds. If the procedure is repeated systematically for other values of E, we will be able to make a table listing values of E and the corresponding values of k. As we proceed with these numerical calculations, we find that there are energy intervals for which no real solution for k exists. These are the values of E for which the left side of Eq. 24.3 is either greater than $+1$ or less than -1. As indicated in the previous section, this

is physically unacceptable and, therefore, these energy values are forbidden. Results from the numerical method just described are illustrated by the solid lines of Fig. 24-10. We note that these solid lines yield the values of E and the corresponding values of k for certain ranges of E. For other energy intervals the value of k is not defined by the solid lines, these are the forbidden energies. Allowed and forbidden energies are projected to the right in Fig. 24-10 to represent the scheme of allowed and forbidden energy bands. Another important result to be noted in Fig. 24-10 is that the curvature of the solid lines is not the same as that of the dashed line, which represents the dispersion relation for the free particle. This has important implications concerning the effective mass of the electrons and will be discussed in detail in Section 24.6.

Although the Kronig-Penney model presented here shows clearly the existence of allowed and forbidden energy bands and at the same time gives us a mathematical expression with which to find the E versus k curves, it does not give much physical insight for the existence of these bands. Moreover, as presented here, it does not answer a question that will be important in the understanding of the difference between conductors, insulators, and semiconductors. The question is: *How many energy states are allowed within a given band?* The answer so far seems to be an infinite number because within a band γ can vary continuously. It would seem, therefore, that E can take an infinite, continuous range of values within an allowed band. The reason for this result is that in the periodic potential model we have assumed that the periodicity is infinitely long. In a real solid we have boundaries and, although we may have 10^{23} potential wells, it is still a finite number of wells. If one introduces the boundary condition that we used in the infinite potential well model of Chapter 23, namely, $\chi = 0$ at the boundaries of the solid, one finds that the continuous spectrum within a band breaks into a quasicontinuous one. Rather than doing this directly, let us look at an alternative method of showing the existence of bands. This method is less quantitative, less mathematical, and less rigorous (although when used to do actual calculations it can be made rigorous), but it clearly shows the physical reason for the bands. This method will also yield the number of energy states allowed within a given band.

24.4 TIGHT-BINDING APPROXIMATION

One useful way to look at the formation of allowed and forbidden energy bands is to start with the energy levels of the individual neutral atoms when they are very far apart and watch the changes in these levels as the atoms get close together and the charge distributions of adjacent atoms begin to overlap. We can gain some insight into what happens by studying a simple one-dimensional quantum mechanical problem: that of two finite square potential wells.

When we examined the one-dimensional infinite potential well of width a, Chapter 20, we found that the eigenfunctions and corresponding energy values that the particle may have are given by Eqs. 20.25 and 20.26.

$$\chi_n = B \sin n \frac{\pi}{a} x \qquad (20.25)$$

$$E_n = n^2 E_0 \qquad n = 1, 2, 3, \ldots \qquad (20.26)$$

where

$$E_0 = \frac{\hbar^2 \pi^2}{2ma^2}$$

In Fig. 24-11 we show the first two eigenfunctions for the infinite potential well (see Fig. 20-4). For the finite well the results are similar but with minor differences. One such difference is that the eigenfunctions do not vanish at the boundary but extend a little bit outside the well (see Fig. 24-12).

Let us consider the two finite potential wells B and C shown in Fig. 24-13*a*. If we have an electron with energy E_1 that we know is definitely in well B, (i.e., well C is treated as though it is absent) the χ describing such an electron will look as in Fig. 24-13*b*; χ_B and therefore χ_B^2 differ from 0 only in the region of the well B. Suppose instead that the electron is definitely in well C with energy E_1. Then the eigenfunction will appear as shown in Fig. 24-13*c*; χ_C and therefore χ_C^2 differ from 0 only in the region of well C. But let us assume that the electron can be found with equal probability in both wells with energy E_1. What eigenfunction do we use to describe the electron? To answer this question, let us put forward the properties that such a wavefunction must have.

1. χ must reflect the fact that the electron can be found with equal probability in both wells. This means that the probability χ^2 must be symmetric with respect to a point halfway between the two wells.
2. The part of χ that reflects the probability that the particle be found in well B with energy E_1 must look like the eigenfunction associated with that energy when the particle is in well B; that is, it must look like Fig. 24-13*b*. And the same applies to the part of the eigenfunction that reflects the probability of finding the particle in well C.

Before we answer the question of eigenfunction selection, let us consider the following. We said that χ_C is an eigenfunction that can represent an electron in well C with energy E_1. Actually $-\chi_C$ could have been used instead. It has the same physical properties as χ_C, and it is a solution to the Schrödinger equation for the same value

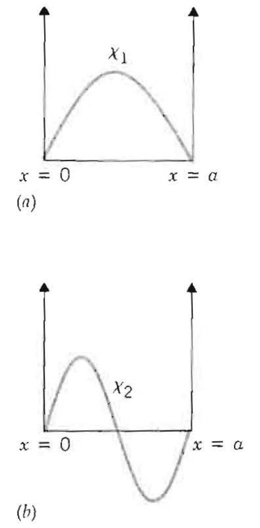


FIGURE 24-11 (*a*) Ground state eigenfunction χ_1 and (*b*) first excited state eigenfunction χ_2 for a particle in an infinite potential well.

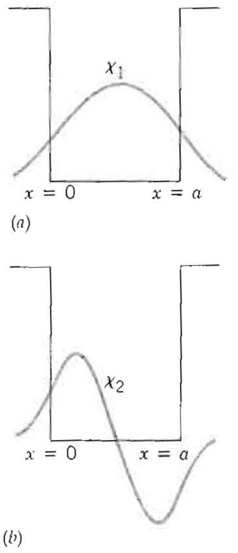


FIGURE 24-12 (*a*) Ground state eigenfunction χ_1 and (*b*) first excited state eigenfunction χ_2 for a particle in a finite potential well.

of the energy E_1,[3] and, moreover, $\chi_C^2 = (-\chi_C)^2$. Therefore, $-\chi$ must be considered equally with $+\chi$.

Let us now answer the question: What eigenfunction do we use to describe an electron with energy E_1 that can be found equally in both wells?

There are two possibilities:

$$\chi_s = a(\chi_B + \chi_C)$$

$$\chi_A = a(\chi_B - \chi_C)$$

The constant a is introduced for normalization purposes; its value is not important to our arguments. It should be clear that χ_s (symmetric) shown in Fig. 24-14*a* satisfies the conditions set forth here. Figure 24-14*b* shows that χ_s^2 is symmetric with respect to the midpoint between the wells, and this reflects the fact that the particle can be found with equal probability in the two wells. The other possibility, χ_A (antisymmetric), Fig. 24-15*a*, also satisfies the conditions set forth, as can be seen by comparing Fig. 24-15*b* for χ_A^2 with Fig. 24-14*b* for χ_s^2.

Let us see what happens to these two eigenfunctions when the separation between the two wells becomes very small. Figure 24-16*a* illustrates the individual eigenfunctions χ_B and χ_C, and Fig. 24-16*b* is the sum $\chi_s = a(\chi_B + \chi_C)$. The eigenfunction χ_s begins to look like the ground state eigenfunction for a well of width $2a$. In fact, in the limit of no separation it becomes the ground state eigenfunction. On the other hand, the antisymmetric eigenfunction appears as in Fig. 24-17. Figure 24-17*a* shows the individual eigenfunctions χ_B and $-\chi_C$, and Fig. 24-17*b* their sum, $\chi_A = a(\chi_B - \chi_C)$. The wavefunction χ_A begins to look like the wavefunction of the first excited state for a well of width $2a$.

We conclude that although when the two wells were far apart both χ_s and χ_A were degenerate eigenfunctions (same energy states), the degeneracy begins to disappear as the two wells get close to each other; χ_A corresponds to a state of higher energy than χ_s.

In our discussion we assumed that the electron can be in either well B or well C. Of course, when the two wells are far apart the question is purely academic. However, when they come close together so that the wavefunctions from the two wells overlap,

---

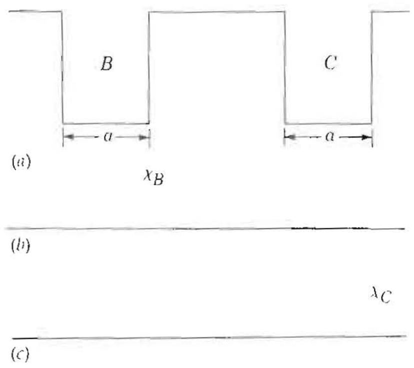
[3] Does $-\chi_C$ satisfy the Schrödinger equation for E_1? Let us substitute it into the Schrödinger equation,

$$-\frac{\hbar^2}{2m}\frac{d^2(-\chi_C)}{dx^2} + E_p(-\chi_C) = E_1(-\chi_C)$$

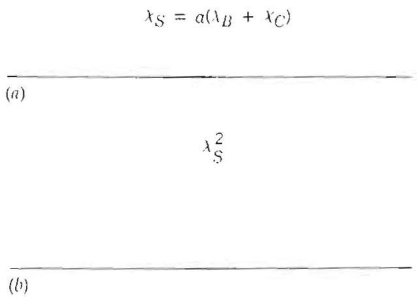
Multiplying both sides of this equation by -1, we get

$$-\frac{\hbar^2}{2m}\frac{d^2\chi_C}{dx^2} + E_p\chi_C = E_1\chi_C$$

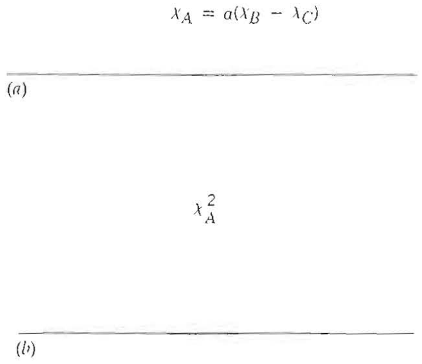
which demonstrates the same equality as does χ_C; so both χ_C and $-\chi_C$ are equivalent solutions to the Schrödinger equations for the same energy E_1.



(*a*) Two finite potential wells *B* and (*C b*) Eigenfunction associated with an electron in well *B* with the ground state energy. (*c*) Eigenfunction associated with an electron in well *C* with the ground state energy.

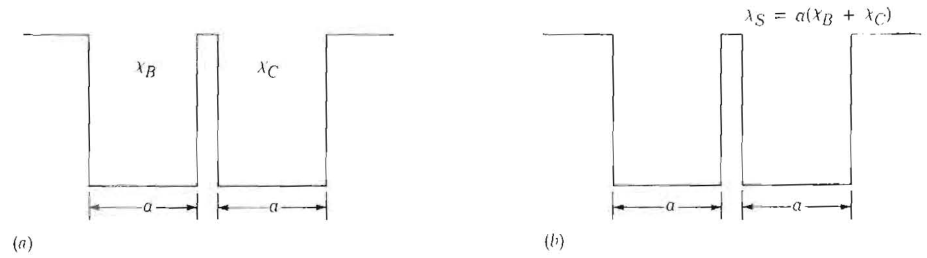


(*a*) Symmetric eigenfunction representing an electron that can be found with equal probability in the two wells of Fig. 24-13 with the ground state energy. (*b*) Probability density associated with the symmetric eigenfunction in (*a*).

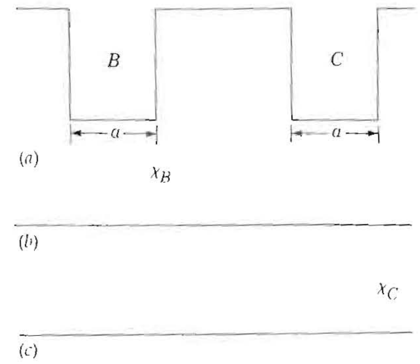


Antisymmetric eigenfunction representing an electron that can be found with equal probability in the two wells of Fig. 24-13 with the ground state energy. (*b*) Probability density associated with the antisymmetric eigenfunction of (*a*).

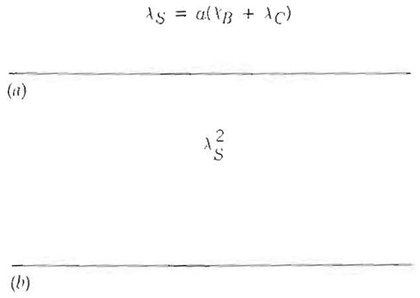
it is possible for the electron that initially was in well *C* to move into well *B*, because



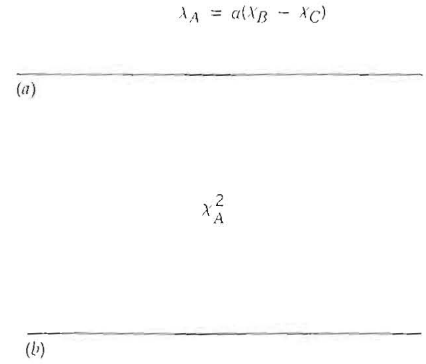
When the two wells are very close together, the symmetric eigenfunction of Fig. 24-14 looks like the ground state eigenfunction for a finite well of width 2*a* (see Fig. 24-12*a*).



