CHAPTER

4

Modern Theory of Solids

One of the great successes of modern physics has been the application of quantum mechanics or the Schrödinger equation to the behavior of molecules and solids. For example, quantum mechanics explains the nature of the bond between atoms, and its consequences. How can carbon bond with four other carbon atoms? What determines the direction and strength of a bond? An intuitively obvious outcome from quantum mechanics is that the energy of the electron is still quantized in the molecule. In addition, the application of quantum mechanics to many atoms, as in a solid, leads to energy bands within which the electron energy levels are almost continuous. The electron energy falls within possible values in a band of energies. It is nearly impossible to comprehend the principles of operation of modern solid-state electronic devices without a good grasp of the band theory of solids. Since we are dealing with a large number of electrons in the solid, we must consider a statistical way of describing their behavior, just as we use the Maxwell distribution of velocities to explain the behavior of gas atoms. An equally important question, therefore, is "What is the probability that an electron is in a state with energy E within an energy band?"

4.1 HYDROGEN MOLECULE: MOLECULAR ORBITAL THEORY OF BONDING

Consider what happens when two hydrogen atoms approach each other to form the hydrogen molecule. This is the H–H (or H_2) system. Let us examine the energy levels of the H–H system as a function of the interatomic distance R. When the atoms are infinitely separated, each atom has its own set of energy levels, labeled $1s$, $2s$, $2p$, etc. The electron energy in each atom is -13.6 eV with respect to the "free" state (electron infinitely separated from the parent nucleus). The energy of the two isolated hydrogen atoms is twice -13.6 eV.

As the atoms approach closer, the electrons interact both with each other and with the other nuclei. To obtain the wavefunctions and the new energy of the electrons, we

need to find the new potential energy function *PE* for the electrons in this new environment and then solve the Schrödinger equation with this new *PE* function. The new energy is actually *lower* than twice −13.6 eV, which means that the H_2 formation is energetically favorable.

The bond formation between two H atoms can be easily explained by describing the behavior of the electron within the molecule. We use a **molecular orbital** ψ, which depends on the interaction of individual atomic wavefunctions and is regarded as an electron wavefunction within the molecule.

In the H_2 molecule, we cannot have two sets of identical atomic ψ_{1s} orbitals, for two reasons. First, this would violate the Pauli exclusion principle, which requires that, in a given system of electrons (those within the H_2 molecule), we cannot have two sets of identical quantum numbers. When the atoms were separated, we did not have this problem, because we had two isolated systems.

Second, as the two atoms approach each other, as shown in Figure 4.1, the atomic ψ_{1s} wavefunctions overlap. This overlap produces two new wavefunctions with different energies and hence different quantum numbers. When the two atomic wavefunctions interfere, they can overlap either in phase (both positive or both negative) or out of phase

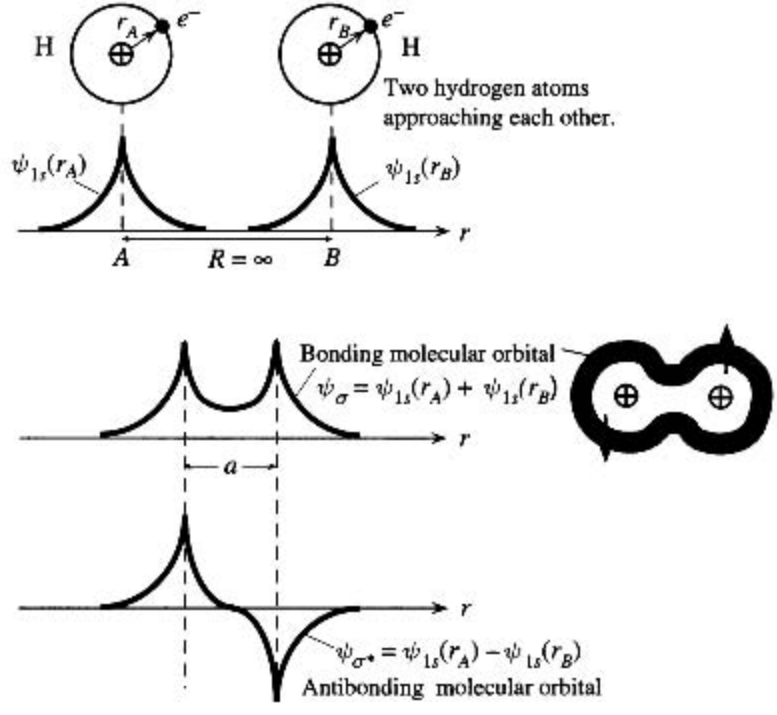


Figure 4.1 Formation of molecular orbitals, bonding, and antibonding (ψ_σ and ψ_{σ^*}) when two H atoms approach each other.
The two electrons pair their spins and occupy the bonding orbital ψ_σ.

(one positive and the other negative), as a result of which two molecular orbitals are formed. These are conventionally labeled ψ_σ and ψ_{σ^*} as illustrated in Figure 4.1. Thus, two of the molecular orbitals in the H–H system are .

$$\psi_\sigma = \psi_{1s}(r_A) + \psi_{1s}(r_B) \quad [4.1]$$

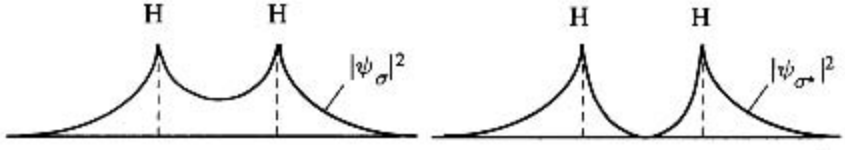
$$\psi_{\sigma^*} = \psi_{1s}(r_A) - \psi_{1s}(r_B) \quad [4.2]$$

where the two hydrogen atoms are labeled A and B, and r_A and r_B are the respective distances of the electrons from their parent nucleus. In generating two separate molecular orbitals ψ_σ and ψ_{σ^*} from a linear combination of two identical atomic orbitals ψ_{1s}, we have used the **linear combination of atomic orbitals (LCAO)** method.

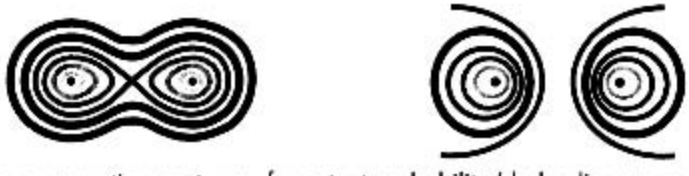
The first molecular orbital ψ_σ is *symmetric* and has considerable magnitude between the nuclei, whereas the second ψ_{σ^*}, is *antisymmetric* and has a node between the nuclei. The resulting electron probability distributions $|\psi_\sigma|^2$ and $|\psi_{\sigma^*}|^2$ are shown in Figure 4.2.

In an analogy to hydrogenic wavefunctions, since ψ_{σ^*} has a node, we would expect it to have a higher energy than the ψ_σ orbital and therefore a different energy quantum number, which means that the Pauli exclusion principle is no longer violated. We can also expect that because $|\psi_\sigma|^2$ has an appreciable electron concentration between the two nuclei, the electrostatic *PE*, and hence the total energy for the wavefunction ψ_σ, will be lower than that for ψ_{σ^*}, as well as those for the individual atomic wavefunctions.

Of course, the true wavefunctions of the electrons in the H_2 system must be determined by solving the Schrödinger equation, but an intelligent guess is that these must look like ψ_σ and ψ_{σ^*}. We can therefore use ψ_σ and ψ_{σ^*} in the Schrödinger equation, with the correct form of the *PE* term V, to evaluate the energies E_σ and E_{σ^*} of ψ_σ and ψ_{σ^*}, respectively, as a function of R. The *PE* function V in the H–H system has positive *PE* contributions arising from electron–electron repulsions and proton–proton



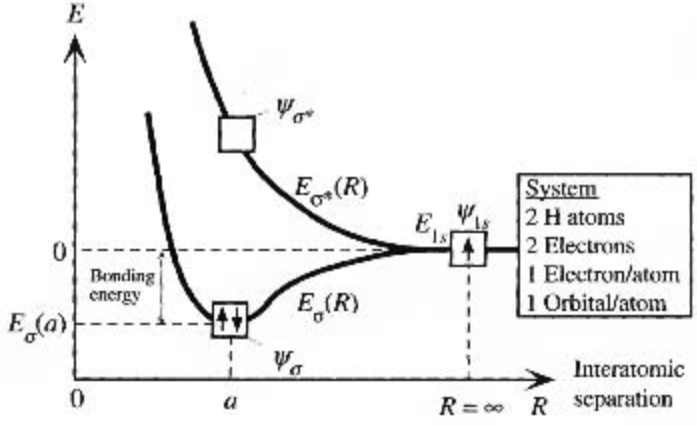
(a) Electron probability distributions for bonding and antibonding orbitals, ψ_σ and ψ_{σ^*}.



(b) Lines representing contours of constant probability (darker lines represent greater relative probability).

Figure 4.2

(a) Energy of ψ_σ and ψ_{σ^*} vs. the interatomic separation R.



(b) Schematic diagram showing the changes in the electron energy as two isolated H atoms, far left and far right, come together to form a hydrogen molecule.

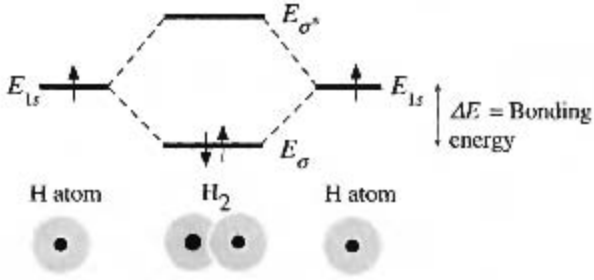
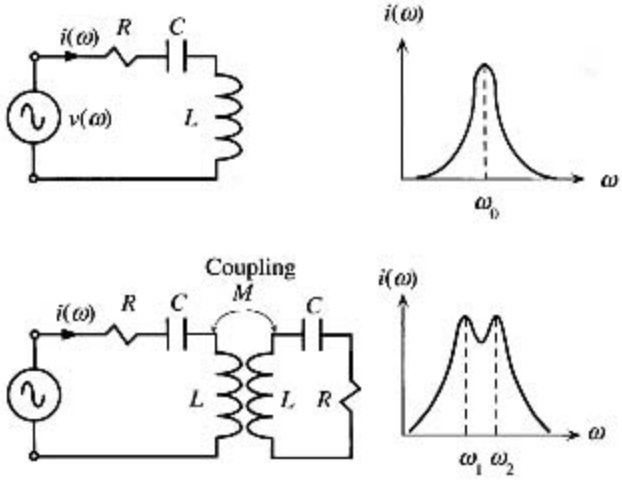


Figure 4.3 Electron energy in the system comprising two hydrogen atoms.

repulsions, but negative PE contributions arising from the attractions of the two electrons to the two protons.

The two energies, E_σ and E_{σ^*}, are widely different, with E_σ below E_{1s} and E_{σ^*} above E_{1s}, as shown schematically in Figure 4.3a. As R decreases and the two H atoms get closer, the energy of the ψ_σ orbital state passes through a minimum at $R = a$. Each orbital state can hold two electrons with spins paired, and within the two hydrogen atoms, we have two electrons. If these enter the ψ_σ orbital and pair their spins, then this new configuration is energetically more favorable than two isolated H atoms. It corresponds to the hydrogen molecule H_2. The energy difference between that of the two isolated H atoms and the E_σ minimum energy at $R = a$ is the bonding energy, as illustrated in Figure 4.3a. When the two electrons in the H_2 molecule occupy the ψ_σ orbital, their probability distribution (and hence, the negative charge distribution) is such that the negative PE, arising from the attractions of these two electrons to the two protons, is stronger in magnitude than the positive PE, arising from electron–electron repulsions and proton–proton repulsions and the kinetic energy of the two electrons. Therefore, the H_2 molecule is energetically stable.

The wavefunction ψ_σ corresponding to the lowest electron energy is called the **bonding orbital,** and ψ_{σ^*} is the **antibonding orbital.** When two atoms are brought together, the two identical atomic wavefunctions combine in two ways to generate two different molecular orbitals, each with a different energy. Effectively, then, an atomic



(a) There is one resonant frequency, ω_0, in an isolated *RLC* circuit.

(b) There are two resonant frequencies in two coupled *RLC* circuits: one below and the other above ω_0.

Figure 4.4

energy level, such as E_{1s}, splits into two, E_σ and E_{σ^*}. The splitting is due to the interaction (or overlap) between the atomic orbitals. Figure 4.3b schematically illustrates the changes in the electron energy levels as two isolated H atoms are brought together to form the H_2 molecule.

The splitting of a one-atom energy level when a molecule is formed is analogous to the splitting of the resonant frequency in an *RLC* circuit when two such circuits are brought together and coupled. Consider the *RLC* circuit shown in Figure 4.4a. The circuit is excited by an ac voltage source. The current peaks at the resonant frequency ω_0, as indicated in Figure 4.4a. When two such identical *RLC* circuits are coupled together and driven by an ac voltage source, the current develops two peaks, at frequencies ω_1 and ω_2, below and above ω_0, as illustrated in Figure 4.4b. The two peaks at ω_1 and ω_2 are due to the mutual inductance that couples the two circuits, allowing them to interact. From this analogy, we can intuitively accept the energy splitting observed in Figure 4.3a.

Consider what happens when two He atoms come together. Recall that the $1s$ orbital has paired electrons and is full. The $1s$ atomic energy level will again split into two levels, E_σ and E_{σ^*}, associated with the molecular orbitals ψ_σ and ψ_{σ^*}, as illustrated in Figure 4.5. However, in the He–He system, there are four electrons, so two occupy the ψ_σ orbital state and two go to the ψ_{σ^*} orbital state. Consequently, the system energy is not lowered by bringing the two He atoms closer. Furthermore, quantum mechanical calculations show that the antibonding energy level E_{σ^*} shifts higher than the bonding level E_σ shifts lower. By the same token, although we could put an additional electron at E_{σ^*} in H_2 to make H_2^-, we could not make H_2^{2-} by placing two electrons at E_{σ^*}.

From the He–He example, we can conclude that, as a general rule, the overlap of full atomic orbital states does not lead to bonding. In fact, full orbitals repel each other, because any overlap results in an increase in the system energy. To form a bond

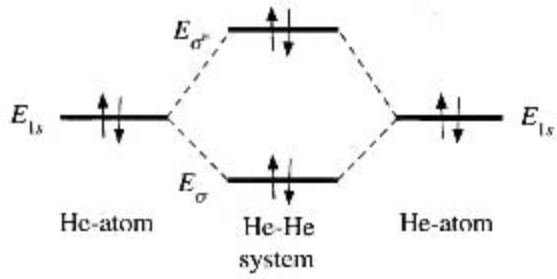


Figure 4.5 Two He atoms have four electrons. When He atoms come together, two of the electrons enter the E_σ level and two the E_{σ^*} level, so the overall energy is greater than two isolated He atoms.

between two atoms, we essentially need an overlap of half-occupied orbitals, as in the H_2 molecule.

EXAMPLE 4.1 **HYDROGEN HALIDE MOLECULE (HF)** We already know that H has a half-occupied $1s$ orbital, which can take part in bonding. Since the F atom has the electronic structure $1s^2 2s^2 p^5$, two of the p orbitals are full and one p orbital, p_x, is half full. This means that only the p_x orbital can participate in bonding. Figure 4.6 shows the electron orbitals in both H and F. When the H atom and the F atom approach each other to form an HF molecule, the ψ_{1s} orbital of H overlaps the p_x orbital of F. There are two possibilities for the overlap. First, ψ_{1s} and p_x can overlap in phase (both positive or both negative), to give a ψ_σ orbital that does not have a node between H and F, as shown in Figure 4.6. Second, they can overlap out of phase (one positive and the other negative), so that the overlap orbital ψ_{σ^*} has a node (similar to ψ_{σ^*} in Figure 4.1). We know from hydrogen atomic wavefunctions in Chapter 3 that orbitals with more nodes have higher energies. The molecular orbital ψ_σ therefore corresponds to a bonding orbital with a lower energy than the ψ_{σ^*} orbital. The two electrons, one from ψ_{1s} and the other from p_x, enter the ψ_σ orbital with spins paired, thereby forming a bond between H and F.

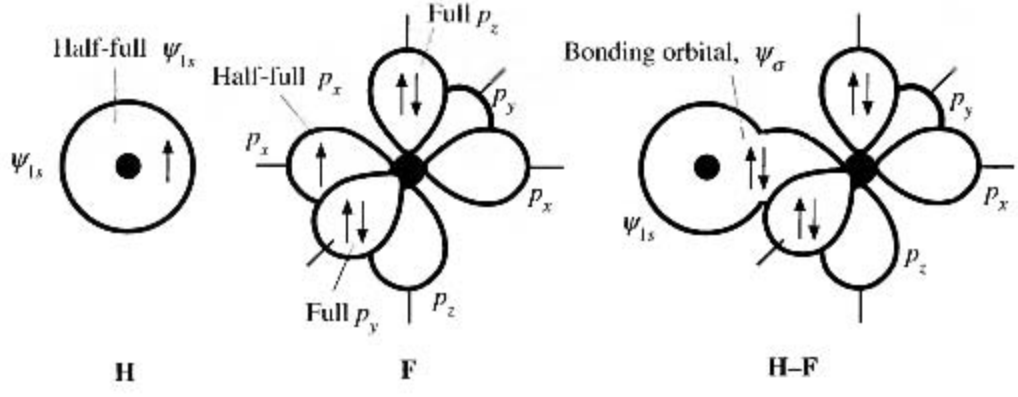


Figure 4.6 H has one half-empty ψ_{1s} orbital.
F has one half-empty p_x orbital but full p_y and p_z orbitals. The overlap between ψ_{1s} and p_x produces a bonding orbital and an antibonding orbital. The two electrons fill the bonding orbital and thereby form a covalent bond between H and F.

4.2 BAND THEORY OF SOLIDS

4.2.1 ENERGY BAND FORMATION

When we bring three hydrogen atoms (labeled A, B, and C) together, we generate three separate molecular orbital states, ψ_a, ψ_b, and ψ_c, from three ψ_{1s} atomic states. Again, this occurs in three different ways, as illustrated in Figure 4.7a. As in the case of the H_2 molecule, each molecular orbital must be either *symmetric* or *antisymmetric* with respect to center atom B.[1] The orbitals that satisfy even and odd requirements are

$$\psi_a = \psi_{1s}(A) + \psi_{1s}(B) + \psi_{1s}(C) \quad \text{[4.3a]}$$

$$\psi_b = \psi_{1s}(A) - \psi_{1s}(C) \quad \text{[4.3b]}$$

$$\psi_c = \psi_{1s}(A) - \psi_{1s}(B) + \psi_{1s}(C) \quad \text{[4.3c]}$$

where $\psi_{1s}(A)$, $\psi_{1s}(B)$, and $\psi_{1s}(C)$ are the $1s$ atomic wavefunctions centered around the atoms A, B, and C, respectively, as shown in Figure 4.7a. For example, the wavefunction $\psi_{1s}(A)$ represents $\psi_{1s}(r_A)$, which is centered around A and has the form $\exp(-r_A/a_o)$, where r_A is the distance from the nucleus of A, and a_o is the Bohr radius. Notice that $\psi_{1s}(B)$ is missing in Equation 4.3b, so ψ_b is antisymmetric.

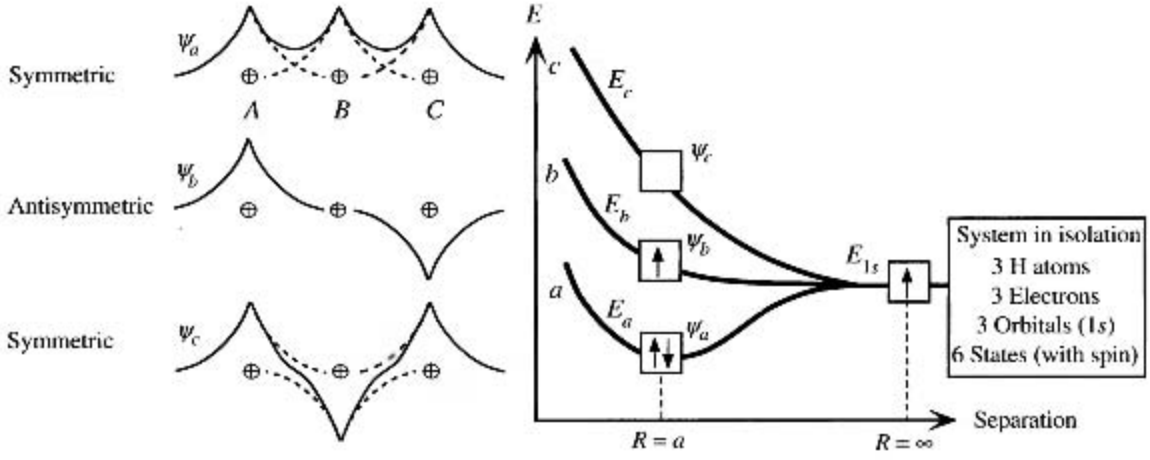
The energies E_a, E_b, and E_c of ψ_a, ψ_b, and ψ_c can be calculated from the Schrödinger equation by using the PE function of this system (the PE also includes proton–proton repulsions). It is clear that since ψ_a, ψ_b, and ψ_c are different, their energies E_a, E_b, and E_c are also different. Consequently, the $1s$ energy level splits into three separate levels, corresponding to the energies of ψ_a, ψ_b, and ψ_c, as depicted by Figure 4.7b. By analogy with the electron wavefunctions in the hydrogen atom, we can argue that if the molecular wavefunction has more nodes, its energy is higher. Thus, ψ_a has the lowest energy E_a, ψ_b has the next higher energy E_b, and ψ_c has the highest energy E_c, as shown in Figure 4.7b. There are three electrons in the three-hydrogen system. The first two pair their spins and enter orbital ψ_a at energy E_a, and the third enters orbital ψ_b at energy E_b. Comparing Figures 4.7 and 4.3, we notice that although H_2 and H_3 both have two electrons in the lowest energy level, H_3 also has an extra electron at the higher energy level (E_b), which tends to increase the net energy of the atom. Thus, the H_3 molecule is much less stable than the H_2 molecule.[2]

Now consider the formation of a solid. Take N Li (lithium) atoms from infinity and bring them together to form the Li metal. Lithium has the electronic configuration $1s^2 2s^1$, which is somewhat like the hydrogen atom, since the K shell is closed and the third electron is alone in the $2s$ orbital.

Based on our previous discussions, we assume that the atomic energy levels will split into N separate energy levels. Since the $1s$ subshell is full and is close to the nucleus, it will not be affected much by the interatomic interactions; consequently, the energy of

[1] The reason is that the molecule *A–B–C*, when *A*, *B*, and *C* are identical atoms, is symmetric with respect to *B*. Thus each wavefunction must have odd or even parity [Chapter 3].

[2] See G. Pimentel and R. Spratley, *Understanding Chemistry*, San Francisco: Holden-Day, Inc., 1972, pp. 682–687 for an excellent discussion.



(a) Three molecular orbitals from three ψ_{1s} atomic orbitals overlapping in three different ways.

(b) The energies of the three molecular orbitals, labeled *a*, *b*, and *c*, in a system with three H atoms.

Figure 4.7

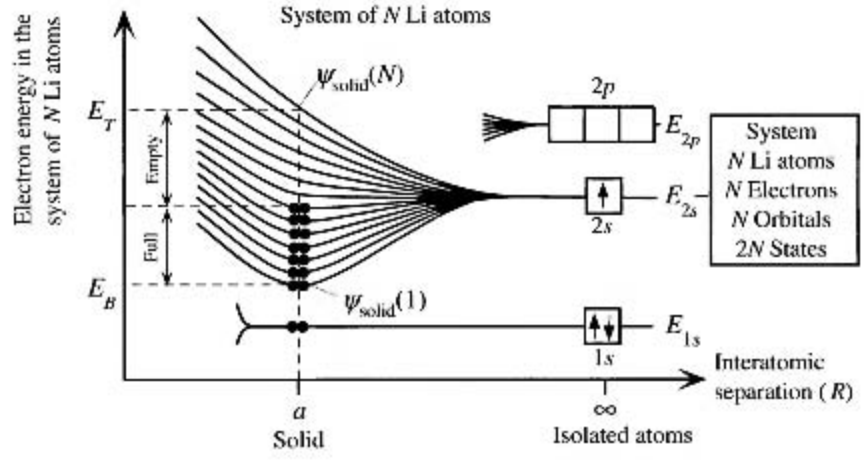


Figure 4.8 The formation of a 2*s* energy band from the 2*s* orbitals when *N* Li atoms come together to form the Li solid.

There are *N* 2*s* electrons, but 2*N* states in the band. The 2*s* band is therefore only half full. The atomic 1*s* orbital is close to the Li nucleus and remains undisturbed in the solid. Thus, each Li atom has a closed *K* shell (full 1*s* orbital).

this state will experience only negligible splitting, if any. Since the $1s$ electrons will stay close to their parent nuclei, we will not consider them during formation of the solid.

In the system of N isolated Li atoms, we have N electrons in N ψ_{2s} orbitals at the energy E_{2s}, as illustrated in Figure 4.8 (at infinite interatomic separation). Let us assume that N is large (typically, $\sim 10^{23}$). As N atoms are brought together to form the solid, the energy level at E_{2s} splits into N finely separated energy levels. The maximum width of the energy splitting depends on the closest interatomic distance a in the solid, as apparent in Figure 4.3a. The atoms separated by a distance greater than $R = a$ give rise to a lesser amount of energy splitting. The interatomic interactions between $N\psi_{2s}$ orbitals thus spread the N energy levels between the bottom and top levels, E_B and E_T, respectively, which are determined by the closest interatomic distance a. Put differently, E_B and E_T are determined by the distance between nearest neighbors. It is obvious that with N very large, the energy separation between two consecutive energy levels is very small; indeed, it is almost infinitesimal and not as exaggerated as in Figure 4.8.

Remember that each energy level E_i in the Li metal of Figure 4.8 is the energy of an electron wavefunction $\psi_{\text{solid}}(i)$ in the solid, where $\psi_{\text{solid}}(i)$ is one particular combination of the N atomic wavefunctions ψ_{2s}. There are N different ways to combine N atomic wavefunctions ψ_{2s}, since each can be added in phase or out of phase, as is apparent in Equations 4.3a to c (see also Figure 4.7a and b). For example, when all N ψ_{2s} are summed in phase, the resulting wavefunction $\psi_{\text{solid}}(1)$ is like ψ_a in Equation 4.3a, and it has the lowest energy. On the other hand, when N ψ_{2s} are summed with alternating phases, $+ - + \cdots$, the resulting wavefunction $\psi_{\text{solid}}(N)$ is like ψ_c, and it has the highest energy. Other combinations of ψ_{2s} give rise to different energy values between E_B and E_T.

The single $2s$ energy level E_{2s} therefore splits into N ($\sim 10^{23}$) finely separated energy levels, forming an **energy band,** as illustrated in Figure 4.8. Consequently, there are N separate energy levels, each of which can take two electrons with opposite spins. The N electrons fill all the levels up to and including the level at $N/2$. Therefore, the band is half full. We do not mean literally that the band is full to the half-energy point. The levels are not spread equally over the band from E_B to E_T, which means that the band cannot be full to the half-energy point. Half filled simply means half the states in the band are filled from the bottom up.

We have generated a half-filled band from a half-filled isolated $2s$ energy level. The energy band resulting from the splitting of the atomic $2s$ energy level is loosely termed the **2s band.** By the same token, the atomic $1s$ levels are full, so any $1s$ band that forms from these $1s$ states will also be full. We can get an idea of the separation of energy levels in the $2s$ band by noting that the maximum separation, $E_T - E_B$, between the top and bottom of the band is on the order of 10 eV, but there are some 10^{23} atoms, giving rise to 10^{23} energy levels between E_B and E_T. Thus, the energy levels are finely separated, forming, for all practical purposes, a continuum of energy levels.

The $2p$ energy level, as well as the higher levels at $3s$ and so on, also split into finely separated energy levels, as shown in Figure 4.9. In fact, some of these energy levels overlap the $2s$ band; hence, they provide further energy levels and "extend" the $2s$ band into higher energy levels, as indicated in Figure 4.10, which shows how energy bands in metals are often represented. The vertical axis is the electron energy. The top of the $2s$ band, which is half full, overlaps the bottom of the $2p$ band, which itself

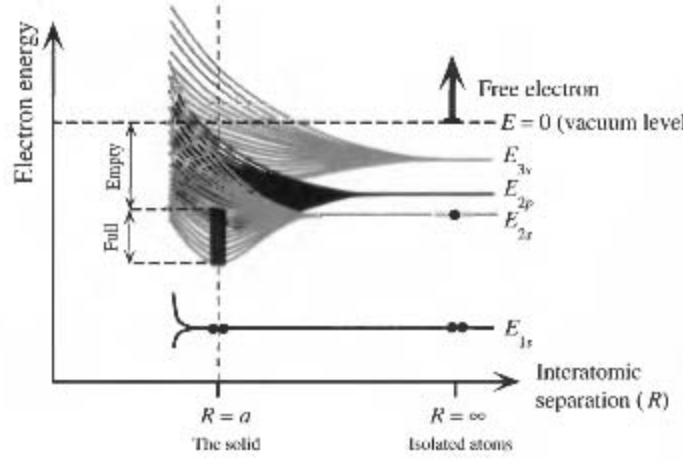


Figure 4.9 As Li atoms are brought together from infinity, the atomic orbitals overlap and give rise to bands.

Outer orbitals overlap first. The 3*s* orbitals give rise to the 3*s* band, 2*p* orbitals to the 2*p* band, and so on. The various bands overlap to produce a single band in which the energy is nearly continuous.

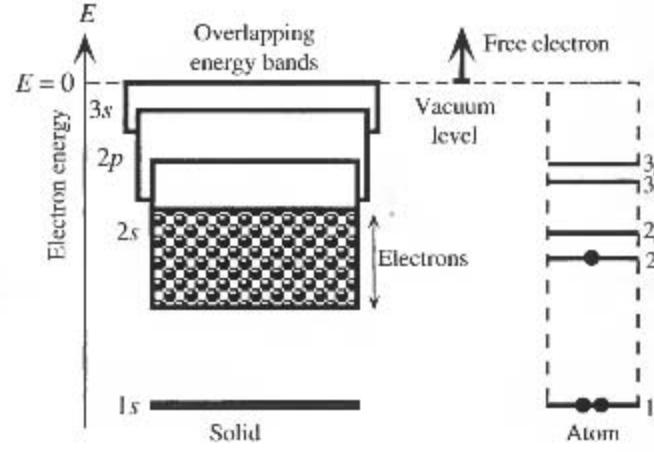


Figure 4.10 In a metal, the various energy bands overlap to give a single energy band that is only partially full of electrons.

There are states with energies up to the vacuum level, where the electron is free.

is overlapped near the top by the 3*s* band. We therefore have a band of energies that stretches from the bottom of the 2*s* band all the way to the vacuum level, as depicted in Figure 4.11. The reader may wonder what happened to the 3*d*, 4*s*, etc., bands. In the solid, the energies of these bands (including the top portion of the 3*s* band) are above the vacuum level, and the electron is free and far from the solid before it can acquire those energies.

At a temperature of absolute zero, or nearly so, the thermal energy is insufficient to excite the electrons to higher energy levels, so all the electrons pair their spins and fill each energy level from E_B up to an energy level E_{FO} that we call the Fermi level at 0 K, as shown in Figure 4.11. The energy value for the Fermi level depends on where we take the reference energy. For example, if we take the vacuum level as the zero reference, then for the Li metal, E_{FO} is at -2.5 eV. The Fermi level is normally measured with respect to the bottom of the band, in which case, it is simply termed the Fermi energy and denoted E_{FO}. For the Li metal, E_{FO} is 4.7 eV, which is with respect to the bottom of the band. The Fermi level has considerable significance, as we will discover later in this chapter.

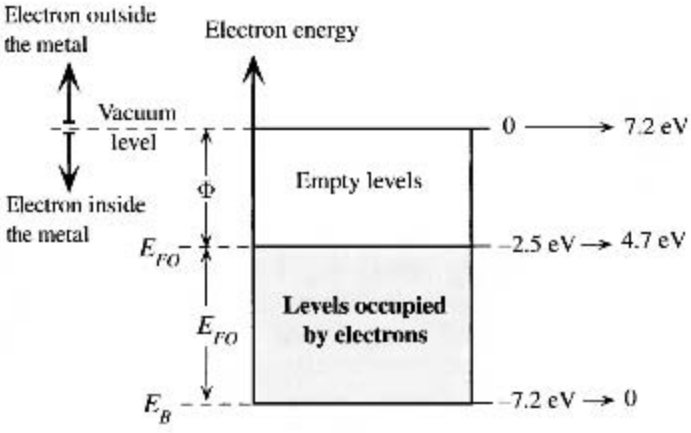


Figure 4.11 Typical electron energy band diagram for a metal.
All the valence electrons are in an energy band, which they only partially fill. The top of the band is the vacuum level, where the electron is free from the solid ($PE = 0$).

At absolute zero, all the energy levels up to the Fermi level are full. The energy required to excite an electron from the Fermi level to the vacuum level, that is, to liberate the electron from the metal, is called the **work function** Φ of the metal. As the temperature increases, some of the electrons get excited to higher energy levels. To determine the probability of finding an electron at an energy level E, we must consider what is called "particle statistics," a topic that is key to understanding the behavior of electronic devices. Clearly, the probability of finding an electron at 0 K at some energy $E < E_{FO}$ is unity, and at $E > E_{FO}$, the probability is zero. Table 4.1 summarizes the Fermi energy and work function of a few selected metals.