(*a*) Two finite potential wells *B* and (*C b*) Eigenfunction associated with an electron in well *B* with the ground state energy. (*c*) Eigenfunction associated with an electron in well *C* with the ground state energy.

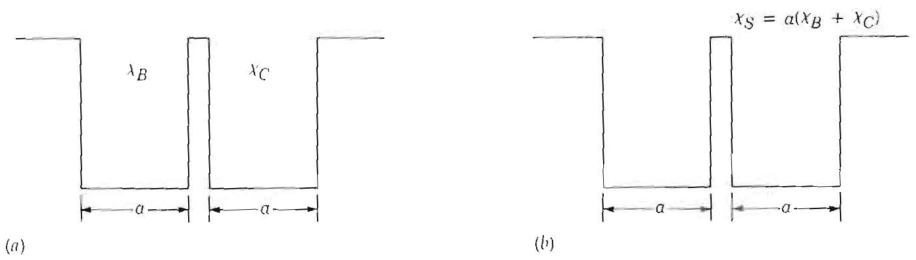


(*a*) Symmetric eigenfunction representing an electron that can be found with equal probability in the two wells of Fig. 24-13 with the ground state energy. (*b*) Probability density associated with the symmetric eigenfunction in (*a*).



Antisymmetric eigenfunction representing an electron that can be found with equal probability in the two wells of Fig. 24-13 with the ground state energy. (*b*) Probability density associated with the antisymmetric eigenfunction of (*a*).

it is possible for the electron that initially was in well *C* to move into well *B*, because

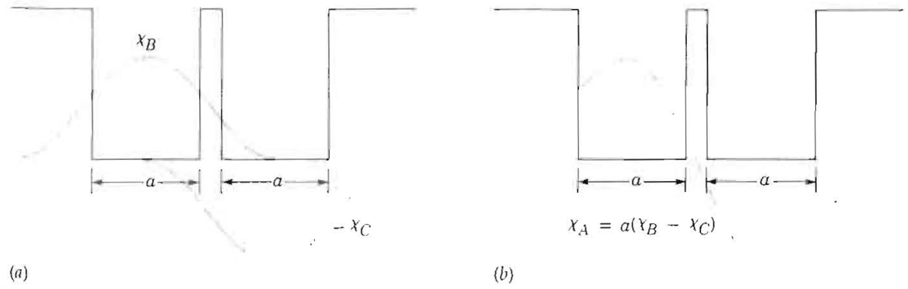


When the two wells are very close together, the symmetric eigenfunction of Fig. 24-14 looks like the ground state eigenfunction for a finite well of width 2*a* (see Fig. 24-12*a*).

its wavefunction is not zero at the location of well B, and the same holds for the electron that initially was in well B; it can move, or *tunnel*, into well C because its wavefunction has a finite value in well C. Thus, even if the electron was initially in one well, after the two wells are brought close together, if we wait long enough, there is no way to predict exactly in which one it will be. Similarly, if we had started with two wells and one electron in each well, after they are brought together, there is no way of saying in which well either of the two electrons will be. The only thing that we can say is that the probability of finding an electron in one well is the same as the probability of finding it in the other. Either χ_A or χ_s will be appropriate to describe its behavior because both χ_A^2 and χ_s^2 are symmetric with respect to the midpoint between the wells.

The important conclusion from this simple artificial example is that *if you start with two identical* χ's *(same energy) in two identical independent systems, when you bring the two together, the two degenerate* χ's *break up into two nondegenerate* χ's. However, this example does not give the reason for this result. What is the physical reason for the breakup of the degeneracy?

Let us look at a real example of two hydrogen atoms, each with its electron in the $1s$ ground state. If one solves the radial part of the Schrödinger equation for the hydrogen atom, Eq. 21.6, one obtains the radial part of the eigenfunction for the $1s$ state as $\chi = e^{-r/r_0}$. In the one-dimensional sketch, the eigenfunction decreases exponentially as the distance r from the nucleus increases. The eigenfunctions χ_B and χ_C in Fig. 24-18 are those associated with the two independent atoms B and C. As the two atoms are brought together, the eigenfunctions overlap, and the electrons from B and C can change places. We are led to consider, just as before, the two possible combinations of χ_B and χ_C shown in Fig. 24-19, where the dashed lines represent how the eigenfunctions of each would appear if the other were not present, and the solid line is the sum or combined eigenfunction. The electron distribution between the two protons can be seen by plotting the probability functions $|\chi_s|^2$ and $|\chi_A|^2$, Fig. 24-20. Both distributions are symmetric with respect to the midpoint between the two protons, and therefore the probability of finding an electron at a certain distance from one proton is the same as the probability of finding it at the same



(*a*) (*b*)

FIG. 24-17 When the wells are very close together, the antisymmetric eigenfunction of Fig. 24-15 looks like the first excited state of a finite well of width $2a$ (see Fig. 24-12*b*).

its wavefunction is not zero at the location of well B, and the same holds for the electron that initially was in well B; it can move, or *tunnel*, into well C because its wavefunction has a finite value in well C. Thus, even if the electron was initially in one well, after the two wells are brought close together, if we wait long enough, there is no way to predict exactly in which one it will be. Similarly, if we had started with two wells and one electron in each well, after they are brought together, there is no way of saying in which well either of the two electrons will be. The only thing that we can say is that the probability of finding an electron in one well is the same as the probability of finding it in the other. Either χ_A or χ_s will be appropriate to describe its behavior because both χ_A^2 and χ_s^2 are symmetric with respect to the midpoint between the wells.

The important conclusion from this simple artificial example is that *if you start with two identical* χ*'s (same energy) in two identical independent systems, when you bring the two together, the two degenerate* χ*'s break up into two nondegenerate* χ*'s*. However, this example does not give the reason for this result. What is the physical reason for the breakup of the degeneracy?

Let us look at a real example of two hydrogen atoms, each with its electron in the $1s$ ground state. If one solves the radial part of the Schrödinger equation for the hydrogen atom, Eq. 21.6, one obtains the radial part of the eigenfunction for the $1s$ state as $\chi = e^{-r/r_0}$. In the one-dimensional sketch, the eigenfunction decreases exponentially as the distance r from the nucleus increases. The eigenfunctions χ_B and χ_C in Fig. 24-18 are those associated with the two independent atoms B and C. As the two atoms are brought together, the eigenfunctions overlap, and the electrons from B and C can change places. We are led to consider, just as before, the two possible combinations of χ_B and χ_C shown in Fig. 24-19, where the dashed lines represent how the eigenfunctions of each would appear if the other were not present, and the solid line is the sum or combined eigenfunction. The electron distribution between the two protons can be seen by plotting the probability functions $|\chi_s|^2$ and $|\chi_A|^2$, Fig. 24-20. Both distributions are symmetric with respect to the midpoint between the two protons, and therefore the probability of finding an electron at a certain distance from one proton is the same as the probability of finding it at the same

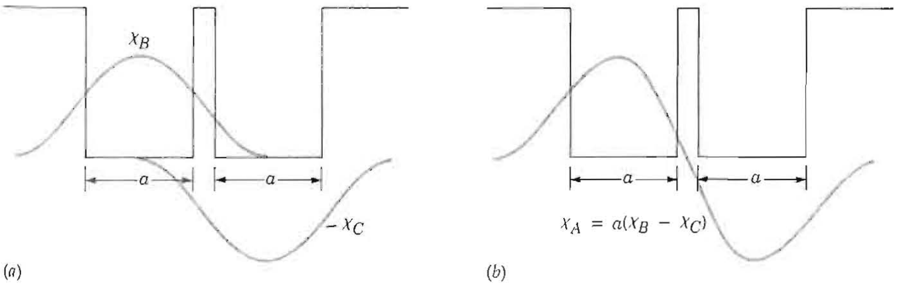


FIGURE 24-17 When the wells are very close together, the antisymmetric eigenfunction of Fig. 24-15 looks like the first excited state of a finite well of width $2a$ (see Fig. 24-12b).

distance from the other proton. Both eigenfunctions are thus suitable to represent the behavior of either electron. However, an electron in a state χ_s has a lower energy than one in χ_A. The reason is that an electron in χ_s is more likely to be between the two protons than being near to just one. Figure 24-20*a* shows that the probability function χ_s^2 in the region between the two protons is greater than on either side of the protons. As a result, the electron spends a considerable amount of time between the two protons. In this region it is under the attractive influence of both protons at once. The binding energy of the electron resulting from the presence of the two protons will be more negative than if it was only under the influence of one of them. On the other hand, an electron in state χ_A spends its time with either one proton or the other. It is hardly ever with both (see Fig. 24-20*b*) and, as a result, this extra contribution to the binding is not there or at least is very small.

Thus, when two atoms are brought together, two separate energy levels are formed from each level of the isolated atom. The physical reason for this effect is the differing ways that the electrons interact with the ions in the symmetric and antisymmetric states.

This is not a band, but two atoms do not make a solid. Suppose that N atoms are brought together to form a solid, then each of the levels of the individual isolated atoms breaks up into N discrete, closely spaced levels and becomes a band of energy levels. This can be illustrated with the 1*s* level of six hydrogen atoms. If we start with six individual 1s states and consider all the possible ways of adding the six individual 1*s* states, we get six types of combinations having different energies (see Fig. 24-21). In the first level, the six individual eigenfunctions all add symmetrically,

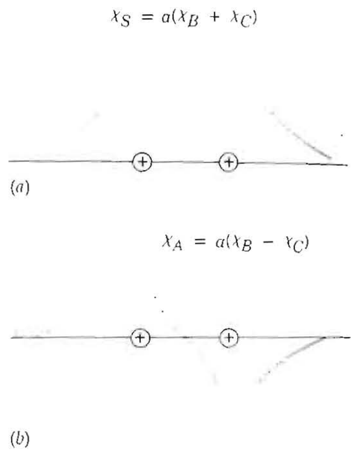
$$\chi_{\text{first level}} = \chi_1 + \chi_2 + \chi_3 + \chi_4 + \chi_5 + \chi_6$$

As a result, there are *five* places along the lattice where the electron will be under the attractive (binding) influence of two nuclei. In the second level, the χ's from the first three atoms add symmetrically among themselves and so do the last three, but the resulting χ from the first three adds antisymmetrically with the resulting χ from the last three:

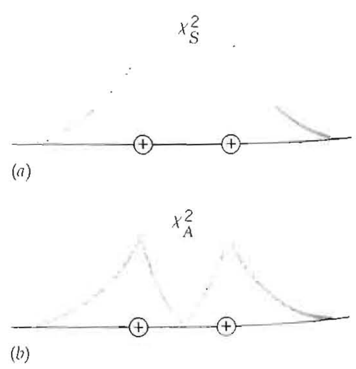
$$\chi_{\text{second level}} = (\chi_1 + \chi_2 + \chi_3) - (\chi_4 + \chi_5 + \chi_6)$$

As a result, there are only *four* places along the lattice where the electron will be under the simultaneous influence of two nuclei that gives the extra negative contribution to the energy. The energy of an electron in this level will be higher than that of an

Radial part of the ground state eigenfunctions of two isolated hydrogen atoms as a function of the distance of the electrons from the respective nuclei.



(*a*) Symmetric combination of the ground state eigenfunctions of the two individual hydrogen atoms of Fig. 24-18. (*b*) Antisymmetric combination of the same two eigenfunctions.



(*a*) Probability density associated with the symmetric eigenfunction of Fig. 24-19*a*. (*b*) Probability density associated with the antisymmetric eigenfunction of Fig. 24-19*b*. Note that χ_s^2 is large in the region between the two nuclei so therefore the electron represented by χ_s spends considerable time between both nuclei. χ_A^2 does not have this feature.

distance from the other proton. Both eigenfunctions are thus suitable to represent the behavior of either electron. However, an electron in a state χ_s has a lower energy than one in χ_A. The reason is that an electron in χ_s is more likely to be between the two protons than being near to just one. Figure 24-20*a* shows that the probability function χ_s^2 in the region between the two protons is greater than on either side of the protons. As a result, the electron spends a considerable amount of time between the two protons. In this region it is under the attractive influence of both protons at once. The binding energy of the electron resulting from the presence of the two protons will be more negative than if it was only under the influence of one of them. On the other hand, an electron in state χ_A spends its time with either one proton or the other. It is hardly ever with both (see Fig. 24-20*b*) and, as a result, this extra contribution to the binding is not there or at least is very small.

Thus, when two atoms are brought together, two separate energy levels are formed from each level of the isolated atom. The physical reason for this effect is the differing ways that the electrons interact with the ions in the symmetric and antisymmetric states.