The electrons in the energy band of a metal are loosely bound valence electrons which become free in the crystal and thereby form a kind of **electron gas.** It is this electron gas that holds the metal ions together in the crystal structure and constitutes the metallic bond. This intuitive interpretation is shown in Figure 4.9. When solid Li is formed from N atoms, the N electrons fill all the lower energy levels up to $N/2$. The energy of the system of N Li atoms, according to Figure 4.9, is therefore much less than that of N isolated Li atoms by virtue of the N electrons taking up lower energy levels. It must be emphasized that the electrons within a band do not belong to any specific atom but to the whole solid. We cannot identify a given electron in the band with a certain Li atom. All the $2s$ electrons essentially form an electron gas and have energies that fall within the energy band. These electrons are constantly moving around in the metal which in terms of quantum mechanics means that their wavefunctions must be of the traveling wave type and not the type that localizes the electron around a given atom (*e.g.*, ψ_{n,ℓ,m_ℓ} in the hydrogen atom). We can represent each electron with a wavevector k so that its momentum p is $\hbar k$.

Table 4.1 Fermi energy and work function of selected metals

	Metal							
	Ag	**Al**	**Au**	**Cs**	**Cu**	**Li**	**Mg**	**Na**
Φ (eV)	4.5	4.28	5.0	2.14	4.65	2.3	3.7	2.75
E_{FO} (eV)	5.5	11.7	5.5	1.58	7.0	4.7	7.1	3.2

4.2.2 PROPERTIES OF ELECTRONS IN A BAND

Since the electrons inside the metal crystal are considered to be "free," their energy is KE. These electrons occupy all the energy levels up to E_{FO} as shown in the band diagram of Figure 4.12a. The energy E of an electron in a metal increases with its momentum p as $p^2/2m_e$. Figure 4.12b shows the energy versus momentum behavior of the electrons in a hypothetical one-dimensional crystal. The energy increases with momentum whether the electron is moving toward the left or right. Electrons take on all available momentum values until their energy reaches E_{FO}. For every electron that is moving right (such as a), there is another (such as b) with the same energy but moving left with the same magnitude of momentum. Thus, the average momentum is zero and there is no net current.

Consider what happens when an electric field $\mathcal{E}_x$ is applied in the $-x$ direction. The electron a at the Fermi level and moving along in the $+x$ direction experiences a force $e\mathcal{E}_x$ along the same direction. It therefore accelerates and gains momentum and hence has the energy as shown in Figure 4.12c. (The actual energy gained from the field is very small compared with E_{FO}, so Figure 4.12c is highly exaggerated.) The electron a at E_{FO} can move to higher energy levels because these adjacent higher levels are empty. The momentum state vacated by a is filled by the electron immediately below which now gains energy and moves up, and so on. An electron that is moving in the $-x$ direction, however, is decelerated (its momentum decreases) and hence loses energy as indicated by b moving to b' in Figure 4.12c. The electrons that are moving in the $+x$ direction gain energy, and those that are moving in the $-x$ direction, lose energy. The whole electron momentum distribution therefore shifts in the $+x$ direction as in Figure 4.12c. Eventually the electron a, now at a', is scattered by a lattice vibration.

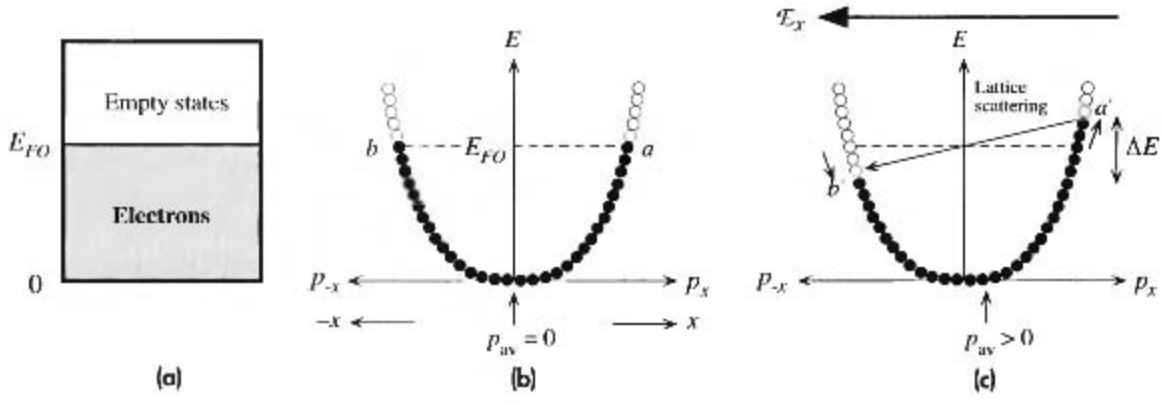


Figure 4.12

(a) Energy band diagram of a metal.

(b) In the absence of a field, there are as many electrons moving right as there are moving left. The motions of two electrons at each energy cancel each other as for a and b.

(c) In the presence of a field in the $-x$ direction, the electron a accelerates and gains energy to a' where it is scattered to an empty state near E_{FO} but moving in the $-x$ direction. The average of all momenta values is along the $+x$ direction and results in a net electric current.

Typically lattice vibrations have small energies but substantial momentum. The scattered electron must find an *unoccupied* momentum state with roughly the same energy, and it must change its momentum substantially. The electron at a' is therefore scattered to an empty state around E_{FO} but with a momentum in the opposite direction. Its momentum is *flipped* as shown in Figure 4.12c. The average momentum of the electrons is no longer zero but finite in the $+x$ direction. Consequently there is a current flow in the $-x$ direction, along the field, as determined by this average momentum p_{av}. Notice that a moves up to a' and b falls down to b'. Under steady-state conduction, lattice scattering simply replenishes the electrons at b' from a'. Notice that for energies below b', for every electron moving right there is another moving left with the same momentum magnitude that cancels it. Thus, electrons below the b' energy level do *not* contribute to conduction and are excluded from further consideration. Notice that electrons above the b' level are only moving right and their momenta are not canceled. Thus, the conductivity is determined by the electrons in the energy range ΔE from b' to a' about the Fermi level as shown in Figure 4.12c. Further, as the energy change from a to a' is orders of magnitude smaller than E_{FO}, we can summarize that conduction occurs by the drift of electrons at the Fermi level.[3] (If we were to calculate ΔE for a typical metal for typical currents, it would be $\sim 10^{-6}$ eV whereas E_{FO} is 1–10 eV. The shift in the distribution in Figure 4.12c is very small indeed; a' and b', for all practical purposes, are at the Fermi level.)

Conduction can be explained very simply and intuitively in terms of a band diagram as shown in Figure 4.13. Notice that the application of the electric field bends the energy band, because the electrostatic PE of the electron is $-eV(x)$ where $V(x)$ is the voltage at position x. However, $V(x)$ changes linearly from 0 to V, by virtue of $dV/dx = -\mathcal{E}_x$. Since $E = -eV(x)$ adds to the energy of the electron, the energy band must bend to account for the additional electrostatic energy. Since only the electrons near E_{FO} contribute to electrical conduction, we can represent this by drifting the electrons at E_{FO} down the potential hill. Although these electrons possess a very high mean velocity ($\sim 10^6$ m s^{-1}), as determined by the Fermi energy, they drift very slowly (10^{-2}–10^{-1} m s^{-1}) with a velocity that is drift mobility × field.

When a metal is illuminated, provided the wavelength of the radiation is correct, it will cause emission of electrons from the metal as in the photoelectric effect. Since Φ is the "minimum energy" required to excite an electron into the vacuum level (out from the metal), the longest wavelength radiation required is $hc/\lambda = \Phi$.

Addition of heat to a metal can excite some of the electrons in the band to higher energy levels. Thus heat can also be absorbed by the conduction electrons of a metal. We also know that the addition of heat increases the amplitude of atomic vibrations. We can therefore guess that the heat capacity of a metal has two terms which are due to energy absorption by the lattice vibrations and energy absorption by conduction electrons. It turns out that at room temperature the energy absorption by lattice vibrations dominates the heat capacity whereas at the lowest temperatures the electronic contribution is important.

[3] In some books (including the first edition of this textbook) it is stated that the electrons at E_{FO} can gain energy from the field and contribute to conduction but not those deep in the band (below b'). This is a simplified statement of the fact that at a level below E_{FO} there is one electron moving along in the $+x$ direction and gaining energy and another one at the same energy but moving along in the $-x$ direction and losing energy so that on average electron at this level does not gain energy.

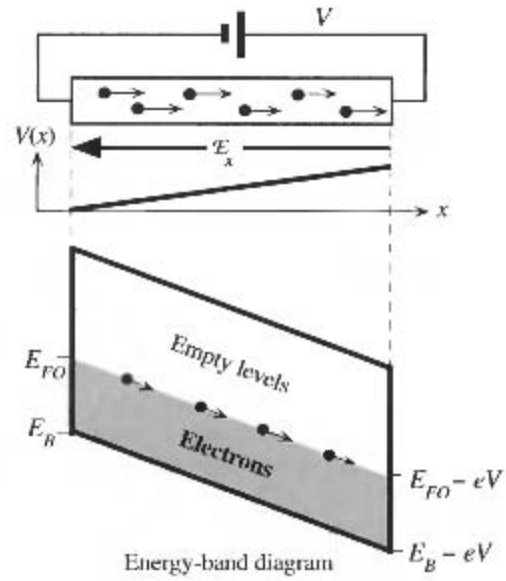


Figure 4.13 Conduction in a metal is due to the drift of electrons around the Fermi level.

When a voltage is applied, the energy band is bent to be lower at the positive terminal so that the electron's potential energy decreases as it moves toward the positive terminal.

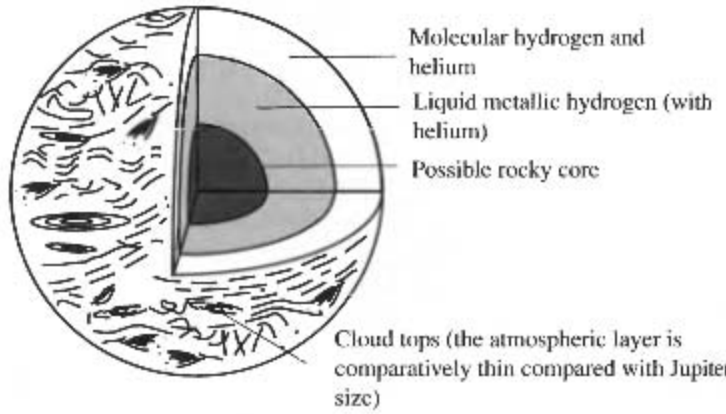


Figure 4.14 The interior of Jupiter is believed to contain liquid hydrogen, which is metallic.

SOURCE: Drawing adapted from T. Hey and P. Walters, *The Quantum Universe*, Cambridge, MA: Cambridge University Press, 1988, p. 96, figure 7.1.

EXAMPLE 4.2 **METALLIC LIQUID HYDROGEN IN JUPITER AND ITS MAGNETIC FIELD** The surface of Jupiter, as visualized schematically in Figure 4.14, mainly consists of a mixture of molecular hydrogen and He gases. Deep in the planet, however, the pressure is so tremendous that the hydrogen molecular bond breaks, leaving a dense ocean of hydrogen atoms. Hydrogen has only one electron in the 1*s* energy level. When atoms are densely packed, the 1*s* energy level forms an energy band, which is then only half filled. This is just like the Li metal, which means we can treat liquid hydrogen as a liquid metal, with electrical properties reminiscent of liquid mercury. Liquid hydrogen can sustain electric currents, which in turn can give rise to the magnetic fields on Jupiter. The origin of the electric currents are not known with certainty. We do know, however, that the core of the planet is hot and emanates heat, which causes convection currents. Temperature differences can readily give rise to electric currents, by virtue of thermoelectric effects, as discussed in Section 4.8.2.

EXAMPLE 4.3

WHAT MAKES A METAL? The Be atom has an electronic structure of $1s^2 2s^2$. Although the Be atom has a full $2s$ energy level, solid Be is a metal. Why?

SOLUTION

We will neglect the K shell ($1s$ state), which is full and very close to the nucleus, and consider only the higher energy states. In the solid, the $2s$ energy level splits into N levels, forming a $2s$ band. With $2N$ electrons, each level is occupied by spin-paired electrons. The $2s$ band is therefore full. However, the empty $2p$ band, from the empty $2p$ energy levels, overlaps the $2s$ band, thereby providing empty energy levels to these $2N$ electrons. Thus, the conduction electrons are in an energy band that is only partially filled; they can gain energy from the field to contribute to electrical conduction. Solid Be is therefore a metal.

EXAMPLE 4.4

FERMI SPEED OF CONDUCTION ELECTRONS IN A METAL In copper, the Fermi energy of conduction electrons is 7.0 eV. What is the speed of the conduction electrons around this energy?

SOLUTION

Since the conduction electrons are not bound to any one atom, their PE must be zero within the solid (but large outside), so all their energy is kinetic. For conduction electrons around the Fermi energy E_{FO} with a speed v_F, we have

$$\frac{1}{2}mv_F^2 = E_{FO}$$

so that

$$v_F = \sqrt{\frac{2E_{FO}}{m_e}} = \sqrt{\frac{2(1.6 \times 10^{-19}\ \text{J/eV})(7.0\ \text{eV})}{(9.1 \times 10^{-31}\ \text{kg})}} = 1.6 \times 10^6\ \text{m s}^{-1}$$

Although the Fermi energy depends on the properties of the energy band, to a good approximation it is only weakly temperature dependent, so v_F will be relatively temperature insensitive, as we will show later in Section 4.7.

4.3 SEMICONDUCTORS

The Si atom has 14 electrons, which distribute themselves in the various atomic energy levels as shown in Figure 4.15. The inner shells ($n = 1$ and $n = 2$) are full and therefore "closed." Since these shells are near the nucleus, when Si atoms come together to form the solid, they are not much affected and they stay around the parent Si atoms. They can therefore be excluded from further discussion. The $3s$ and $3p$ subshells are farther away from the nucleus. When two Si atoms approach, these electrons strongly interact with each other. Therefore, in studying the formation of bands in the Si solid, we will only consider the $3s$ and $3p$ levels.

The first task is to examine why Si actually bonds with four neighbors, since the $3s$ orbital is full and there are only two electrons in the $3p$ orbitals. The full $3s$ orbital should not overlap a neighbor and become involved in bonding. Since only two $3p$ orbitals are half full, bonds should be formed with two neighboring Si atoms. In reality,

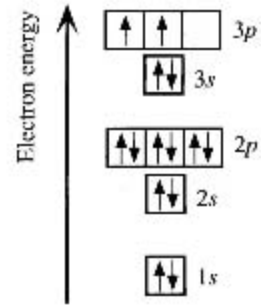
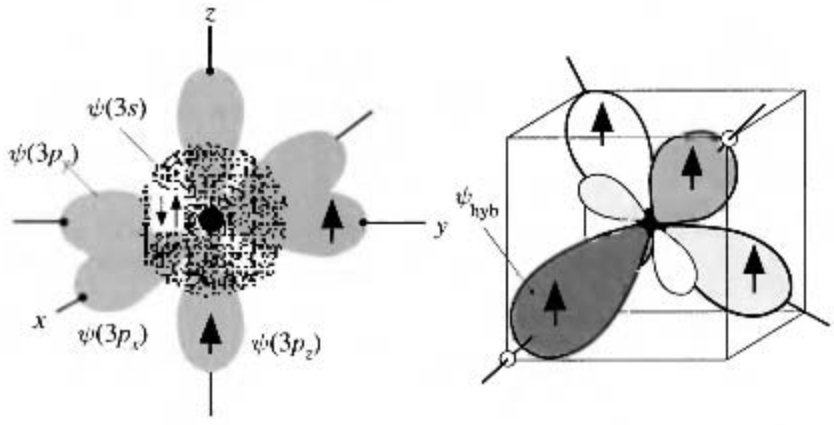


Figure 4.15 The electronic structure of Si.



(a) Isolated Si (b) Si just before bonding

Figure 4.16

(a) Si is in Group IV in the Periodic Table. An isolated Si atom has two electrons in the 3*s* and two electrons in the 3*p* orbitals.

(b) When Si is about to bond, the one 3*s* orbital and the three 3*p* orbitals become perturbed and mixed to form four hybridized orbitals, ψ_{hyb}, called sp^3 orbitals, which are directed toward the corners of a tetrahedron. The ψ_{hyb} orbital has a large major lobe and a small back lobe. Each ψ_{hyb} orbital takes one of the four valence electrons.

the 3*s* and 3*p* energy levels are quite close, and when five Si atoms approach each other, the interaction results in the four orbitals $\psi(3s)$, $\psi(3p_x)$, $\psi(3p_y)$, and $\psi(3p_z)$ mixing together to form four new **hybrid orbitals,** which are directed in tetrahedral directions; that is, each one is aimed as far away from the others as possible, as illustrated in Figure 4.16. We call this process sp^3 **hybridization,** since one *s* orbital and three *p* orbitals are mixed. (The superscript 3 on *p* has nothing to do with the number of electrons; it refers to the number of *p* orbitals used in the hybridization.)

The four sp^3 hybrid orbitals, ψ_{hyb}, each have one electron, so they are half occupied. This means that four Si atoms can have their orbitals ψ_{hyb} overlap to form bonds with one Si atom, which is what actually happens; thus, one Si atom bonds with four other Si atoms in tetrahedral directions.

In the same way, one Si atom bonds with four H atoms to form the important gas SiH_4, known as silane, which is widely used in the semiconductor technology to fabricate Si devices. In SiH_4, four hybridized orbitals of the Si atom overlap with the $1s$ orbitals of four H atoms. In exactly the same way, one carbon atom bonds with four hydrogen atoms to form methane, CH_4.

There are two ways in which the hybrid orbital ψ_{hyb} can overlap with that of the neighboring Si atom to form two molecular orbitals. They can add in phase (both positive or both negative) or out of phase (one positive and the other negative) to produce a bonding or an antibonding molecular orbital ψ_B and ψ_A, respectively, with energies E_B and E_A. Each Si–Si bond thus corresponds to two paired electrons in a bonding molecular orbital ψ_B. In the solid, there are $N(\sim 5 \times 10^{22}\,\text{cm}^{-3})$ Si atoms, and there are nearly as many such ψ_B bonds. The interactions between the ψ_B orbitals (*i.e.*, the Si–Si bonds) lead to the splitting of the E_B energy level to N levels, thereby forming an energy band labeled the **valence band** (VB) by virtue of the valence electrons it contains. Since the energy level E_B is full, so is the valence band. Figure 4.17 illustrates the formation of the VB from E_B.

In the solid, the interactions between the N number of ψ_A orbitals result in the splitting of the energy level E_A to N levels and the formation of an energy band that is

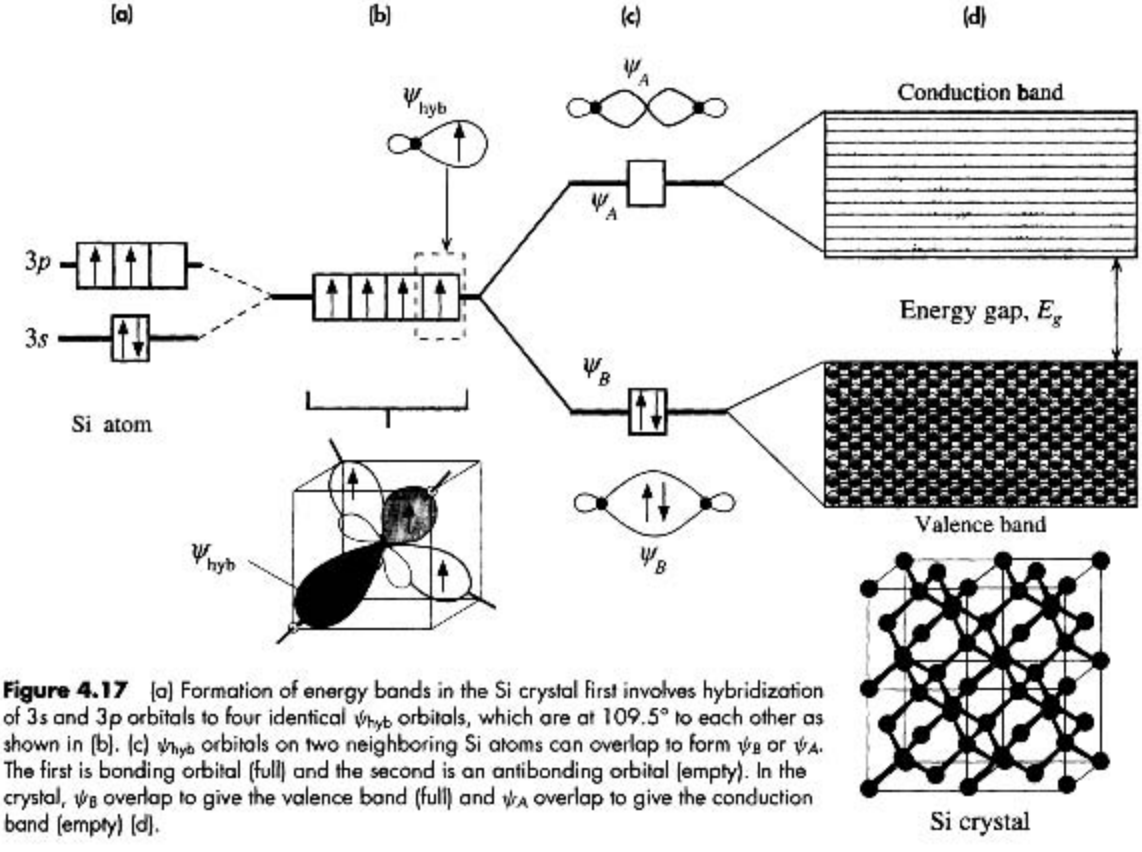


Figure 4.17 (a) Formation of energy bands in the Si crystal first involves hybridization of 3s and 3p orbitals to four identical ψ_{hyb} orbitals, which are at 109.5° to each other as shown in (b). (c) ψ_{hyb} orbitals on two neighboring Si atoms can overlap to form ψ_B or ψ_A. The first is bonding orbital (full) and the second is an antibonding orbital (empty). In the crystal, ψ_B overlap to give the valence band (full) and ψ_A overlap to give the conduction band (empty) (d).

completely empty and separated from the full valence band by a definite energy gap E_g. In this energy region, there are no states; therefore, the electron cannot have energy with a value within E_g. The energy band formed from $N\psi_A$ orbitals is a **conduction band** (CB), as also indicated in Figure 4.17.

The electronic states in the VB (and also in the CB) extend throughout the whole solid, because they result from $N\psi_B$ orbitals interfering and overlapping each other. As before $N\psi_B$, orbitals can overlap in N different ways to produce N distinct wavefunctions ψ_{vb} that extend throughout the solid. We cannot relate a particular electron to a particular bond or site because the wavefunctions ψ_{vb} corresponding to the VB energies are not concentrated at a single location. The electrical properties of solids are based on the fact that in solids, such as semiconductors and insulators, there are certain bands of allowed energies for the electrons, and these bands are separated by energy gaps, that is, bandgaps. The valence and conduction bands for the ideal Si crystal shown in Figure 4.17 are separated by an **energy gap,** or a **bandgap,** E_g, in which there are no allowed electron energy levels.

At temperatures above absolute zero, the atoms in a solid vibrate due to their thermal energy. Some of the atoms can acquire a sufficiently high energy from thermal fluctuations to strain and rupture their bonds. Physically, there is a possibility that the atomic vibration will impart sufficient energy to the electron for it to surmount the bonding energy and leave the bond. The electron must then enter a higher energy state. In the case of Si, this means entering a state in the CB, as shown in Figure 4.18. If there is an applied electric field $\mathcal{E}_x$ in the $+x$ direction, then the excited electron will be acted on by a force $-e\mathcal{E}_x$ and it will try to move in the $-x$ direction. For it to do so, there must be empty higher energy levels, so that as the electron accelerates and gains energy, it moves up in the band. When an electron collides with a lattice vibration, it loses the energy acquired from the field and drops down within the CB. Again, it should be emphasized that states in an energy band are extended; that is, the electron is not localized to any one atom.

Note also that the thermal generation of an electron from the VB to the CB leaves behind a VB state with a missing electron. This unoccupied electron state has an apparent positive charge, because this crystal region was neutral prior to the removal of the electron. The VB state with the missing electron is called a **hole** and is denoted h^+. The hole can "move" in the direction of the field by exchanging places with a

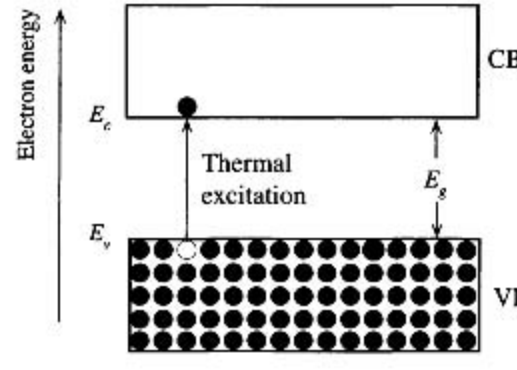


Figure 4.18 Energy band diagram of a semiconductor.
CB is the conduction band and VB is the valence band. At 0 K, the VB is full with all the valence electrons.

neighboring valence electron hence it contributes to conduction, as will be discussed in Chapter 5.

EXAMPLE 4.5

CUTOFF WAVELENGTH OF A Si PHOTODETECTOR What wavelengths of light can be absorbed by a Si photodetector given $E_g = 1.1$ eV? Can such a photodetector be used in fiber-optic communications at light wavelengths of 1.31 μm and 1.55 μm?

SOLUTION

The energy bandgap E_g of Si is 1.1 eV. A photon must have at least this much energy to excite an electron from the VB to the CB, where the electron can drift. Excitation corresponds to the breaking of a Si–Si bond. A photon of less energy does not get absorbed, because its energy will put the electron in the bandgap where there are no states. Thus, $hc/\lambda > E_g$ gives

$$\lambda < \frac{hc}{E_g} = \frac{(6.6 \times 10^{-34}\ \text{J s})(3 \times 10^{8}\ \text{m s}^{-1})}{(1.1\ \text{eV})(1.6 \times 10^{-19}\ \text{J/eV})}$$

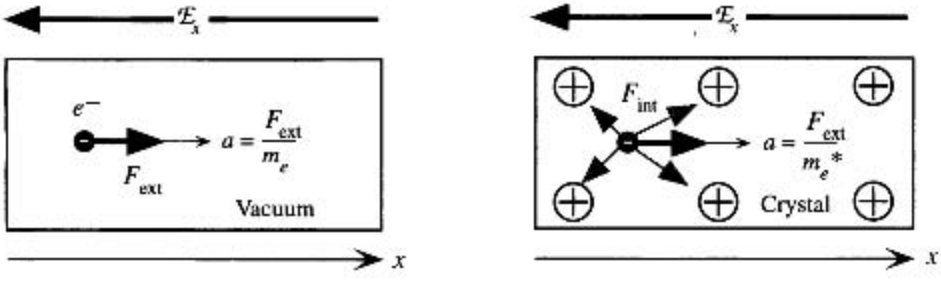
$$= 1.13 \times 10^{-6}\ \text{m} \quad \text{or} \quad 1.1\ \mu\text{m}$$

Since optical communications networks use wavelengths of 1.3 and 1.55 μm, these light waves will not be absorbed by Si and thus cannot be detected by a Si photodetector.

4.4 ELECTRON EFFECTIVE MASS

When an electric field $\mathcal{E}_x$ is applied to a metal, an electron near the Fermi level can gain energy from the field and move to higher energy levels, as shown in Figure 4.12. The external force $F_{ext} = e\mathcal{E}_x$ is in the x direction, and it drives the electron along x. The acceleration of the electron is still given by $a = F_{ext}/m_e$, where m_e is the mass of the electron in vacuum.

The law $F_{ext} = m_e a$ cannot strictly be valid for the electron inside a solid, because the electron interacts with the host ions and experiences internal forces F_{int} as it moves around, as depicted in Figure 4.19. The electron therefore has a PE that varies with distance. Recall that we interpret mass as inertial resistance against acceleration per unit



(a) An external force F_{ext} applied to an electron in a vacuum results in an acceleration $a_{vac} = F_{ext}/m_e$.

(b) An external force F_{ext} applied to an electron in a crystal results in an acceleration $a_{cryst} = F_{ext}/m_e^*$.

Figure 4.19

applied force. When an external force F_{ext} is applied to an electron in the vacuum level, as in Figure 4.19a, the electron will accelerate by an amount

$$a_{vac} = \frac{F_{ext}}{m_e} \quad [4.4]$$

as determined by its mass m_e in vacuum.

When the same force F_{ext} is applied to the electron inside a crystal, the acceleration of the electron will be different, because it will also experience internal forces, as shown in Figure 4.19b. Its acceleration in the crystal will be

$$a_{cryst} = \frac{F_{ext} + F_{int}}{m_e} \quad [4.5]$$

where F_{int} is the sum of all the internal forces acting on the electron, which is quite different than Equation 4.4. To the outside agent applying the force F_{ext}, the electron will appear to be exhibiting a different inertial mass, since its acceleration will be different. It would be most useful for the external agent if the effect of the internal forces in F_{int} could be accounted for in a simple way, and if the acceleration could be calculated from the external force F_{ext} alone, through something like Equation 4.4. This is indeed possible.

In a crystalline solid, the atoms are arranged periodically, and the variation of F_{int}, and hence the *PE*, or $V(x)$, of the electron with distance along x, is also periodic. In principle, then, the effect on the electron motion can be predicted and accounted for. When we solve the Schrödinger equation with the periodic *PE*, or $V(x)$, we essentially obtain the effect of these internal forces on the electron motion. It has been found that when the electron is in a band that is not full, we can still use Equation 4.4, but instead of the mass in vacuum m_e, we must use the effective mass m_e^* of the electron in that particular crystal. The effective mass is a quantum mechanical quantity that behaves in the same way as the inertial mass in classical mechanics. The acceleration of the electron in the crystal is then simply

$$a_{cryst} = \frac{F_{ext}}{m_e^*} \quad [4.6]$$

The effects of all internal forces are incorporated into m_e^*. It should be emphasized that m_e^*, is obtained theoretically from the solution of the Schrödinger equation for the electron in a particular crystal, a task that is by no means trivial. However, the effective mass can be readily measured. For some of the familiar metals, m_e^* is very close to m_e. For example, in copper, $m_e^* = m_e$ for all practical purposes, whereas in lithium $m_e^* = 1.28m_e$, as shown in Table 4.2. On the other hand, m_e^* for many metals and

Table 4.2 Effective mass m_e^* of electrons in some metals

Metal	Ag	Au	Bi	Cu	K	Li	Na	Ni	Pt	Zn
$\frac{m_e^*}{m_e}$	0.99	1.10	0.047	1.01	1.12	1.28	1.2	28	13	0.85

semiconductors is appreciably different than the electron mass in vacuum and can even be negative. (m_e^* depends on the properties of the band that contains the electron. This is further discussed in Section 5.11.)

4.5 DENSITY OF STATES IN AN ENERGY BAND

Although we know there are many energy levels (perhaps $\sim 10^{23}$) in a given band, we have not yet considered how many states (or electron wavefunctions) there are per unit energy per unit volume in that band. Consider the following *intuitive* argument. The crystal will have N atoms and there will be N electron wavefunctions $\psi_1, \psi_2, \ldots, \psi_N$ that represent the electron within the whole crystal. These wavefunctions are constructed from N different combinations of atomic wavefunctions, $\psi_A, \psi_B, \psi_C, \ldots$ as schematically illustrated in Figure 4.20a,[4] starting with

$$\psi_1 = \psi_A + \psi_B + \psi_C + \psi_D + \cdots$$

all the way to alternating signs

$$\psi_N = \psi_A - \psi_B + \psi_C - \psi_D + \cdots$$

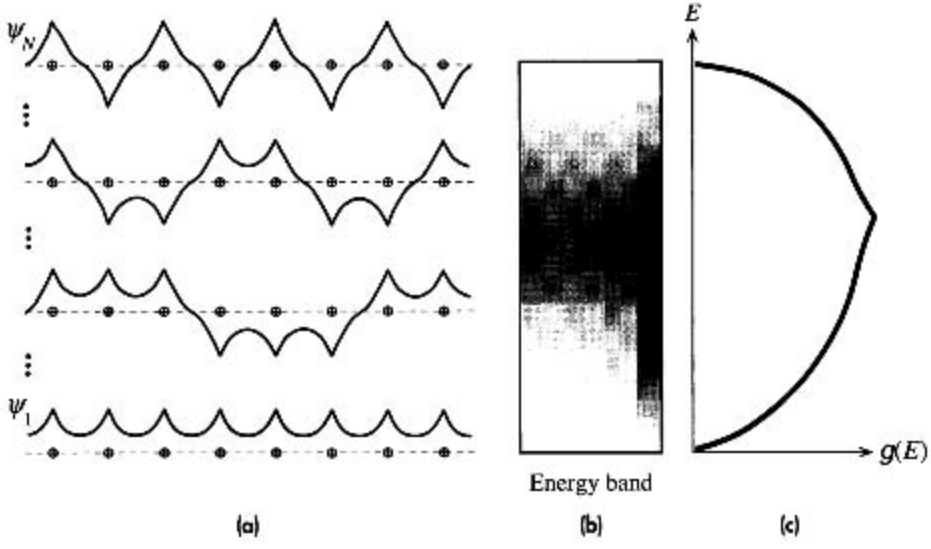


Figure 4.20

(a) In the solid there are N atoms and N extended electron wavefunctions from ψ_1 all the way to ψ_N. There are many wavefunctions, states, that have energies that fall in the central regions of the energy band.

(b) The distribution of states in the energy band; darker regions have a higher number of states.

(c) Schematic representation of the density of states $g(E)$ versus energy E.

[4] This intuitive argument, as schematically depicted in Figure 4.20a, is obviously highly simplified because the solid is three-dimensional (3-D) and we should combine the atomic wavefunctions not on a linear chain but on a 3-D lattice. In the 3-D case there are large numbers of wavefunctions with energies that fall in the central regions of the band.

and there are $N(\sim 10^{23})$ combinations. The lowest-energy wavefunction will be ψ_1 constructed by adding all atomic wavefunctions (all in phase), and the highest-energy wavefunction will be ψ_N from alternating the signs of the atomic wavefunctions, which will have the highest number of nodes. Between these two extremes, especially around $N/2$, there will be many combinations that will have comparable energies and fall near the middle of the band. (By analogy, if we arrange $N = 10$ coins by heads and tails, there will be many combinations of coins in which there are 5 heads and 5 tails, and only one combination in which there are 10 heads or 10 tails.) We therefore expect the number of energy levels, each corresponding to an electron wavefunction in the crystal, in the central regions of the band to be very large as depicted in Figure 4.20b and c.

Figure 4.20c illustrates schematically how the energy and volume density of electronic states change across an energy band. We define the **density of states** $g(E)$ such that $g(E)\,dE$ is the number of states (*i.e.*, wavefunctions) in the energy interval E to $(E + dE)$ per unit volume of the sample. Thus, the number of states per unit volume up to some energy E' is

$$S_v(E') = \int_0^{E'} g(E)\,dE \qquad [4.7]$$

which is called the total number of states per unit volume with energies less than E'. This is denoted $S_v(E')$.

To determine the density of states function $g(E)$, we must first determine the number of states with energies less than E' in a given band. This is tantamount to calculating $S_v(E')$ in Equation 4.7. Instead, we will improvise and use the energy levels for an electron in a three-dimensional potential well. Recall that the energy of an electron in a cubic *PE* well of size L is given by

$$E = \frac{h^2}{8m_e L^2}\left(n_1^2 + n_2^2 + n_3^2\right) \qquad [4.8]$$

where n_1, n_2, and n_3 are integers 1, 2, 3, ... The spatial dimension L of the well now refers to the size of the entire solid, as the electron is confined to be somewhere inside that solid. Thus, L is very large compared to atomic dimensions, which means that the separation between the energy levels is very small. We will use Equation 4.8 to describe the energies of **free electrons** inside the solid (as in a metal).

Each combination of n_1, n_2, and n_3 is one electron orbital state. For example, $\psi_{n_1,n_2,n_3} = \psi_{1,1,2}$ is one possible orbital state. Suppose that in Equation 4.8 E is given as E'. We need to determine how many combinations of n_1, n_2, n_3 (*i.e.*, how many ψ) have energies less than E', as given by Equation 4.8. Assume that $(n_1^2 + n_2^2 + n_3^2) = n'^2$. The object is to enumerate all possible choices of integers for n_1, n_2, and n_3 that satisfy $n_1^2 + n_2^2 + n_3^2 \leq n'^2$.

The two-dimensional case is easy to solve. Consider $n_1^2 + n_2^2 \leq n'^2$ and the two-dimensional ***n*-space** where the axes are n_1 and n_2, as shown in Figure 4.21. The two-dimensional space is divided by lines drawn at $n_1 = 1, 2, 3, \ldots$ and $n_2 = 1, 2, 3, \ldots$ into infinitely many boxes (squares), each of which has a unit area and represents a possible state ψ_{n_1,n_2}. For example, the state $n_1 = 1, n_2 = 3$ is shaded, as is that for $n_1 = 2, n_2 = 2$.

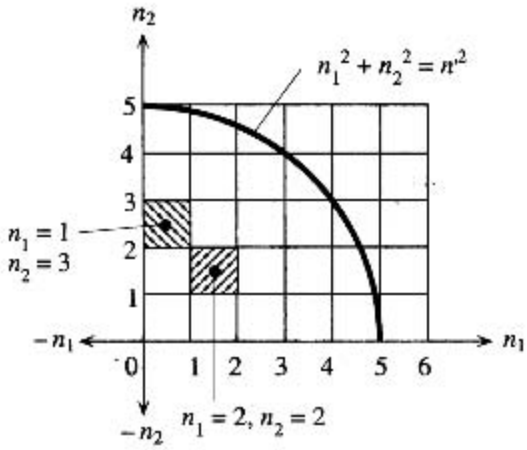


Figure 4.21 Each state, or electron wavefunction in the crystal, can be represented by a box at n_1, n_2.

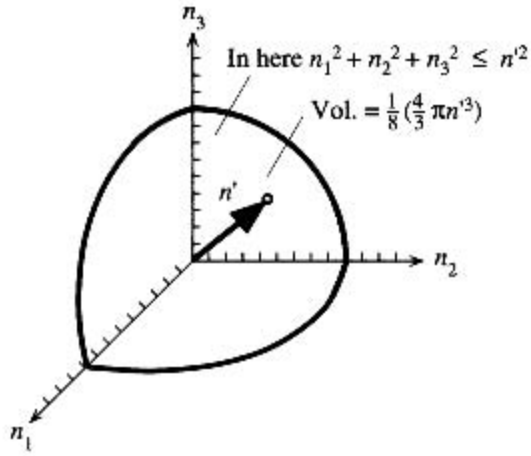


Figure 4.22 In three dimensions, the volume defined by a sphere of radius n' and the positive axes n_1, n_2, and n_3, contains all the possible combinations of positive n_1, n_2, and n_3 values that satisfy $n_1^2 + n_2^2 + n_3^2 \leq n'^2$.

Clearly, the area contained by n_1, n_2 and the circle defined by $n'^2 = n_1^2 + n_2^2$ (just like $r^2 = x^2 + y^2$) is the number of states that satisfy $n_1^2 + n_2^2 \leq n'^2$. This area is $\frac{1}{4}(\pi n'^2)$.

In the three-dimensional case, $n_1^2 + n_2^2 + n_3^2 \leq n'^2$ is required, as indicated in Figure 4.22. This is the volume contained by the positive n_1, n_2, and n_3 axes and the surface of a sphere of radius n'. Each state has a unit volume, and within the sphere, $n_1^2 + n_2^2 + n_3^2 \leq n'^2$ is satisfied. Therefore, the number of orbital states $S_{\text{orb}}(n')$ within this volume is given by

$$S_{\text{orb}}(n') = \frac{1}{8}\left(\frac{4}{3}\pi n'^3\right) = \frac{1}{6}\pi n'^3$$

Each orbital state can take two electrons with opposite spins, which means that the number of states, including spin, is given by

$$S(n') = 2S_{\text{orb}}(n') = \frac{1}{3}\pi n'^3$$

We need this expression in terms of energy. Substituting $n'^2 = 8m_e L^2 E'/h^2$ from Equation 4.8 in $S(n')$, we get

$$S(E') = \frac{\pi L^3 (8m_e E')^{3/2}}{3h^3}$$

Since L^3 is the physical volume of the solid, the number of states per unit volume $S_v(E')$ with energies $E \leq E'$ is

$$S_v(E') = \frac{\pi (8m_e E')^{3/2}}{3h^3} \qquad [4.9]$$

Furthermore, from Equation 4.7, $dS_v/dE = g(E)$. By differentiating Equation 4.9 with respect to energy, we get

Density of states

$$g(E) = (8\pi 2^{1/2})\left(\frac{m_e}{h^2}\right)^{3/2} E^{1/2} \qquad [4.10]$$

Equation 4.10 shows that the density of states $g(E)$ increases with energy as $E^{1/2}$ from the bottom of the band. As we approach the top of the band, according to our understanding in Figure 4.20d, $g(E)$ should decrease with energy as $(E_{top} - E)^{1/2}$, where E_{top} is the top of the band, so that as $E \rightarrow E_{top}$, $g(E) \rightarrow 0$. The electron mass m_e in Equation 4.10 should be the *effective mass* m_e^* as in Equation 4.6. Further, Equation 4.10 strictly applies only to *free electrons* in a crystal. However, we will frequently use it to approximate the true $g(E)$ versus E behavior near the band edges for both metals and semiconductors.

Having found the distribution of the electron energy states, Equation 4.10, we now wish to determine the number of states that actually contain electrons; that is, the probability of finding an electron at an energy level E. This is given by the Fermi–Dirac statistics.

As an example, one convenient way of calculating the population of a city is to find the density of houses in that city (*i.e.*, the number of houses per unit area), multiply that by the probability of finding a human in a house, and finally, integrate the result over the area of the city. The problem is working out the chances of actually finding someone at home, using a mathematical formula. For those who like analogies, if $g(A)$ is the density of houses and $f(A)$ is the probability that a house is occupied, then the population of the city is

$$n = \int_{\text{City}} f(A)g(A)\, dA$$

where the integration is done over the entire area of the city. This equation can be used to find the number of electrons per unit volume within a band. If E is the electron energy and $f(E)$ is the probability that a state with energy E is occupied, then

$$n = \int_{\text{Band}} f(E)g(E)\, dE$$

where the integration is done over all the energies of the band.

EXAMPLE 4.6 **X-RAY EMISSION AND THE DENSITY OF STATES IN A METAL** Consider what happens when a metal such as Al is bombarded with high-energy electrons. The inner atomic energy levels are not disturbed in the solid, so these inner levels remain as distinct single levels, each one localized to the parent atom. When an energetic electron hits an electron in one of the inner atomic energy levels, it knocks out this electron from the metal leaving behind a vacancy in the inner core as depicted in Figure 4.23a. An electron in the energy band of the solid can then fall down to occupy this empty state and emit a photon in the process. The energy difference between the energies in the band and the inner atomic level is in the X-ray range, so the emitted photon is an X-ray photon. Since electrons occupy the band from the bottom E_B to the Fermi level E_F, the

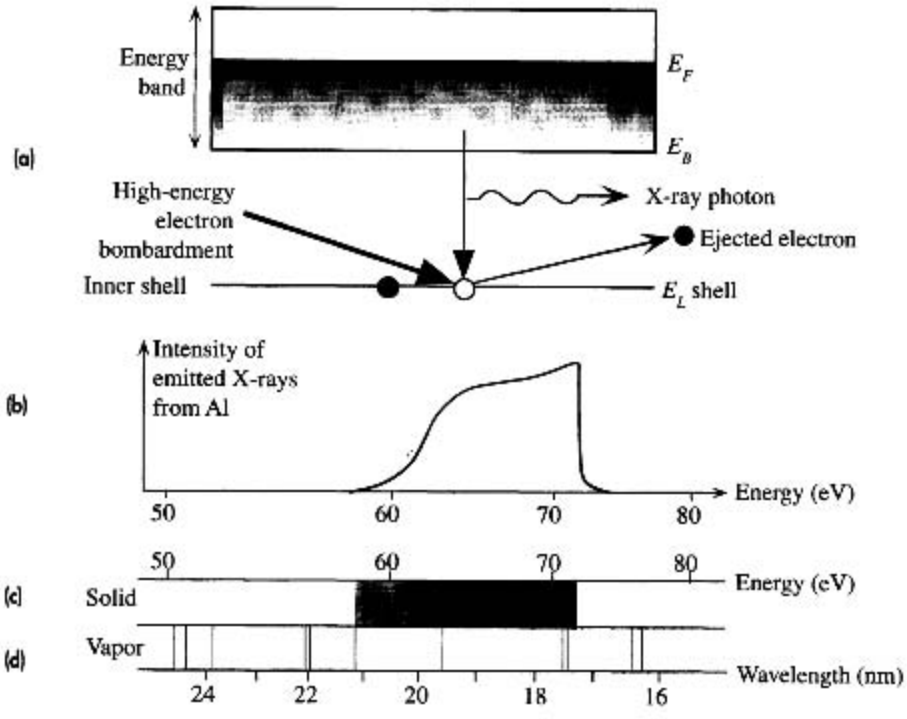


Figure 4.23

(a) High-energy electron bombardment knocks out an electron from the closed inner L shell leaving an empty state. An electron from the energy band of the metal drops into the L shell to fill the vacancy and emits a soft X-ray photon in the process.

(b) The spectrum (intensity versus photon energy) of soft X-ray emission from a metal involves a range of energies corresponding to transitions from the bottom of the band and from the Fermi level to the L shell. The intensity increases with energy until around E_F where it drops sharply.

(c) and (d) contrast the emission spectra from a solid and vapor (isolated gas atoms).

emitted X-ray photons have a range of energies corresponding to transitions from E_B and E_F to the inner atomic level as shown in Figure 4.23b. These energies are in the soft X-ray spectrum. We assumed that the levels above E_F are almost empty, though, undoubtedly, there is no sharp transition from full to empty levels at E_F. Further, since the density of states increases from E_B toward E_F, there are more and more electrons that can fall down to the atomic level as we move from E_B toward E_F. Therefore the intensity of the emitted X-ray radiation increases with energy until the energy reaches the Fermi level beyond which there are only a small number of electrons available for the transit. Figure 4.23c and d contrasts the emission spectra from an aluminum crystal (solid) and its vapor. The line spectra from a vapor become an emission band in the spectrum of the solid.

The X-ray intensity emitted from Al in Figure 4.23 starts to rise at around 60 eV and then sharply falls around 72 eV. Thus the energy range is 12 eV, which represents approximately the Fermi energy with respect to the bottom of the band, that is, $E_F \approx 72 - 60 = 12$ eV with respect to E_B.

EXAMPLE 4.7 **DENSITY OF STATES IN A BAND** Given that the width of an energy band is typically ~10 eV, calculate the following, in per cm^3 and per eV units:

a. The density of states at the center of the band.

b. The number of states per unit volume within a small energy range kT about the center.

c. The density of states at kT above the bottom of the band.

d. The number of states per unit volume within a small energy range of kT to $2kT$ from the bottom of the band.

SOLUTION

The density of states, or the number of states per unit energy range per unit volume $g(E)$, is given by

$$g(E) = (8\pi 2^{1/2})\left(\frac{m_e}{h^2}\right)^{3/2} E^{1/2}$$

which gives the number of states per cubic meter per Joule of energy. Substituting $E = 5$ eV, we have

$$g_{\text{center}} = (8\pi 2^{1/2})\left[\frac{9.1 \times 10^{-31}}{(6.626 \times 10^{-34})^2}\right]^{3/2} (5 \times 1.6 \times 10^{-19})^{1/2} = 9.50 \times 10^{46}\ \text{m}^{-3}\,\text{J}^{-1}$$

Converting to cm^{-3} and eV^{-1}, we get

$$\begin{aligned} g_{\text{center}} &= (9.50 \times 10^{46}\ \text{m}^{-3}\,\text{J}^{-1})(10^{-6}\ \text{m}^3\,\text{cm}^{-3})(1.6 \times 10^{-19}\ \text{J eV}^{-1}) \\ &= 1.52 \times 10^{22}\ \text{cm}^{-3}\,\text{eV}^{-1} \end{aligned}$$

If δE is a small energy range (such as kT), then, by definition, $g(E)\,\delta E$ is the number of states per unit volume in δE. To find the number of states per unit volume within kT at the center of the band, we multiply g_{center} by kT or $(1.52 \times 10^{22}\ \text{cm}^{-3}\,\text{eV}^{-1})(0.026\ \text{eV})$ to get $3.9 \times 10^{20}\ \text{cm}^{-3}$. This is not a small number!

At kT above the bottom of the band, at 300 K ($kT = 0.026$ eV), we have

$$\begin{aligned} g_{0.026} &= (8\pi 2^{1/2})\left[\frac{9.1 \times 10^{-31}}{(6.626 \times 10^{-34})^2}\right]^{3/2} (0.026 \times 1.6 \times 10^{-19})^{1/2} \\ &= 6.84 \times 10^{45}\ \text{m}^{-3}\,\text{J}^{-1} \end{aligned}$$

Converting to cm^{-3} and eV^{-1} we get

$$\begin{aligned} g_{0.026} &= (6.84 \times 10^{45}\ \text{m}^{-3}\,\text{J}^{-1})(10^{-6}\ \text{m}^3\,\text{cm}^{-3})(1.6 \times 10^{-19}\ \text{J eV}^{-1}) \\ &= 1.10 \times 10^{21}\ \text{cm}^{-3}\,\text{eV}^{-1} \end{aligned}$$

Within kT, the volume density of states is

$$(1.10 \times 10^{21}\ \text{cm}^{-3}\,\text{eV}^{-1})(0.026\ \text{eV}) = 2.8 \times 10^{19}\ \text{cm}^{-3}$$

This is very close to the bottom of the band and is still very large.

TOTAL NUMBER OF STATES IN A BAND

EXAMPLE 4.8

a. Based on the overlap of atomic orbitals to form the electron wavefunction in the crystal, how many states should there be in a band?

b. Consider the density of states function

$$g(E) = (8\pi 2^{1/2})\left(\frac{m_e}{h^2}\right)^{3/2} E^{1/2}$$

By integrating $g(E)$, estimate the total number of states in a band per unit volume, and compare this with the atomic concentration for silver. For silver, we have $E_{FO} = 5.5$ eV and $\Phi = 4.5$ eV. (Note that "state" means a distinct wavefunction, including spin.)

SOLUTION

a. We know that when N atoms come together to form a solid, N atomic orbitals can overlap N different ways to produce N orbitals or $2N$ states in the crystal, since each orbital has two states, spin up and spin down. These states form the band.

b. For silver, $E_{FO} = 5.5$ eV and $\Phi = 4.5$ eV, so the width of the energy band is 10 eV. To estimate the total volume density of states, we assume that the density of states $g(E)$ reaches its maximum at the center of the band $E = E_{\text{center}} = 5$ eV. Integrating $g(E)$ from the bottom of the band, $E = 0$, to the center, $E = E_{\text{center}}$, yields the number of states per unit volume up to the center of the band. This is half the total number of states in the whole band, that is, $\frac{1}{2}S_{\text{band}}$, where S_{band} is the number of states per unit volume in the band and is determined by

$$\frac{1}{2}S_{\text{band}} = \int_0^{E_{\text{center}}} g(E)\,dE = \frac{16\pi 2^{1/2}}{3}\left(\frac{m_e}{h^2}\right)^{3/2} E_{\text{center}}^{3/2}$$

or

$$\frac{1}{2}S_{\text{band}} = \frac{16\pi 2^{1/2}}{3}\left[\frac{9.1 \times 10^{-31}\ \text{kg}}{(6.626 \times 10^{-34}\ \text{J s})^2}\right]^{3/2} (5\ \text{eV} \times 1.6 \times 10^{-19}\ \text{J/eV})^{3/2}$$

$$= 5.08 \times 10^{28}\ \text{m}^{-3} = 5.08 \times 10^{22}\ \text{cm}^{-3}$$

Thus

$$S_{\text{band}} = 10.16 \times 10^{22}\ \text{states cm}^{-3}$$

We must now calculate the number of atoms per unit volume in silver. Given the density $d = 10.5$ g cm^{-3} and the atomic mass $M_{\text{at}} = 107.9$ g mol^{-1} of silver, the atomic concentration is

$$n_{\text{Ag}} = \frac{dN_A}{M_{\text{at}}} = 5.85 \times 10^{22}\ \text{atoms cm}^{-3}$$

As expected, the density of states is almost twice the atomic concentration, even though we used a crude approximation to estimate the density of states.

4.6 STATISTICS: COLLECTIONS OF PARTICLES

4.6.1 BOLTZMANN CLASSICAL STATISTICS

Given a collection of particles in random motion and colliding with each other,[5] we need to determine the concentration of particles in the energy range E to $(E + dE)$. Consider the process shown in Figure 4.24, in which two electrons with energies E_1 and E_2 interact and then move off in different directions, with energies E_3 and E_4. Let the probability of an electron having an energy E be $P(E)$, where $P(E)$ is the fraction of electrons with an energy E. Assume there are no restrictions to the electron energies, that is, we can ignore the Pauli exclusion principle. The probability of this event is then $P(E_1)P(E_2)$. The probability of the reverse process, in which electrons with energies E_3 and E_4 interact, is $P(E_3)P(E_4)$. Since we have thermal equilibrium, that is, the system is in equilibrium, the forward process must be just as likely as the reverse process, so

$$P(E_1)P(E_2) = P(E_3)P(E_4) \qquad [4.11]$$

Furthermore, the energy in this collision must be conserved, so we also need

$$E_1 + E_2 = E_3 + E_4 \qquad [4.12]$$

We therefore need to find the $P(E)$ that satisfies both Equations 4.11 and 4.12. Based on our experience with the distribution of energies among gas molecules, we can guess that the solution for Equations 4.11 and 4.12 would be

Boltzmann probability function

$$P(E) = A \exp\left(-\frac{E}{kT}\right) \qquad [4.13]$$

where k is the Boltzmann constant, T is the temperature, and A is a constant. We can show that Equation 4.13 is a solution to Equations 4.11 and 4.12 by a simple substitution. Equation 4.13 is the **Boltzmann probability function** and is shown in Figure 4.25. The probability of finding a particle at an energy E therefore decreases exponentially with energy. We assume, of course, that any number of particles may have a given energy E. In other words, there is no restriction such as permitting only one particle per state at an energy E, as in the Pauli exclusion principle. The term kT appears in Equation 4.13 because the average energy as calculated by using $P(E)$ then agrees with experiments. (There is no kT in Equations 4.11 and 4.12.)

Suppose that we have N_1 particles at energy level E_1 and N_2 particles at a higher energy E_2. Then, by Equation 4.13, we have

Boltzmann statistics

$$\frac{N_2}{N_1} = \exp\left(-\frac{E_2 - E_1}{kT}\right) \qquad [4.14]$$

[5] From Chapter 1, we can associate this with the kinetic theory of gases. The energies of the gas molecules, which are moving around randomly, are distributed according to the Maxwell–Boltzmann statistics.

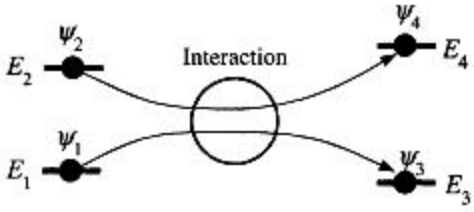


Figure 4.24 Two electrons with initial wavefunctions ψ_1 and ψ_2 at E_1 and E_2 interact and end up at different energies E_3 and E_4. Their corresponding wavefunctions are ψ_3 and ψ_4.

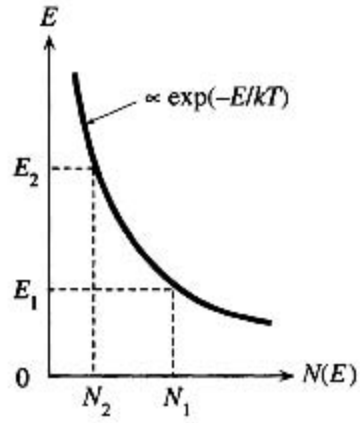


Figure 4.25 The Boltzmann energy distribution describes the statistics of particles, such as electrons, when there are many more available states than the number of particles.

If $E_2 - E_1 \gg kT$, then N_2 can be orders of magnitude smaller than N_1. As the temperature increases, N_2/N_1 also increases. Therefore, increasing the temperature populates the higher energy levels.

Classical particles obey the Boltzmann statistics. Whenever there are many more states (by orders of magnitude) than the number of particles, the likelihood of two particles having the same set of quantum numbers is negligible and we do not have to worry about the Pauli exclusion principle. In these cases, we can use the Boltzmann statistics. An important example is the statistics of electrons in the conduction band of a semiconductor where, in general, there are many more states than electrons.

4.6.2 FERMI–DIRAC STATISTICS

Now consider the interaction for which no two electrons can be in the same quantum state, which is essentially obedience to the Pauli exclusion principle, as shown in Figure 4.24. We assume that we can have only one electron in a particular quantum state ψ (including spin) associated with the energy value E. We therefore need those states that have energies E_3 and E_4 to be not occupied. Let $f(E)$ be the probability that an electron is in such a state, with energy E in this new interaction environment. The probability of the forward event in Figure 4.24 is

$$f(E_1)f(E_2)[1 - f(E_3)][1 - f(E_4)]$$

The square brackets represent the probability that the states with energies E_3 and E_4 are empty. In thermal equilibrium, the reverse process, the electrons with E_3 and E_4 interacting to transfer to E_1 and E_2, has just as equal a likelihood as the forward process.

Paul Adrien Maurice Dirac (1902–1984) received the 1933 Nobel prize for physics with Erwin Schrödinger. His first degree was in electrical engineering from Bristol University. He obtained his PhD in 1926 from Cambridge University under Ralph Fowler.

SOURCE: Courtesy of AIP Emilio Segrè Visual Archives.

Thus, $f(E)$ must satisfy the equation

$$f(E_1)f(E_2)[1 - f(E_3)][1 - f(E_4)] = f(E_3)f(E_4)[1 - f(E_1)][1 - f(E_2)] \quad [4.15]$$

In addition, for energy conservation, we must have

$$E_1 + E_2 = E_3 + E_4 \quad [4.16]$$

By an "intelligent guess," the solution to Equations 4.15 and 4.16 is

$$f(E) = \frac{1}{1 + A\exp\left(\dfrac{E}{kT}\right)} \quad [4.17]$$

where A is a constant. You can check that this is a solution by substituting Equation 4.17 into 4.15 and using Equation 4.16. The reason for the term kT in Equation 4.17 is not obvious from Equations 4.15 and 4.16. It appears in Equation 4.17 so that the mean properties of this system calculated by using $f(E)$ agree with experiments. Letting $A = \exp(-E_F/kT)$, we can write Equation 4.17 as

Fermi–Dirac statistics

$$f(E) = \frac{1}{1 + \exp\left(\dfrac{E - E_F}{kT}\right)} \quad [4.18]$$

where E_F is a constant called the **Fermi energy.** The probability of finding an electron in a state with energy E is given by Equation 4.18, which is called the **Fermi–Dirac function.**

The behavior of the Fermi–Dirac function is shown in Figure 4.26. Note the effect of temperature. As T increases, $f(E)$ extends to higher energies. At energies of a few kT (0.026 eV) above E_F, $f(E)$ behaves almost like the Boltzmann function

$$f(E) = \exp\left[-\frac{(E - E_F)}{kT}\right] \qquad (E - E_F) \gg kT \quad [4.19]$$

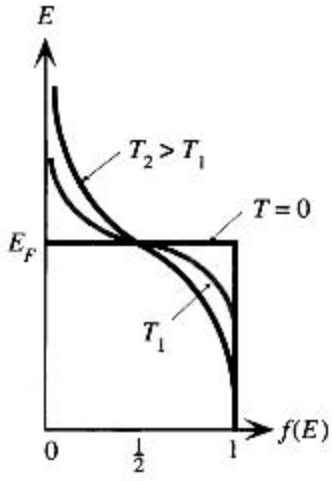


Figure 4.26
The Fermi–Dirac function $f(E)$ describes the statistics of electrons in a solid. The electrons interact with each other and the environment, obeying the Pauli exclusion principle.

Above absolute zero, at $E = E_F$, $f(E_F) = \frac{1}{2}$. We define the Fermi energy as that energy for which the probability of occupancy $f(E_F)$ equals $\frac{1}{2}$. The approximation to $f(E)$ in Equation 4.19 at high energies is often referred to as the **Boltzmann tail** to the Fermi–Dirac function.

4.7 QUANTUM THEORY OF METALS

4.7.1 FREE ELECTRON MODEL[6]

We know that the number of states $g(E)$ for an electron, per unit energy per unit volume, increases with energy as $g(E) \propto E^{1/2}$. We have also calculated that the probability of an electron being in a state with an energy E is the Fermi–Dirac function $f(E)$. Consider the energy band diagram for a metal and the density of states $g(E)$ for that band, as shown in Figure 4.27a and b, respectively.

At absolute zero, all the energy levels up to E_F are full. At 0 K, $f(E)$ has the step form at E_F (Figure 4.26). This clarifies why E_F in $f(E)$ is termed the Fermi energy. At 0 K, $f(E) = 1$ for $E < E_F$, and $f(E) = 0$ for $E > E_F$, so at 0 K, E_F separates the empty and full energy levels. This explains why we restricted ourselves to 0 K or thereabouts when we introduced E_F in the band theory of metals.

At some finite temperature, $f(E)$ is *not* zero beyond E_F, as indicated in Figure 4.27c. This means that some of the electrons are excited to, and thereby occupy, energy levels above E_F. If we multiply $g(E)$, by $f(E)$, we obtain the number of electrons per unit energy per unit volume, denoted n_E. The distribution of electrons in the energy levels is described by $n_E = g(E)\ f(E)$.

Since $f(E) = 1$ for $E \ll E_F$, the states near the bottom of the band are all occupied; thus, $n_E \propto E^{1/2}$ initially. As E passes through E_F, $f(E)$ starts decreasing

[6] The free electron model of metals is also known as the Sommerfeld model.

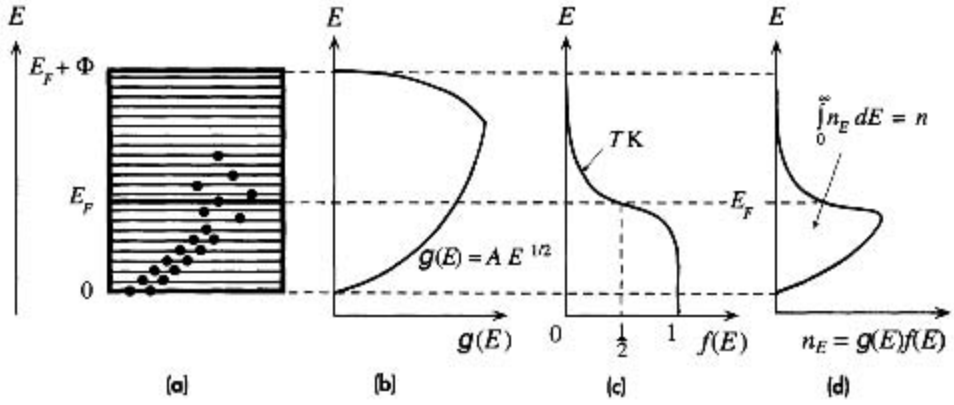


Figure 4.27

(a) Above 0 K, due to thermal excitation, some of the electrons are at energies above E_F.

(b) The density of states, $g(E)$ versus E in the band.

(c) The probability of occupancy of a state at an energy E is $f(E)$.

(d) The product $g(E)f(E)$ is the number of electrons per unit energy per unit volume, or the electron concentration per unit energy. The area under the curve on the energy axis is the concentration of electrons in the band.

sharply. As a result, n_E takes a turn and begins to decrease sharply as well, as depicted in Figure 4.27d.

In the small energy range E to $(E + dE)$, there are $n_E\,dE$ electrons per unit volume. When we sum all $n_E\,dE$ from the bottom to the top of the band ($E = 0$ to $E = E_F + \Phi$), we get the total number of valence electrons per unit volume, n, in the metal, as follows:

$$n = \int_0^{\text{Top of band}} n_E\,dE = \int_0^{\text{Top of band}} g(E)\,f(E)\,dE \qquad \textbf{[4.20]}$$

Since $f(E)$ falls very sharply when $E > E_F$, we can carry the integration to $E = \infty$, rather than to $(E_F + \Phi)$, because $f \to 0$ when $E \gg E_F$. Putting in the functional forms of $g(E)$ and $f(E)$ (*e.g.*, from Equations 4.10 and 4.18), we obtain

$$n = \frac{8\pi 2^{1/2} m_e^{3/2}}{h^3} \int_0^\infty \frac{E^{1/2}\,dE}{1 + \exp\left(\dfrac{E - E_F}{kT}\right)} \qquad \textbf{[4.21]}$$

If we could integrate this, we would obtain an expression relating n and E_F. At 0 K, however, $E_F = E_{FO}$ and the integrand exists only for $E < E_{FO}$. If we integrate at 0 K, Equation 4.21 yields

Fermi energy at $T = 0$ K

$$E_{FO} = \left(\frac{h^2}{8m_e}\right)\left(\frac{3n}{\pi}\right)^{2/3} \qquad \textbf{[4.22]}$$

It may be thought that E_F is temperature independent, since it was sketched that way in Figure 4.26. However, in our derivation of the Fermi–Dirac statistics, there was no restriction that demanded this. Indeed, since the number of electrons in a band is fixed, E_F at a temperature T is implicitly determined by Equation 4.21, which can be solved to express E_F in terms of n and T. It turns out that at 0 K, E_F is given by Equation 4.22, and it changes very little with temperature. In fact, by utilizing various mathematical approximations, it is not too difficult to integrate Equation 4.21 to obtain the **Fermi energy** at a temperature T, as follows:

$$E_F(T) = E_{FO}\left[1 - \frac{\pi^2}{12}\left(\frac{kT}{E_{FO}}\right)^2\right] \qquad [4.23]$$

Fermi energy at T (K)

which shows that $E_F(T)$ is only weakly temperature dependent, since $E_{FO} \gg kT$.

The Fermi energy has an important significance in terms of the average energy E_{av} of the conduction electrons in a metal. In the energy range E to $(E + dE)$, there are $n_E\,dE$ electrons with energy E. The average energy of an electron will therefore be

$$E_{av} = \frac{\int E n_E\,dE}{\int n_E\,dE} \qquad [4.24]$$

If we substitute $g(E)f(E)$ for n_E and integrate, the result at 0 K is

$$E_{av}(0) = \frac{3}{5}E_{FO} \qquad [4.25]$$

Average energy per electron at 0 K

Above absolute zero, the **average energy** is approximately

$$E_{av}(T) = \frac{3}{5}E_{FO}\left[1 + \frac{5\pi^2}{12}\left(\frac{kT}{E_{FO}}\right)^2\right] \qquad [4.26]$$

Average energy per electron at T (K)

Since $E_{FO} \gg kT$, the second term in the square brackets is much smaller than unity, and $E_{av}(T)$ shows only a very weak temperature dependence. Furthermore, in our model of the metal, the electrons are free to move around within the metal, where their potential energy PE is zero, whereas outside the metal, it is $E_F + \Phi$ (Figure 4.11). Therefore, their energy is purely kinetic. Thus, Equation 4.26 gives the average KE of the electrons in a metal

$$\frac{1}{2}m_e v_e^2 = E_{av} \approx \frac{3}{5}E_{FO}$$

where v_e is the root mean square (rms) speed of the electrons, which is simply called the **effective speed.** The effective speed v_e depends on the Fermi energy E_{FO} and is relatively insensitive to temperature. Compare this with the behavior of molecules in an ideal gas. In that case, the average $KE = \frac{3}{2}kT$, so $\frac{1}{2}mv^2 = \frac{3}{2}kT$. Clearly, the average speed of molecules in a gas increases with temperature.