This is not a band, but two atoms do not make a solid. Suppose that N atoms are brought together to form a solid, then each of the levels of the individual isolated atoms breaks up into N discrete, closely spaced levels and becomes a band of energy levels. This can be illustrated with the 1*s* level of six hydrogen atoms. If we start with six individual 1s states and consider all the possible ways of adding the six individual 1*s* states, we get six types of combinations having different energies (see Fig. 24-21). In the first level, the six individual eigenfunctions all add symmetrically,

$$\chi_{\text{first level}} = \chi_1 + \chi_2 + \chi_3 + \chi_4 + \chi_5 + \chi_6$$

As a result, there are *five* places along the lattice where the electron will be under the attractive (binding) influence of two nuclei. In the second level, the χ's from the first three atoms add symmetrically among themselves and so do the last three, but the resulting χ from the first three adds antisymmetrically with the resulting χ from the last three:

$$\chi_{\text{second level}} = (\chi_1 + \chi_2 + \chi_3) - (\chi_4 + \chi_5 + \chi_6)$$

As a result, there are only *four* places along the lattice where the electron will be under the simultaneous influence of two nuclei that gives the extra negative contribution to the energy. The energy of an electron in this level will be higher than that of an

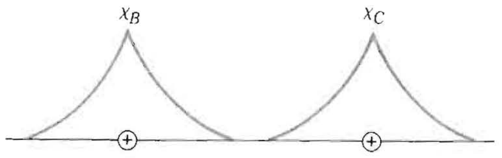


FIGURE 24-18 Radial part of the ground state eigenfunctions of two isolated hydrogen atoms as a function of the distance of the electrons from the respective nuclei.

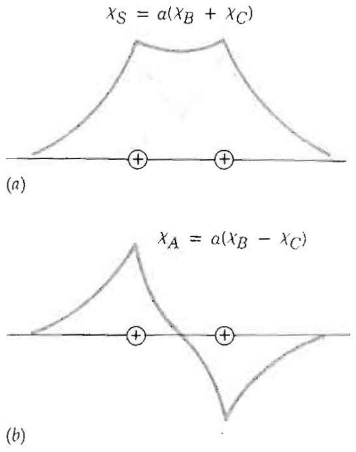


FIGURE 24-19 (*a*) Symmetric combination of the ground state eigenfunctions of the two individual hydrogen atoms of Fig. 24-18. (*b*) Antisymmetric combination of the same two eigenfunctions.

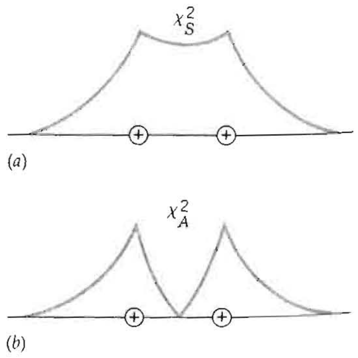


FIGURE 24-20 (*a*) Probability density associated with the symmetric eigenfunction of Fig. 24-19*a*. (*b*) Probability density associated with the antisymmetric eigenfunction of Fig. 24-19*b*. Note that χ_s^2 is large in the region between the two nuclei so therefore the electron represented by χ_s spends considerable time between both nuclei. χ_A^2 does not have this feature.

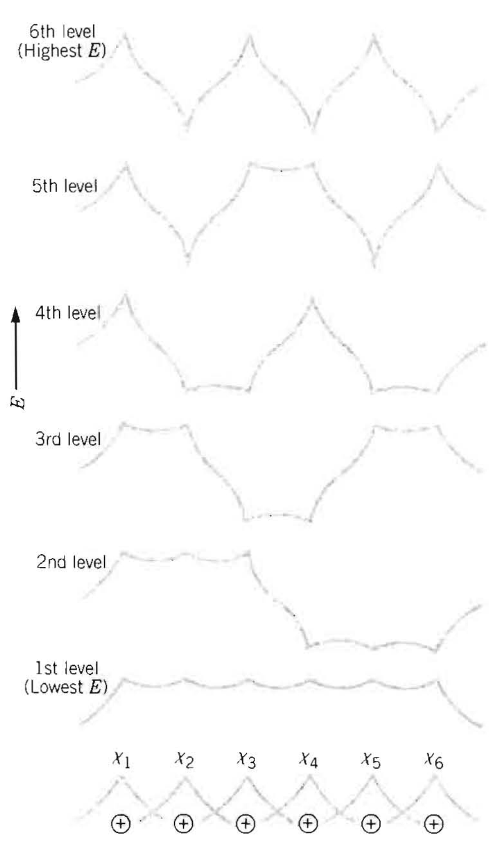


FIGURE 24-21 Six possible combinations of the ground state eigenfunctions of six hydrogen atoms, each corresponding to a different energy state. At the bottom of the figure the six individual eigenfunctions are sketched. The sketches above them represent six possible ways of adding them either symmetrically or antisymmetrically.

electron in the first level. The combination of the individual χ's to form the rest of the levels is done in a similar way as the first two examples. The result concerning the energy of each can also be reasoned in a similar way: In the third level, there are only ***three*** places where the electron would be under the simultaneous influence of two nuclei and, consequently, the energy of the level will be higher than the previous two. This goes on until the sixth level, where this extra contribution to the binding is nowhere in the lattice. Therefore, this state corresponds to the highest energy.

It should be apparent that when we extend this analysis to a lattice of N atoms, the individual identical states of the atoms will give rise to N *different energy states*. However, regardless of the number, the two extremes will look like the first and sixth level of our example. It should also be apparent that the energy difference between the two extremes and, therefore the width of the band, should not depend appreciably on N. Increasing N increases the number of sites where the extra contribution to the energy can take place. But it simultaneously decreases the amount of time that the electron spends in any one site, making the total time in all such sites constant. What affects the width of the band is how close any two atoms are to each other. The closer, the greater the overlap of the wavefunctions and therefore the larger χ^2 will

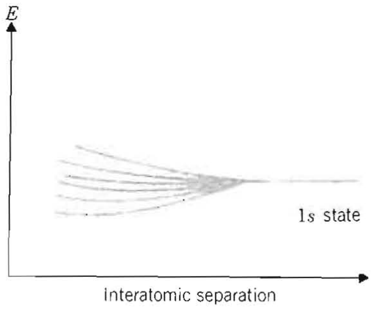
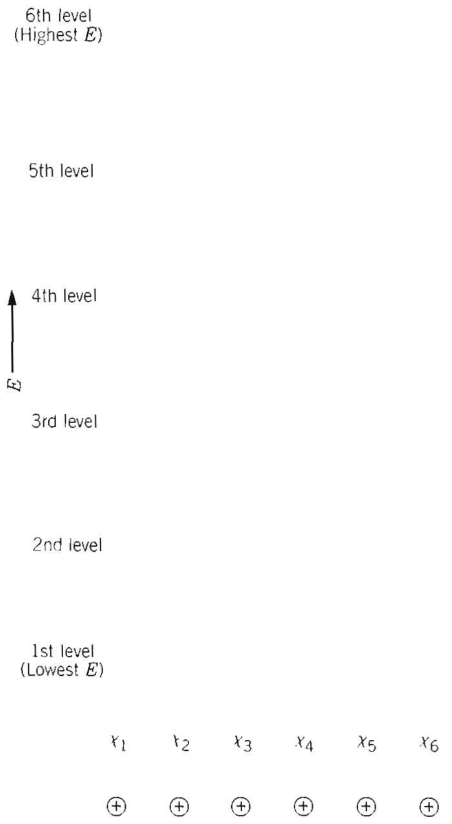


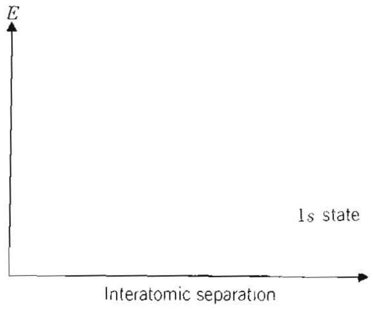
FIGURE 24-22 Splitting of the 1s state of six hydrogen atoms into a band of six energy levels as the separation between the atoms decreases. Note the increase in the bandwidth with decreasing interatomic separation.



Six possible combinations of the ground state eigenfunctions of six hydrogen atoms, each corresponding to a different energy state. At the bottom of the figure the six individual eigenfunctions are sketched. The sketches above them represent six possible ways of adding them either symmetrically or antisymmetrically.

electron in the first level. The combination of the individual χ's to form the rest of the levels is done in a similar way as the first two examples. The result concerning the energy of each can also be reasoned in a similar way: In the third level, there are only *three* places where the electron would be under the simultaneous influence of two nuclei and, consequently, the energy of the level will be higher than the previous two. This goes on until the sixth level, where this extra contribution to the binding is nowhere in the lattice. Therefore, this state corresponds to the highest energy.

It should be apparent that when we extend this analysis to a lattice of N atoms, the individual identical states of the atoms will give rise to N *different energy states*. However, regardless of the number, the two extremes will look like the first and sixth level of our example. It should also be apparent that the energy difference between the two extremes and, therefore the width of the band, should not depend appreciably on N. Increasing N increases the number of sites where the extra contribution to the energy can take place. But it simultaneously decreases the amount of time that the electron spends in any one site, making the total time in all such sites constant. What affects the width of the band is how close any two atoms are to each other. The closer, the greater the overlap of the wavefunctions and therefore the larger χ^2 will



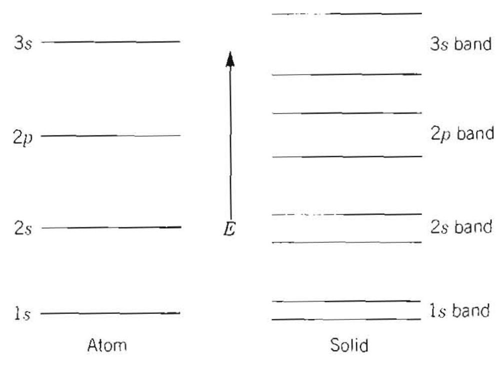
Splitting of the 1s state of six hydrogen atoms into a band of six energy levels as the separation between the atoms decreases. Note the increase in the bandwidth with decreasing interatomic separation.

be between the two nuclei for symmetric states, with the consequence of stronger binding. Figure 24-22 shows how the wavefunction for six different atoms separates into six different energy levels when the atoms are brought close enough together. And the closer they are brought, the greater is the separation of the energy levels. For the usual interatomic separation found in a solid, the width of the band is typically a few electron volts. For a macroscopic solid with $N \sim 10^{23}$, the separation between adjacent levels will be $\sim 10^{-23}$ eV, an insignificant amount.

We thus have shown that when one brings a large number N of atoms together to form a solid, the individual atomic energy levels of the atom break up into a quasicontinuous energy band. Within the band there are N distinct but very close energy levels.

So far, we have limited ourselves to one-electron atoms where the electron is in the $1s$ state. When the analysis is extended to multielectron atoms where we have electrons in other states, one finds that each of the individual atomic states breaks up into similar bands of quasicontinuous states. Thus, if we consider sodium (Na) with an electronic configuration $1s^2\,2s^2\,2p^6\,3s^1$ (Section 21.7b), we may expect the band structure illustrated schematically in Fig. 24-23. Notice that the bandwidth for the low-lying levels is smaller than for the higher energy ones. The reason is that electrons in the lower levels are electrons in the inner subshells of the atoms; these electrons are not influenced much by the presence of other atoms because their wavefunctions do not overlap significantly with those of the electrons of other atoms. Therefore, they give rise to narrower bands.

The situation depicted in Fig. 24-23 is what one may expect in general. The situation for real solids is somewhat more complicated. We must realize that we have been using qualitative arguments and one-dimensional models for the sake of mathematical as well as conceptual simplicity. These models have educational value because they bring out the main features of band theory. However, if we want to get theoretical results that can be compared with those of the experiments, we must face the three-dimensional world, and our qualitative arguments must become quantitative. When this is done, we should not be surprised that the simple pictures may have to be modified somewhat. The main features that we have found—namely,



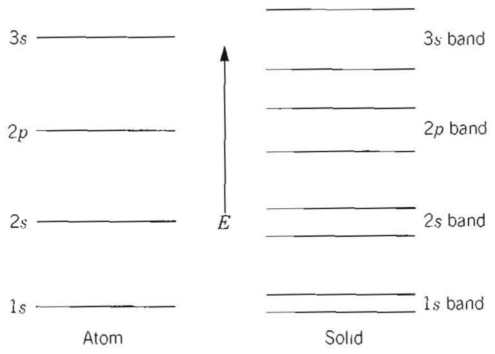
Expected splitting of the first four atomic levels of sodium into four energy bands in a sodium crystal.

be between the two nuclei for symmetric states, with the consequence of stronger binding. Figure 24-22 shows how the wavefunction for six different atoms separates into six different energy levels when the atoms are brought close enough together. And the closer they are brought, the greater is the separation of the energy levels. For the usual interatomic separation found in a solid, the width of the band is typically a few electron volts. For a macroscopic solid with $N \sim 10^{23}$, the separation between adjacent levels will be $\sim 10^{-23}$ eV, an insignificant amount.