The relationship $\frac{1}{2}mv_e^2 \approx \frac{3}{5}E_{FO}$ is an important conclusion that comes from the application of quantum mechanical concepts, ideas that lead to $g(E)$ and $f(E)$ and so on. It cannot be proved without invoking quantum mechanics. The fact that the average electronic speed is nearly constant is the only way to explain the observation that the resistivity of a metal is proportional to T (and not $T^{3/2}$), as we saw in Chapter 2.

4.7.2 CONDUCTION IN METALS

We know from our energy band discussions that in metals only those electrons in a small range ΔE around the Fermi energy E_F contribute to electrical conduction as shown in Figure 4.12c. The concentration n_F of these electrons is approximately $g(E_F)\,\Delta E$ inasmuch as ΔE is very small. The electron a moves to a', as shown in Figure 4.12b and c, and then it is scattered to an empty state above b'. In steady conduction, all the electrons in the energy range ΔE that are moving to the right are not canceled by any moving to the left and hence contribute to the current. An electron at the bottom of the ΔE range gains energy ΔE to move a' in a time interval Δt that corresponds to the scattering time τ. It gains a momentum Δp_x. Since $\Delta p_x/\Delta t =$ external force $= e\mathcal{E}_x$, we have $\Delta p_x = \tau e\mathcal{E}_x$. The electron a has an energy $E = p_x^2/(2m_e^*)$ which we can differentiate to obtain ΔE when the momentum changes by Δp_x,

$$\Delta E = \frac{p_x}{m_e^*}\Delta p_x = \frac{(m_e^* v_F)}{m_e^*}(\tau e\mathcal{E}_x) = e v_F \tau \mathcal{E}_x$$

The current J_x is due to all the electrons in the range ΔE which are moving toward the right in Figure 4.12c,

$$J_x = e n_F v_F = e[g(E_F)\,\Delta E]v_F = e[g(E_F) e v_F \tau \mathcal{E}_x]v_F = e^2 v_F^2 \tau g(E_F)\mathcal{E}_x$$

The conductivity is therefore

$$\sigma = e^2 v_F^2 \tau g(E_F)$$

However, the numerical factor is wrong because Figure 4.12c considers only a hypothetical one-dimensional crystal. In a three-dimensional crystal, the conductivity is one-third of the conductivity value just determined:

Conductivity of Fermi-level electrons

$$\sigma = \frac{1}{3}e^2 v_F^2 \tau g(E_F) \qquad [4.27]$$

This conductivity expression is in sharp contrast with the classical expression in which all the electrons contribute to conduction. According to Equation 4.27, what is important is the density of states at the Fermi energy $g(E_F)$. For example, Cu and Mg are metals with valencies I and II. Classically, Cu and Mg atoms each contribute one and two conduction electrons, respectively, into the crystal. Thus, we would expect Mg to have higher conductivity. However, the Fermi level in Mg is where the top tail of the $3s$ band overlaps the bottom tail of the $3p$ band where the density of states is small. In Cu, on the other hand, E_F is nearly in the middle of the $4s$ band where the density of states is high. Thus, Mg has a lower conductivity than Cu.

The scattering time τ in Equation 4.27 assumes that the scattered electrons at E_F remain in the same energy band. In certain metals, there are two different energy bands that overlap at E_F. For example, in Ni (see Figure 4.61), $3d$ and $4s$ bands overlap at E_F. An electron can be scattered from the $4s$ to the $3d$ band, and vice versa. Electrons in the $3d$ band have very low drift mobilities and effectively do not contribute to conduction, so only $g(E_F)$ of the $4s$ band operates in Equation 4.27.

Since 4*s* to 3*d* band scattering is an additional scattering mechanism, by virtue of Matthiessen's rule, the scattering time τ for the 4*s* band electrons is shortened. Thus, Ni has poorer conductivity than Cu.

In deriving Equation 4.27 we did not assume a particular density of states model. If we now apply the *free electron model* for $g(E_F)$ as in Equation 4.10, and also relate E_F to the total number of conduction electrons per unit volume n as in Equation 4.22, we would find that the conductivity is the same as the **Drude model,** that is,

$$\sigma = \frac{e^2 n \tau}{m_e} \qquad [4.28]$$

Drude model and free electrons

EXAMPLE 4.9

MEAN SPEED OF CONDUCTION ELECTRONS IN A METAL Calculate the Fermi energy E_{FO} at 0 K for copper and estimate the average speed of the conduction electrons in Cu. The density of Cu is 8.96 g cm^{-3} and the relative atomic mass (atomic weight) is 63.5.

SOLUTION

Assuming each Cu atom donates one free electron, we can find the concentration of electrons from the density d, atomic mass M_{at}, and Avogadro's number N_A, as follows:

$$n = \frac{dN_A}{M_{at}} = \frac{8.96 \times 6.02 \times 10^{23}}{63.5}$$

$$= 8.5 \times 10^{22}\ \text{cm}^{-3} \quad \text{or} \quad 8.5 \times 10^{28}\ \text{m}^{-3}$$

The Fermi energy at 0 K is given by Equation 4.22:

$$E_{FO} = \left(\frac{h^2}{8m_e}\right)\left(\frac{3n}{\pi}\right)^{2/3}$$

Substituting $n = 8.5 \times 10^{28}$ m^{-3} and the values for h and m_e, we obtain

$$E_{FO} = 1.1 \times 10^{-18}\ \text{J} \quad \text{or} \quad 7\ \text{eV}$$

To estimate the mean speed of the electrons, we calculate the rms speed v_e from $\frac{1}{2}m_e v_e^2 = \frac{3}{5}E_{FO}$. The mean speed will be close to the rms speed. Thus, $v_e = (6E_{FO}/5m_e)^{1/2}$. Substituting for E_{FO} and m_e, we find $v_e = 1.2 \times 10^6$ m s^{-1}.

EXAMPLE 4.10

CONDUCTION IN SILVER Consider silver whose density of states $g(E)$ was calculated in Example 4.8, assuming a *free electron model* for $g(E)$ as in Equation 4.10. For silver, $E_F = 5.5$ eV, so from Equation 4.10, the density of states at E_F is $g(E_F) = 1.60 \times 10^{28}$ m^{-3} eV^{-1}. The velocity of Fermi electrons, $v_F = (2E_F/m_e)^{1/2} = 1.39 \times 10^6$ m s^{-1}. The conductivity σ of Ag at room temperature is $62.5 \times 10^6\ \Omega^{-1}\ \text{m}^{-1}$. Substituting for σ, $g(E_F)$, and v_F in Equation 4.27,

$$\sigma = 62.5 \times 10^6 = \frac{1}{3}e^2 v_F^2 \tau g(E_F) = \frac{1}{3}(1.6 \times 10^{-19})^2 (1.39 \times 10^6)^2 \tau \left(\frac{1.60 \times 10^{28}}{1.6 \times 10^{-19}}\right)$$

we find $\tau = 3.79 \times 10^{-14}$ s. The *mean free path* $\ell = v_F \tau = 53$ nm. The *drift mobility* of E_F electrons is $\mu = e\tau/m_e = 67\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$.

From Example 4.8, since Ag has a valency of I, the concentration of conduction electrons is $n = n_{Ag} = 5.85 \times 10^{28}\ m^{-3}$. Substituting for n and σ in Equation 4.28 gives

$$\sigma = 62.5 \times 10^6 = \frac{e^2 n \tau}{m_e} = \frac{(1.6 \times 10^{-19})^2 (5.85 \times 10^{28})\tau}{(9.1 \times 10^{-31})}$$

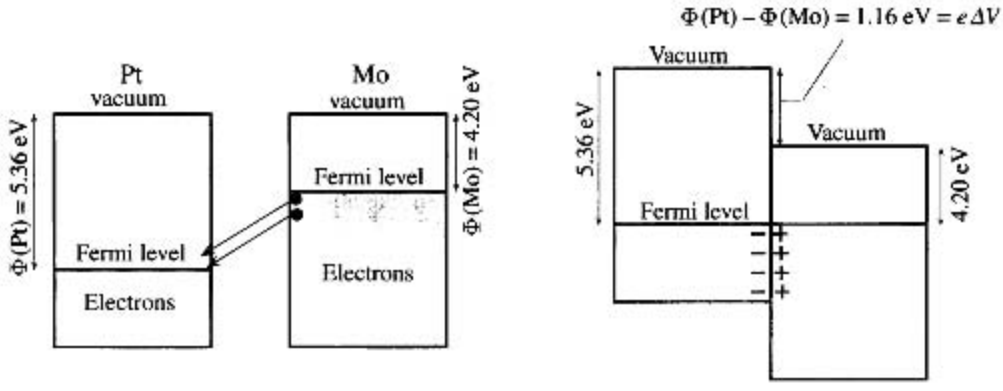
we find $\tau = 3.79 \times 10^{-14}$ s as expected because we have used the free electron model.

4.8 FERMI ENERGY SIGNIFICANCE

4.8.1 METAL–METAL CONTACTS: CONTACT POTENTIAL

Suppose that two metals, platinum (Pt) with a work function 5.36 eV and molybdenum (Mo) with a work function 4.20 eV, are brought together, as shown in Figure 4.28a. We know that in metals, all the energy levels up to the Fermi level are full. Since the Fermi level is higher in Mo (due to a smaller Φ), the electrons in Mo are more energetic. They therefore immediately go over to the Pt surface (by tunneling), where there are empty states at lower energies, which they can occupy. This electron transfer from Mo to the Pt surface reduces the total energy of the electrons in the Pt–Mo system, but at the same time, the Pt surface becomes negatively charged with respect to the Mo surface. Consequently, a contact voltage (or a potential difference) develops at the junction between Pt and Mo, with the Mo side being positive.

The electron transfer from Mo to Pt continues until the contact potential is large enough to prevent further electron transfer: the system reaches equilibrium. It should be apparent that the transfer of energetic electrons from Mo to Pt continues until the two Fermi levels are lined up, that is, until the Fermi level is uniform and the same in both metals, so that no part of the system has more (or less) energetic electrons, as



(a) Electrons are more energetic in Mo, so they tunnel to the surface of Pt.

(b) Equilibrium is reached when the Fermi levels are lined up.

Figure 4.28 When two metals are brought together, there is a contact potential ΔV.

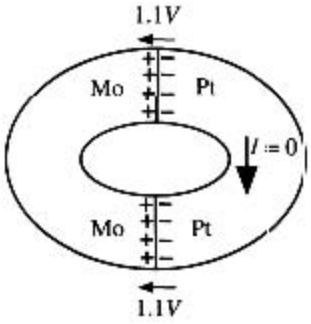


Figure 4.29 There is no current when a closed circuit is formed by two different metals, even though there is a contact potential at each contact.
The contact potentials oppose each other.

illustrated in Figure 4.28b. Otherwise, the energetic electrons in one part of the system will flow toward a region with lower energy states. Under these conditions, the Pt–Mo system is in equilibrium. The contact voltage ΔV is determined by the difference in the work functions, that is,

$$e\,\Delta V = \Phi(\text{Pt}) - \Phi(\text{Mo}) = 5.36\,\text{eV} - 4.20\,\text{eV} = 1.16\,\text{eV}$$

We should note that away from the junction on the Mo side, we must still provide an energy of $\Phi = 4.20\,\text{eV}$ to free an electron, whereas away from the junction on the Pt side, we must provide $\Phi = 5.36\,\text{eV}$ to free an electron. This means that the vacuum energy level going from Mo to Pt has a step $\Delta\Phi$ at the junction. Since we must do work equivalent to $\Delta\Phi$ to get a free electron (*e.g.*, on the metal surface) from the Mo surface to the Pt surface, this represents a voltage of $\Delta\Phi/e$ or 1.16 V.

From the second law of thermodynamics,[7] this contact voltage cannot do work; that is, it cannot drive current in an external circuit. To see this, we can close the Pt metal–Mo metal circuit to form a ring, as depicted in Figure 4.29. As soon as we close the circuit, we create another junction with a contact voltage that is equal and opposite to that of the first junction. Consequently, going around the circuit, the net voltage is zero and the current is therefore zero.

There is a deep significance to the Fermi energy E_F, which should at least be mentioned. For a given metal the Fermi energy represents the free energy per electron called the **electrochemical potential** μ. In other words, the Fermi energy is a measure of the potential of an electron to do electrical work ($e \times V$) or nonmechanical work, through chemical or physical processes.[8] In general, when two metals are brought into contact, the Fermi level (with respect to a vacuum) in each will be different. This difference means a difference in the chemical potential $\Delta\mu$, which in turn means that the system will do external work, which is obviously not possible. Instead, electrons are immediately transferred from one metal to the other, until the free energy per electron μ for the whole system is minimized and is uniform across the two metals, so that

[7] By the way, the second law of thermodynamics simply says that you cannot extract heat from a system in thermal equilibrium and do work (i.e., charge × voltage).

[8] A change in any type of *PE* can, in principle, be used to do work, that is, $\Delta(PE)$ = work done. Chemical *PE* is the potential to do nonmechanical work (*e.g.*, electrical work) by virtue of physical or chemical processes. The chemical *PE* per electron is E_F and ΔE_F = electrical work per electron.

$\Delta\mu = 0$. We can guess that if the Fermi level in one metal could be maintained at a higher level than the other, by using an external energy source (*e.g.*, light or heat), for example, then the difference could be used to do electrical work.

4.8.2 THE SEEBECK EFFECT AND THE THERMOCOUPLE

Consider a conductor such as an aluminum rod that is heated at one end and cooled at the other end as depicted in Figure 4.30. The electrons in the hot region are more energetic and therefore have greater velocities than those in the cold region.[9]

Consequently there is a net diffusion of electrons from the hot end toward the cold end which leaves behind exposed positive metal ions in the hot region and accumulates electrons in the cold region. This situation prevails until the electric field developed between the positive ions in the hot region and the excess electrons in the cold region prevents further electron motion from the hot to the cold end. A voltage therefore develops between the hot and cold ends, with the hot end at positive potential. The potential difference ΔV across a piece of metal due to a temperature difference ΔT is called the **Seebeck effect.**[10] To gauge the magnitude of this effect we introduce a special coefficient which is defined as the potential difference developed per unit temperature difference, or

Thermoelectric power or Seebeck coefficient

$$S = \frac{dV}{dT} \qquad [4.29]$$

By convention, the sign of S represents the potential of the cold side with respect to the hot side. If electrons diffuse from the hot end to the cold end as in Figure 4.30, then the cold side is negative with respect to the hot side and the Seebeck coefficient is *negative* (as for aluminum).

In some metals, such as copper, this intuitive explanation fails to explain why electrons actually diffuse from the cold to the hot region, giving rise to *positive* Seebeck coefficients; the polarity of the voltage in Figure 4.30 is actually reversed for copper. The reason is that the net diffusion process depends on how the mean free path ℓ and the mean free time (due to scattering from lattice vibrations) change with the electron energy, which can be quite complicated. Typical Seebeck coefficients for various selected metals are listed in Table 4.3.

Consider two neighboring regions H (hot) and C (cold) with widths corresponding to the mean free paths ℓ and ℓ' in H and C as depicted in Figure 4.31a. Half the electrons in H would be moving in the $+x$ direction and the other half in the $-x$ direction. Half of the electrons in H therefore cross into C, and half in C cross into H. Suppose that, very roughly, the electron concentration n in H and C is about the same. The number of electrons crossing from H to C is $\frac{1}{2}n\ell$, and the number crossing from C to H is $\frac{1}{2}n\ell'$. Then,

$$\text{Net diffusion from H to C} \propto \tfrac{1}{2}n(\ell - \ell') \qquad [4.30]$$

[9] The conduction electrons around the Fermi energy have a mean speed that has only a small temperature dependence. This small change in the mean speed with temperature is, nonetheless, intuitively significant in appreciating the thermoelectric effect. The actual effect, however, depends on the mean free path as discussed later.

[10] Thomas Seebeck observed the thermoelectric effect in 1821 using two different metals as in the thermocouple, which is the only way to observe the phenomenon. It was William Thomson (Lord Kelvin) who explained the observed effect.

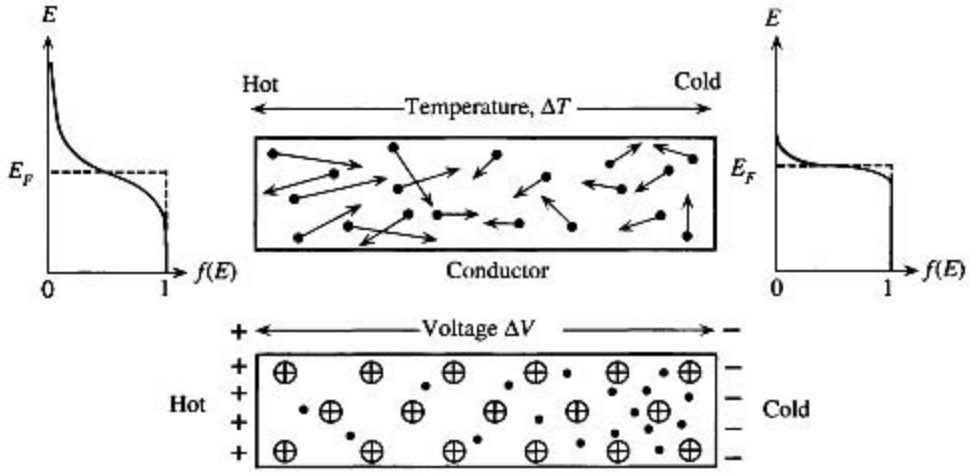


Figure 4.30 The Seebeck effect.
A temperature gradient along a conductor gives rise to a potential difference.

Suppose that the scattering of electrons is such that ℓ increases strongly with the electron energy. Then electrons in H, which are more energetic, have a longer mean free path, that is, $\ell > \ell'$ as shown in Figure 4.31a. This means that the net migration is from H to C and S is negative, as in aluminum. In those metals such as copper in which ℓ decreases strongly with the energy, electrons in the cold region have a longer mean free path, $\ell' > \ell$ as shown in Figure 4.31b. The net electron migration is then from C to H and S is positive. Even this qualitative explanation is not quite correct because n is not the same in H and C (diffusion changes n) and, further, we neglected the change in the mean scattering time with the electron energy.

The coefficient S is widely referred to as the **thermoelectric power** even though this term is misleading, as it refers to a voltage difference rather than power. A more appropriate recent term is the **Seebeck coefficient.** S is a material property that depends on temperature, $S = S(T)$, and is tabulated for many materials as a function of

Table 4.3 Seebeck coefficients of selected metals (from various sources)

Metal	S at 0 °C ($\mu V\ K^{-1}$)	S at 27 °C ($\mu V\ K^{-1}$)	E_F (eV)	x
Al	−1.6	−1.8	11.6	2.78
Au	+1.79	+1.94	5.5	−1.48
Cu	+1.70	+1.84	7.0	−1.79
K		−12.5	2.0	3.8
Li	+14		4.7	−9.7
Mg	−1.3		7.1	1.38
Na		−5	3.1	2.2
Pd	−9.00	−9.99		
Pt	−4.45	−5.28		

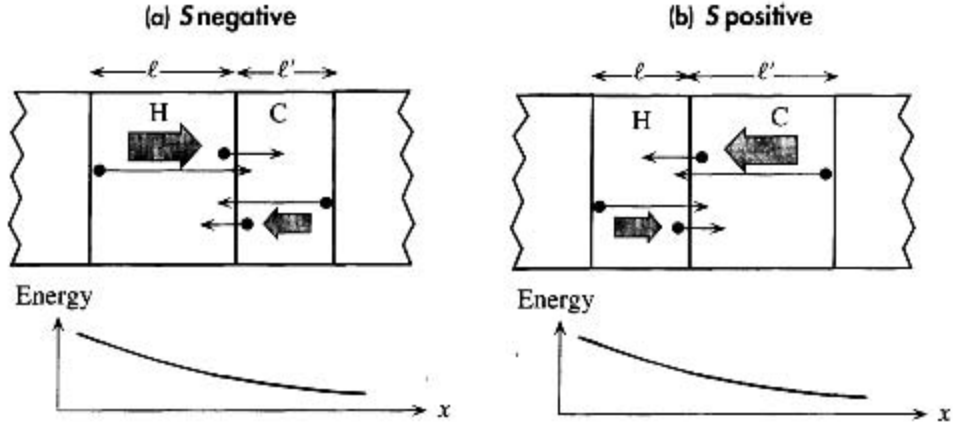


Figure 4.31 Consider two neighboring regions H (hot) and C (cold) with widths corresponding to the mean free paths ℓ and ℓ' in H and C.
Half the electrons in H would be moving in the $+x$ direction and the other half in the $-x$ direction. Half of the electrons in H therefore cross into C, and half in C cross into H.

temperature. Given the Seebeck coefficient $S(T)$ for a material, Equation 4.29 yields the voltage difference between two points where temperatures are T_o and T as follows:

$$\Delta V = \int_{T_o}^{T} S\,dT \qquad [4.31]$$

A proper explanation of the Seebeck effect has to consider how electrons around the Fermi energy E_F, which contribute to electrical conduction, are scattered by lattice vibrations, impurities, and crystal defects. This scattering process controls the mean free path and hence the Seebeck coefficient (Figure 4.31). The scattered electrons need empty states, which in turn requires that we consider how the density of states changes with the energy as well. Moreover, in certain metals such as Ni, there are overlapping partially filled bands and the Fermi electron can be scattered from one electronic band to another, for example from the $4s$ band to the $3d$ band, which must also be considered (see Question 4.25). The Seebeck coefficient for many metals is given by the **Mott and Jones equation,**

Mott and Jones thermoelectric power

$$S \approx -\frac{\pi^2 k^2 T}{3eE_{FO}}x \qquad [4.32]$$

where x is a numerical constant that takes into account how various charge transport parameters (such as ℓ) depend on the electron energy. A few examples for x are given in Table 4.3. The reason for the kT/E_{FO} factor in Equation 4.32 is that only those electrons about a kT around the Fermi level E_{FO} are involved in the transport and scattering processes. Equation 4.32 does not apply directly to transition metals (Ni, Pd, Pt) that have overlapping bands. These metals have a negative Seebeck coefficient that is proportional to temperature as in Equation 4.32, but the exact expression depends on the band structure.

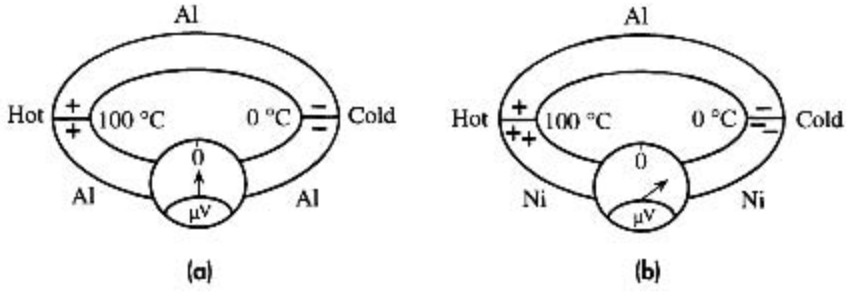


Figure 4.32

(a) If Al wires are used to measure the Seebeck voltage across the Al rod, then the net emf is zero.

(b) The Al and Ni have different Seebeck coefficients. There is therefore a net emf in the Al–Ni circuit between the hot and cold ends that can be measured.

Suppose that we try to measure the voltage difference ΔV across the aluminum rod by using aluminum connecting wires to a voltmeter as indicated in Figure 4.32a. The same temperature difference now also exists across the aluminum connecting wires; therefore an identical voltage also develops across the connecting wires, opposing that across the aluminum rod. Consequently no net voltage will be registered by the voltmeter. It is, however, possible to read a net voltage difference, if the connecting wires are of different material, that is, have a different Seebeck coefficient from that of aluminum. Then the thermoelectric voltage across this material is different than that across the aluminum rod, as in Figure 4.32b.

The Seebeck effect is fruitfully utilized in the thermocouple (TC), shown in Figure 4.32b, which uses two different metals with one junction maintained at a reference temperature T_o and the other used to sense the temperature T. The voltage across each metal element depends on its Seebeck coefficient. The potential difference between the two wires will depend on $S_A - S_B$. By virtue of Equation 4.31, the electromotive force (emf) between the two wires, $V_{AB} = \Delta V_A - \Delta V_B$, is then given by

Thermocouple emf between metals A and B

$$V_{AB} = \int_{T_o}^{T} (S_A - S_B)\, dT = \int_{T_o}^{T} S_{AB}\, dT \qquad [4.33]$$

where $S_{AB} = S_A - S_B$ is defined as the thermoelectric power for the thermocouple pair A-B. For the chromel-alumel (K-type) TC, for example, $S_{AB} \approx 40\ \mu\text{V K}^{-1}$ at 300 K.

The output voltage from a TC pair obviously depends on the two metals used. Instead of tabulating the emf from all possible pairs of materials in the world, which would be a challenging task, engineers have tabulated the emfs available when a given material is used with a reference metal which is chosen to be platinum. The reference junction is kept at 0 °C (273.16 K) which corresponds to a mixture of ice and water. Some typical materials and their emfs are listed in Table 4.4.

Thermocouple equation

Using the expression for the Seebeck coefficient, Equation 4.32, in Equation 4.33, and then integrating, leads to the familiar thermocouple equation,

$$V_{AB} = a\ \Delta T + b(\Delta T)^2 \qquad [4.34]$$

Table 4.4 Thermoelectric emf for metals at 100 and 200 °C with respect to Pt and the reference junction at 0 °C

	emf (mV)	
Material	**At 100 °C**	**At 200 °C**
Copper, Cu	0.76	1.83
Aluminum, Al	0.42	1.06
Nickel, Ni	−1.48	−3.10
Palladium, Pd	−0.57	−1.23
Platinum, Pt	0	0
Silver, Ag	0.74	1.77
Alumel	−1.29	−2.17
Chromel	2.81	5.96
Constantan	−3.51	−7.45
Iron, Fe	1.89	3.54
90% Pt–10% Rh (platinum-rhodium)	0.643	1.44

where a and b are the thermocouple coefficients and $\Delta T = T - T_o$ is the temperature with respect to the reference temperature T_o (273.16 K). The inference from Equation 4.34 is that the emf output from the thermocouple wires does not depend linearly on the temperature difference ΔT. Figure 4.33 shows the emf output versus temperature for various thermocouples. It should be immediately obvious that the voltages are small, typically a few tens of a microvolt per degree temperature difference. At

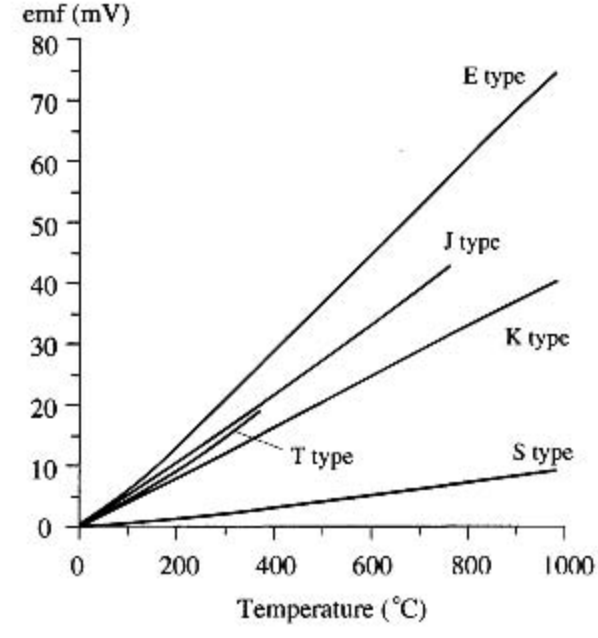


Figure 4.33 Output emf versus temperature (°C) for various thermocouples between 0 to 1000 °C.

0 °C, by definition, the TC emf is zero. The K-type thermocouple, the chromel-alumel pair, is a widely employed general-purpose thermocouple sensor up to about 1200 °C.

EXAMPLE 4.11

THE THERMOCOUPLE EMF Consider a thermocouple pair from Al and Cu which have Fermi energies and x as in Table 4.3. Estimate the emf available from this thermocouple if one junction is held at 0 °C and the other at 100 °C.

SOLUTION

We essentially have the arrangement shown in Figure 4.32b but with Cu replacing Ni and Cu having the cold end positive (S is positive). For each metal there will be a voltage across it, given by integrating the Seebeck coefficient from T_o (at the low temperature end) to T. From the Mott and Jones equation,

$$\Delta V = \int_{T_o}^{T} S\,dT = \int_{T_o}^{T} -\frac{x\pi^2 k^2 T}{3eE_{FO}}\,dT = -\frac{x\pi^2 k^2}{6eE_{FO}}(T^2 - T_o^2)$$

The available emf (V_{AB}) is the difference in ΔV for the two metals (A and B), so

$$V_{AB} = \Delta V_A - \Delta V_B = -\frac{\pi^2 k^2}{6e}\left[\frac{x_A}{E_{FAO}} - \frac{x_B}{E_{FBO}}\right](T^2 - T_o^2)$$

where in this example $T = 373$ K and $T_o = 273$ K.

For Al (A), $E_{FAO} = 11.6$ eV, $x_A = 2.78$, and for copper (B), $E_{FBO} = 7.0$ eV, $x_B = -1.79$. Thus,

$$V_{AB} = -189\ \mu\text{V} - (+201\ \mu\text{V}) = -390\ \mu\text{V}$$

Thermocouple emf calculations that closely represent experimental observations require thermocouple voltages for various metals listed against some reference metal. The reference is usually Pt with the reference junction at 0 °C. From Table 4.4 we can read Al–Pt and Cu–Pt emfs as $V_{\text{Al-Pt}} = 0.42$ mV and $V_{\text{Cu-Pt}} = 0.76$ mV at 100 °C with the experimental error being around ±0.01 mV, so that for the Al–Cu pair,

$$V_{\text{Al-Cu}} = V_{\text{Al-Pt}} - V_{\text{Cu-Pt}} = 0.42\ \text{mV} - 0.76\ \text{mV} = -0.34\ \text{mV}$$

There is a reasonable agreement with the calculation using the Mott and Jones equation.

EXAMPLE 4.12

THE THERMOCOUPLE EQUATION We know that we can only measure differences between thermoelectric powers of materials. When two different metals A and B are connected to make a thermocouple, as in Figure 4.32b, then the net emf is the voltage difference between the two elements. From Example 4.11,

$$\Delta V_{AB} = \Delta V_A - \Delta V_B = \int_{T_o}^{T} (S_A - S_B)\,dT = \int_{T_o}^{T} S_{AB}\,dT$$

$$= -\frac{\pi^2 k^2}{6e}\left[\frac{x_A}{E_{FAO}} - \frac{x_B}{E_{FBO}}\right](T^2 - T_o^2)$$

$$= C(T^2 - T_o^2)$$

where C is a constant that is independent of T but dependent on the material properties (x, E_{FO} for the metals).

We can now expand V_{AB} about T_o by using Taylor's expansion

$$F(T) \approx F(T_o) + \Delta T\ (dF/dT)_o + \tfrac{1}{2}(\Delta T)^2(d^2F/dT^2)_o$$

where the function $F = V_{AB}$ and $\Delta T = T - T_o$ and the derivatives are evaluated at T_o. The result is the thermocouple equation:

$$V_{AB}(T) = a(\Delta T) + b(\Delta T)^2$$

where the coefficients a and b are $2CT_o$ and C, respectively.

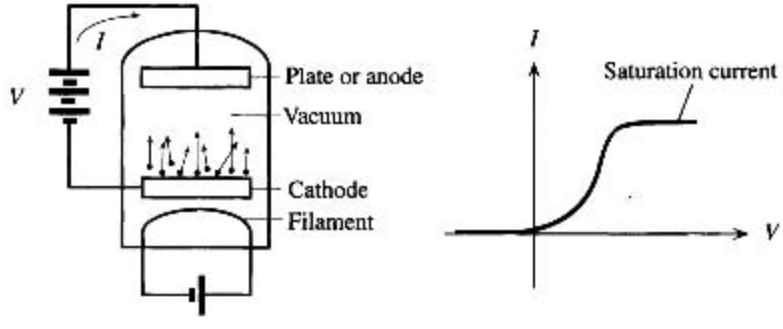
It is clear that the magnitude of the emf produced depends on C or $S_A - S_B$, which we can label as S_{AB}. The greater the thermoelectric power difference S_{AB} for the TC, the larger the emf produced. For the copper constantan TC, S_{AB} is about 43 μV K^{-1}.

4.9 THERMIONIC EMISSION AND VACUUM TUBE DEVICES

4.9.1 THERMIONIC EMISSION: RICHARDSON–DUSHMAN EQUATION

Even though most of us view vacuum tubes as electrical antiques, their basic principle of operation (electrons emitted from a heated cathode) still finds application in cathode ray and X-ray tubes and various RF microwave vacuum tubes, such as triodes, tetrodes, klystrons, magnetrons, and traveling wave tubes and amplifiers. Therefore, it is useful to examine how electrons are emitted when a metal is heated.

When a metal is heated, the electrons become more energetic as the Fermi–Dirac function extends to higher temperatures. Some of the electrons have sufficiently large energies to leave the metal and become free. This situation is self-limiting because as the electrons accumulate outside the metal, they prevent more electrons from leaving the metal. (Put differently, emitted electrons leave a net positive charge behind, which pulls the electrons in.) Consequently, we need to replenish the "lost" electrons and collect the emitted ones, which is done most conveniently using the vacuum tube arrangement in a closed circuit, as shown in Figure 4.34a. The cathode, heated by a filament, emits electrons. A battery connected between the cathode and the anode replenishes



(a) Thermionic electron emission in a vacuum tube.