We thus have shown that when one brings a large number N of atoms together to form a solid, the individual atomic energy levels of the atom break up into a quasicontinuous energy band. Within the band there are N distinct but very close energy levels.

So far, we have limited ourselves to one-electron atoms where the electron is in the $1s$ state. When the analysis is extended to multielectron atoms where we have electrons in other states, one finds that each of the individual atomic states breaks up into similar bands of quasicontinuous states. Thus, if we consider sodium (Na) with an electronic configuration $1s^2\,2s^2\,2p^6\,3s^1$ (Section 21.7b), we may expect the band structure illustrated schematically in Fig. 24-23. Notice that the bandwidth for the low-lying levels is smaller than for the higher energy ones. The reason is that electrons in the lower levels are electrons in the inner subshells of the atoms; these electrons are not influenced much by the presence of other atoms because their wavefunctions do not overlap significantly with those of the electrons of other atoms. Therefore, they give rise to narrower bands.

The situation depicted in Fig. 24-23 is what one may expect in general. The situation for real solids is somewhat more complicated. We must realize that we have been using qualitative arguments and one-dimensional models for the sake of mathematical as well as conceptual simplicity. These models have educational value because they bring out the main features of band theory. However, if we want to get theoretical results that can be compared with those of the experiments, we must face the three-dimensional world, and our qualitative arguments must become quantitative. When this is done, we should not be surprised that the simple pictures may have to be modified somewhat. The main features that we have found—namely,



Expected splitting of the first four atomic levels of sodium into four energy bands in a sodium crystal.

that each atomic level breaks up into a band and that in each band there are N energy levels—are retained. These two features will enable us to understand the differences between conductors, insulators, and semiconductors.

24.5 CONDUCTORS, INSULATORS, AND SEMICONDUCTORS

We are now in a position to understand why some solids are good conductors, and some are not. We must keep in mind two facts.

As we explained in Section 23.3f, for electrons to experience an acceleration in the presence of an electric field $\mathcal{E}$ and therefore to contribute to the current, they must be able to move into new, slightly higher energy states. This means that the states that are available for the electrons must be both *empty* and *allowed*. For example, if relatively few electrons reside in an otherwise empty band, a large number of unoccupied states are available into which the electrons can move; these electrons can acquire energy from the electric field and contribute to the current. On the other hand, if a band is full, then the electrons in that band cannot contribute to the current because they cannot move into slightly higher energy states. They therefore cannot be accelerated by the electric field.

There is a limit to the number of electrons that can be placed in a given band. We know that there are N different energy states in each band. If the band is an s band (one formed from atomic s states) then, the orbital quantum number $l = 0$ and therefore $m_l = 0$ and $m_s = \pm 1/2$. We can place two electrons in each of the N states without violating Pauli's exclusion principle. In a p band $l = 1$, $m_l = 0, 1, -1$ and for each value of m_l, $m_s = \pm 1/2$. In each of the N energy levels we can put six electrons. In general, because for a given l there are $(2l+1)$ values of m_l and for each m_l there are two values of m_s, we have $2(2l+1)N$ openings available to the electrons in a given band; for example, in a d band the number is $2(2 \times 2 + 1)N = 10N$.

Let us consider some hypothetical examples on the basis of these two facts. Consider a solid of N atoms with each atom having 11 electrons. Altogether there are $11N$ electrons. $2N$ electrons may be put in the $1s$ band, $2N$ in the $2s$ band, $6N$ in the $2p$ band. There remain N electrons that may be placed in the next available band, the $3s$. But the $3s$ band has room for $2N$ electrons, and therefore it will be only half full (see Fig. 24-24). As a consequence, half the states in the $3s$ band (the unoccupied ones) are available to the N electrons in that band. Because there are available states, the N electrons in this band can be accelerated by an electric field and move into higher energy states. *This solid would be an electrical conductor*. Note that

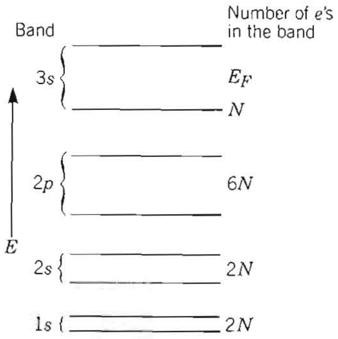


FIGURE 24-24: Schematic of the occupation of the bands by electrons in a sodium crystal of N atoms and having, therefore, $11N$ electrons. The highest energy band with electrons ($3s$ band) is only half full with N electrons, and thus sodium is a monovalent metal. The $3s$ band is the conduction band of sodium.

that each atomic level breaks up into a band and that in each band there are N energy levels—are retained. These two features will enable us to understand the differences between conductors, insulators, and semiconductors.

24.5 CONDUCTORS, INSULATORS, AND SEMICONDUCTORS

We are now in a position to understand why some solids are good conductors, and some are not. We must keep in mind two facts.

> As we explained in Section 23.3f, for electrons to experience an acceleration in the presence of an electric field $\mathcal{E}$ and therefore to contribute to the current, they must be able to move into new, slightly higher energy states. This means that the states that are available for the electrons must be both *empty* and *allowed*. For example, if relatively few electrons reside in an otherwise empty band, a large number of unoccupied states are available into which the electrons can move; these electrons can acquire energy from the electric field and contribute to the current. On the other hand, if a band is full, then the electrons in that band cannot contribute to the current because they cannot move into slightly higher energy states. They therefore cannot be accelerated by the electric field.
>
> There is a limit to the number of electrons that can be placed in a given band. We know that there are N different energy states in each band. If the band is an s band (one formed from atomic s states) then, the orbital quantum number $l = 0$ and therefore $m_l = 0$ and $m_s = \pm 1/2$. We can place two electrons in each of the N states without violating Pauli's exclusion principle. In a p band $l = 1$, $m_l = 0, 1, -1$ and for each value of m_l, $m_s = \pm 1/2$. In each of the N energy levels we can put six electrons. In general, because for a given l there are $(2l + 1)$ values of m_l and for each m_l there are two values of m_s, we have $2(2l + 1)N$ openings available to the electrons in a given band; for example, in a d band the number is $2(2 \times 2 + 1)N = 10N$.

Let us consider some hypothetical examples on the basis of these two facts. Consider a solid of N atoms with each atom having 11 electrons. Altogether there are $11N$ electrons. $2N$ electrons may be put in the $1s$ band, $2N$ in the $2s$ band, $6N$ in the $2p$ band. There remain N electrons that may be placed in the next available band, the $3s$. But the $3s$ band has room for $2N$ electrons, and therefore it will be only half full (see Fig. 24-24). As a consequence, half the states in the $3s$ band (the unoccupied ones) are available to the N electrons in that band. Because there are available states, the N electrons in this band can be accelerated by an electric field and move into higher energy states. *This solid would be an electrical conductor*. Note that

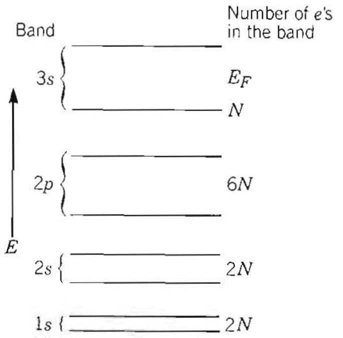


FIGURE 24-24 Schematic of the occupation of the bands by electrons in a sodium crystal of N atoms and having, therefore, $11N$ electrons. The highest energy band with electrons ($3s$ band) is only half full with N electrons, and thus sodium is a monovalent metal. The $3s$ band is the conduction band of sodium.

because the N electrons in the 3*s* band obey the Pauli exclusion principle, they obey Fermi-Dirac statistics; that is, at $T = 0$ K they all reside in the lowest energy levels of the band and the highest occupied level becomes the Fermi level. For $T > 0$, a few (as we saw in the previous chapter) can be above the Fermi level. The N electrons in the 3*s* band behave as predicted by the QMFE model with one minor modification that we will discuss later. We can see that the effect of the periodic potential on the motion of electrons in a solid with a partially filled band is unimportant, and that is the reason why the QMFE model was so successful in predicting the properties of conductors. This hypothetical solid is of course Na ($Z = 11$). The only difference between the hypothetical solid and the real Na is that in sodium metal the next band, the 3*p*, overlaps the 3*s* band. But all that this does is to provide additional empty energy levels for the N electrons in the 3*s* band.

The same arguments used for Na apply to lithium ($1s^2\,2s^1$), potassium ($1s^2\,2s^2\,2p^6\,3s^2\,3p^6\,4s^1$), as well as to rubidium and cesium. (See the periodic table in Chapter 21.)

It should be noted that the highest energy band containing electrons is called the *valence band*. If, as in the case of sodium, this band is only partially filled, it is also called the *conduction band* because electrons in that band are responsible for conduction processes.

Next consider magnesium ($Z = 12$) with electronic configuration $1s^2\,2s^2\,2p^6\,3s^2$. In a solid with N atoms, there are $12N$ electrons. Following the previous scheme, $2N$ electrons go into the 1*s* band, $2N$ into the 2*s*, $6N$ into the 2*p*, and the remaining $2N$ into the 3*s*. All $12N$ electrons have been accounted for in the process. They have completely filled the 3*s* band. According to our previous argument, there are no empty energy states available for the electrons in the 3*s* band to move into; therefore, they cannot contribute to conduction. With no empty states through which charged particles may contribute to conduction, Mg should be an insulator. But it is not. The reason is that as in the case of Na, the 3*p* and 3*s* bands overlap (see Fig. 24-25). Because the 3*p* band has $6N$ empty states, the $2N$ electrons have available to them $2N + 6N$ states of which only $2N$ are occupied. Mg is therefore a conductor. Similar arguments apply to beryllium, calcium, zinc, and barium, all of which are in the same group in the periodic table.

Now consider carbon ($1s^2\,2s^2\,2p2$) in its diamond structure. As N atoms of C are brought together, they have $6N$ electrons. $2N$ fill the 1*s* band, $2N$ fill the 2*s* band, and there are $2N$ electrons left to place in the next available band, the 2*p*, which has room for $6N$ electrons. The 2*p* band would be a partially filled band with plenty of empty states available (see Fig. 24-26). Diamond should therefore be a conductor. But it is not. It is an excellent insulator. As we mentioned before, the qualitative arguments elucidate the main features of band structure. However, when dealing with a specific crystalline material, the arguments must become quantitative. When

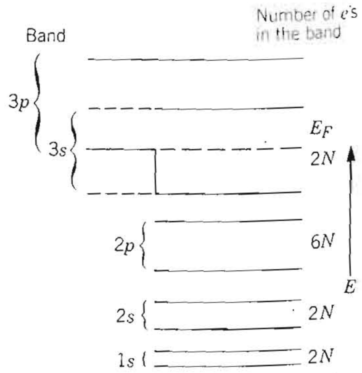


FIGURE 24-25 Occupation of the energy bands in a magnesium crystal of N atoms ($12N$ electrons). The overlap between the 3*s* and the 3*p* bands make magnesium a metal.

because the N electrons in the $3s$ band obey the Pauli exclusion principle, they obey Fermi-Dirac statistics; that is, at $T = 0$ K they all reside in the lowest energy levels of the band and the highest occupied level becomes the Fermi level. For $T > 0$, a few (as we saw in the previous chapter) can be above the Fermi level. The N electrons in the $3s$ band behave as predicted by the QMFE model with one minor modification that we will discuss later. We can see that the effect of the periodic potential on the motion of electrons in a solid with a partially filled band is unimportant, and that is the reason why the QMFE model was so successful in predicting the properties of conductors. This hypothetical solid is of course Na ($Z = 11$). The only difference between the hypothetical solid and the real Na is that in sodium metal the next band, the $3p$, overlaps the $3s$ band. But all that this does is to provide additional empty energy levels for the N electrons in the $3s$ band.

The same arguments used for Na apply to lithium ($1s^2\,2s^1$), potassium ($1s^2\,2s^2\,2p^6\,3s^2\,3p^6\,4s^1$), as well as to rubidium and cesium. (See the periodic table in Chapter 21.)

It should be noted that the highest energy band containing electrons is called the *valence band*. If, as in the case of sodium, this band is only partially filled, it is also called the *conduction band* because electrons in that band are responsible for conduction processes.

Next consider magnesium ($Z = 12$) with electronic configuration $1s^2\,2s^2\,2p^6\,3s^2$. In a solid with N atoms, there are $12N$ electrons. Following the previous scheme, $2N$ electrons go into the $1s$ band, $2N$ into the $2s$, $6N$ into the $2p$, and the remaining $2N$ into the $3s$. All $12N$ electrons have been accounted for in the process. They have completely filled the $3s$ band. According to our previous argument, there are no empty energy states available for the electrons in the $3s$ band to move into; therefore, they cannot contribute to conduction. With no empty states through which charged particles may contribute to conduction, Mg should be an insulator. But it is not. The reason is that as in the case of Na, the $3p$ and $3s$ bands overlap (see Fig. 24-25). Because the $3p$ band has $6N$ empty states, the $2N$ electrons have available to them $2N + 6N$ states of which only $2N$ are occupied. Mg is therefore a conductor. Similar arguments apply to beryllium, calcium, zinc, and barium, all of which are in the same group in the periodic table.

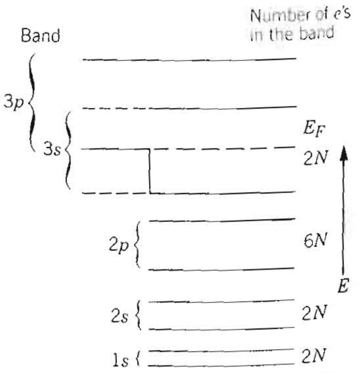


FIGURE 24-25 Occupation of the energy bands in a magnesium crystal of N atoms ($12N$ electrons). The overlap between the $3s$ and the $3p$ bands make magnesium a metal.

Now consider carbon ($1s^2\,2s^2\,2p2$) in its diamond structure. As N atoms of C are brought together, they have $6N$ electrons. $2N$ fill the $1s$ band, $2N$ fill the $2s$ band, and there are $2N$ electrons left to place in the next available band, the $2p$, which has room for $6N$ electrons. The $2p$ band would be a partially filled band with plenty of empty states available (see Fig. 24-26). Diamond should therefore be a conductor. But it is not. It is an excellent insulator. As we mentioned before, the qualitative arguments elucidate the main features of band structure. However, when dealing with a specific crystalline material, the arguments must become quantitative. When

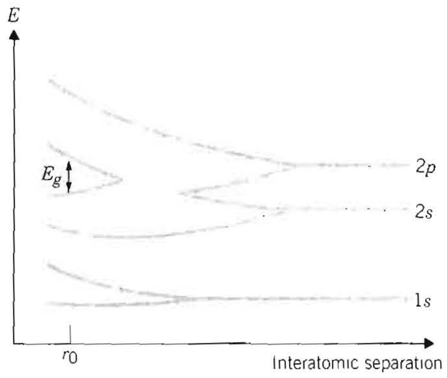


FIGURE 24-27 The splitting of the atomic energy levels of carbon into energy bands is followed by the merging of the 2*s* and 2*p* bands and a subsequent splitting of these bands as the interatomic spacing decreases. At the equilibrium inter-atomic spacing r_0, an energy gap E_g separates two hybrid '2*s* 2*p*' energy bands in a diamond crystal.

this is done, interesting features occur, such as band overlap. In the case of diamond, germanium, and silicon, an even more interesting feature is revealed.

As the carbon atoms are brought together to form diamond, the energy levels begin to split into bands starting with the outermost shell, $n = 2$ (2*s* and 2*p* levels) (Fig. 24-27). As the interatomic spacing decreases farther, the 2*s* and 2*p* bands begin to overlap and merge into a single '2*s* 2*p*' band with 8*N* states available. As the separation decreases even farther, approaching the interatomic equilibrium spacing r_0, the '2*s* 2*p*' band splits again into two hybrid bands separated by an energy gap E_g, which increases with decreasing separation. The value of E_g is about 6 eV for the equilibrium distance of $r_0 \approx 1.5 \times 10^{-10}$ m. However, each of these two bands now contains 4*N* states. The result: Of the total 6*N* electrons, 2*N* go into the 1*s* band and the remaining 4*N* into the lower hybrid '2*s* 2*p*' band and fill it. Thus, at $T = 0$ K the valence band (the lower '2*s* 2*p*' band) is full (Fig. 24-28), and diamond is an insulator. Note that this is only true for the diamond structure of carbon, not for graphite. The bands of germanium (Ge) and silicon (Si) show a similar behavior. In the case of Si, the mixing and subsequent splitting occurs between the 3*s* and 3*p*, whereas in the case of Ge it occurs between the 4*s* and the 4*p*. There is, however, an important *quantitative* difference between diamond and Si and Ge. The energy gap E_g between the filled valence band and the next empty band for Ge and Si is much smaller than for C: E_g (Ge) = 0.7 eV, E_g (Si) = 1.1 eV. At $T = 0$ K, pure C, Si, and Ge behave identically. They are perfect insulators because the valence band is filled. However, as T increases, some of the electrons in the valence band can be thermally excited across the energy gap into the next band, which now becomes the *conduction band*, and as a result electrical conduction can take place. How many electrons can be excited depends on how big E_g is and, of course, on T. The higher T, the greater the thermal

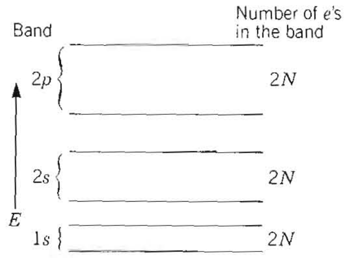
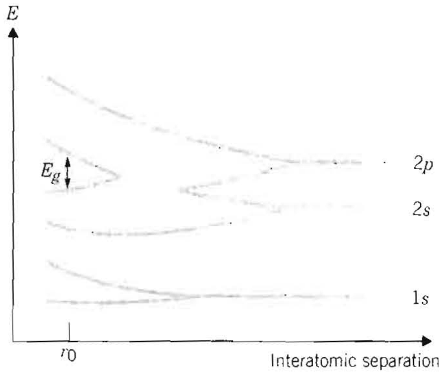


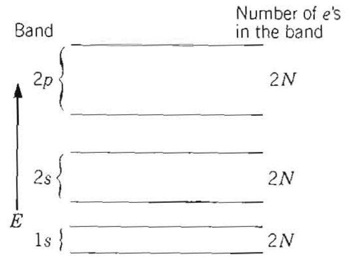
FIGURE 24-26 Expected occupation of the energy bands in a diamond crystal of *N* atoms.



The splitting of the atomic energy levels of carbon into energy bands is followed by the merging of the 2s and 2p bands and a subsequent splitting of these bands as the interatomic spacing decreases. At the equilibrium inter-atomic spacing r_0, an energy gap E_g separates two hybrid '2s 2p' energy bands in a diamond crystal.

this is done, interesting features occur, such as band overlap. In the case of diamond, germanium, and silicon, an even more interesting feature is revealed.

As the carbon atoms are brought together to form diamond, the energy levels begin to split into bands starting with the outermost shell, $n = 2$ (2s and 2p levels) (Fig. 24-27). As the interatomic spacing decreases farther, the 2s and 2p bands begin to overlap and merge into a single '2s 2p' band with 8N states available. As the separation decreases even farther, approaching the interatomic equilibrium spacing r_0, the '2s 2p' band splits again into two hybrid bands separated by an energy gap E_g, which increases with decreasing separation. The value of E_g is about 6 eV for the equilibrium distance of $r_0 \approx 1.5 \times 10^{-10}$ m. However, each of these two bands now contains 4N states. The result: Of the total 6N electrons, 2N go into the 1s band and the remaining 4N into the lower hybrid '2s 2p' band and fill it. Thus, at $T = 0$ K the valence band (the lower '2s 2p' band) is full (Fig. 24-28), and diamond is an insulator. Note that this is only true for the diamond structure of carbon, not for graphite. The bands of germanium (Ge) and silicon (Si) show a similar behavior. In the case of Si, the mixing and subsequent splitting occurs between the 3s and 3p, whereas in the case of Ge it occurs between the 4s and the 4p. There is, however, an important *quantitative* difference between diamond and Si and Ge. The energy gap E_g between the filled valence band and the next empty band for Ge and Si is much smaller than for C: E_g (Ge) = 0.7 eV, E_g (Si) = 1.1 eV. At $T = 0$ K, pure C, Si, and Ge behave identically. They are perfect insulators because the valence band is filled. However, as T increases, some of the electrons in the valence band can be thermally excited across the energy gap into the next band, which now becomes the *conduction band*, and as a result electrical conduction can take place. How many electrons can be excited depends on how big E_g is and, of course, on T. The higher T, the greater the thermal



Expected occupation of the energy bands in a diamond crystal of N atoms.

[redacted]

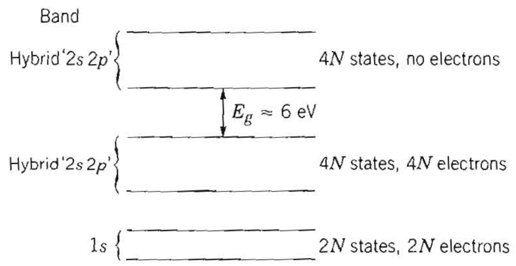


FIGURE 24-28 Actual occupation of the energy bands in a diamond crystal with N atoms ($6N$ electrons). The lower hybrid '$2s\,2p$' band is full and separated by an energy gap E_g from the higher hybrid '$2s\,2p$' band, which has no electrons.

energy, and therefore the greater the number of electrons that will be able to make the jump across the energy gap and, naturally, the greater the electrical conductivity. This is the reason why the conductivity of an insulator and of a semiconductor increases with T (as opposed to a metallic conductor). We will discuss this further in the next chapter.

The number of electrons in the conduction band at a given temperature will depend on E_g, the magnitude of the energy gap between the valence and the conduction bands. The smaller E_g, the greater the probability that the electrons at the top of the valence band will jump into the conduction band. Thus, at a given temperature, we expect that Si ($E_g \approx 1$ eV) will have more electrons in the conduction band and therefore, be a better conductor than diamond ($E_g = 6$ eV). In fact, as we will show in Chapter 25, the probability of transition across the energy gap is very sensitive to the magnitude of E_g. A doubling of E_g will reduce the number of conduction electrons by several orders of magnitude (powers of 10). It is the magnitude of E_g that determines whether a solid is an insulator (diamond) or a semiconductor (Si, Ge) at ambient temperatures. Detailed calculations of the number of conduction electrons and its dependence on E_g and T are presented in Chapter 25.

24.6 EFFECTIVE MASS

When an electric field $\mathcal{E}$ acts on a *free* electron, it exerts a force $e\mathcal{E}$ that, from Newton's law, will produce an acceleration inversely proportional to its mass, $a = e\mathcal{E}/m$. What happens when the electron to be accelerated is not free but happens to be in a crystal under the influence of the potential of the lattice ions? The answer is that it will still accelerate according to Newton's law; however, the electron responds as if it had some *effective mass*, which is different from its true mass. As we will show, this is because $\mathcal{E}$ is not the only electric field acting on the electron inside the crystal.

We will introduce this concept by using a semiclassical picture: an argument that is half classical and half quantum mechanical. The quantum mechanical part lies in the fact that the motion of an electron is governed by a wave, and that the velocity

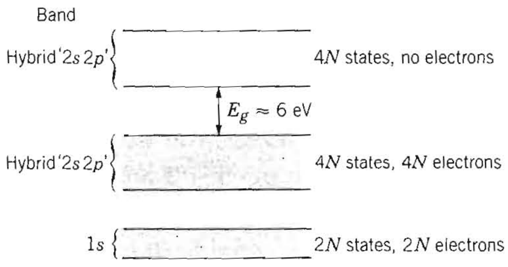


FIGURE 24-28 Actual occupation of the energy bands in a diamond crystal with N atoms ($6N$ electrons). The lower hybrid '$2s\,2p'$ band is full and separated by an energy gap E_g from the higher hybrid '$2s\,2p'$ band, which has no electrons.

energy, and therefore the greater the number of electrons that will be able to make the jump across the energy gap and, naturally, the greater the electrical conductivity. This is the reason why the conductivity of an insulator and of a semiconductor increases with T (as opposed to a metallic conductor). We will discuss this further in the next chapter.

The number of electrons in the conduction band at a given temperature will depend on E_g, the magnitude of the energy gap between the valence and the conduction bands. The smaller E_g, the greater the probability that the electrons at the top of the valence band will jump into the conduction band. Thus, at a given temperature, we expect that Si ($E_g \approx 1$ eV) will have more electrons in the conduction band and therefore, be a better conductor than diamond ($E_g = 6$ eV). In fact, as we will show in Chapter 25, the probability of transition across the energy gap is very sensitive to the magnitude of E_g. A doubling of E_g will reduce the number of conduction electrons by several orders of magnitude (powers of 10). It is the magnitude of E_g that determines whether a solid is an insulator (diamond) or a semiconductor (Si, Ge) at ambient temperatures. Detailed calculations of the number of conduction electrons and its dependence on E_g and T are presented in Chapter 25.