(b) Current-voltage characteristics of a vacuum diode.

Figure 4.34

the cathode electrons and provides a positive bias to the anode to collect the thermally emitted electrons from the cathode. The vacuum inside the tube ensures that the electrons do not collide with the air molecules and become dispersed, with some even being returned to the cathode by collisions. Therefore, the vacuum is essential. The current due to the flow of emitted electrons from the cathode to the anode depends on the anode voltage as indicated in Figure 4.34b. The current increases with the anode voltage until, at sufficiently high voltages, all the emitted electrons are collected by the anode and the current *saturates*. The **saturation current** of the vacuum diode depends on the rate of thermionic emission of electrons which we will derive below. The vacuum tube in Figure 4.34a acts as a **rectifier** because there is no current flow when the anode voltage becomes negative; the anode then repels the electrons.

We know that only those electrons with energies greater than $E_F + \Phi$ (Fermi energy + work function) which are moving toward the surface can leave the metal. Their number depends on the temperature, by virtue of the Fermi–Dirac statistics. Figure 4.35 shows how the concentration of conduction electrons with energies above $E_F + \Phi$ increases with temperature. We know that conduction electrons behave as if they are free within the metal. We can therefore take the PE to be zero within the metal, but $E_F + \Phi$ outside the metal. The energy E of the electron within the metal is then purely kinetic, or

$$E = \frac{1}{2}m_e v_x^2 + \frac{1}{2}m_e v_y^2 + \frac{1}{2}m_e v_z^2 \qquad [4.35]$$

Suppose that the surface of the metal is perpendicular to the direction of emission, say along x. For an electron to be emitted from the surface, its $KE = \frac{1}{2}mv_x^2$ along x must be greater than the potential energy barrier $E_F + \Phi$, that is,

$$\frac{1}{2}mv_x^2 > E_F + \Phi \qquad [4.36]$$

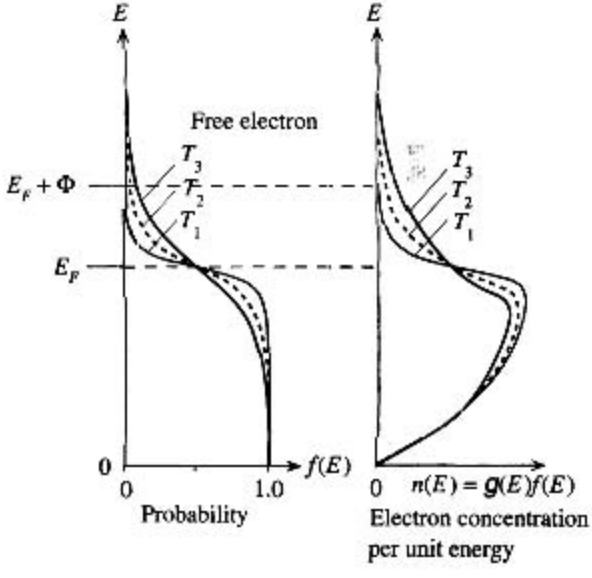


Figure 4.35 Fermi–Dirac function $f(E)$ and the energy density of electrons $n(E)$ (electrons per unit energy and per unit volume) at three different temperatures.
The electron concentration extends more and more to higher energies as the temperature increases. Electrons with energies in excess of $E_F + \Phi$ can leave the metal (thermionic emission).

Left to right: Owen Williams Richardson, Robert Andrews Millikan, and Arthur Holly Compton at an international conference on nuclear physics, Rome, 1931. Richardson won the physics Nobel prize in 1928 for thermionic emission.

SOURCE: Amaldi Archives, Dipartimento di Fisica, Università La Sapienza, Rome; courtesy of AIP Emilio Segrè Visual Archives.

Let $dn(v_x)$ be the number of electrons moving along x with velocities in the range v_x to $(v_x + dv_x)$, with v_x satisfying emission in Equation 4.36. These electrons will be emitted when they reach the surface. Their number $dn(v_x)$ can be determined from the density of states and the Fermi–Dirac statistics, since energy and velocity are related through Equation 4.35. Close to $E_F + \Phi$, the Fermi–Dirac function will approximate the Boltzmann distribution, $f(E) = \exp[-(E - E_F)/kT]$. The number $dn(v_x)$ is therefore at least proportional to this exponential energy factor.

The emission of $dn(v_x)$ electrons will give a thermionic current density $dJ_x = ev_x\, dn(v_x)$. This must be integrated (summed) for all velocities satisfying Equation 4.36 to obtain the total current density J_x, or simply J. Since $dn(v_x)$ includes an exponential energy function, the integration also leads to an exponential. The final result is

Richardson–Dushman thermionic emission equation

$$J = B_o T^2 \exp\left(-\frac{\Phi}{kT}\right) \qquad [4.37]$$

where $B_o = 4\pi e m_e k^2/h^3$. Equation 4.37 is called the **Richardson–Dushman equation,** and B_o is the Richardson–Dushman constant, whose value is $1.20 \times 10^6\ \mathrm{A\,m^{-2}\,K^{-2}}$. We see from Equation 4.37 that the emitted current from a heated cathode varies exponentially with temperature and is sensitive to the work function Φ of the cathode material. Both factors are apparent in Equation 4.37.

The wave nature of electrons means that when an electron approaches the surface, there is a probability that it may be reflected back into the metal, instead of being emitted over the potential barrier. As the potential energy barrier becomes very large, $\Phi \to \infty$, the electrons are totally reflected and there is no emission. Taking into account that waves can be reflected, the thermionic emission equation is appropriately modified to

Thermionic emission

$$J = B_e T^2 \exp\left(-\frac{\Phi}{kT}\right) \qquad [4.38]$$

where $B_e = (1 - R)B_o$ is the **emission constant** and R is the reflection coefficient. The value of R will depend on the material and the surface conditions. For most metals, B_e is about half of B_o, whereas for some oxide coatings on Ni cathodes used in thermionic tubes, B_e can be as low as 1×10^2 A m^{-2} K^{-2}.

Equation 4.37 was derived by neglecting the effect of the applied field on the emission process. Since the anode is positively biased with respect to the cathode, the field will not only collect the emitted electrons (by drifting them to the anode), but will also enhance the process of thermal emission by lowering the potential energy barrier Φ.

There are many thermionic emission–based vacuum tubes that find applications in which it is not possible or practical to use semiconductor devices, especially at high-power and high-frequency operation at the same time, such as in radio and TV broadcasting, radars, microwave communications; for example, a tetrode vacuum tube in radio broadcasting equipment has to handle hundreds of kilowatts of power. X-ray tubes operate on the thermionic emission principle in which electrons are thermally emitted, and then accelerated and impacted on a metal target to generate X-ray photons.

EXAMPLE 4.13

VACUUM TUBES It is clear from the Richardson–Dushman equation that to obtain an efficient thermionic cathode, we need high temperatures and low work functions. Metals such as tungsten (W) and tantalum (Ta) have high melting temperatures but high work functions. For example, for W, the melting temperature T_m is 3680 °C and its work function is about 4.5 eV. Some metals have low work functions, but also low melting temperatures, a typical example being Cs with $\Phi = 1.8$ eV and $T_m = 28.5$ °C. If we use a thin film coating of a low Φ material, such as ThO or BaO, on a high-melting-temperature base metal such as W, we can maintain the high melting properties and obtain a lower Φ. For example, Th on W has a $\Phi = 2.6$ eV and $T_m = 1845$ °C. Most vacuum tubes use indirectly heated cathodes that consist of the oxides of B, Sr, and Ca on a base metal of Ni. The operating temperatures for these cathodes are typically 800 °C.

A certain transmitter-type vacuum tube has a cylindrical Th-coated W (thoriated tungsten) cathode, which is 4 cm long and 2 mm in diameter. Estimate the saturation current if the tube is operated at a temperature of 1600 °C, given that the emission constant is $B_e = 3.0 \times 10^4$ A m^{-2} K^{-2} for Th on W.

SOLUTION

We apply the Richardson–Dushman equation with $\Phi = 2.6$ eV, $T = (1600 + 273)$ K $= 1873$ K, and $B_e = 3.0 \times 10^4$ A m^{-2} K^{-2}, to find the maximum current density that can be obtained from the cathode at 1873 K, as follows:

$$J = (3.0 \times 10^4 \text{ A m}^{-2} \text{ K}^{-2})(1873 \text{ K})^2 \exp\left[-\frac{(2.6 \times 1.6 \times 10^{-19})}{(1.38 \times 10^{-23} \times 1873)}\right]$$

$$= 1.08 \times 10^4 \text{ A m}^{-2}$$

The emission surface area is

$$A = \pi(\text{diameter})(\text{length}) = \pi(2 \times 10^{-3})(4 \times 10^{-2}) = 2.5 \times 10^{-4} \text{ m}^2$$

so the saturation current, which is the maximum current obtainable (*i.e.*, the thermionic current), is

$$I = JA = (1.08 \times 10^4 \text{ A m}^{-2})(2.5 \times 10^{-4} \text{ m}^2) = 2.7 \text{ A}$$

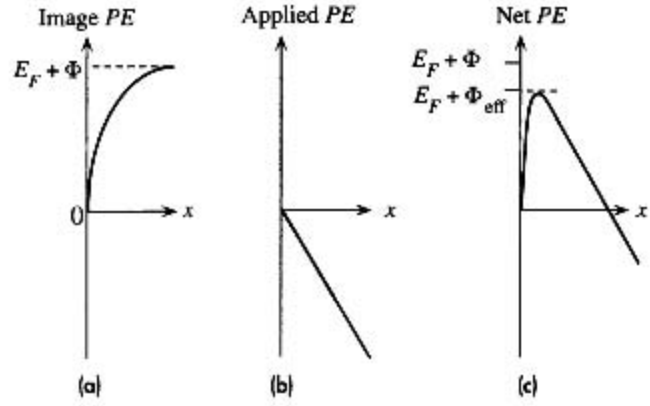


Figure 4.36
(a) *PE* of the electron near the surface of a conductor.
(b) Electron *PE* due to an applied field, that is, between cathode and anode.
(c) The overall *PE* is the sum.

4.9.2 SCHOTTKY EFFECT AND FIELD EMISSION

When a positive voltage is applied to the anode with respect to the cathode, the electric field at the cathode helps the thermionic emission process by lowering the *PE* barrier Φ. This is called the **Schottky effect.** Consider the *PE* of the electron just outside the surface of the metal. The electron is pulled in by the effective positive charge left in the metal. To represent this attractive *PE* we use the **theorem of image charges** in electrostatics,[11] which says that an electron at a distance x from the surface of a conductor possesses a potential energy that is

$$PE_{\text{image}}(x) = -\frac{e^2}{16\pi\varepsilon_o x} \qquad [4.39]$$

where ε_o is the absolute permittivity.

This equation is valid for x much greater than the atomic separation a; otherwise, we must consider the interaction of the electron with the individual ions. Further, Equation 4.39 has a reference level of zero *PE* at infinity ($x = \infty$), but we defined $PE = 0$ to be inside the metal. We must therefore modify Equation 4.39 to conform to our definition of zero *PE* as a reference. Figure 4.36a shows how this "image *PE*" varies with x in this system. In the region $x < x_o$, we artificially bring $PE_{\text{image}}(x)$ to zero at $x = 0$, so our definition $PE = 0$ within the metal is maintained. Far away from the surface, the *PE* is expected to be $(E_F + \Phi)$ (and not zero, as in Equation 4.39), so we modify Equation 4.39 to read

$$PE_{\text{image}}(x) = (E_F + \Phi) - \frac{e^2}{16\pi\varepsilon_o x} \qquad [4.40]$$

The present model, which takes $PE_{\text{image}}(x)$ from 0 to $(E_F + \Phi)$ along Equation 4.40, is in agreement with the thermionic emission analysis, since the electron must still overcome a *PE* barrier of $E_F + \Phi$ to escape.

[11] An electron at a distance x from the surface of a conductor experiences a force as if there were a positive charge of $+e$ at a distance $2x$ from it. The force is $e^2/[4\pi\varepsilon_o(2x)^2]$ or $e^2/[16\pi\varepsilon_o x^2]$. The result is called the image charge theorem. Integrating the force gives the potential energy in Equation 4.39.

From the definition of potential, which is potential energy per unit charge, when a voltage difference is applied between the anode and cathode, there is a PE gradient just outside the surface of the metal, given by $eV(x)$, or

$$PE_{\text{applied}}(x) = -ex\mathcal{E} \qquad [4.41]$$

where $\mathcal{E}$ is the applied field and is assumed, for all practical purposes, to be uniform. The variation of $PE_{\text{applied}}(x)$ with x is depicted in Figure 4.36b. The total $PE(x)$ of the electron outside the metal is the sum of Equations 4.40 and 4.41, as sketched in Figure 4.36c,

$$PE(x) \quad (E_F + \Phi) \quad \frac{}{16\pi\varepsilon_o x} \quad ex\mathcal{E} \qquad [4.42]$$

Note that the $PE(x)$ outside the metal no longer goes up to $(E_F + \Phi)$, and the PE barrier against thermal emission is effectively reduced to $(E_F + \Phi_{\text{eff}})$, where Φ_{eff} is a new effective work function that takes into account the effect of the applied field. The new barrier $(E_F + \Phi_{\text{eff}})$ can be found by locating the maximum of $PE(x)$, that is, by differentiating Equation 4.42 and setting it to zero. The **effective work function** in the presence of an applied field is therefore

$$\Phi_{\text{eff}} \quad \Phi \quad \left(\frac{e^3\mathcal{E}}{4\pi\varepsilon_o}\right) \qquad [4.43]$$

This lowering of the work function by the applied field, as predicted by Equation 4.43, is the **Schottky effect.** The current density is given by the Richardson–Dushman equation, but with Φ_{eff} instead of Φ,

$$J = B_e T^2 \exp\left[-\frac{(\Phi - \beta_S \mathcal{E}^{1/2})}{kT}\right] \qquad [4.44]$$

Field-assisted thermionic emission

where $\beta_S = [e^3/4\pi\varepsilon_o]^{1/2}$ is **the Schottky coefficient,** whose value is 3.79×10^{-5} $(\text{eV}/\sqrt{\text{V m}^{-1}})$.

When the field becomes very large, for example, $\mathcal{E} > 10^7$ V cm^{-1}, the $PE(x)$ outside the metal surface may bend sufficiently steeply to give rise to a narrow PE barrier. In this case, there is a distinct probability that an electron at an energy E_F will tunnel through the barrier and escape into vacuum, as depicted in Figure 4.37. The likelihood of tunneling depends on the effective height Φ_{eff} of the PE barrier above E_F, as well as the width x_F of the barrier at energy level E_F. Since tunneling is temperature independent, the emission process is termed **field emission.** The tunneling probability P was calculated in Chapter 3, and depends on Φ_{eff} and x_F through the equation[12]

$$P \approx \exp\left[\frac{-2(2m_e\Phi_{\text{eff}})^{1/2}x_F}{\hbar}\right]$$

We can easily find x_F by noting that when $x = x_F$, $PE(x_F)$ is level with E_F, as shown in Figure 4.37. From Equation 4.42, when the field is very strong, then around

[12] In Chapter 3 we showed that the transmission probability $T = T_o \exp(-2\alpha a)$ where $\alpha^2 = 2m(V_o - E)/\hbar^2$ and a is the barrier width. The pre-exponential constant T_o can be taken to be ~ 1. Clearly $V_o - E = \Phi_{\text{eff}}$ since electrons with $E = E_F$ are tunneling and $a = x_F$.

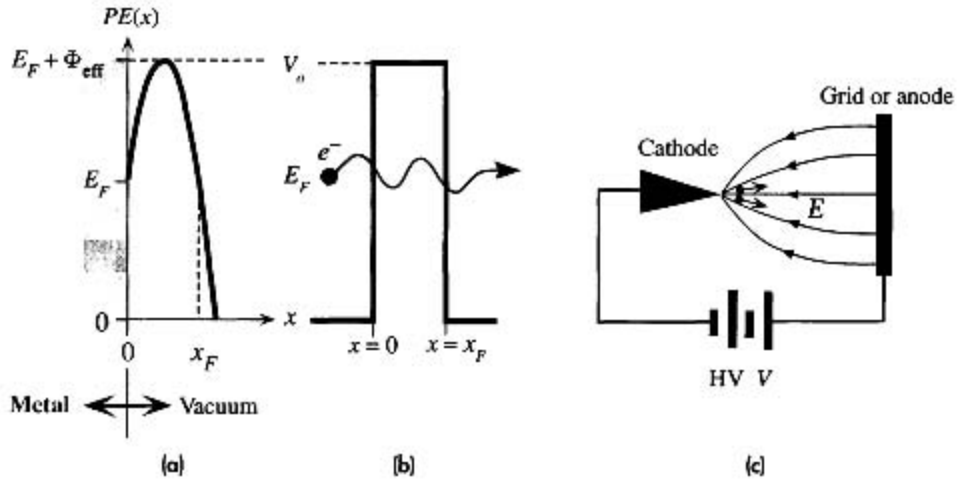


Figure 4.37

(a) Field emission is the tunneling of an electron at an energy E_F through the narrow PE barrier induced by a large applied field.

(b) For simplicity, we take the barrier to be rectangular.

(c) A sharp point cathode has the maximum field at the tip where the field emission of electrons occurs.

$x \approx x_F$ the second term is negligible compared to the third, so putting $x = x_F$ and $PE(x_F) = E_F$ in Equation 4.42 yields $\Phi = e\mathcal{E}x_F$. Substituting $x_F = \Phi/e\mathcal{E}$ in Equation 4.45, we can obtain the tunneling probability P

Field-assisted tunneling probability

$$P \approx \exp\left[-\frac{2(2m_e\Phi_{eff})^{1/2}\Phi}{e\hbar\mathcal{E}}\right] \qquad [4.45]$$

Equation 4.45 represents the probability P that an electron in the metal at E_F will tunnel out from the metal, as in Figure 4.37a and b, and become field-emitted. In a more rigorous analysis we have to consider that electrons not just at E_F but at energies below E_F can also tunnel out (though with lower probability) and we have to abandon the rough rectangular $PE(x)$ approximation in Figure 4.37b.

To calculate the current density J we have to consider how many electrons are moving toward the surface per second and per unit area, the electron flux, and then multiply this flow by the probability that they will tunnel out. The final result of the calculations is the **Fowler–Nordheim equation,** which still has the exponential field dependence in Equation 4.45,

Field-assisted tunneling: Fowler–Nordheim equation

$$J_{\text{field-emission}} \approx C\mathcal{E}^2 \exp\left(-\frac{\mathcal{E}_c}{\mathcal{E}}\right) \qquad [4.46a]$$

in which C and $\mathcal{E}_c$ are temperature-independent constants

$$C = \frac{e^3}{8\pi h\Phi} \quad \text{and} \quad \mathcal{E}_c = \frac{8\pi(2m_e\Phi^3)^{1/2}}{3eh} \qquad [4.46b]$$

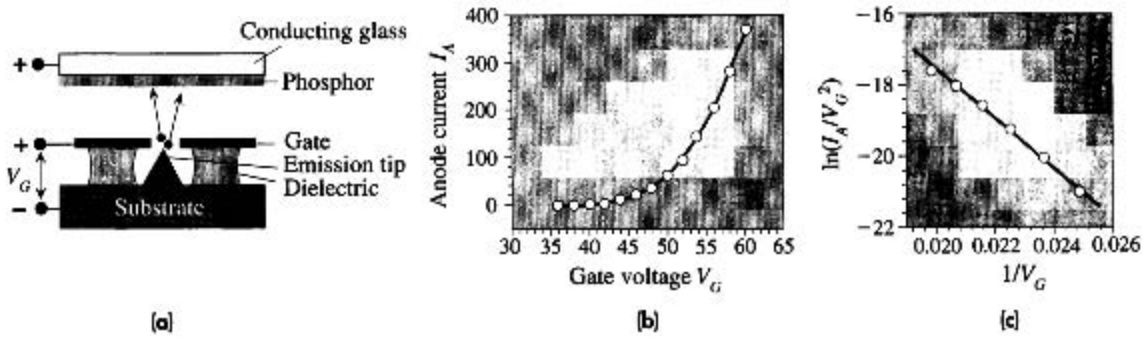


Figure 4.38

(a) Spindt-type cathode and the basic structure of one of the pixels in the FED.

(b) Emission (anode) current versus gate voltage.

(c) Fowler–Nordheim plot that confirms field emission.

that depend on the work function Φ of the metal. Equation 4.46a can also be used for field emission of electrons from a metal into an insulating material by using the electron *PE* barrier Φ_B from metal's E_F into the insulator's conduction band (where the electron is free) instead of Φ.

Notice that the field $\mathcal{E}$ in Equation 4.46a has taken over the role of temperature in thermionic emission in Equation 4.38. Since field-assisted emission depends exponentially on the field via Equation 4.46a, it can be enhanced by shaping the cathode into a cone with a sharp point where the field is maximum and the electron emission occurs from the tip as depicted in Figure 4.37c. The field $\mathcal{E}$ in Equation 4.46a is the *effective field* at the tip of the cathode that emits the electrons.

A popular field-emission tip design is based on the **Spindt tip cathode,** named after its originator. As shown in Figure 4.38a, the emission cathode is an iceberg-type sharp cone and there is a positively biased **gate** above it with a hole to extract the emitted electrons. A positively biased **anode** draws and accelerates the electrons passing through the gate toward it, which impinge on a phosphor screen to generate light by **cathodoluminescence,** a process in which light is emitted from a material when it is bombarded with electrons. Arrays of such electron field-emitters are used in field emission displays (FEDs) to generate bright images with vivid colors. Color is obtained by using red, green, and blue phosphors. The field at the tip is controlled by the potential difference between the gate and the cathode, the gate voltage V_G, which therefore controls field emission. Since $\mathcal{E} \propto V_G$, Equation 4.46a can be written to obtain the emission current or the anode current I_A as

Fowler–Nordheim anode current in a field emission device

$$I_A = aV_G^2 \exp\left(-\frac{b}{V_G}\right) \qquad [4.47]$$

where a and b are constants that depend on the particular field-emitting structure and cathode material. Figure 4.38b shows the dependence of I_A on V_G. There is a very sharp increase with the voltage once the threshold voltages (around ~45 V in Figure 4.38b) are reached to start the electron emission. Once the emission is fully operating,

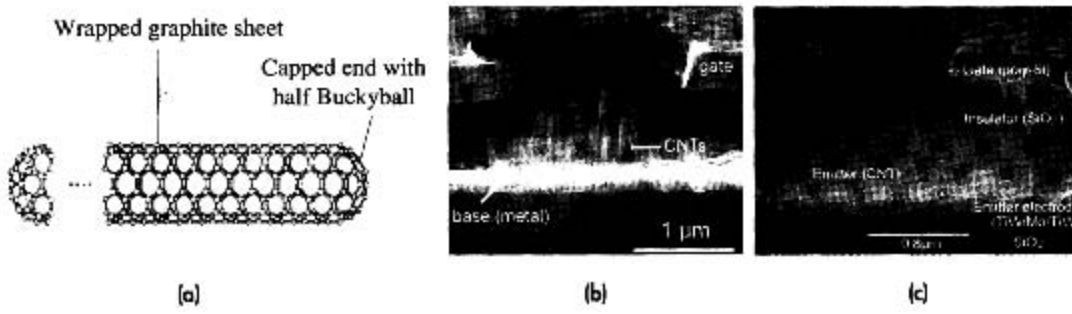


Figure 4.39

(a) A carbon nanotube (CNT) is a whisker-like, very thin and long carbon molecule with rounded ends, almost the perfect shape to be an electron field-emitter.

(b) Multiple CNTs as electron emitters.

(c) A single CNT as an emitter.

SOURCE: Courtesy of Professor W. I. Milne, University of Cambridge; G. Pirio *et al.*, *Nanotechnology*, **13**, 1, 2002.

I_A versus V_G follows the Fowler–Nordheim emission. A plot of $\ln(I_A/V_G^2)$ versus $1/V_G$ is a straight line as shown in Figure 4.38c.

Field emission has a number of distinct advantages. It is much more power efficient than thermionic emission which requires heating the cathode to high temperatures. In principle, field emission can be operated at high frequencies (fast switching times) by reducing various capacitances in the emission device or controlling the electron flow with a grid. Field emission has a number of important realized and potential applications: field emission microscopy, microwave amplifiers (high power and wide bandwidth), parallel electron beam microscopy, nanolithography, portable X-ray generators, and FEDs. For example, FEDs are thin flat displays (~2 mm thick), that have a low power consumption, quick start, and most significantly, a wide viewing angle of about 170°. Monochrome FEDs are already on the market, and color FEDs are expected to be commercialized soon, probably before the fourth edition of this text.

Typically molybdenum, tungsten, and hafnium have been used as the field-emission tip materials. Micromachining (microfabrication) has lead to the use of Si emission tips as well. Good electron emission characteristics have been also reported for diamond-like carbon films. Recently there has been a particular interest in using carbon nanotubes as emitters. A **carbon nanotube** (CNT) is a very thin filament-like carbon molecule whose diameter is in the nanometer range but whose length can be quite long, *e.g.*, 10–100 microns, depending on how it is grown or prepared. A CNT is made by rolling a graphite sheet into a tube and then capping the ends with hemispherical buckminsterfullerene molecules (a half Buckyball) as shown in Figure 4.39a. Depending on how the graphite sheet is rolled up, the CNT may be a metal or a semiconductor[13]. The high aspect ratio (length/diameter) of the CNT makes it an efficient

[13] Carbon nanotubes can be single-walled or multiwalled (when the graphite sheets are wrapped more than once) and can have quite complicated structures. There is no doubt that they possess some remarkable properties, so it is likely that CNTs will eventually be used in various engineering applications. See, for example, M. Baxendale, *J. Mater. Sci.: Mater Electron*, **14**, 657, 2003.

electron emitter. If one were to wonder what is the best shape for an efficient field emission tip, one might guess that it should be a sharp cone with some suitable apex angle. However, it turns out that the best emitter is actually a whisker-type thin filament with a rounded tip, much like a CNT. It is as if the CNT has been designed by nature to be the best field emitter. Figure 4.39b and c shows SEM photographs of two CNT Spindt-type emitters. Figure 4.39b has several CNTs, and Figure 4.39c just one CNT for electron emission. (Which is more efficient?)

EXAMPLE 4.14

FIELD EMISSION Field emission displays operate on the principle that electrons can be readily emitted from a microscopic sharp point source (*cathode*) that is biased negatively with respect to a neighboring electrode (*gate* or *grid*) as depicted in Figure 4.38a. Emitted electrons impinge on colored phosphors on a screen and cause light emission by cathodoluminescence. There are millions of these microscopic field emitters to constitute the image. A particular field emission cathode in a field-emission-type flat panel display gives a current of 61.0 μA when the voltage between the cathode and the grid is 50 V. The current is 279 μA when the voltage is 58.2 V. What is the current when the voltage is 56.2 V?

SOLUTION

Equation 4.47 related I_A to V_G,

$$I_A = aV_G^2 \exp\left(-\frac{b}{V_G}\right)$$

where a and b are constants that can be determined from the two sets of data given. Thus,

$$61.0\ \mu\text{A} = a50^2 \exp\left(-\frac{b}{50}\right) \qquad \text{and} \qquad 279\ \mu\text{A} = a58.2^2 \exp\left(-\frac{b}{58.2}\right)$$

Dividing the first by the second gives

$$\frac{61.0}{279} = \frac{50^2}{58.2^2} \exp\left[-b\left(\frac{1}{50} - \frac{1}{58.2}\right)\right]$$

which can be solved to obtain $b = 431.75$ V and hence $a = 137.25\ \mu\text{A/V}^2$. At $V = 58.2$ V,

$$I = (137.25)(56.2)^2 \exp\left(-\frac{431.75}{56.2}\right) = 200\ \mu\text{A}$$

The experimental value for this device was 202 μA, which happens to be the device in Figure 4.37b (close).

4.10 PHONONS

4.10.1 HARMONIC OSCILLATOR AND LATTICE WAVES

Quantum Harmonic Oscillator In the classical picture of a solid, the constituent atoms are held together by bonds which can be represented by springs. According to the kinetic molecular theory, the atoms in a solid are constantly vibrating about their equilibrium positions by stretching and compressing their springs. The oscillations are

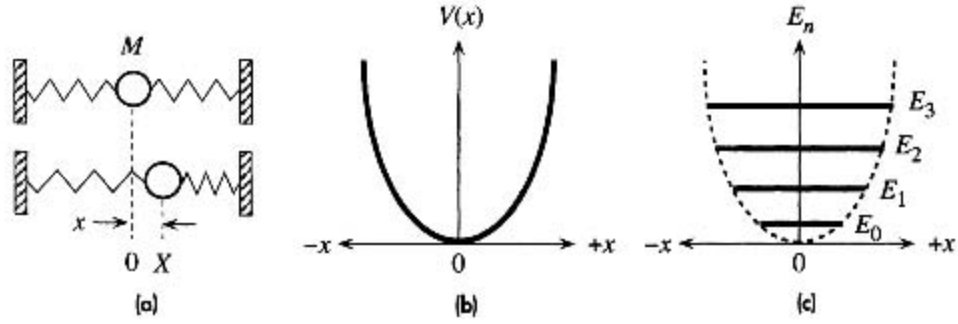


Figure 4.40

(a) Harmonic vibrations of an atom about its equilibrium position assuming its neighbors are fixed.
(b) The *PE* curve $V(x)$ versus displacement from equilibrium, x.
(c) The energy is quantized.

assumed to be simple harmonic so that the average kinetic and potential energies are the same. Figure 4.40a shows a one-dimensional independent simple harmonic oscillator that represents an atom of mass M attached by springs to fixed neighbors. The potential energy $V(x)$ is a function of displacement x from equilibrium. For small displacements, $V(x)$ is parabolic in x, as indicated in Figure 4.40b, that is,

Harmonic potential energy

$$V(x) = \frac{1}{2}\beta x^2 \qquad [4.48]$$

where β is a spring constant. The instantaneous energy, in principle, can be of any value. Equation 4.48 neglects the cubic term and is therefore symmetric about the equilibrium position at $x = 0$. It is called a **harmonic** approximation to the *PE* curve.

In modern physics, the energy of such a harmonic oscillator must be calculated using the *PE* in Equation 4.48 in the Schrödinger equation so that

Schrödinger equation: harmonic oscillator

$$\frac{d^2\psi}{dx^2} + \frac{2M}{\hbar^2}\left(E - \frac{1}{2}\beta x^2\right)\psi = 0 \qquad [4.49]$$

The solution of Equation 4.49 shows that the energy E_n of such a harmonic oscillator is quantized,

Energy of a harmonic oscillator

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega \qquad [4.50]$$

where ω is the angular frequency of the vibrations[14] and n is a quantum number $0, 1, 2, 3, \ldots$. The oscillation frequency is determined by the spring constant β and the mass M through $\omega = (\beta/M)^{1/2}$. Figure 4.40c shows the allowed energies of the quantum mechanical harmonic oscillator.

[14] Henceforth frequency will imply ω.

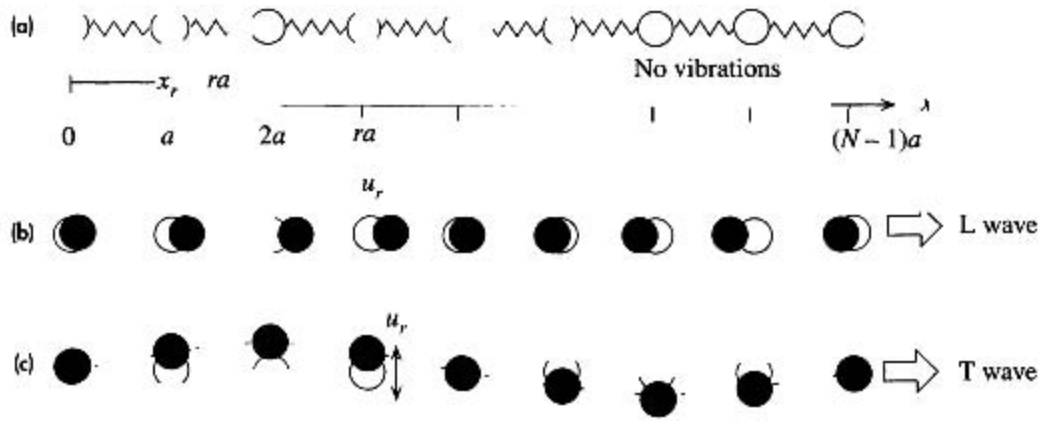


Figure 4.41

(a) A chain of N atoms through a crystal in the absence of vibrations.

(b) Coupled atomic vibrations generate a traveling longitudinal (L) wave along x. Atomic displacements (u_r) are parallel to x.

(c) A transverse (T) wave traveling along x. Atomic displacements (u_r) are perpendicular to the x axis. (b) and (c) are snapshots at one instant.

It is apparent that the minimum energy of the oscillator can never be zero but must be a finite value that is $E_0 = \frac{1}{2}\hbar\omega$. This energy is called the **zero-point energy.** As the temperature approaches 0 K, the harmonic oscillator would have an energy of E_0 and not zero. The energy levels are equally spaced by an amount $\hbar\omega$, which represents the amount of energy absorbed or emitted by the oscillator when it is excited and de-excited to a neighboring energy level. The vibrational energies of a molecule due to its atoms vibrating relative to each other, *e.g.*, the vibrations of the Cl_2 molecule in which the Cl–Cl bond is stretched and compressed, can also be described by Equation 4.50.

Phonons Atoms in a solid are coupled to each other by bonds. Atomic vibrations are therefore also coupled. These coupled vibrations lead to waves that involve cooperative vibrations of many atoms and cannot be represented by independent vibrations of individual atoms. Figure 4.41a shows a chain of atoms in a crystal. As an atom vibrates it transfers its energy to neighboring vibrating atoms and the coupled vibrations produce traveling wave-trains in the crystal.[15] (Consider grabbing and strongly vibrating the first atom in the atomic chain in Figure 4.41a. Your vibrations will be coupled and transferred by the springs to neighboring atoms in the chain along x.) Two examples are shown in Figure 4.41b and c. In the first, the atomic vibrations are parallel to the direction of propagation x and the wave is a **longitudinal wave.** In the second, the vibrations are transverse to the direction of propagation and the corresponding wave is a **transverse wave.** Suppose that x_r is the position of the rth atom in the absence of vibrations, that is, $x_r = ra$, where r is an integer from 0 to N, the number of atoms in the chain, as indicated in Figure 4.41a. By writing the mechanical equations (Newton's

[15] In the presence of coupling, the individual atoms do not execute simple harmonic motion.

second law) for the coupled atoms in Figure 4.41a, we can show that the displacement u_r from equilibrium at a location x_r is given by a **traveling-wave**-like behavior,[16]

Traveling-wave-type lattice vibrations

$$u_r = A\exp[j(Kx_r - \omega t)] \quad [4.51]$$

where A is the amplitude, K is a wavevector, and ω is the angular frequency. Notice that the Kx_r term is very much like the usual kx phase term of a traveling wave propagating in a continuous medium; the only difference is that Kx_r exists at discrete x_r locations. The wave-train described by Equation 4.51 in the crystal is called a **lattice wave.** Along the x direction it has a **wavelength** $\Lambda = 2\pi/K$ over which the longitudinal (or transverse) displacement u_r repeats itself. The displacement u_r repeats itself at one location over a time period $2\pi/\omega$. A wave traveling in the opposite direction to Equation 4.51 is of course also possible. Indeed, two oppositely traveling waves of the same frequency can interfere to set up a stationary wave which is also a lattice wave.

The lattice wave described by Equation 4.51 is a *harmonic oscillation* with a frequency ω that itself has no coupling to another lattice wave. The energy possessed by this lattice vibration is *quantized* in much the same way as the energy of the quantized harmonic oscillator in Equation 4.50. The energy of a lattice vibration therefore can only be multiples of $\hbar\omega$ above the zero-point energy, $\frac{1}{2}\hbar\omega$. The quantum of energy $\hbar\omega$ is therefore the smallest unit of lattice vibrational energy that can be added or subtracted from a lattice wave. The quantum of lattice vibration $\hbar\omega$ is called a **phonon** in analogy with the quantum of electromagnetic radiation, the photon. Whenever a lattice vibration interacts with another lattice vibration, an electron or a photon, in the crystal, it does so as if it had possessed a momentum of $\hbar K$. Thus,

Phonon energy

$$E_{\text{phonon}} = \hbar\omega = h\nu \quad [4.52]$$

Phonon momentum

$$p_{\text{phonon}} = \hbar K \quad [4.53]$$

The frequency of vibrations ω and the wavevector K of a lattice wave are related. If we were to use Equation 4.51 in the mechanical equations that describe the coupled atomic vibrations, we would find that

Dispersion relation

$$\omega = 2\left(\frac{\beta}{M}\right)^{1/2}\left|\sin\left(\frac{1}{2}Ka\right)\right| \quad [4.54]$$

which relates ω and K and is called the **dispersion relation.** Figure 4.42 shows how the frequency ω of the lattice waves increases with increasing wavevector K, or decreasing wavelength Λ. From Equation 4.54, there can be no frequencies higher than $\omega_{\max} = 2(\beta/M)^{1/2}$, which is the **lattice cut-off frequency.** Both longitudinal and transverse waves exhibit this type of dispersion relationship shown in Figure 4.42a though their exact ω–K curves would be different depending on the nature of interatomic bonding and the crystal structure. The dispersion relation in Equation 4.54 is periodic in K with a period $2\pi/a$. Only values of K in the range $-\pi/a < K < \pi/a$ are physically meaningful. A point A with K_A is the same as a point B with K_B because we can shift K by the period, $2\pi/a$ as shown in Figure 4.42a.

[16] The exponential notation for a wave is convenient, but we have to consider only the real part to actually represent the wave in the physical world.

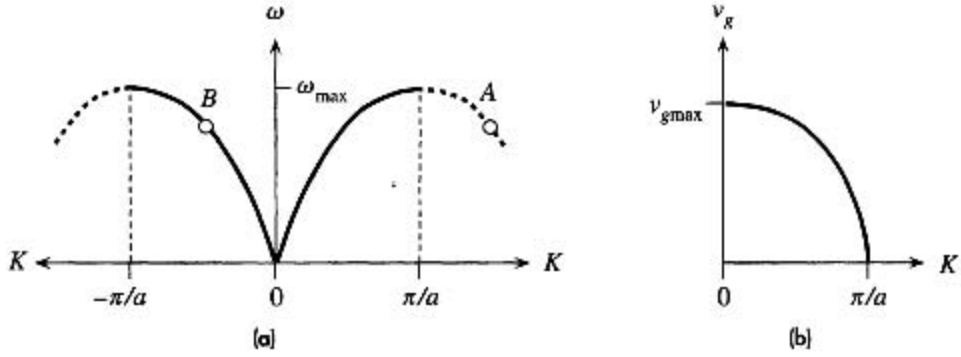


Figure 4.42
(a) Frequency ω versus wavevector K relationship for lattice waves.
(b) Group velocity v_g versus wavevector K.

The velocity at which traveling waves carry energy is called the **group velocity** v_g of the wave.[17] It depends on the slope $d\omega/dK$ of the ω–K dispersion curve, so for lattice waves,

$$v_g = \frac{d\omega}{dK} = \left(\frac{\beta}{M}\right)^{1/2} a\cos\left(\frac{1}{2}Ka\right) \qquad [4.55]$$

Group velocity

which is shown in Figure 4.42b. Points A and B in Figure 4.42a have the same group velocity and are equivalent.

The number of distinct or independent lattice waves, with different wavevectors, in a crystal is not infinite but depends on the number of atoms N. Consider a linear crystal as in Figure 4.43 with many atoms. We will take N to be large and ignore the difference between N and $N-2$. The lattice waves in this crystal would be standing waves represented by two oppositely traveling waves. The crystal length $L = Na$ can support multiples of the half-wavelength $\frac{1}{2}\Lambda$ as indicated in Figure 4.43,

$$q\frac{\Lambda}{2} = L = Na \qquad q = 1, 2, 3, \ldots \qquad [4.56a]$$

Vibrational modes

or

$$K = \frac{q\pi}{L} = \frac{q\pi}{Na} \qquad q = 1, 2, 3, \ldots \qquad [4.56b]$$

Vibrational modes

where q is an integer. Each particular K value K_q represents one distinct lattice wave with a particular frequency as determined by the dispersion relation. Four examples are shown in Figure 4.43. Each of these K_q values defines a **mode or state of lattice vibration.** Each mode is an independent lattice vibration. Its energy can be increased or decreased only by a quantum amount of $\hbar\omega$. Since K_q values outside the range $-\pi/a < K < \pi/a$ are the same as those in that range (A and B are the same

[17] For those readers who are not familiar with the group velocity concept, this is discussed in Chapter 9 without prerequisite material.

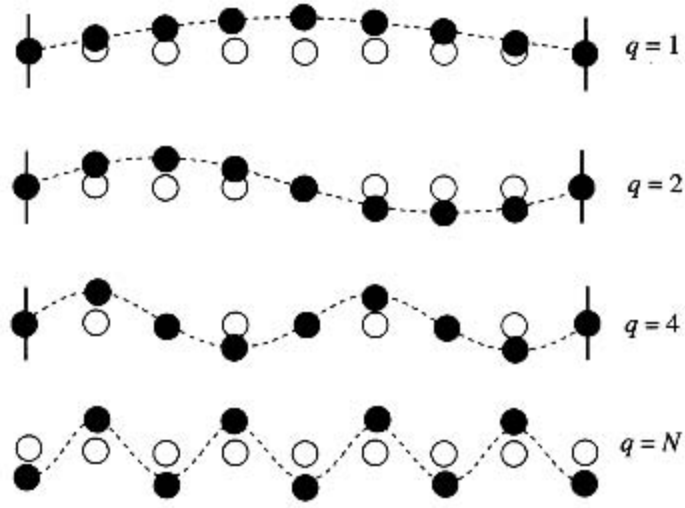


Figure 4.43 Four examples of standing waves in a linear crystal corresponding to $q = 1, 2, 4$, and N.
q is maximum when alternating atoms are vibrating in opposite directions. A portion from a very long crystal is shown.

in Figure 4.42a), it is apparent that the maximum value of q is N and thus the **number of modes** is also N. Notice that as q increases, Λ decreases. The smallest Λ occurs when alternating atoms in the crystal are moving in opposite directions which corresponds to $\frac{1}{2}\Lambda = a$, that is, $q = N$, as shown in Figure 4.43. In terms of the wavevector, $K = 2\pi/\Lambda = \pi/a$. Smaller wavelengths or longer wavevectors are meaningless and correspond to shifting K by a multiple of $2\pi/a$. Since N is large, the ω versus K curve in Figure 4.42a consists of very finely separated distinct points, each corresponding to a particular q, analogous to the energy levels in an energy band.

The above ideas for the linear chain of atoms can be readily extended to a three-dimensional crystal. If L_x, L_y, and L_z are the sides of the solid along the x, y, and z axes, with N_x, N_y, and N_z number of atoms, respectively, then the wavevector components along x, y, and z are

Lattice vibrational modes in 3-D

$$K_x = \frac{q_x \pi}{L_x} \qquad K_y = \frac{q_y \pi}{L_y} \qquad K_z = \frac{q_z \pi}{L_z} \tag{4.57}$$

where the integers q_x, q_y, and q_z run from 1 to N_x, N_y, and N_z, respectively. The total number of permitted modes is $N_x N_y N_z$ or N, the total number of atoms in the solid. Vibrations however can be set up independently along the x, y, and z directions so that the actual *number of independent modes* is $3N$.

4.10.2 DEBYE HEAT CAPACITY

The heat capacity of a solid represents the increase in the internal energy of the crystal per unit increase in the temperature. The increase in the internal energy is due to an increase in the energy of lattice vibrations. This is generally true for all the solids except metals at very low temperatures where the heat capacity is due to the electrons

near the Fermi level becoming excited to higher energies. For most practical temperature ranges of interest, the heat capacity of solids is determined by the excitation of lattice vibrations. The **molar heat capacity** C_m is the increase in the internal energy U_m of a crystal of N_A atoms per unit increase in the temperature at constant volume,[18] that is, $C_m = dU_m/dT$.

The simplest approach to calculating the average energy is first to assume that all the lattice vibrational modes have the same frequency ω. (We will account for different modes having different frequencies later.) If E_n is the energy of a harmonic oscillator such as a lattice vibration, then the average energy, by definition, is given by

$$\overline{E} = \frac{\sum_{n=0}^{\infty} E_n P(E_n)}{\sum_{n=0}^{\infty} P(E_n)} \qquad [4.58]$$

Average energy of oscillators

where $P(E_n)$ is the probability that the vibration has the energy E_n which is proportional to the Boltzmann factor. Thus we can use $P(E_n) \propto \exp(-E_n/kT)$ and $E_n = (n + \frac{1}{2})\hbar\omega$ in Equation 4.58. We can drop the zero-point energy as this does not affect the heat capacity (which deals with energy *changes*). The substitution and calculation of Equation 4.58 yields the vibrational mean energy at a frequency ω,

$$\overline{E}(\omega) = \frac{\hbar\omega}{\exp\left(\frac{\hbar\omega}{kT}\right) - 1} \qquad [4.59]$$

Average energy of oscillators at ω

This energy increases with temperature. Each phonon has an energy of $\hbar\omega$. Thus, the *phonon concentration in the crystal increases with temperature*; increasing the temperature creates more phonons.

To find the internal energy due to *all* the lattice vibrations we must also consider how many modes there are at various frequencies, that is, the distribution of the modes over the possible frequencies, the spectrum of the vibrations. Suppose that $g(\omega)$ is the number of modes per unit frequency, that is, $g(\omega)$ is the **density of vibrational states** or modes. Then $g(\omega)\,d\omega$ is the number of states in the range $d\omega$. The internal energy U_m of all lattice vibrations for 1 mole of solid is

$$U_m = \int_0^{\omega_{max}} \overline{E}(\omega) g(\omega)\, d\omega \qquad [4.60]$$

Internal energy of all lattice vibrations

The integration is up to a certain allowed maximum frequency ω_{max} (Figure 4.42a). The density of states $g(\omega)$ for the lattice vibrations can be found in a similar fashion to the density of states for electrons in an energy band, and we will simply quote the result,

$$g(\omega) \approx \frac{3V}{2\pi^2}\frac{\omega^2}{v^3} \qquad [4.61]$$

Density of states for lattice vibrations

[18] Constant volume in the definition means that the heat added to the system increases the internal energy without doing mechanical work by changing the volume.

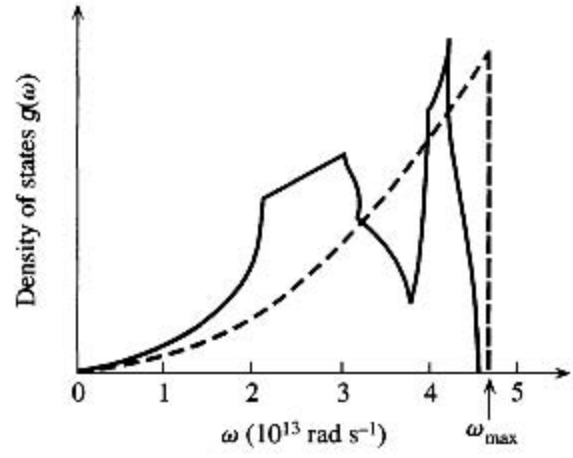


Figure 4.44 Density of states for phonons in copper. The solid curve is deduced from experiments on neutron scattering. The broken curve is the three-dimensional Debye approximation, scaled so that the areas under the two curves are the same.
This requires that $\omega_{max} \approx 4.5 \times 10^{13}$ rad s^{-1}, or a Debye characteristic temperature $T_D = 344$ K.

where v is the mean velocity of longitudinal and transverse waves in the solid and V is the volume of the crystal. Figure 4.44 shows the spectrum $g(\omega)$ for a real crystal such as Cu and the expression in Equation 4.61. The maximum frequency is ω_{max} and is determined by the fact that the total number of modes up to ω_{max} must be $3N_A$. It is called the **Debye frequency.** Thus, integrating $g(\omega)$ up to ω_{max} we find,

Debye frequency

$$\omega_{max} \approx v(6\pi^2 N_A/V)^{1/3} \qquad [4.62]$$

This maximum frequency ω_{max} corresponds to an energy $\hbar\omega_{max}$ and to a temperature T_D defined by,

Debye temperature

$$T_D = \frac{\hbar\omega_{max}}{k} \qquad [4.63]$$

and is called the **Debye temperature.** Qualitatively, it represents the temperature above which all vibrational frequencies are executed by the lattice waves.

Thus, by using Equations 4.59 to 4.63 in Equation 4.60 we can evaluate U_m and hence differentiate U_m with respect to temperature to obtain the molar heat capacity at constant volume,

Heat capacity: lattice vibrations

$$C_m = 9R\left(\frac{T}{T_D}\right)^3 \int_0^{T_D/T} \frac{x^4 e^x\, dx}{(e^x - 1)^2} \qquad [4.64]$$

which is the Debye heat capacity expression.

Figure 4.45 represents the constant-volume molar heat capacity C_m of nearly all crystals, Equation 4.64, as a function of temperature, normalized with respect to the Debye temperature. The **Dulong–Petit rule** of $C_m = 3R$ is only obeyed when $T > T_D$. Notice that C_m at $T = 0.5T_D$ is $0.825(3R)$ whereas at $T = T_D$ it is $0.952(3R)$. For most practical purposes, C_m is to within 6 percent of $3R$ when the temperature is at $0.9T_D$. For example, for copper $T_D = 315$ K and above

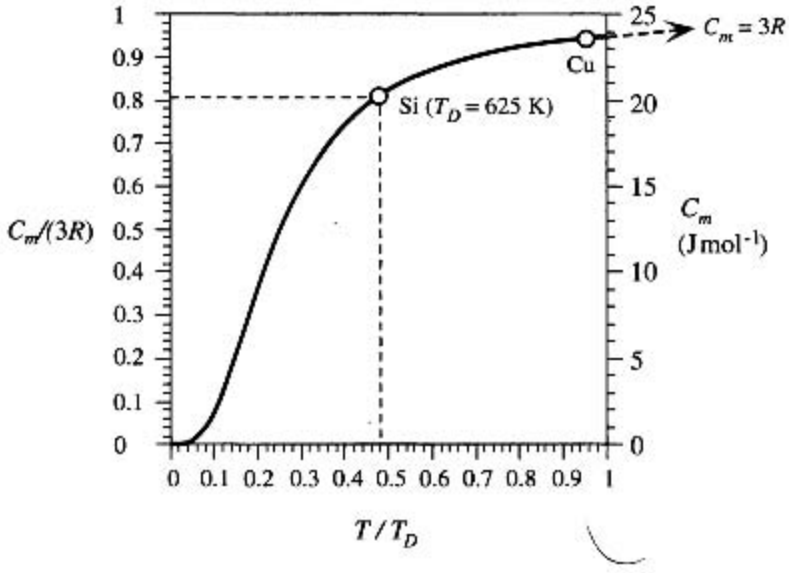


Figure 4.45 Debye constant-volume molar heat capacity curve.
The dependence of the molar heat capacity C_m on temperature with respect to the Debye temperature: C_m versus T/T_D. For Si, $T_D = 625$ K, so at room temperature (300 K), $T/T_D = 0.48$ and C_m is only 0.81 (3R).

about $0.9T_D$, that is, above 283 K (or 10 °C), $C_m \approx 3R$, as borne out by experiments.[19] Table 4.5 provides typical values for T_D, and heat capacities for a few selected elements. It is left as an exercise to check the accuracy of Equation 4.64 for predicting the heat capacity given the T_D values. At the lowest temperatures when $T \ll T_D$, Equation 4.64 predicts that $C_m \propto T^3$, and this is indeed observed in low-temperature heat capacity experiments on a variety of crystals.[20]

It is useful to provide a physical picture of the Debye model inherent in Equation 4.64. As the temperature increases from near zero, the increase in the crystal's vibrational energy is due to *more* phonons being created and *higher* frequencies being excited. The phonon concentration increases as T^3, and the mean phonon energy increases as T. Thus, the internal energy increases as T^4. At temperatures above T_D, increasing the temperature creates *more* phonons but does not increase the mean phonon energy and does not excite higher frequencies. All frequencies up to ω_{max} have now been excited. The internal energy increases only due to more phonons being created. The phonon concentration and hence the internal energy increase as T; the heat capacity is constant as expected from Equation 4.64.

[19] Sometimes it is stated that the Debye temperature is a characteristics temperature for each material at which all the atoms are able to possess vibrational kinetic energies in accordance with the Maxwell equipartition of energy principle; that is, the average vibrational kinetic energy will be $\frac{3}{2}kT$ per atom and average potential energy will also be $\frac{3}{2}kT$. This means that the average energy per atom is $3kT$, and hence the heat capacity is $3kN_A$ or $3R$ per mole which is the *Dulong–Petit rule*.

[20] Well-known exceptions are glasses, noncrystalline solids, whose heat capacity is proportional to $a_1T + a_2T^3$, where a_1 and a_2 are constants.

Table 4.5 Debye temperatures T_D, heat capacities, and thermal conductivities of selected elements

	Crystal							
	Ag	**Be**	**Cu**	**Diamond**	**Ge**	**Hg**	**Si**	**W**
T_D (K)*	215	1000	315	1860	360	100	625	310
C_m (J K^{-1} mol^{-1})†	25.6	16.46	24.5	6.48	23.38	27.68	19.74	24.45
c_v (J K^{-1} g^{-1})†	0.237	1.825	0.385	0.540	0.322	0.138	0.703	0.133
κ (W m^{-1} K^{-1})†	429	183	385	1000	60	8.65	148	173

*T_D is obtained by fitting the Debye curve to the experimental molar heat capacity data at the point $C_m = \frac{1}{2}$ (3R).
†C_m, c_v, and κ are at 25 °C.
SOURCE: T_D data from J. De Launay, *Solid State Physics*, vol. 2, F. Seitz and D. Turnbull, eds., Academic Press, New York, 1956.

It is apparent that, above the Debye temperature, the increase in temperature leads to the creation of more phonons. In Chapters 1 and 2, using classical concepts only, we had mentioned that increasing the temperature increases the magnitude of atomic vibrations. This simple and intuitive classical concept in terms of modern physics corresponds to creating more phonons with temperature. We can use the photon analogy from Chapter 3. When we increase the intensity of light of a given frequency, classically we simply increase the electric field (magnitude of the vibrations), but in modern physics we have to increase the number of photons flowing per unit area.

EXAMPLE 4.15 **SPECIFIC HEAT CAPACITY OF Si** Find the specific heat capacity c_s of a silicon crystal at room temperature given $T_D = 625$ K for Si.

SOLUTION

At room temperature, $T = 300$ K, $(T/T_D) = 0.48$, and, from Figure 4.45, the molar heat capacity is

$$C_m = 0.81(3R) = 20.2 \text{ J K}^{-1} \text{ mol}^{-1}$$

The specific heat capacity c_s from the Debye curve is

$$c_s = \frac{C_m}{M_{at}} \approx \frac{(0.81 \times 25 \text{ J K}^{-1} \text{ mol}^{-1})}{(28.09 \text{ g mol}^{-1})} = 0.72 \text{ J K}^{-1} \text{ g}^{-1}$$

The experimental value of 0.70 J K^{-1} g^{-1} is very close to the Debye value.

EXAMPLE 4.16 **SPECIFIC HEAT CAPACITY OF GaAs** Example 4.15 applied Equation 4.64, the Debye molar heat capacity C_m, to the silicon crystal in which all atoms are of the same type. It was relatively simple to calculate the specific heat capacity c_s (what is really used in engineering) from the molar heat capacity C_m by using $c_s = C_m/M_{at}$ where M_{at} is the atomic mass of the type of atom (only one) in the crystal. When the crystal has two types of atoms, we must modify the specific heat capacity derivation. We can still keep the symbol C_m to represent the Debye molar heat capacity given in Equation 4.64. Consider a GaAs crystal that has N_A units of GaAs, that is,

1 mole of GaAs. There will be 1 mole (N_A atoms) of Ga and 1 mole of As atoms. To a reasonable approximation we can assume that each mole of Ga and As contributes a C_m amount of heat capacity so that the total heat capacity of 1 mole GaAs will be $C_m + C_m$ or $2C_m$, a maximum of 50 J K^{-1} mol^{-1}. The total mass of this 1 mole of GaAs is $M_{Ga} + M_{As}$. Thus, the specific heat capacity of GaAs is

Specific heat capacity of GaAs

$$c_s = \frac{C_{total}}{M_{total}} = \frac{C_m + C_m}{M_{Ga} + M_{As}} = \frac{2C_m}{M_{Ga} + M_{As}}$$

which can alternatively be written as

Specific heat capacity of a polyatomic crystal

$$c_s = \frac{C_m}{\frac{1}{2}(M_{Ga} + M_{As})} = \frac{C_m}{\overline{M}}$$

where $\overline{M} = (M_{Ga} + M_{As})/2$ is the average atomic mass of the constituent atoms. Although we derived c_s for GaAs, it can also be applied to other compounds by suitably calculating an average atomic mass $\overline{M}$. GaAs has a Debye temperature $T_D = 344$ K, so that at a room temperature of 300 K, $T/T_D = 0.87$, and from Figure 4.45, $C_m/(3R) = 0.94$. Therefore,

$$c_s = \frac{C_m}{\overline{M}} = \frac{(0.94)(25 \text{ J K}^{-1} \text{ mol}^{-1})}{\frac{1}{2}(69.72 \text{ g mol}^{-1} + 74.92 \text{ g mol}^{-1})} = 0.325 \text{ J K}^{-1} \text{ g}^{-1}$$

At −40 °C, $T/T_D = 0.68$, and $C_m/(3R) = 0.90$, so the new $c_s = (0.90/0.94)(0.325) = 0.311$ J K^{-1} g^{-1}, which is not a large change in c_s.

The heat capacity per unit volume C_v can be found from $C_v = c_s\rho$, where ρ is the density. Thus, at 300 K, $C_v = (0.325 \text{ J K}^{-1} \text{ g}^{-1})(5.32 \text{ g cm}^{-3}) = 1.73 \text{ J K}^{-1} \text{ cm}^{-3}$. The calculated c_s match the reported experimental values very closely.

EXAMPLE 4.17

LATTICE WAVES AND SOUND VELOCITY Consider *longitudinal* waves in a linear crystal and three atoms at $r-1$, r, and $r+1$ as in Figure 4.46. The displacement of each atom from equilibrium in the $+x$ direction is u_{r-1}, u_r, and u_{r+1}, respectively. Consider the rth atom. Its bond with the left neighbor stretches by $(u_r - u_{r-1})$. Its bond with the right neighbor stretches by $(u_{r+1} - u_r)$. The left spring exerts a force $\beta(u_r - u_{r-1})$, and the right spring exerts a force $\beta(u_{r+1} - u_r)$. The net force on the rth atom is mass × acceleration,

$$\text{Net force} = \beta(u_{r+1} - u_r) - \beta(u_r - u_{r-1}) = M\frac{d^2u_r}{dt^2}$$

so

$$M\frac{d^2u_r}{dt^2} = \beta(u_{r+1} - 2u_r + u_{r-1}) \qquad [4.65]$$

Wave equation

This is the **wave equation** that describes the coupled longitudinal vibrations of the atoms in the crystal. A similar expression can also be derived for transverse vibrations. We can substitute Equation 4.51 in Equation 4.65 to show that Equation 4.51 is indeed a solution of the wave

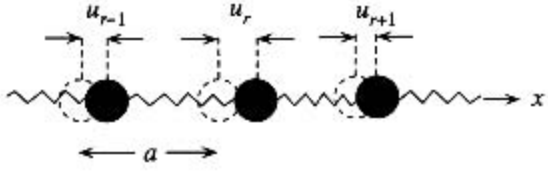


Figure 4.46 Atoms executing longitudinal vibrations parallel to x.

equation. It is assumed that the crystal response is **linear,** that is, the net force is proportional to net displacement.

The **group velocity** of lattice waves is given by Equation 4.55. For sufficiently small K, or long wavelengths, such that $\frac{1}{2}Ka \ll 1$,

Long-wavelength group velocity

$$v_g = \left(\frac{\beta}{M}\right)^{1/2} a\cos\left(\frac{1}{2}Ka\right) \approx \left(\frac{\beta}{M}\right)^{1/2} a \qquad [4.66]$$

which is a constant. It is the slope of the straight-line region of ω versus K curve for small K values in Figure 4.42. Furthermore, the elastic modulus Y depends on the slope of the net force versus displacement curve as derived in Example 1.5. From Equation 4.48 $F_N = dV/dx = \beta x$ and hence $Y = \beta/a$. Moreover, each atom occupies a volume of a^3, so the density ρ is M/a^3. Substituting both of these results in Equation 4.66 yields

Longitudinal elastic wave velocity

$$v_g \approx \left(\frac{Y}{\rho}\right)^{1/2} \qquad [4.67]$$

The relationship has to be modified for an actual crystal incorporating a small numerical factor multiplying Y. Aluminum has a density of 2.7 g cm^{-3} and $Y = 70$ GPa, so the long-wavelength longitudinal velocity from Equation 4.67 is 5092 m s^{-1}. The sound velocity in Al is 5100 m s^{-1}, which is very close.

4.10.3 THERMAL CONDUCTIVITY OF NONMETALS

In nonmetals the heat transfer involves lattice vibrations, that is, phonons. The heat absorbed in the hot region increases the amplitudes of the lattice vibrations, which is the same as generating more phonons. These new phonons travel toward the cold regions and thereby transport the lattice energy from the hot to cold end. The **thermal conductivity** κ measures the rate at which heat can be transported through a medium per unit area per unit temperature gradient. It is proportional to the rate at which a medium can absorb energy; that is, κ is proportional to the heat capacity. κ is also proportional to the rate at which phonons are transported which is determined by their mean velocity v_{ph}. In addition, of course, κ is proportional to the *mean free path* ℓ_{ph} that a phonon has to travel before losing its momentum just as the electrical conductivity is proportional to the electron's mean free path. A rigorous classical treatment gives κ as

Thermal conductivity due to phonons

$$\kappa = \tfrac{1}{3}C_v v_{ph}\ell_{ph} \qquad [4.68]$$

where C_v is the heat capacity per unit volume. The mean free path ℓ_{ph} depends on various processes that can scatter the phonons and *hinder* their propagation along the direction of heat flow. Phonons collide with other phonons, crystal defects, impurities, and crystal surfaces.

The mean phonon velocity v_{ph} is constant and approximately independent of temperature. At temperatures above the Debye temperature, C_v is constant and, thus, $\kappa \propto \ell_{ph}$. The mean free path of phonons at these temperatures is determined by phonon–phonon collisions, that is, phonons interacting with other phonons as depicted in Figure 4.47. Since the phonon concentration n_{ph} increases with temperature, $n_{ph} \propto T$, the mean free path decreases as $\ell_{ph} \propto 1/T$. Thus, κ decreases with increasing temperature as observed for most crystals at sufficiently high temperatures.

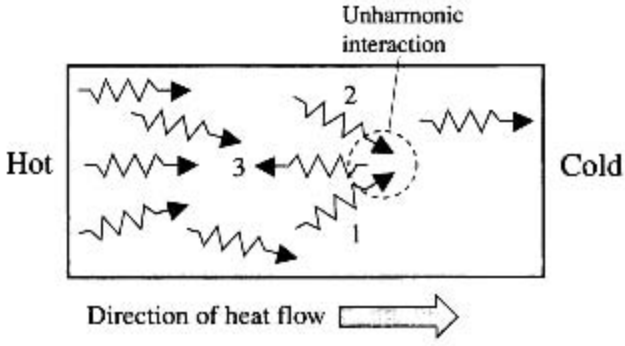


Figure 4.47 Phonons generated in the hot region travel toward the cold region and thereby transport heat energy. Phonon–phonon unharmonic interaction generates a new phonon whose momentum is toward the hot region.

The phonon–phonon collisions that are responsible for limiting the thermal conductivity, that is, scattering the phonon momentum in the opposite direction to the heat flow, are due to the **unharmonicity (asymmetry)** of the interatomic potential energy curve. Stated differently, the net force F acting on an atom is not simply βx but also has an x^2 term; it is **nonlinear.** The greater the asymmetry or nonlinearity, the larger is the effect of such momentum flipping collisions. The same asymmetry that is responsible for thermal expansion of solids is also responsible for determining the thermal conductivity. When two phonons 1 and 2 interact in a crystal region as in Figure 4.47, the *nonlinear* behavior and the *periodicity* of the lattice cause a new phonon 3 to be generated. This new phonon 3 has the same energy as the sum of 1 and 2, but it is traveling in the wrong direction! (The frequency of 3 is the sum of the frequencies of 1 and 2.)

At low temperatures there are two factors. The phonon concentration is too low for phonon–phonon collisions to be significant. Instead, the mean free path ℓ_{ph} is determined by phonon collisions with crystal imperfections, most significantly, crystal surfaces and grain boundaries. Thus, ℓ_{ph} depends on the sample geometry and crystallinity. Further, as we expect from the Debye model, C_v depends on T^3, so κ has the same temperature dependence as C_v, that is, $\kappa \propto T^3$. Between the two temperature regimes κ exhibits a peak as shown in Figure 4.48 for sapphire (crystalline Al_2O_3) and

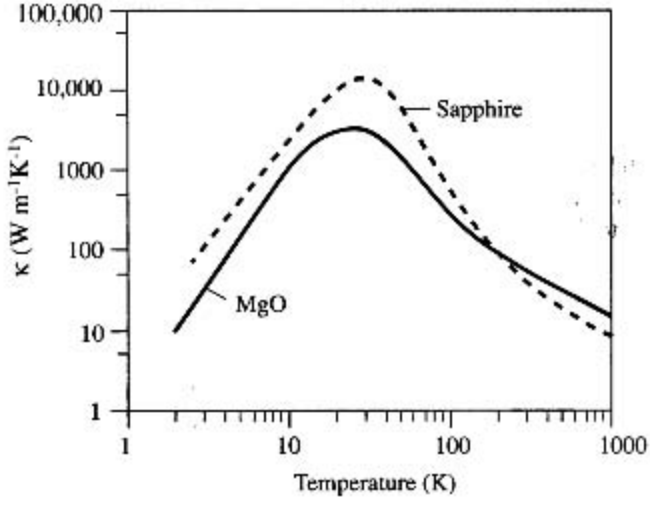


Figure 4.48 Thermal conductivity of sapphire and MgO as a function of temperature.

MgO crystals. Even though there are no conduction electrons in these two example crystals, they nonetheless exhibit substantial thermal conductivity.

EXAMPLE 4.18 **PHONONS IN GaAs** Estimate the phonon mean free path in GaAs at room temperature 300 K and at 20 K from its κ, C_v, and v_{ph}, using Equation 4.68. At room temperature, semiconductor data handbooks list the following for GaAs: $\kappa = 45\ \text{W m}^{-1}\ \text{K}^{-1}$, elastic modulus $Y = 85$ GPa, density $\rho = 5.32\ \text{g cm}^{-3}$, and specific heat capacity $c_s = 0.325\ \text{J K}^{-1}\ \text{g}^{-1}$. At 20 K, $\kappa = 4000\ \text{W m}^{-1}\ \text{K}^{-1}$ and $c_s = 0.0052\ \text{J K}^{-1}\ \text{g}^{-1}$. Y and ρ and hence v_{ph} do not change significantly with temperature compared with the changes in κ and C_v with temperature.