24.6 EFFECTIVE MASS

When an electric field $\mathcal{E}$ acts on a *free* electron, it exerts a force $e\mathcal{E}$ that, from Newton's law, will produce an acceleration inversely proportional to its mass, $a = e\mathcal{E}/m$. What happens when the electron to be accelerated is not free but happens to be in a crystal under the influence of the potential of the lattice ions? The answer is that it will still accelerate according to Newton's law; however, the electron responds as if it had some *effective mass*, which is different from its true mass. As we will show, this is because $\mathcal{E}$ is not the only electric field acting on the electron inside the crystal.

We will introduce this concept by using a semiclassical picture: an argument that is half classical and half quantum mechanical. The quantum mechanical part lies in the fact that the motion of an electron is governed by a wave, and that the velocity of the electron is equal to the group velocity v_{group} of the wave, that is, the velocity of the envelope, of the wave packet. In our treatment of matter waves and wave packets (Section 19.7), we saw that the group velocity is given by Eq. 19.10:

$$v_{\text{group}} = \frac{dE}{dp} \tag{19.10}$$

where E is the energy of the particle and p is its momentum.

In the case of a free particle, we can readily show that the group velocity is equal to the particle velocity. For a free particle, the energy

$$E + \frac{1}{2} m v_{\text{particle}}^2 = \frac{p^2}{2m}$$

Therefore

$$v_{\text{group}} = \frac{dE}{dp} = \frac{d}{dp}\left(\frac{p^2}{2m}\right) = \frac{p}{m} = \frac{m v_{\text{particle}}}{m} = v_{\text{particle}}$$

Although we have shown that group $v_{\text{group}} = v_{\text{particle}}$ for the free particle case only, it can be shown that the relation holds even when the particle is not free, such as the case of an electron in a lattice.

Equation 19.10 defines the group velocity in terms of the energy E and the momentum p of the particle. As we have seen in the Kronig-Penney model, the energy is often expressed in terms of the wave vector k. It is convenient, therefore, to define the group velocity in terms of E and k. This can be done by using de Broglie's relation $p = h/\lambda$ (Eq. 19.1) and the fact that $\lambda = 2\pi/k$ (Eq. 11.12). Combining Eqs. 19.1 and 11.12 we have $p = \hbar k$ and hence $dp = \hbar dk$. Substituting this result for dp in Eq. 19.10, we obtain

$$v_{\text{group}} = \frac{1}{\hbar}\frac{dE}{dk} \tag{24.5}$$

The classical part of the argument uses the definition from mechanics that if a force does work dW on a particle, the energy of that particle increases by the same amount: $dE = dW$. Applying this to the present case, we have

$$dE = dW = e\mathcal{E}\,dx = e\mathcal{E}\frac{dx}{dt}\,dt = e\mathcal{E}\,v_g\,dt$$

The rate at which the energy of the particle is changing is therefore

$$\frac{dE}{dt} = e\mathcal{E}\,v_g \tag{24.6}$$

We also know from Newton's law that when a force acts on a particle, it will be accelerated. By definition, the acceleration a is

$$a = \frac{dv_{\text{particle}}}{dt} = \frac{dv_g}{dt}$$

of the electron is equal to the group velocity v_{group} of the wave, that is, the velocity of the envelope, of the wave packet. In our treatment of matter waves and wave packets (Section 19.7), we saw that the group velocity is given by Eq. 19.10:

$$v_{\text{group}} = \frac{dE}{dp} \tag{19.10}$$

where E is the energy of the particle and p is its momentum.

In the case of a free particle, we can readily show that the group velocity is equal to the particle velocity. For a free particle, the energy

$$E + \frac{1}{2}mv^2_{\text{particle}} = \frac{p^2}{2m}$$

Therefore

$$v_{\text{group}} = \frac{dE}{dp} = \frac{d}{dp}\left(\frac{p^2}{2m}\right) = \frac{p}{m} = \frac{mv_{\text{particle}}}{m} = v_{\text{particle}}$$

Although we have shown that group $v_{\text{group}} = v_{\text{particle}}$ for the free particle case only, it can be shown that the relation holds even when the particle is not free, such as the case of an electron in a lattice.

Equation 19.10 defines the group velocity in terms of the energy E and the momentum p of the particle. As we have seen in the Kronig-Penney model, the energy is often expressed in terms of the wave vector k. It is convenient, therefore, to define the group velocity in terms of E and k. This can be done by using de Broglie's relation $p = h/\lambda$ (Eq. 19.1) and the fact that $\lambda = 2\pi/k$ (Eq. 11.12). Combining Eqs. 19.1 and 11.12 we have $p = \hbar k$ and hence $dp = \hbar dk$. Substituting this result for dp in Eq. 19.10, we obtain

$$v_{\text{group}} = \frac{1}{\hbar}\frac{dE}{dk} \tag{24.5}$$

The classical part of the argument uses the definition from mechanics that if a force does work dW on a particle, the energy of that particle increases by the same amount: $dE = dW$. Applying this to the present case, we have

$$dE = dW = e\mathcal{E}\,dx = e\mathcal{E}\frac{dx}{dt}\,dt = e\mathcal{E}\,v_g\,dt$$

The rate at which the energy of the particle is changing is therefore

$$\frac{dE}{dt} = e\mathcal{E}\,v_g \tag{24.6}$$

We also know from Newton's law that when a force acts on a particle, it will be accelerated. By definition, the acceleration a is

$$a = \frac{dv_{\text{particle}}}{dt} = \frac{dv_g}{dt}$$

Substituting Eq. 24.5 for v_g we obtain

$$a = \frac{1}{\hbar}\frac{d}{dt}\frac{dE}{dk}$$

Interchanging the order of the differentiation of E, we may write

$$a = \frac{1}{\hbar}\frac{d}{dk}\frac{dE}{dt} \tag{24.7}$$

Substituting Eq. 24.6 for dE/dt in Eq. 24.7 yields

$$a = \frac{e\mathcal{E}}{\hbar}\frac{dv_g}{dk}$$

From Eq. 24.5, it follows that

$$a = \frac{1}{\hbar^2}\frac{d^2E}{dk^2}e\mathcal{E}$$

Rearranging terms, we get

$$e\mathcal{E} = \frac{\hbar^2}{\dfrac{d^2E}{dk^2}}\,a \tag{24.8}$$

Noting that $e\mathcal{E}$ is the force of the externally applied electric field, we conclude that Eq. 24.8 has the form $F = m^*a$, where

$$m^* = \frac{\hbar^2}{\dfrac{d^2E}{dk^2}} \tag{24.9}$$

The response of the electron in the solid to an externally applied electric field is as if it had an effective mass m^* given by the expression in Eq. 24.9. Let us see if Eq. 24.9 gives the correct result for the free electron case. From Eq. 24.4, for the free electron

$$E = \frac{\hbar^2k^2}{2m} \tag{24.4}$$

and

$$\frac{dE}{dk} = \frac{\hbar^2k}{m}; \quad \frac{d^2E}{dk^2} = \frac{\hbar^2}{m}$$

Substitute this result into Eq. 24.9 and obtain

$$m^* = \frac{\hbar^2}{\dfrac{\hbar^2}{m}} = m$$

When the electron is free, the effective mass is the true mass, as it should be. However, when the electron is in a crystal, m^* is different from m because the energy is not proportional to k^2, as we saw in the Kronig-Penney model.

Substituting Eq. 24.5 for v_g we obtain

$$a = \frac{1}{\hbar}\frac{d}{dt}\frac{dE}{dk}$$

Interchanging the order of the differentiation of E, we may write

$$a = \frac{1}{\hbar}\frac{d}{dk}\frac{dE}{dt} \tag{24.7}$$

Substituting Eq. 24.6 for dE/dt in Eq. 24.7 yields

$$a = \frac{e\mathcal{E}}{\hbar}\frac{dv_g}{dk}$$

From Eq. 24.5, it follows that

$$a = \frac{1}{\hbar^2}\frac{d^2E}{dk^2}e\mathcal{E}$$

Rearranging terms, we get

$$e\mathcal{E} = \frac{\hbar^2}{\frac{d^2E}{dk^2}}\,a \tag{24.8}$$

Noting that $e\mathcal{E}$ is the force of the externally applied electric field, we conclude that Eq. 24.8 has the form $F = m^*a$, where

$$m^* = \frac{\hbar^2}{\frac{d^2E}{dk^2}} \tag{24.9}$$

The response of the electron in the solid to an externally applied electric field is as if it had an effective mass m^* given by the expression in Eq. 24.9. Let us see if Eq. 24.9 gives the correct result for the free electron case. From Eq. 24.4, for the free electron

$$E = \frac{\hbar^2k^2}{2m} \tag{24.4}$$

and

$$\frac{dE}{dk} = \frac{\hbar^2k}{m}; \quad \frac{d^2E}{dk^2} = \frac{\hbar^2}{m}$$

Substitute this result into Eq. 24.9 and obtain

$$m^* = \frac{\hbar^2}{\frac{\hbar^2}{m}} = m$$

When the electron is free, the effective mass is the true mass, as it should be. However, when the electron is in a crystal, m^* is different from m because the energy is not proportional to k^2, as we saw in the Kronig-Penney model.

The physical reason is the following. The electron in the crystal moves under the influence of *internal forces* exerted by the electric fields of the ions of the lattice and the *external force* resulting from the externally applied electric field. If we choose to use Eq. 24.8 to describe the motion of the electrons, we describe the motion in terms of the *external force alone*. However, the effect of the internal forces is hidden in m^*.

Let us recall the results that we obtained from the Kronig-Penney model concerning the relation between E and k (Fig. 24-10). The relation for the first allowed band is represented by the solid line in Fig. 24-30*a*. For small k's both curves, the free electron one (the dashed line) and the one obtained from the model, are quite similar. Both the first and the second derivatives are almost the same; therefore, m is about equal to m^*. Notice, however, that in the Kronig-Penney curve the first derivative (the slope) first increases for small values of k and then decreases as k approaches π/d. We illustrate a qualitative plot of the slope dE/dk versus k in Fig. 24-30*b*. We may also draw a qualitative illustration of d^2E/dk^2 from Fig. 24.30*b*. Its appearance is that of Fig. 24-30*c*. The reciprocal of d^2E/dk^2 is proportional to the effective mass, from Eq. 24.9. These reciprocal curves are shown in Fig. 24-30d. We see that for small values of k the effective mass m^* is essentially equal to the mass of a free electron m. As dE/dk approaches the maximum, d^2E/dk^2 begins to decrease and m^* increases. When dE/dk reaches the maximum, $d^2E/dk^2 = 0$ and m^* becomes infinite. Subsequently, d^2E/dk^2 becomes negative and m^* is negative.

We may draw the following conclusions about the effective mass m^* of an electron moving in a periodic lattice (see Fig. 24-30d).

1. m^* is not always equal to m.
2. m^* can be greater than m and, in fact, infinite.
3. m^* can be less than m or even negative.

We can gain some understanding of the behavior of m^* if we consider the way an electron wave is reflected by the lattice ions. We know from Section 12.7 that a wave is totally Bragg reflected if

$$2d \sin\theta = n\lambda \tag{12.11}$$

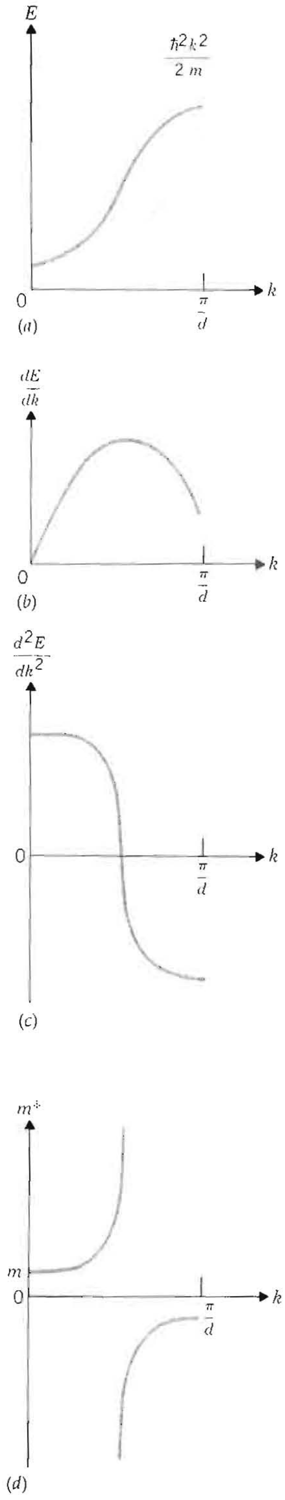


FIGURE 24-29 (*a*) Dependence of the energy E of an electron on its wave vector k, and hence its momentum p (solid line) for the first allowed energy band of the Kronig-Penney model (see Fig. 24-10). The dashed line illustrates the relation between E and k for the free electron case. (*b*) Qualitative plot of the derivative of E with respect to k as a function of k for the solid line of (*a*). (*c*) Qualitative plot of the second derivative of E with respect to k as a function of k. (d) Dependence of the effective mass m^* of the electron on the wave vector k for the situation depicted by the solid line of Figure (*a*). Note that the effective mass is negative near the top of the band, that is, as k approaches π/d.