SOLUTION

The phonon velocity v_{ph} from Equation 4.67 is approximately

$$v_{ph} \approx \sqrt{\frac{Y}{\rho}} = \sqrt{\frac{85 \times 10^9\ \text{N m}^{-2}}{5.32 \times 10^3\ \text{kg m}^{-3}}} = 4000\ \text{m s}^{-1}$$

Heat capacity per unit volume $C_v = c_s\rho = (325\ \text{J K}^{-1}\ \text{kg}^{-1})(5320\ \text{kg m}^{-3}) = 1.73 \times 10^6\ \text{J K}^{-1}\ \text{m}^{-3}$. From Equation 4.68, $\kappa = \frac{1}{3}C_v v_{ph}\ell_{ph}$,

$$\ell_{ph} = \frac{3\kappa}{C_v v_{ph}} = \frac{(3)(45\ \text{W m}^{-1}\ \text{K}^{-1})}{(1.73 \times 10^6\ \text{J K}^{-1}\ \text{m}^{-3})(4000\ \text{m s}^{-1})} = 2.0 \times 10^{-8}\text{m} \quad \text{or} \quad 20\ \text{nm}$$

We can easily repeat the calculation at 20 K, given $\kappa \approx 4000\ \text{W m}^{-1}\ \text{K}^{-1}$ and $c_s = 5.2\ \text{J K}^{-1}\ \text{kg}^{-1}$, so $C_v = c_s\rho \approx (5.2\ \text{J K}^{-1}\ \text{kg}^{-1})(5320\ \text{kg m}^{-3}) = 2.77 \times 10^4\ \text{J K}^{-1}\ \text{m}^{-3}$. Y and ρ and hence $v_{ph}(\approx 4000\ \text{m s}^{-1})$, do not change significantly with temperature compared with κ and C_v. Thus,

$$\ell_{ph} = \frac{3\kappa}{C_v v_{ph}} \approx \frac{(3)(4 \times 10^3\ \text{W m}^{-1}\ \text{K}^{-1})}{(2.77 \times 10^4\ \text{J K}^{-1}\ \text{m}^{-3})(4000\ \text{m s}^{-1})} = 1.1 \times 10^{-4}\text{m} \quad \text{or} \quad 0.011\ \text{cm}$$

For small specimens, the above phonon mean free path will be comparable to the sample size, which means that ℓ_{ph} will actually be limited by the sample size. Consequently κ will depend on the sample dimensions, being smaller for smaller samples, similar to the dependence of the electrical conductivity of thin films on the film thickness.

4.10.4 ELECTRICAL CONDUCTIVITY

Except at low temperatures, the electrical conductivity of metals is primarily controlled by scattering of electrons around E_F by lattice vibrations, that is, phonons. These electrons have a speed $v_F = (2E_F/m_e)^{1/2}$ and a momentum of magnitude $m_e v_F$. We know that the electrical conductivity σ is proportional to the mean collision time τ of the electrons, that is, $\sigma \propto \tau$. This scattering time assumes that each scattering process is 100 percent efficient in randomizing the electron's momentum, that is, destroying the momentum gained from the field, which may not be the case. If it takes on average N collisions to randomize the electron's momentum, and τ is the mean time between the scattering events, then the *effective* scattering time is simply $N\tau$ and $\sigma \propto N\tau$. ($1/N$ indicates the efficiency of each scattering process in randomizing the velocity.)

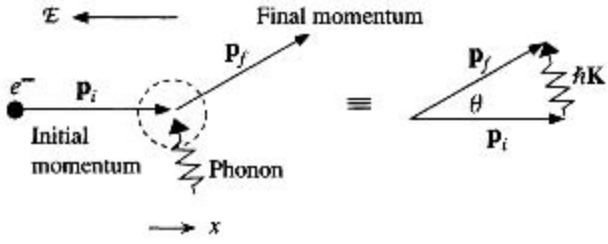


Figure 4.49 Low-angle scattering of a conduction electron by a phonon.

Figure 4.49 shows an example in which an electron with an initial momentum $\mathbf{p}_i$ collides with a lattice vibration of momentum $\hbar\mathbf{K}$. The result of the interaction is that the electron's momentum is deflected through a small angle θ to $\mathbf{p}_f$ which still has a component along the original direction x. This is called a low-angle scattering process. It will take many such collisions to reverse the electron's momentum which corresponds to flipping the momentum along the $+x$ direction to the $-x$ direction. Recall that the momentum gained from the field is actually very small compared with the momentum of the electron which is $m_e v_F$. A scattered electron must have an energy close to E_F because lower energy states are filled. Thus, $\mathbf{p}_i$ and $\mathbf{p}_f$ have approximately the same magnitude $p_i = p_f = m_e v_F$ as shown in Figure 4.49.

At temperatures above the Debye temperature, we can assume that most of the phonons are vibrating with the Debye frequency ω_{max} and the phonon concentration n_{ph} increases as T. These phonons have sufficient energies and momenta to fully scatter the electron on impact. Thus,

$$\sigma \propto \tau \propto \frac{1}{n_{ph}} \propto \frac{1}{T} \qquad [4.69a]$$

Electrical conductivity $T > T_D$

When $T < T_D$, the phonon concentration follows $n_{ph} \propto T^3$, and the mean phonon energy $\overline{E}_{ph} \propto T$, because, as the temperature is raised, higher frequencies are excited. However, these phonons have low energy and small momenta, thus they only cause small-angle scattering processes as in Figure 4.49. The average phonon momentum $\hbar K$ is also proportional to the temperature (recall that at low frequencies Figure 4.42a shows that $\hbar\omega \propto \hbar K$). It will take many such collisions, say N, to flip the electron's momentum by $2m_e v_F$ from $+m_e v_F$ to $-m_e v_F$. During each collision, a phonon of momentum $\hbar K$ is absorbed as shown in Figure 4.49. Thus, if all phonons deflected the electron in the same angular direction, the collisions would sequentially add to θ in Figure 4.49, and we will need $(2m_e v_F)/(\hbar K)$ number of steps to flip the electron's momentum. The actual collisions add θ's randomly and the process is similar to particle diffusion, random walk, in Example 1.12 ($L^2 = Na^2$, where L = displaced distance after N jumps and a = jump step). Thus,

$$N = \frac{(2m_e v_F)^2}{(\hbar K)^2} \propto \frac{1}{T^2}$$

The conductivity is therefore given by

$$\sigma \propto N\tau \propto \frac{N}{n_{ph}} \propto \frac{1}{T^5} \qquad [4.69b]$$

Electrical conductivity $T < T_D$

which is indeed observed for Cu in Figure 2.8 when $T < T_D$ over the range where impurity scattering is negligible.

ADDITIONAL TOPICS

4.11 BAND THEORY OF METALS: ELECTRON DIFFRACTION IN CRYSTALS

A rigorous treatment of the band theory of solids involves extensive quantum mechanical analysis and is beyond the scope of this book. However, we can attain a satisfactory understanding through a semiquantitative treatment.

We know that the wavefunction of the electron moving freely along x in space is a traveling wave of the spatial form $\psi_k(x) = \exp(jkx)$, where k is the wavevector $k = 2\pi/\lambda$ of the electron and $\hbar k$ is its momentum. Here, $\psi_k(x)$ represents a traveling wave because it must be multiplied by $\exp(-j\omega t)$, where $\omega = E/\hbar$, to get the total wavefunction $\Psi(x, t) = \exp[j(kx - \omega t)]$.

We will assume that an electron moving freely within the crystal and within a given energy band should also have a traveling wave type of wavefunction,

$$\psi_k(x) = A \exp(jkx) \qquad [4.70]$$

where k is the electron wavevector in the crystal and A is the amplitude. This is a reasonable expectation, since, to a first order, we can take the *PE* of the electron inside a solid as zero, $V = 0$. Yet, the *PE* must be large outside, so the electron is contained within the crystal. When the *PE* is zero, Equation 4.70 is a solution to the Schrödinger equation. The momentum of the electron described by the traveling wave Equation 4.70 is then $\hbar k$ and its energy is

$$E_k = \frac{(\hbar k)^2}{2m_e} \qquad [4.71]$$

The electron, as a traveling wave, will freely propagate through the crystal. However, not all traveling waves, can propagate in the lattice. The electron cannot have any k value in Equation 4.70 and still move through the crystal. Waves can be reflected and diffracted, whether they are electron waves, X-rays, or visible light. Diffraction occurs when reflected waves interfere constructively. Certain k values will cause the electron wave to be diffracted, preventing the wave from propagating.

The simplest illustration that certain k values will result in the electron wave being diffracted is shown in Figure 4.50 for a hypothetical linear lattice in which diffraction is simply a reflection (what we call diffraction becomes Bragg reflection). The electron is assumed to be propagating in the forward direction along x with a traveling wave function of the type in Equation 4.70. At each atom, some of this wave will be reflected. At A, the reflected wave is A' and has a magnitude A'. If the reflected waves A', B', and C' will reinforce each other, a full reflected wave will be created, traveling in the backward direction. The reflected waves A', B', C', ... will reinforce each other if the path difference between A', B', C', ... is $n\lambda$, where λ is the wavelength and $n = 1, 2, 3, \ldots$ is an integer. When wave B' reaches A', it has traveled an additional

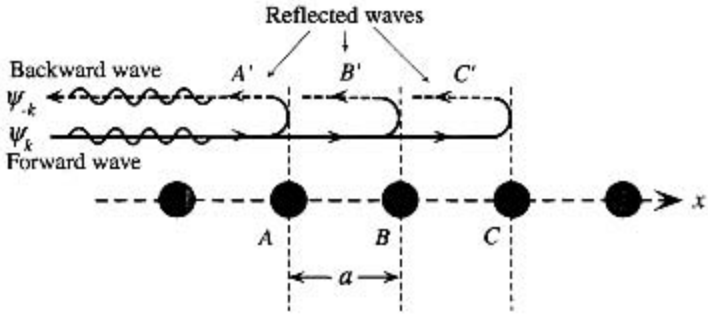


Figure 4.50 An electron wave propagation through a linear lattice.
For certain k values, the reflected waves at successive atomic planes reinforce each other, giving rise to a reflected wave traveling in the backward direction. The electron cannot then propagate through the crystal.

distance of $2a$. The path difference between A' and B' is therefore $2a$. For A' and B' to reinforce each other, we need

$$2a = n\lambda \qquad n = 1, 2, 3, \ldots$$

Substituting $\lambda = 2\pi/k$, we obtain the condition in terms of k

$$k = \frac{n\pi}{a} \qquad n = 1, 2, 3, \ldots \qquad [4.72]$$

Thus, whenever k is such that it satisfies the condition in Equation 4.72, all the reflected waves reinforce each other and produce a backward-traveling, reflected wave of the following form (with a negative k value):

$$\psi_{-k}(x) = A\exp(-jkx) \qquad [4.73]$$

This wave will also probably suffer a reflection, since its k satisfies Equation 4.72, and the reflections will continue. The crystal will then contain waves traveling in the forward and backward directions. These waves will interfere to give **standing waves** inside the crystal. Hence, whenever the k value satisfies Equation 4.72, traveling waves cannot propagate through the lattice. Instead, there can only be standing waves. For k satisfying Equation 4.72, the electron wavefunction consists of waves ψ_k and ψ_{-k} interfering in two possible ways to give two possible standing waves:

$$\psi_c(x) = A\exp(jkx) + A\exp(-jkx) = A_c\cos\left(\frac{n\pi x}{a}\right) \qquad [4.74]$$

$$\psi_s(x) = A\exp(jkx) - A\exp(-jkx) = A_s\sin\left(\frac{n\pi x}{a}\right) \qquad [4.75]$$

The probability density distributions $|\psi_c(x)|^2$ and $|\psi_s(x)|^2$ for the two standing waves are shown in Figure 4.51. The first standing wave $\psi_c(x)$ is at a maximum on the ion cores, and the other $\psi_s(x)$ is at a maximum between the ion cores. Note also that both the standing waves $\psi_c(x)$ and $\psi_s(x)$ are solutions to the Schrödinger equation.

The closer the electron is to a positive nucleus, the lower is its electrostatic PE, by virtue of $-e^2/4\pi\varepsilon_o r$. The PE of the electron distribution in $\psi_c(x)$ is lower than that in $\psi_s(x)$, because the maxima for $\psi_c(x)$ are nearer the positive ions. Therefore, the energy of the electron in $\psi_c(x)$ is lower than that of the electron in $\psi_s(x)$, or $E_c < E_s$.

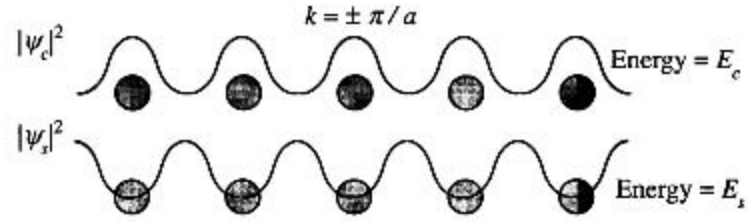


Figure 4.51 Forward and backward waves in the crystal with $k = \pm \pi/a$ give rise to two possible standing waves ψ_c and ψ_s. Their probability density distributions $|\psi_c|^2$ and $|\psi_s|^2$ have maxima either at the ions or between the ions, respectively.

It is not difficult to evaluate the energies E_c and E_s. The kinetic energy of the electron is the same in both $\psi_c(x)$ and $\psi_s(x)$, because these wavefunctions have the same k value and KE is given by $(\hbar k)^2/2m_e$. However, there is an electrostatic PE arising from the interaction of the electron with the ion cores, and this PE is different for the two wavefunctions. Suppose that $V(x)$ is the electrostatic PE of the electron at position x. We then must find the average, using the probability density distribution. Given that $|\psi_c(x)|^2\,dx$ is the probability of finding the electron at x in dx, the potential energy V_c of the electron is simply $V(x)$ averaged over the entire linear length L of the crystal. Thus, the potential energy V_c for $\psi_c(x)$ is

$$V_c = \frac{1}{L}\int_0^L V(x)|\psi_c(x)|^2\,dx = -V_n \qquad [4.76]$$

where V_n is the numerical result of the integration, which depends on $k = n\pi/a$ or n, by virtue of Equation 4.74. The integration in Equation 4.76 is a negative number that depends on n. We do not need to evaluate the integral, as we only need its final numerical result.

Using $|\psi_s(x)|^2$, we can also find V_s, the PE associated with $\psi_s(x)$. The result is that V_s is a positive quantity given by $+V_n$, where V_n is again the numerical result of the integration in Equation 4.76, which depends on n. The energies of the wavefunctions ψ_c and ψ_s whenever $k = n\pi/a$ are

$$E_c = \frac{(\hbar k)^2}{2m_e} - V_n \qquad k = \frac{n\pi}{a} \qquad [4.77]$$

$$E_s = \frac{(\hbar k)^2}{2m_e} + V_n \qquad k = \frac{n\pi}{a} \qquad [4.78]$$

Clearly, whenever k has the critical values $n\pi/a$, there are only two possible values for the energies E_c and E_s as determined by Equations 4.77 and 4.78; no other energies are allowed in between. These two energies are separated by $2V_n$.

Away from the critical k values determined by $k = n\pi/a$, the electron simply propagates as a traveling wave; the wave does not get reflected. The energy is then given by the free-running wave solution to the Schrödinger equation, that is, Equation 4.71,

$$E_k = \frac{(\hbar k)^2}{2m_e} \qquad \text{Away from } k = \frac{n\pi}{a} \qquad [4.79]$$

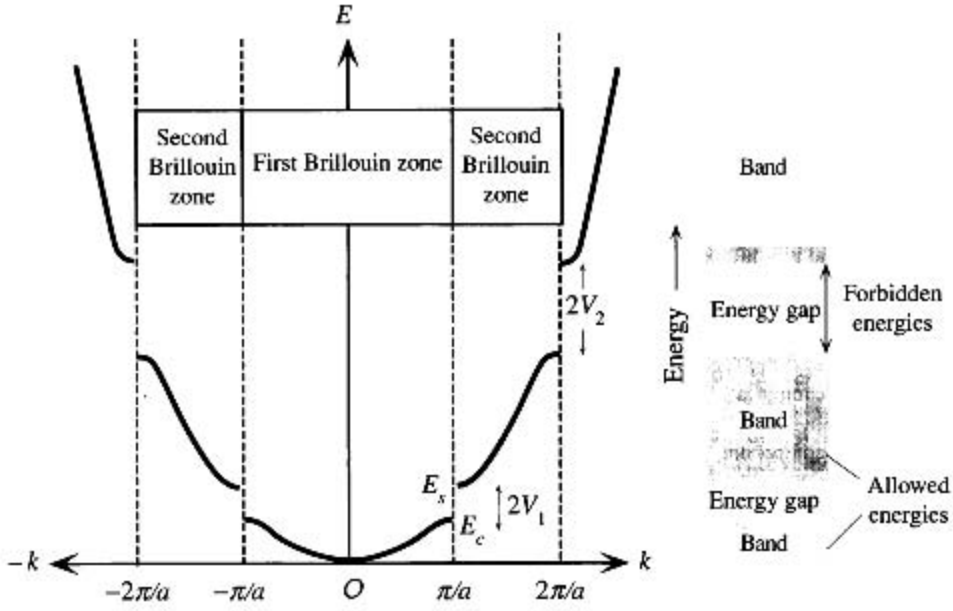


Figure 4.52 The energy of the electron as a function of its wavevector k inside a one-dimensional crystal.

There are discontinuities in the energy at $k = \pm n\pi/a$, where the waves suffer Bragg reflections in the crystal. For example, there can be no energy value for the electron between E_c and E_s. Therefore, $E_s - E_c$ is an energy gap at $k = \pm\pi/a$. Away from the critical k values, the E–k behavior is like that of a free electron, with E increasing with k as $E = (\hbar k)^2/2m_e$. In a solid, these energies fall within an energy band.

It seems that the energy of the electron increases parabolically with k along Equation 4.79 and then suddenly, at $k = n\pi/a$, it suffers a sharp discontinuity and increases parabolically again. Although the discontinuities at the critical points $k = n\pi/a$ are expected, by virtue of the Bragg reflection of waves, reflection effects will still be present to a certain extent, even within a small region around $k = n\pi/a$. The individual reflections shown in Figure 4.50 do not occur exactly at the origins of the atoms at $x = a, 2a, 3a, \ldots$. Rather, they occur over some distance, since the wave must interact with the electrons in the ion cores to be reflected. We therefore expect E–k behavior to deviate from Equation 4.79 in the neighborhood of the critical points, even if k is not exactly $n\pi/a$. Figure 4.52 shows the E–k behavior we expect, based on these arguments.

In Figure 4.52, we notice that there are certain energy ranges occurring at $k = \pm(n\pi/a)$ in which there are no allowed energies for the electron. As we saw previously, the electron cannot possess an energy between E_c and E_s at $k = \pi/a$. These energy ranges form **energy gaps** at the critical points $k = \pm(n\pi/a)$.

The range of k values from zero to the first energy gap at $k = \pm(\pi/a)$ defines a zone of k values called the **first Brillouin zone.** The zone between the first and second energy gap defines the **second Brillouin zone,** and so on. The Brillouin zone boundaries therefore identify where the energy discontinuities, or gaps, occur along the k axis.

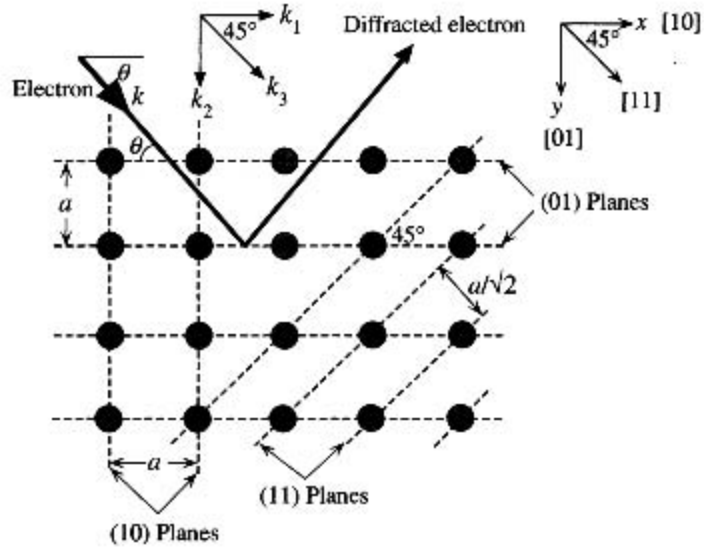


Figure 4.53 Diffraction of the electron in a two-dimensional crystal.
Diffraction occurs whenever k has a component satisfying $k_1 = \pm n\pi/a$, $k_2 = \pm n\pi/a$, or $k_3 = \pm n\pi\sqrt{2}/a$. In general terms, diffraction occurs when $k \sin\theta = n\pi/a$.

Electron motion in the three-dimensional crystal can be readily understood based on the concepts described here. For simplicity, we consider an electron propagating in a two-dimensional crystal, which is analogous, for example, to propagation in the xy plane of a crystal, as depicted in Figure 4.53. For certain k values and in certain directions, the electron will suffer diffraction and will be unable to propagate in the crystal.

Suppose that the electron's k vector along x is k_1. Whenever $k_1 = \pm n\pi/a$, the electron will be diffracted by the planes perpendicular to x, that is, the (10) planes.[21] Similarly, it will be diffracted by the (01) planes whenever its k vector along y is $k_2 = \pm n\pi/a$. The electron can also be diffracted by the (11) planes, whose separation is $a/\sqrt{2}$. If the component of k perpendicular to the (11) plane is k_3, then whenever $k_3 = \pm n\pi(\sqrt{2}/a)$, the electron will experience diffraction. These diffraction conditions can all be expressed through the **Bragg diffraction condition** $2d \sin\theta = n\lambda$, or

Bragg diffraction condition

$$k \sin\theta = \frac{n\pi}{d} \qquad [4.80]$$

where d is the interplanar separation and n is an integer; $d = a$ for (10) planes, and $d = a/\sqrt{2}$ for (11) planes.

When we plot the energy of the electron as a function of k, we must consider the direction of k, since the diffraction behavior in Equation 4.80 depends on $\sin\theta$. Along x, at $\theta = 0$, the energy gap occurs at $k = \pm(n\pi/a)$. Along $\theta = 45°$, it is at $k = \pm n\pi(\sqrt{2}/a)$, which is farther away. The E–k behavior for the electron in the two-dimensional lattice is shown in Figure 4.54 for the [10] and [11] directions. The figure shows that the first energy gap along x, in the [10] direction, is at $k = \pi/a$. Along the [11] direction, which is at 45° to the x axis, the first gap is at $k = \pi\sqrt{2}/a$.

[21] We use Miller indices in two dimensions by dropping the third digit but keeping the same interpretation. The direction along x is [10] and the plane perpendicular to x is (10).

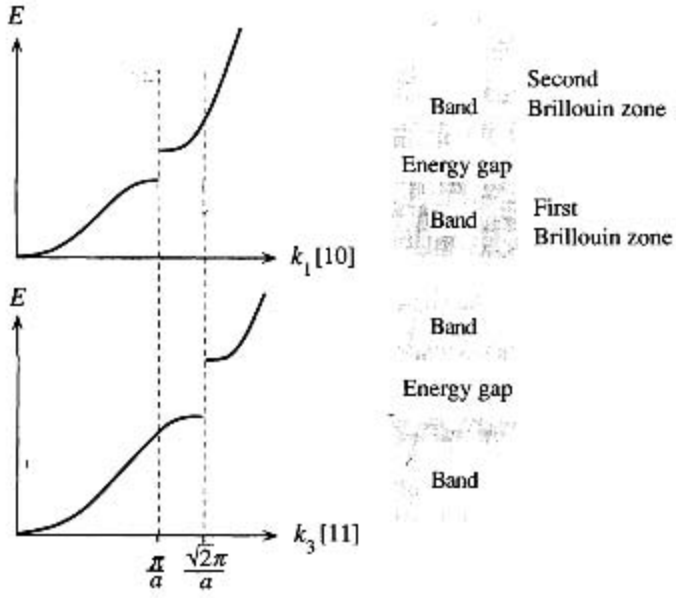


Figure 4.54 The E–k behavior for the electron along different directions in the two-dimensional crystal.

The energy gap along [10] is at π/a whereas it is at $\sqrt{2}\pi/a$ along [11].

When we consider the overlap of the energy bands along [10] and [11], in the case of a metal, there is no apparent energy gap. The electron can always find any energy simply by changing its direction.

The effects of overlap between energy bands and of energy gaps in different directions are illustrated in Figure 4.55. In the case of a semiconductor, the energy gap along [10] overlaps that along [11], so there is an overall energy gap. The electron in the semiconductor cannot have an energy that falls into this energy gap.

The first and second Brillouin zones for the two-dimensional lattice of Figure 4.53 are shown in Figure 4.56. The zone boundaries mark the occurrences of energy gaps in k space (space defined by k axes along the x and y directions). When we look at the E–k behavior, we must consider the crystal directions. This is most conveniently done by plotting energy contours in k space, as in Figure 4.57. Each contour connects all those values of k that possess the same energy. A point such as P on an energy contour gives the value of k for that energy along the direction OP. Initially, the energy contours are circles, as the energy follows $(\hbar k)^2/2m_e$ behavior, whatever the direction of k. However, near the critical values, that is, near the Brillouin zone boundaries, E increases more slowly than the parabolic relationship, as is apparent in Figure 4.52. Therefore, the circles begin to bulge as critical k values are approached. In Figure 4.57, the high-energy contours are concentrated in the corners of the zone, simply because the critical value is reached last along [11]. The energy contours do not continue smoothly across the zone boundary, because of the energy discontinuity in the E–k relationship at the boundary. Indeed, Figure 4.54 shows that the lowest energy in the second Brillouin zone may be lower than the highest energy in the first Brillouin zone.

There are two cases of interest. In the first, there is no apparent energy gap, as in Figure 4.57a, which corresponds to Figure 4.55a. The electron can have any energy

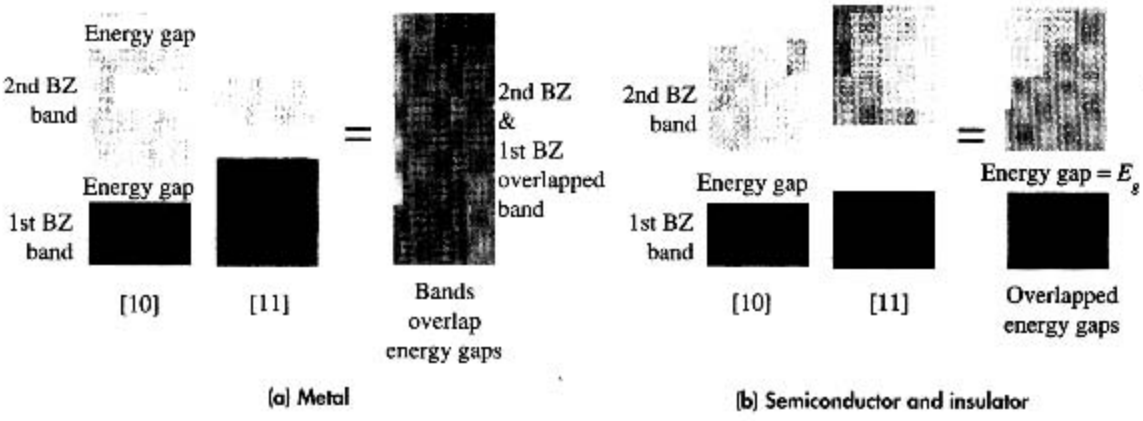


Figure 4.55

(a) For the electron in a metal, there is no apparent energy gap because the second BZ (Brillouin zone) along [10] overlaps the first BZ along [11]. Bands overlap the energy gaps. Thus, the electron can always find any energy by changing its direction.

(b) For the electron in a semiconductor, there is an energy gap arising from the overlap of the energy gaps along the [10] and [11] directions. The electron can never have an energy within this energy gap E_g.

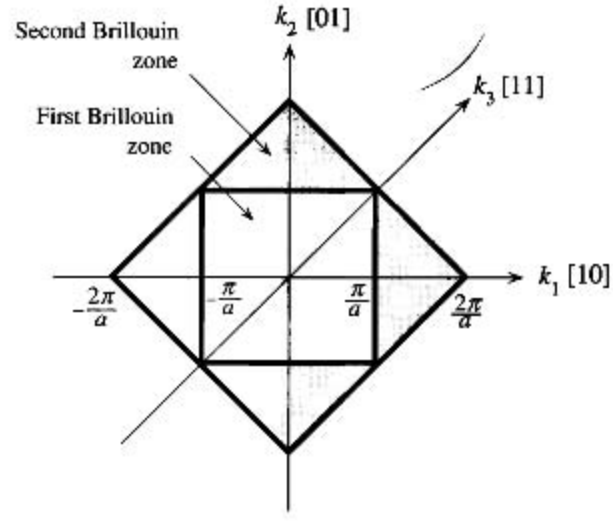


Figure 4.56 The Brillouin zones in two dimensions for the cubic lattice.

The Brillouin zones identify the boundaries where there are discontinuities in the energy (energy gaps).

value. In the second case, there is a range of energies that are not allowed, as shown in Figure 4.57b, which corresponds to Figure 4.55b.

In three dimensions, the E–k energy contour in Figure 4.57 becomes a surface in three-dimensional k space. To understand the use of such E–k contours or surfaces, consider that an E–k contour (or a surface) is made of many finely separated individual points, each representing a possible electron wavefunction ψ_k with a possible

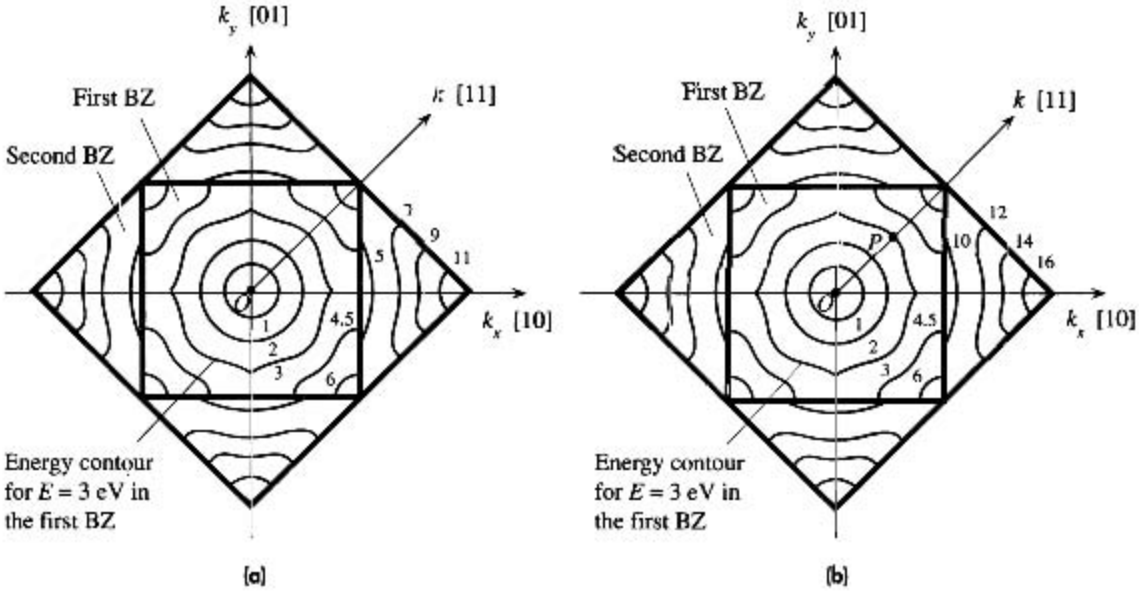


Figure 4.57 Energy contours in k space (space defined by k_x, k_y).

Each contour represents the same energy value. Any point P on the contour gives the values of k_x and k_y for that energy in that direction from O. For point P, $E = 3$ eV and OP along [11] is k.

(a) In a metal, the lowest energy in the second zone (5 eV) is lower than the highest energy (6 eV) in the first zone. There is an overlap of energies between the Brillouin zones.

(b) In a semiconductor or an insulator, there is an energy gap between the highest energy contour (6 eV) in the first zone and the lowest energy contour (10 eV) in the second zone.

energy E. At absolute zero, all the energies up to the Fermi energy are taken by the valence electrons. In k space, the energy surface, corresponding to the Fermi energy is termed the **Fermi surface.** The shape of this Fermi surface provides a means of interpreting the electrical and magnetic properties of solids.