The physical reason is the following. The electron in the crystal moves under the influence of *internal forces* exerted by the electric fields of the ions of the lattice and the *external force* resulting from the externally applied electric field. If we choose to use Eq. 24.8 to describe the motion of the electrons, we describe the motion in terms of the *external force alone*. However, the effect of the internal forces is hidden in m^*.

Let us recall the results that we obtained from the Kronig-Penney model concerning the relation between E and k (Fig. 24-10). The relation for the first allowed band is represented by the solid line in Fig. 24-30*a*. For small k's both curves, the free electron one (the dashed line) and the one obtained from the model, are quite similar. Both the first and the second derivatives are almost the same; therefore, m is about equal to m^*. Notice, however, that in the Kronig-Penney curve the first derivative (the slope) first increases for small values of k and then decreases as k approaches π/d. We illustrate a qualitative plot of the slope dE/dk versus k in Fig. 24-30*b*. We may also draw a qualitative illustration of d^2E/dk^2 from Fig. 24.30*b*. Its appearance is that of Fig. 24-30*c*. The reciprocal of d^2E/dk^2 is proportional to the effective mass, from Eq. 24.9. These reciprocal curves are shown in Fig. 24-30d. We see that for small values of k the effective mass m^* is essentially equal to the mass of a free electron m. As dE/dk approaches the maximum, d^2E/dk^2 begins to decrease and m^* increases. When dE/dk reaches the maximum, $d^2E/dk^2 = 0$ and m^* becomes infinite. Subsequently, d^2E/dk^2 becomes negative and m^* is negative.

We may draw the following conclusions about the effective mass m^* of an electron moving in a periodic lattice (see Fig. 24-30d).

1. m^* is not always equal to m.
2. m^* can be greater than m and, in fact, infinite.
3. m^* can be less than m or even negative.

We can gain some understanding of the behavior of m^* if we consider the way an electron wave is reflected by the lattice ions. We know from Section 12.7 that a wave is totally Bragg reflected if

$$2d \sin\theta = n\lambda \tag{12.11}$$

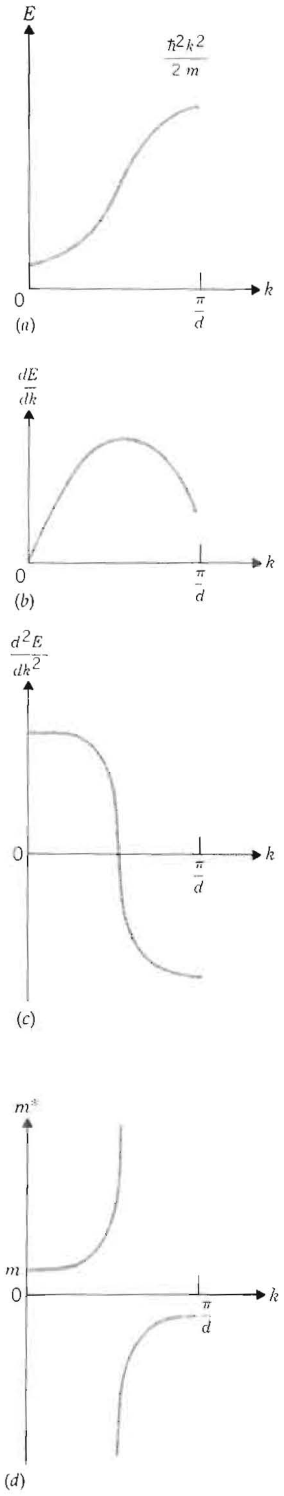


FIGURE 24-29 (*a*) Dependence of the energy E of an electron on its wave vector k, and hence its momentum p (solid line) for the first allowed energy band of the Kronig-Penney model (see Fig. 24-10). The dashed line illustrates the relation between E and k for the free electron case. (*b*) Qualitative plot of the derivative of E with respect to k as a function of k for the solid line of (*a*). (*c*) Qualitative plot of the second derivative of E with respect to k as a function of k. (d) Dependence of the effective mass m^* of the electron on the wave vector k for the situation depicted by the solid line of Figure (*a*). Note that the effective mass is negative near the top of the band, that is, as k approaches π/d.

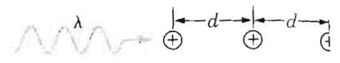


FIGURE 24-30 Bragg scattering of an electron wave by a one-dimensional array of ions.

In the case of a one-dimensional lattice that we have been considering, $\theta = 90°$ (Fig. 23-21) and Eq. 12.11 becomes

$$2d = n\lambda$$

This is sketched in Fig. 24-31. We can express the Bragg equation in terms of the wave vector $k = 2\pi/\lambda$

$$2d = n\lambda = n\frac{2\pi}{k}$$

and

$$k = n\frac{\pi}{d}$$

At the bottom of a band where $k \approx 0$, there is practically no reflection because the Bragg condition is far from being satisfied. The lattice ions will have little effect on the electron when it is accelerated by the external field $\mathcal{E}$. Higher up in the band, k will get closer to the critical value π/d and reflection starts to become appreciable. In this region, as the external field $\mathcal{E}$ accelerates the electron, the momentum increases and gets closer to the critical value. The external field increases the forward momentum, but at the same time enhances reflection, and the reflection corresponds to reversing the sign of the momentum. At the point where $1/m^* = 0$ (that is, $m^* = \infty$), the gain in the forward momentum resulting from the applied $\mathcal{E}$ is exactly compensated by the resulting enhancement in reflection by the lattice ions. The net change in the forward momentum is zero. Thus the overall response of the electron to the field $\mathcal{E}$ is as if it had an infinite mass; that is, it cannot be accelerated. At the top of the band, even closer to the critical value for total reflection, the second effect (enhanced reflection) is more important than the direct action of the applied field. The net result is that the electron responds with a change in momentum that is in the opposite direction to what the free electron would have acquired: the electron responds as if it had a negative mass.

The concept of effective mass has many uses. For example, in the free electron models we derived an expression for the electrical conductivity σ by considering the

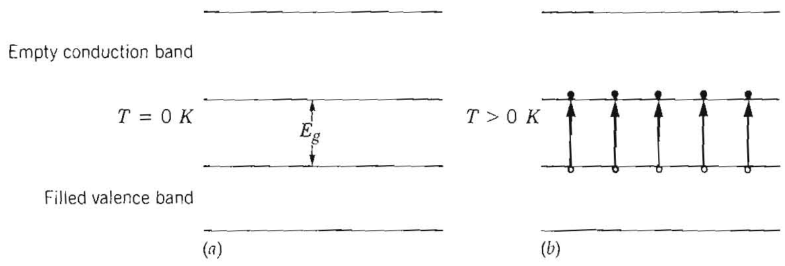


FIGURE 24-31 (*a*) At $T = 0$ K, the valence band of a semiconductor is filled with electrons and separated by an energy gap E_g from an empty conduction band. (*b*) At $T > 0$ K electrons are thermally excited into the conduction band, leaving behind in the valence band unoccupied energy states called *holes*, which behave as mobile positive charge carriers.

In the case of a one-dimensional lattice that we have been considering, $\theta = 90°$ (Fig. 23-21) and Eq. 12.11 becomes

$$2d = n\lambda$$

This is sketched in Fig. 24-31. We can express the Bragg equation in terms of the wave vector $k = 2\pi/\lambda$

$$2d = n\lambda = n\frac{2\pi}{k}$$

and

$$k = n\frac{\pi}{d}$$

At the bottom of a band where $k \approx 0$, there is practically no reflection because the Bragg condition is far from being satisfied. The lattice ions will have little effect on the electron when it is accelerated by the external field $\mathcal{E}$. Higher up in the band, k will get closer to the critical value π/d and reflection starts to become appreciable. In this region, as the external field $\mathcal{E}$ accelerates the electron, the momentum increases and gets closer to the critical value. The external field increases the forward momentum, but at the same time enhances reflection, and the reflection corresponds to reversing the sign of the momentum. At the point where $1/m^* = 0$ (that is, $m^* = \infty$), the gain in the forward momentum resulting from the applied $\mathcal{E}$ is exactly compensated by the resulting enhancement in reflection by the lattice ions. The net change in the forward momentum is zero. Thus the overall response of the electron to the field $\mathcal{E}$ is as if it had an infinite mass; that is, it cannot be accelerated. At the top of the band, even closer to the critical value for total reflection, the second effect (enhanced reflection) is more important than the direct action of the applied field. The net result is that the electron responds with a change in momentum that is in the opposite direction to what the free electron would have acquired: the electron responds as if it had a negative mass.

The concept of effective mass has many uses. For example, in the free electron models we derived an expression for the electrical conductivity σ by considering the

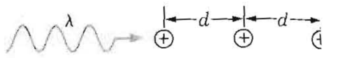
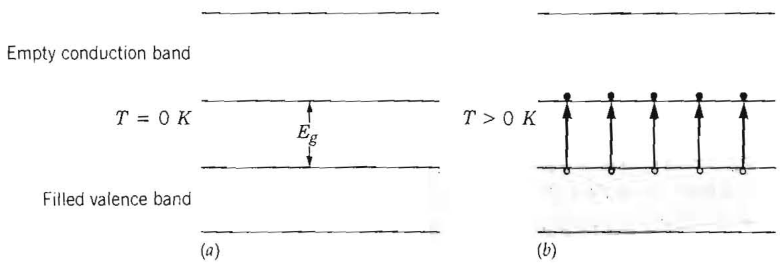


FIGURE 24-30 Bragg scattering of an electron wave by a one-dimensional array of ions.



(a) (b)

FIGURE 24-31 (*a*) At $T = 0\,\mathrm{K}$, the valence band of a semiconductor is filled with electrons and separated by an energy gap E_g from an empty conduction band. (*b*) At $T > 0\,\mathrm{K}$ electrons are thermally excited into the conduction band, leaving behind in the valence band unoccupied energy states called *holes*, which behave as mobile positive charge carriers.

response of the electrons to an electric field, and we found the conductivity $\sigma \propto 1/m$ (Eq. 23.14). We can easily make the correction to the model, that is, account for the fact that the electrons are not really free but instead they move in the electric potential of the ions. All that has to be done is to replace the true mass m by the effective mass m^*. In many metals this has little effect because $m^* \approx m$ (for example, Cu, Na, Al, and K). But in some metals it has a significant effect. As an example, the average m^* for iron (Fe) is about 10 times the free electron mass m, and this is one reason why iron is not a very good electrical conductor.

What about negative mass? Are there any cases where the electrons travel in the same direction as the electric field? These questions lead us to our next topic.

24.7 HOLES

The concept that electrons near the top of the band have negative effective mass leads to a very interesting feature that has a tremendous importance in the operation of all semiconductor devices.

At $T = 0$ K, the band structure of a semiconductor is characterized by a fully occupied valence band and a completely empty conduction band (see Fig. 24-32*a*). The semiconductor ideally is an insulator with zero conductivity at $T = 0$ K. As the temperature is raised, some electrons in the valence band can receive enough thermal energy and be excited into the conduction band because the energy gap between the two bands is rather narrow. The result is that there are some electrons in an otherwise empty conduction band and some unoccupied states in an otherwise filled valence band, see Fig. 24-32*b*. An empty state in the valence band is called a *hole*.

The electrons in the conduction band can move under the influence of an external electric field because they have available to them many empty higher energy states, and they can contribute to the current density J. Similarly, the electrons in the valence band can move into the empty states (holes) left by the electrons that were excited into the conduction band. We will assume, based on our previous discussion, that the empty states at the top of the valence band are *negative effective mass states*.

The interesting and important feature that we mentioned before is that *the conduction by the electrons in the valence band as they move into the empty negative mass states is completely equivalent to the conduction by particles of positive charge and positive mass. The number of such $+q$, $+m$ particles is equal to the number of available empty states, that is, the number of "holes."* Basically, what we are saying is that when considering the contribution to the electric current from the valence band, we ignore the electrons, and instead we treat it as if conduction took place via positively charged holes.

response of the electrons to an electric field, and we found the conductivity $\sigma \propto 1/m$ (Eq. 23.14). We can easily make the correction to the model, that is, account for the fact that the electrons are not really free but instead they move in the electric potential of the ions. All that has to be done is to replace the true mass m by the effective mass m^*. In many metals this has little effect because $m^* \approx m$ (for example, Cu, Na, Al, and K). But in some metals it has a significant effect. As an example, the average m^* for iron (Fe) is about 10 times the free electron mass m, and this is one reason why iron is not a very good electrical conductor.

What about negative mass? Are there any cases where the electrons travel in the same direction as the electric field? These questions lead us to our next topic.

24.7 HOLES

The concept that electrons near the top of the band have negative effective mass leads to a very interesting feature that has a tremendous importance in the operation of all semiconductor devices.

At $T = 0$ K, the band structure of a semiconductor is characterized by a fully occupied valence band and a completely empty conduction band (see Fig. 24-32*a*). The semiconductor ideally is an insulator with zero conductivity at $T = 0$ K. As the temperature is raised, some electrons in the valence band can receive enough thermal energy and be excited into the conduction band because the energy gap between the two bands is rather narrow. The result is that there are some electrons in an otherwise empty conduction band and some unoccupied states in an otherwise filled valence band, see Fig. 24-32*b*. An empty state in the valence band is called a *hole*.

The electrons in the conduction band can move under the influence of an external electric field because they have available to them many empty higher energy states, and they can contribute to the current density J. Similarly, the electrons in the valence band can move into the empty states (holes) left by the electrons that were excited into the conduction band. We will assume, based on our previous discussion, that the empty states at the top of the valence band are *negative effective mass states*.

The interesting and important feature that we mentioned before is that *the conduction by the electrons in the valence band as they move into the empty negative mass states is completely equivalent to the conduction by particles of positive charge and positive mass. The number of such $+q$, $+m$ particles is equal to the number of available empty states, that is, the number of "holes."* Basically, what we are saying is that when considering the contribution to the electric current from the valence band, we ignore the electrons, and instead we treat it as if conduction took place via positively charged holes.

convention is that it is negative if the predominant type of carrier is the electron, and positive if it is the hole. Regardless of the complexities involved in the calculation of R_H when both types of carriers are present, it should be clear that R_H could never be positive if holes were not a reality. Some examples of the value of the Hall coefficient are listed in Table 24-2.

TABLE 24-2

Solid	R_H (m^3/C)
Lithium	-17×10^{-11}
Sodium	-25×10^{-11}
Beryllium	$+24 \times 10^{-11}$
Zinc	$+3 \times 10^{-11}$
Cadmium	$+6 \times 10^{-11}$

It can be seen in this table that some well-known metals conduct electricity with holes as the predominant charge carrier.

We conclude this chapter with a comment about the behavior of holes. Consider the situation depicted in Fig. 24-32*b*, where a few electrons have been excited into the conduction band of the semiconductor leaving behind an equal number of holes at the top of the valence band. As we indicated earlier in this section, if additional energy is provided to the charge carriers of the semiconductor, such as by the application of an external electric field, the electrons in the valence band will move up into the empty states (the holes). When this happens, the holes move down from their initial position in the valence band. Because we have decided to ignore the electrons in the valence band and instead consider only the holes, we conclude that when energy is provided to the carriers in the valence band, the holes move down, or putting it differently, the energy of the holes *increases downward* from the top of the valence band. This fact will be used in Section 25.2a to calculate the hole concentration in the valence band of a semiconductor.

PROBLEMS

24.1 In the Kronig-Penney model, the relation between the energy E and the momentum p of the electrons is given by the relation

$$P\frac{\sin \gamma d}{\gamma d} + \cos \gamma d = \cos kd$$

where

$$P = \frac{mE_{p0}bd}{\hbar^2}, \quad \gamma = \left(\frac{2mE}{\hbar^2}\right)^{1/2}, \quad \text{and} \quad k = \frac{p}{\hbar}$$

Show that if $E_{p0} = 0$, the energy spectrum becomes continuous and it is that of the free particle,

$$E = \frac{\hbar^2 k^2}{2m} = \frac{p^2}{2m}$$

convention is that it is negative if the predominant type of carrier is the electron, and positive if it is the hole. Regardless of the complexities involved in the calculation of R_H when both types of carriers are present, it should be clear that R_H could never be positive if holes were not a reality. Some examples of the value of the Hall coefficient are listed in Table 24-2.

TABLE 24-2

Solid	R_H (m^3/C)
Lithium	-17×10^{-11}
Sodium	-25×10^{-11}
Beryllium	$+24 \times 10^{-11}$
Zinc	$+3 \times 10^{-11}$
Cadmium	$+6 \times 10^{-11}$

It can be seen in this table that some well-known metals conduct electricity with holes as the predominant charge carrier.

We conclude this chapter with a comment about the behavior of holes. Consider the situation depicted in Fig. 24-32*b*, where a few electrons have been excited into the conduction band of the semiconductor leaving behind an equal number of holes at the top of the valence band. As we indicated earlier in this section, if additional energy is provided to the charge carriers of the semiconductor, such as by the application of an external electric field, the electrons in the valence band will move up into the empty states (the holes). When this happens, the holes move down from their initial position in the valence band. Because we have decided to ignore the electrons in the valence band and instead consider only the holes, we conclude that when energy is provided to the carriers in the valence band, the holes move down, or putting it differently, the energy of the holes *increases downward* from the top of the valence band. This fact will be used in Section 25.2a to calculate the hole concentration in the valence band of a semiconductor.

PROBLEMS

24.1 In the Kronig-Penney model, the relation between the energy E and the momentum p of the electrons is given by the relation

$$P\frac{\sin \gamma d}{\gamma d} + \cos \gamma d = \cos kd$$

where

$$P = \frac{mE_{p0}bd}{\hbar^2}, \quad \gamma = \left(\frac{2mE}{\hbar^2}\right)^{1/2}, \quad \text{and} \quad k = \frac{p}{\hbar}$$

Show that if $E_{p0} = 0$, the energy spectrum become continuous and it is that of the free particle,

$$E = \frac{\hbar^2 k^2}{2m} = \frac{p^2}{2m}$$

Use physical arguments to justify this result.

Sketch the possible wavefunctions, as in Fig. 24-21, for a 3 atom chain of hydrogen atoms. How many energy levels are formed?

The solution of the Kronig Penney model shows that the width of the allowed bands increases as the energy increases (see Fig. 24-9). Use the arguments presented in Section 24-4 concerning the overlap of the atomic wavefunctions to explain this result.

Use the arguments presented in Section 24-5 to predict the band structure of (a) Li, and (b) Al. (c) Given the results predicted by these simplified arguments, how do you explain that aluminum is trivalent?

Use the concept of band theory to explain the following observations: Most insulators are transparent to visible light, semiconductors are transparent to infrared light but opaque to visible light, all metals are opaque to light of all wavelengths.

The energy gaps of some alkali halides are KCl = 7.6 eV, KBr = 6.3 eV, KI = 5.6 eV. Which of these are transparent to visible light? At what wavelength does each become opaque?

The experimentally determined value of the Fermi energy in Na is 2.50 eV. In Chapter 23 we showed that

$$E_F = \frac{\hbar^2}{2m}[3N\pi^2]^{2/3}$$

Use the measured value of E_F to calculate the effective mass of the electrons. Na is monovalent. Its atomic weight and density are 22.99 g/mole and 0.97 g/cm^3, respectively.

Answer: 1.26 m, where m is the free electron mass.

The density of aluminum is 2.70 g/cm^3 and its molecular weight is 26.98 g/mole. (a) Calculate the Fermi energy. (b) If the experimental value of E_F is 12 eV, what is the electron effective mass in aluminum? Aluminum is trivalent.

Answer: (a) 11.6 eV, (b) 0.97 m, where m is the free electron mass.

The experimentally measured specific heat of the conduction electrons in beryllium is $C_v = 0.54 \times 10^{-4}T$ calories/mole-K. (a) What is the Fermi energy for beryllium? (b) What is the electron effective mass in beryllium? The density of beryllium is 1.86 g/cm^3 and its molecular weight is 9.01 g/mole.

Answer: (a) 15.7 eV, (b) 0.92 m, where m is the free electron mass.

For a free electron, the relation between its energy and its momentum is given by

$$E = \frac{p^2}{2m}$$

Because $p = \hbar k$, the relation can be written as

$$E = \frac{\hbar^2 k^2}{2m}$$

The dependence of E on k is parabolic. For an electron moving in the periodic potential of the lattice ions, the relation between E and k is more complicated. As we have seen, the solution of the Schrödinger equation for an electron moving in a one-dimensional periodic potential consisting of equally spaced potential wells yields the following relation between E and k

$$P\frac{\sin \gamma d}{\gamma d} + \cos \gamma d = \cos kd$$

where

$$P = \frac{mE_{p0}bd}{\hbar^2}$$

is a constant, b is the separation between the wells, d is the periodicity of the wells (see Fig. 24-4), and $\gamma = [2mE/\hbar^2]^{1/2}$. This equation cannot be solved explicitly for $E(k)$. It can be solved numerically. That is, we pick a value of E, plug it into the equation, and find k. The procedure is repeated for other values of E. This numerical solution yields a rather interesting result. There are ranges of E for which k is a real number. These ranges are separated by ranges of E for which k is imaginary. This happens because the left side of the equation is greater than 1 or smaller than -1. Because the right side is $\cos kd$, and the cosine of a real number varies between 1 and -1, k is imaginary for those values of E for which the left side of the equation is greater than 1 or less than -1. The momentum of a particle cannot be imaginary; therefore, the electron cannot possess those values of E for which k is imaginary. This leads to the existence of allowed and forbidden energy bands in solids.

Use physical arguments to justify this result.

Sketch the possible wavefunctions, as in Fig. 24-21, for a 3 atom chain of hydrogen atoms. How many energy levels are formed?

The solution of the Kronig Penney model shows that the width of the allowed bands increases as the energy increases (see Fig. 24-9). Use the arguments presented in Section 24-4 concerning the overlap of the atomic wavefunctions to explain this result.

Use the arguments presented in Section 24-5 to predict the band structure of (a) Li, and (b) Al. (c) Given the results predicted by these simplified arguments, how do you explain that aluminum is trivalent?

Use the concept of band theory to explain the following observations: Most insulators are transparent to visible light, semiconductors are transparent to infrared light but opaque to visible light, all metals are opaque to light of all wavelengths.

The energy gaps of some alkali halides are KCl = 7.6 eV, KBr = 6.3 eV, KI = 5.6 eV. Which of these are transparent to visible light? At what wavelength does each become opaque?

The experimentally determined value of the Fermi energy in Na is 2.50 eV. In Chapter 23 we showed that

$$E_F = \frac{\hbar^2}{2m}[3N\pi^2]^{2/3}$$

Use the measured value of E_F to calculate the effective mass of the electrons. Na is monovalent. Its atomic weight and density are 22.99 g/mole and 0.97 g/cm^3, respectively.

Answer: 1.26 m, where m is the free electron mass.

The density of aluminum is 2.70 g/cm^3 and its molecular weight is 26.98 g/mole. (a) Calculate the Fermi energy. (b) If the experimental value of E_F is 12 eV, what is the electron effective mass in aluminum? Aluminum is trivalent.

Answer: (a) 11.6 eV, (b) 0.97 m, where m is the free electron mass.

The experimentally measured specific heat of the conduction electrons in beryllium is $C_v = 0.54 \times 10^{-4} T$ calories/mole-K. (a) What is the Fermi energy for beryllium? (b) What is the electron effective mass in beryllium? The density of beryllium is 1.86 g/cm^3 and its molecular weight is 9.01 g/mole.

Answer: (a) 15.7 eV, (b) 0.92 m, where m is the free electron mass.

For a free electron, the relation between its energy and its momentum is given by

$$E = \frac{p^2}{2m}$$

Because $p = \hbar k$, the relation can be written as

$$E = \frac{\hbar^2 k^2}{2m}$$

The dependence of E on k is parabolic. For an electron moving in the periodic potential of the lattice ions, the relation between E and k is more complicated. As we have seen, the solution of the Schrödinger equation for an electron moving in a one-dimensional periodic potential consisting of equally spaced potential wells yields the following relation between E and k

$$P\frac{\sin \gamma d}{\gamma d} + \cos \gamma d = \cos kd$$

where

$$P = \frac{mE_{p0}bd}{\hbar^2}$$

is a constant, b is the separation between the wells, d is the periodicity of the wells (see Fig. 24-4), and $\gamma = [2mE/\hbar^2]^{1/2}$. This equation cannot be solved explicitly for $E(k)$. It can be solved numerically. That is, we pick a value of E, plug it into the equation, and find k. The procedure is repeated for other values of E. This numerical solution yields a rather interesting result. There are ranges of E for which k is a real number. These ranges are separated by ranges of E for which k is imaginary. This happens because the left side of the equation is greater than 1 or smaller than -1. Because the right side is $\cos kd$, and the cosine of a real number varies between 1 and -1, k is imaginary for those values of E for which the left side of the equation is greater than 1 or less than -1. The momentum of a particle cannot be imaginary; therefore, the electron cannot possess those values of E for which k is imaginary. This leads to the existence of allowed and forbidden energy bands in solids.