For example, Na has one $3s$ electron per atom. In the solid, the $3s$ band is half full. The electrons take energies up to E_F, which corresponds to a spherical Fermi surface within the first Brillouin zone, as indicated in Figure 4.58a. We can then say that all the valence electrons (or nearly all) in this alkali solid exhibit an $E = (\hbar k)^2/2m_e$ type of behavior, as if they were free. When an external force is applied, such as an electric or magnetic field, we can treat the electron behavior as if it were free inside the metal with a constant mass. This is a desirable simplification for studying such metals. We can illustrate this desirability with an example. The Hall coefficient R_H derived in Chapter 2 was based on treating the electron as if it were a free particle inside the metal, or

$$R_H = -\frac{1}{en} \qquad [4.81]$$

For Na, the experimental value of R_H is -2.50×10^{-10} m^3 C^{-1}. Using the density (0.97 g cm^{-3}) and atomic mass (23) of Na and one valence electron per atom, we can

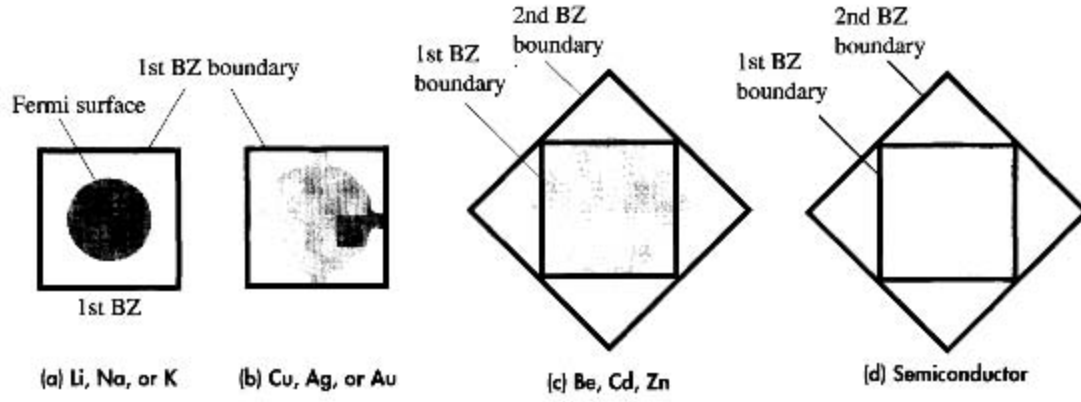


Figure 4.58 Schematic sketches of Fermi surfaces in two dimensions, representing various materials qualitatively.
(a) Monovalent group IA metals.
(b) Group IB metals.
(c) Be (Group IIA), Zn, and Cd (Group IIB).
(d) A semiconductor.

calculate $n = 2.54 \times 10^{28}$ m^{-3} and $R_H = -2.46 \times 10^{-10}$ m^3 C^{-1}, which is very close to the experimental value.

In the case of Cu, Ag, and Au (the IB metals in the Periodic Table), the Fermi surface is inside the first Brillouin zone, but it is not spherical as depicted in Figure 4.58b. Also, it touches the centers of the zone boundaries. Some of those electrons near the zone boundary behave quite differently than $E = (\hbar k)^2/2m_e$, although the majority of the electrons in the sphere do exhibit this type of behavior. To an extent, we can expect the free electron derivations to hold. The experimental value of R_H for Cu is -0.55×10^{-10} m^3 C^{-1}, whereas the expected value, based on Equation 4.81 with one electron per atom, is -0.73×10^{-10} m^3 C^{-1}, which is noticeably greater than the experimental value.

The divalent metals Be, Mg, and Ca have closed outer s subshells and should have a full s band in the solid. Recall that electrons in a full band cannot respond to an applied field and drift. We also know that there should be an overlap between the s and p bands, forming one partially filled continuous energy band, so these metals are indeed conductors. In terms of Brillouin zones, their structure is based on Figure 4.55a, which has the second zone overlapping the first Brillouin zone. The Fermi surface extends into the second zone and the corners of the first zone are empty, as depicted in Figure 4.58c. Since there are empty energy levels next to the Fermi surface, the electrons can gain energy and drift in response to an applied field. But the surface is not spherical; indeed, near the corners of the first zone, it even has the wrong curvature. Therefore, it is no longer possible to describe these electrons on the Fermi surface as obeying $E = (\hbar k)^2/2m_e$. When a magnetic field is applied to a drifting electron to bend its trajectory, its total behavior is different than that expected when it is acting as a free particle. The external force changes the momentum $\hbar k$ and the corresponding

change in the energy depends on the Fermi surface and can be quite complicated. To finish the example on the Hall coefficient, we note that based on two valence electrons per atom (Group IIA), the Hall coefficient for Be should be -0.25×10^{-10} m^3 C^{-1}, but the measured value is a positive coefficient of $+2.44 \times 10^{-10}$ m^3 C^{-1}. Equation 4.81 is therefore useless. It seems that the electrons moving at the Fermi surface of Be are equivalent to the motion of positive charges (like holes), so the Hall effect registers a positive coefficient.

The Fermi surface of a semiconductor is simply the boundary of the first Brillouin zone, because there is an energy gap between the first and the second Brillouin zones, as depicted in Figure 4.55b. In a semiconductor, all the energy levels up to the energy gap are taken up by the valence electrons. The first Brillouin zone forms the valence band and the second forms the conduction band.

4.12 GRÜNEISEN'S MODEL OF THERMAL EXPANSION

We considered thermal expansion in Section 1.4.2 where the principle is illustrated in Figure 1.18, which shows the potential energy curve $U(r)$ for two atoms separated by a distance r in a crystal. At temperature T_1 we know that the atoms will be vibrating about their equilibrium positions between positions B and C, compressing (B) and stretching (C) the bond between them. The line BC corresponds to the total energy E of the pair of atoms. The average separation at T_1 is at A, halfway between B and C. We also know that the PE curve $U(r)$ is *asymmetric*, and it is this asymmetry that leads to the phenomenon of thermal expansion. When the temperature increases from T_1 to T_2, the atoms vibrate between B' and C' and the average separation between the atoms also increases, from A to A', which we identified as *thermal expansion*. If the PE curve were symmetric, then there would be no thermal expansion.

Since the linear expansion coefficient λ is related to the shape of the PE curve, $U(r)$, it is also related to the elastic bulk modulus K that measures how difficult it is to stretch or compress the bonds. K depends on $U(r)$ in the same way that the elastic modulus Y depends on $U(r)$ as explained in Example 1.5.[22] Further, λ also depends on the amount of increase from BC to $B'C'$ per degree of increase in the temperature. λ must therefore also depend on the heat capacity. When the temperature increases by a small amount δT, the energy per atom increases by $(C_v\, \delta T)/N$ where C_v is the heat capacity per unit volume and N is the number of atoms per unit volume. If $C_v\, \delta T$ is large, then the line $B'C'$ in Figure 1.18 will be higher up on the energy curve and the average separation A' will therefore be larger. Thus, the larger is the heat capacity, the greater is the interatomic separation, which means $\lambda \propto C_v$. Further, the average separation, point A, depends on how much the bonds are stretched and compressed. For large

[22] K is a measure of the elastic change in the volume of a body in response to an applied pressure; large K means a small change in volume for a given pressure. Y is a measure of the elastic change in the length of the body in response to an applied stress; large Y means a small change in length. Both involve stretching or compressing bonds.

amounts of displacement from equilibrium, the average A will be greater as more asymmetry of the PE curve is used. Thus, the smaller is the elastic modulus K, the greater is λ; we see that $\lambda \propto C_v/K$.

Asymmetric potential energy curve

If we were to expand $U(r)$ about its minimum value $U_{\min}$ at $r = r_o$, we would obtain the Taylor expansion,

$$U(r) = U_{\min} + a_2(r - r_o)^2 + a_3(r - r_o)^3 + \cdots$$

where a_2 and a_3 are coefficients related to the second and third derivatives of U at r_o. The term $(r - r_o)$ is missing because we are expanding a series about $U_{\min}$ where $dU/dr = 0$. The $U_{\min}$ and the $a_2(r - r_o)^2$ term give a parabola about $U_{\min}$ which is a symmetric curve around r_o and therefore does not lead to thermal expansion. It is the a_3 term that gives the expansion because it leads to asymmetry. Thus the amount of expansion λ also depends on the amount of asymmetry with respect to symmetry, that is a_3/a_2. Thus,

Linear expansion coefficient

$$\lambda \propto \frac{a_3}{a_2}\frac{C_v}{K}$$

The ratio of a_3 and a_2 depends on the nature of the bond. A simplified analytical treatment (beyond the scope of this book) gives λ as

Grüneisen's law

$$\lambda \approx 3\gamma \frac{C_v}{K} \qquad [4.82]$$

where γ is a "constant" called the *Grüneisen parameter*. The Grüneisen constant γ is approximately $-(r_o a_3)/(2a_2)$ where r_o is the equilibrium atomic separation, and thus γ represents the asymmetry of the energy curve. The approximate equality simply emphasizes the number of assumptions that are typically made in deriving Equation 4.82. The Grüneisen parameter γ is of the order of unity for many materials; experimentally, $\gamma = 0.1 - 1$. We can also write the Grüneisen law in terms of the molar heat capacity C_m (heat capacity per mole) or the specific heat capacity c_s (heat capacity per unit mass). If ρ is the density, and M_{at} is the atomic mass of the constituent atoms of the crystal, then

Grüneisen's law

$$\lambda = 3\gamma \frac{\rho C_m}{M_{at} K} = 3\gamma \frac{\rho c_s}{K} \qquad [4.83]$$

We can calculate the Grüneisen parameter γ for materials that possess different types of interatomic bonding and thereby obtain typical values for γ. This would also expose the extent of unharmonicity in the bonding. Given the experimental values for λ, K, ρ and c_s, the Grüneisen parameters have been calculated from Equation 4.83 and are listed in Table 4.6. An interesting feature of the results is that the experimental γ values, within a factor of 2–3, are about the same, at least to an order of magnitude. Equation 4.83 also indicates that the λ versus T behavior should resemble the C_v versus T dependence, which is approximately the case if one compares Figure 1.20 with Figure 4.45. (K does not change much with temperature.) There is one notable difference. At very low temperatures λ can change sign and become negative for certain crystals, whereas C_v cannot.

Table 4.6 The Grüneisen parameter for some selected materials with different types of interatomic bonding

Material	ρ (g cm^{-3})	λ ($\times 10^{-6}$ K^{-1})	K (GPa)	c_v (J kg^{-1} K^{-1})	γ
Iron (metallic, BCC)	7.9	12.1	170	444	0.20
Copper (metallic, FCC)	8.96	17	140	380	0.23
Germanium (covalent)	5.32	6	77	322	0.09
Glass (covalent-ionic)	2.45	8	70	800	0.10
NaCl (ionic)	2.16	39.5	28	880	0.19
Tellurium (mixed)	6.24	18.2	40	202	0.19
Polystyrene (van der Waals)	1.05	100	3	1200	0.08

CD Selected Topics and Solved Problems

Selected Topics	Solved Problems
Hall Effect	The Water Molecule
Thermal Conductivity	
Thermoelectric Effects in Metals: Thermocouples	
Thermal Expansion (Grüneisen's Law)	

DEFINING TERMS

Average energy E_{av} of an electron in a metal is determined by the Fermi–Dirac statistics and the density of states. It increases with the Fermi energy and also with the temperature.

Boltzmann statistics describes the behavior of a collection of particles (*e.g.*, gas atoms) in terms of their energy distribution. It specifies the number of particles $N(E)$ with given energy, through $N(E) \propto \exp(-E/kT)$, where k is the Boltzmann constant. The description is nonquantum mechanical in that there is no restriction on the number of particles that can have the same state (the same wavefunction) with an energy E. Also, it applies when there are only a few particles compared to the number of possible states, so the likelihood of two particles having the same state becomes negligible. This is generally the case for thermally excited electrons in the conduction band of a semiconductor, where there are many more states than electrons. The kinetic energy distribution of gas molecules in a tank obeys the Boltzmann statistics.

Cathode is a negative electrode. It emits electrons or attracts positive charges, that is, cations.

Debye frequency is the maximum frequency of lattice vibrations that can exist in a particular crystal. It is the cut-off frequency for lattice vibrations.

Debye temperature is a characteristic temperature of a particular crystal above which nearly all the atoms are vibrating in accordance with the kinetic molecular theory, that is, each atom has an average energy (potential + kinetic) of $3kT$ due to atomic vibrations, and the heat capacity is determined by the Dulong–Petit rule.

Density of states $g(E)$ is the number of electron states [*e.g.*, wavefunctions, $\psi(n, \ell, m_\ell, m_s)$] per unit energy per unit volume. Thus, $g(E)\,dE$ is the number of states in the energy range E to $(E + dE)$ per unit volume.

Density of vibrational states is the number of lattice vibrational modes per unit angular frequency range.

Dispersion relation relates the angular frequency ω and the wavevector K of a wave. In a crystal lattice, the coupling of atomic oscillations leads to a particular relationship between ω and K which determines the allowed lattice waves and their group velocities. The dispersion relation is specific to the crystal structure, that is, it depends on the lattice, basis, and bonding.

Effective electron mass m_e^* represents the inertial resistance of an electron inside a crystal against an acceleration imposed by an external force, such as the applied electric field. If $F_{ext} = eE_x$ is the external applied force due to the applied field $\mathcal{E}_x$, then the effective mass m_e^* determines the acceleration a of the electron by $eE_x = m_e^* a$. This takes into account the effect of the internal fields on the motion of the electron. In vacuum where there are no internal fields, m_e^* is the mass in vacuum m_e.

Fermi–Dirac statistics determines the probability of an electron occupying a state at an energy level E. This takes into account that a collection of electrons must obey the Pauli exclusion principle. The Fermi–Dirac function quantifies this probability via $f(E) = 1/\{1 + \exp[(E - E_F)/kT]\}$, where E_F is the Fermi energy.

Fermi energy is the maximum energy of the electrons in a metal at 0 K.

Field emission is the tunneling of an electron from the surface of a metal into vacuum, due to the application of a strong electric field (typically $\mathcal{E} > 10^9$ V m^{-1}).

Group velocity is the velocity at which traveling waves carry energy. If ω is the angular frequency and K is the wavevector of a wave, then the group velocity $v_g = d\omega/dK$.

Harmonic oscillator is an oscillating system, for example, two masses joined by a spring, that can be described by *simple harmonic motion*. In quantum mechanics, the energy of a harmonic oscillator is quantized and can only increase or decrease by a discrete amount $\hbar\omega$. The minimum energy of a harmonic oscillator is not zero but $\frac{1}{2}\hbar\omega$ (see **zero-point energy**).

Lattice wave is a wave in a crystal due to coupled oscillations of the atoms. Lattice waves may be traveling or stationary waves.

Linear combination of atomic orbitals (LCAO) is a method for obtaining the electron wavefunction in the molecule from a linear combination of individual atomic wavefunctions. For example, when two H atoms A and B come together, the electron wavefunctions, based on LCAO, are

$$\psi_a = \psi_{1s}(A) + \psi_{1s}(B)$$
$$\psi_b = \psi_{1s}(A) - \psi_{1s}(B)$$

where $\psi_{1s}(A)$ and $\psi_{1s}(B)$ are atomic wavefunctions centered around the H atoms A and B, respectively. The ψ_a and ψ_b represent molecular orbital wavefunctions for the electron; they reflect the behavior of the electron, or its probability distribution, in the molecule.

Mode or **state of lattice vibration** is a distinct, independent way in which a crystal lattice can vibrate with its own particular frequency ω and wavevector K. There are only a finite number of vibrational modes in a crystal.

Molecular orbital wavefunction, or simply molecular orbital, is a wavefunction for an electron within a system of two or more nuclei (*e.g.*, molecule). A molecular orbital determines the probability distribution of the electron within the molecule, just as the atomic orbital determines the electron's probability distribution within the atom. A molecular orbital can take two electrons with opposite spins.

Orbital is a region of space in an atom or molecule where an electron with a given energy may be found. An orbit, which is a well-defined path for an electron, cannot be used to describe the whereabouts of the electron in an atom or molecule because the electron has a probability distribution. Orbitals are generally represented by a surface within which the total probability is high, for example, 90 percent.

Orbital wavefunction, or simply orbital, describes the spatial dependence of the electron. The orbital is $\psi(r, \theta, \phi)$, which depends on n, ℓ, and m_ℓ, and the spin dependence m_s is excluded.

Phonon is a quantum of lattice vibrational energy of magnitude $\hbar\omega$, where ω is the vibrational angular frequency. A phonon has a momentum $\hbar K$ where K is the wavevector of the lattice wave.

Seebeck effect is the development of a built-in potential difference across a material as a result of a temperature gradient. If dV is the built-in potential across a

temperature difference dT, then the Seebeck coefficient S is defined as $S = dV/dT$. The coefficient gauges the magnitude of the Seebeck effect. Only the net Seebeck voltage difference between different metals can be measured. The principle of the thermocouple is based on the Seebeck effect.

State is a possible wavefunction for the electron that defines its spatial (orbital) and spin properties, for example, $\psi(n, \ell, m_\ell, m_s)$ is a state of the electron. From the Schrödinger equation, each state corresponds to a certain electron energy E. We thus speak of a state with energy E, state of energy E, or even an energy state. Generally there may be more than one state ψ with the same energy E.

Thermionic emission is the emission of electrons from the surface of a heated metal.

Work function is the minimum energy needed to free an electron from the metal at a temperature of absolute zero. It is the energy separation of the Fermi level from the vacuum level.

Zero-point energy is the minimum energy of a harmonic oscillator $\frac{1}{2}\hbar\omega$. Even at 0 K, an oscillator in quantum mechanics will have a finite amount of energy which is its zero-point energy. Heisenberg's uncertainty principle does not allow a harmonic oscillator to have zero energy because that would mean no uncertainty in the momentum and consequently an infinite uncertainty in space ($\Delta p_x\, \Delta x > \hbar$).

QUESTIONS AND PROBLEMS

4.1 **Phase of an atomic orbital**

a. What is the functional form of a $1s$ wavefunction $\psi(r)$? Sketch schematically the atomic wavefunction $\psi_{1s}(r)$ as a function of distance from the nucleus.

b. What is the total wavefunction $\Psi_{1s}(r, t)$?

c. What is meant by two wavefunctions $\Psi_{1s}(A)$ and $\Psi_{1s}(B)$ that are out of phase?

d. Sketch schematically the two wavefunctions $\Psi_{1s}(A)$ and $\Psi_{1s}(B)$ at one instant.

4.2 **Molecular orbitals and atomic orbitals** Consider a linear chain of four identical atoms representing a hypothetical molecule. Suppose that each atomic wavefunction is a $1s$ wavefunction. This system of identical atoms has a center of symmetry C with respect to the center of the molecule (midway between the second and the third atom), and all molecular wavefunctions must be either symmetric or antisymmetric about C.

a. Using the LCAO principle, sketch the possible molecular orbitals.

b. Sketch the probability distributions $|\psi|^2$.

c. If more nodes in the wavefunction lead to greater energies, order the energies of the molecular orbitals.

Note: The electron wavefunctions, and the related probability distributions, in a simple potential energy well that are shown in Figure 3.15 can be used as a rough *guide* toward finding the appropriate molecular wavefunctions in the four-atom symmetric molecule. For example, if we were to smooth the electron potential energy in the four-atom molecule into a constant potential energy, that is, generate a potential energy well, we should be able to modify or distort, without flipping, the molecular orbitals to somewhat resemble ψ_1 to ψ_4 sketched in Figure 3.15. Consider also that the number of nodes increases from none for ψ_1 to three for ψ_4 in Figure 3.15.

4.3 **Diamond and tin** Germanium, silicon, and diamond have the same crystal structure, that of diamond. Bonding in each case involves sp^3 hybridization. The bonding energy decreases as we go from C to Si to Ge, as noted in Table 4.7.

a. What would you expect for the bandgap of diamond? How does it compare with the experimental value of 5.5 eV?

b. Tin has a tetragonal crystal structure, which makes it different than its group members, diamond, silicon, and germanium.

1. Is it a metal or a semiconductor?
2. What experiments do you think would expose its semiconductor properties?

Table 4.7

Property	Diamond	Silicon	Germanium	Tin
Melting temperature, °C	3800	1417	937	232
Covalent radius, nm	0.077	0.117	0.122	0.146
Bond energy, eV	3.60	1.84	1.7	1.2
First ionization energy, eV	11.26	8.15	7.88	7.33
Bandgap, eV	?	1.12	0.67	?

4.4 **Compound III–V Semiconductors** Indium as an element is a metal. It has a valency of III. Sb as an element is a metal and has a valency of V. InSb is a semiconductor, with each atom bonding to four neighbors, just like in silicon. Explain how this is possible and why InSb is a semiconductor and not a metal alloy. (Consider the electronic structure and sp^3 hybridization for each atom.)

4.5 **Compound II–VI semiconductors** CdTe is a semiconductor, with each atom bonding to four neighbors, just like in silicon. In terms of covalent bonding and the positions of Cd and Te in the Periodic Table, explain how this is possible. Would you expect the bonding in CdTe to have more ionic character than that in III–V semiconductors?

***4.6** **Density of states for a two-dimensional electron gas** Consider a two-dimensional electron gas in which the electrons are restricted to move freely within a square area a^2 in the xy plane. Following the procedure in Section 4.5, show that the density of states $g(E)$ is constant (independent of energy).

4.7 **Fermi energy of Cu** The Fermi energy of electrons in copper at room temperature is 7.0 eV. The electron drift mobility in copper, from Hall effect measurements, is 33 $cm^2\ V^{-1}\ s^{-1}$.

a. What is the speed v_F of conduction electrons with energies around E_F in copper? By how many times is this larger than the average thermal speed $v_{thermal}$ of electrons, if they behaved like an ideal gas (Maxwell–Boltzmann statistics)? Why is v_F much larger than $v_{thermal}$?

b. What is the De Broglie wavelength of these electrons? Will the electrons get diffracted by the lattice planes in copper, given that interplanar separation in Cu = 2.09 Å? (Solution guide: Diffraction of waves occurs when $2d \sin\theta = \lambda$, which is the Bragg condition. Find the relationship between λ and d that results in $\sin\theta > 1$ and hence no diffraction.)

c. Calculate the mean free path of electrons at E_F and comment.

4.8 **Free electron model, Fermi energy, and density of states** Na and Au both are valency I metals; that is, each atom donates one electron to the sea of conduction electrons. Calculate the Fermi energy (in eV) of each at 300 K and 0 K. Calculate the mean speed of all the conduction electrons and also the speed of electrons at E_F for each metal. Calculate the density of states as states per eV cm^{-3} at the Fermi energy and also at the center of the band, to be taken at $(E_F + \Phi)/2$. (See Table 4.1 for Φ.)

4.9 **Fermi energy and electron concentration** Consider the metals in Table 4.8 from Groups I, II, and III in the Periodic Table. Calculate the Fermi energies at absolute zero, and compare the values with the experimental values. What is your conclusion?

Table 4.8

Metal	Group	M_M	Density (g cm^{-3})	E_F (eV) [Calculated]	E_F (eV) [Experiment]
Cu	I	63.55	8.96	—	6.5
Zn	II	65.38	7.14	—	11.0
Al	III	27	2.70	—	11.8

4.10 Temperature dependence of the Fermi energy

a. Given that the Fermi energy for Cu is 7.0 eV at absolute zero, calculate the E_F at 300 K. What is the percentage change in E_F and what is your conclusion?

b. Given the Fermi energy for Cu at absolute zero, calculate the average energy and mean speed per conduction electron at absolute zero and 300 K, and comment.

4.11 X-ray emission spectrum from sodium Structure of the Na atom is [Ne]$3s^1$. Figure 4.59a shows the formation of the $3s$ and $3p$ energy bands in Na as a function of internuclear separation. Figure 4.59b shows the X-ray emission spectrum (called the L-band) from crystalline sodium in the soft X-ray range as explained in Example 4.6.

a. From Figure 4.59a, estimate the nearest neighbor equilibrium separation between Na atoms in the crystal if some electrons in the $3s$ band spill over into the states in the $3p$ band.

b. Explain the origin of the X-ray emission band in Figure 4.59b and the reason for calling it the L-band.

c. What is the Fermi energy of the electrons in Na from Figure 4.59b?

d. Taking the valency of Na to be I, what is the expected Fermi energy and how does it compare with that in part (c)?

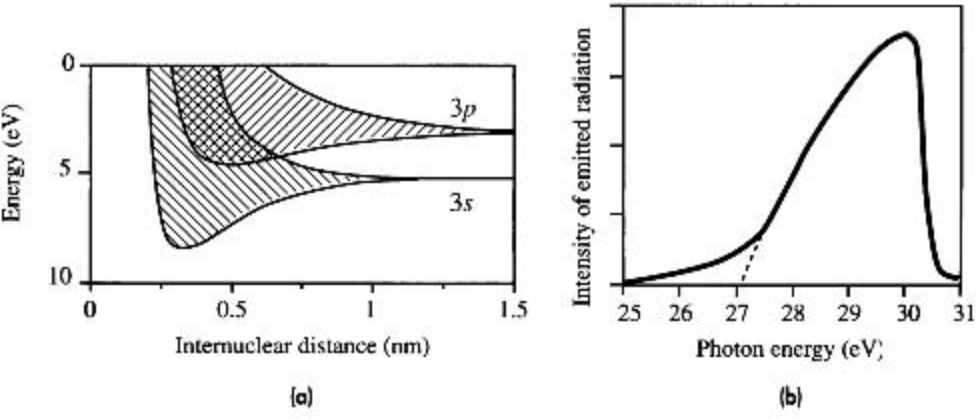


Figure 4.59
(a) Energy band formation in sodium.
(b) L-emission band of X-rays from sodium.
SOURCE: (b) Data extracted from W. M. Cadt and D. H. Tomboulian, *Phys. Rev.*, **59**, 1941, p. 381.

4.12 Conductivity of metals in the free electron model Consider the general expression for the conductivity of metals in terms of the density of states $g(E_F)$ at E_F given by

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

Show that within the free electron theory, this reduces to $\sigma = e^2 n\tau/m_e$, the Drude expression.

Mean free path of conduction electrons in a metal Show that within the free electron theory, the mean free path ℓ and conductivity σ are related by

$$\frac{e^2}{3^{1/3}\pi^{2/3}\hbar}\ell n^{2/} \qquad 87 \times 10^{-3}\ell n^{4/3}$$

Mean free path and conductivity in the free electron model

Calculate ℓ for Cu and Au, given each metal's resistivity of 17 nΩ m and 22 nΩ m, respectively, and that each has a valency of I. We are used to seeing $\sigma \propto n$. Can you explain why $\sigma \propto n^{2/3}$?

4.14 Low-temperature heat capacity of metals The heat capacity of conduction electrons in a metal is proportional to the temperature. The overall heat capacity of a metal is determined by the lattice heat capacity, except at the lowest temperatures. If δE_t is the increase in the total energy of the conduction electrons (per unit volume) and δT is the increase in the temperature of the metal as a result of heat addition, E_t has been calculated as follows:

$$\int_0^\infty Eg(E)f(E)\,dE \qquad E_t(0) + \left(\frac{\pi^2}{4}\right)\frac{n(kT)^2}{E_{FO}}$$

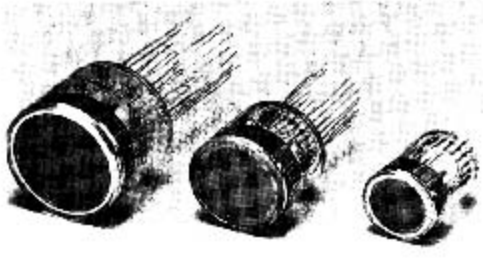
where $E_t(0)$ is the total energy per unit volume at 0 K, n is the concentration of conduction electrons, and E_{FO} is the Fermi energy at 0 K. Show that the heat capacity per unit volume due to conduction electrons in the free electron model of metals is

Heat capacity of conduction electrons

$$\frac{\pi^2}{2}\left(\frac{nk^2}{E_{FO}}\right)T \qquad \gamma T \tag{4.84}$$

where $\gamma = (\pi^2/2)(nk^2/E_{FO})$. Calculate C_e for Cu, and then using the Debye equation for the lattice heat capacity, find C_v for Cu at 10 K. Compare the two values and comment. What is the comparison at room temperature? (Note: $C_{\text{volume}} = C_{\text{molar}}(\rho/M_{\text{at}})$, where ρ is the density in g cm^{-3}, C_{volume} is in J K^{-1} cm^{-3}, and M_{at} is the atomic mass in g mol^{-1}.)

4.15 Secondary emission and photomultiplier tubes When an energetic (high velocity) projectile electron collides with a material with a low work function, it can cause electron emission from the surface. This phenomenon is called **secondary emission.** It is fruitfully utilized in photomultiplier tubes as illustrated in Figure 4.60. The tube is evacuated and has a photocathode for receiving photons as a signal. An incoming photon causes photoemission of an electron from the photocathode material. The electron is then accelerated by a positive voltage applied to an electrode called a dynode which has a work function that easily allows secondary emission. When the accelerated electron strikes dynode D_1, it can release several electrons. All these electrons, the original and the secondary electrons, are then accelerated by the more positive voltage applied to dynode D_2. On impact with D_2, further electrons are released by secondary emission. The secondary emission process continues at each dynode stage until the final electrode, called the anode, is reached whereupon all the electrons are collected which results in a signal. Typical applications for photomultiplier tubes are in X-ray and nuclear medical instruments



Photomultiplier tubes.
| SOURCE: Courtesy of Hamamatsu.

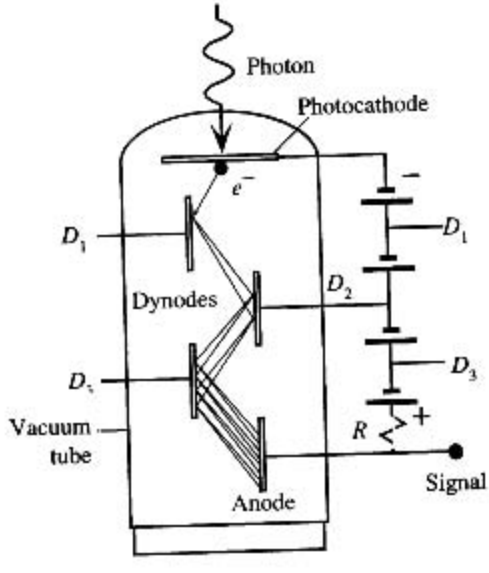


Figure 4.60 The photomultiplier tube.

(X-ray CT scanner, positron CT scanner, gamma camera, etc.), radiation measuring instruments (*e.g.*, radon counter), X-ray diffractometers, and radiation measurement in high-energy physics research.

A particular photomultiplier tube has the following properties. The photocathode is made of a semiconductor-type material with $E_g \approx 1$ eV, an electron affinity χ of 0.4 eV, and a quantum efficiency of 20 percent at 400 nm. *Quantum efficiency* is defined as the number of photoemitted electrons per absorbed photon. The diameter of the photocathode is 18 mm. There are 10 dynode electrodes and an applied voltage of 1250 V between the photocathode and anode. Assume that this voltage is equally distributed among all the electrodes.

a. What is the longest threshold wavelength for the phototube?

b. What is the maximum kinetic energy of the emitted electron if the photocathode is illuminated with a 400 nm radiation?

c. What is the emission current from the photocathode at 400 nm illumination?

d. What is the *KE* of the electron as it strikes the first dynode electrode?

e. It has been found that the tube has a gain of 10^6 electrons per incident photon. What is the average number of secondary electrons released at each dynode?

4.16 **Thermoelectric effects and E_F** Consider a thermocouple pair that consists of gold and aluminum. One junction is at 100 °C and the other is at 0 °C. A voltmeter (with a very large input resistance) is inserted into the aluminum wire. Use the properties of Au and Al in Table 4.3 to estimate the emf registered by the voltmeter and identify the positive end.

4.17 **The thermocouple equation** Although inputting the measured emf for V in the thermocouple equation $V = a\Delta T + b(\Delta T)^2$ leads to a quadratic equation, which in principle can be solved for ΔT, in general ΔT is related to the measured emf via

$$\Delta T = a_1 V + a_2 V^2 + a_3 V^3 + \cdots$$

with the coefficients a_1, a_2, etc., determined for each pair of TCs. By carrying out a Taylor's expansion of the TC equation, find the first two coefficients a_1 and a_2. Using an emf table for the K-type thermocouple or Figure 4.33, evaluate a_1 and a_2.

4.18 **Thermionic emission** A vacuum tube is required to have a cathode operating at 800 °C and providing an emission (saturation) current of 10 A. What should be the surface area of the cathode for the two materials in Table 4.9? What should be the operating temperature for the Th on W cathode, if it is to have the same surface area as the oxide-coated cathode?

Table 4.9

	B_r (A m^{-2} K^{-2})	Φ (eV)
Th on W	3×10^4	2.6
Oxide coating	100	1

4.19 **Field-assisted emission in MOS devices** Metal-oxide-semiconductor (MOS) transistors in microelectronics have a metal gate on an SiO_2 insulating layer on the surface of a doped Si crystal. Consider this as a parallel plate capacitor. Suppose the gate is an Al electrode of area 50 μm × 50 μm and has a voltage of 10 V with respect to the Si crystal. Consider two thicknesses for the SiO_2, (a) 100 Å and (b) 40 Å, where (1 Å = 10^{-10} m). The work function of Al is 4.2 eV, but this refers to electron emission into vacuum, whereas in this case, the electron is emitted into the oxide. The potential energy barrier Φ_B between Al and SiO_2 is about 3.1 eV, and the field-emission current density is given by Equation 4.46a and b. Calculate the field-emission current for the two cases. For simplicity, take m_e to be the electron mass in free space. What is your conclusion?

4.20 **CNTs and field emission** The electric field at the tip of a sharp emitter is much greater than the "applied field," $\mathcal{E}_o$. The applied field is simply defined as V_G/d where d is the distance from the cathode tip to the gate or the grid; it represents the average nearly uniform field that would exist if the tip were replaced by a flat surface so that the cathode and the gate would almost constitute a parallel plate capacitor. The tip experiences an effective field $\mathcal{E}$ that is much greater than $\mathcal{E}_o$, which is expressed by a **field enhancement factor** β that depends on the geometry of the cathode–gate emitter, and the shape of the emitter; $\mathcal{E} = \beta\mathcal{E}_o$. Further, we can take $\Phi_{\text{eff}}^{1/2}\Phi \approx \Phi^{3/2}$ in Equation 4.46. The final expression for the field-emission current density then becomes

Fowler–Nordheim field emission current

$$J = \frac{1.5 \times 10^6}{\Phi}\beta^2\mathcal{E}_o^2 \exp\left(\frac{10.4}{\Phi^{1/2}}\right)\exp\left(-\frac{6.44 \times 10^7\Phi^{3/2}}{\beta\mathcal{E}_o}\right) \qquad \textbf{[4.85]}$$

where Φ is in eV. For a particular CNT emitter, $\Phi = 4.9$ eV. Estimate the applied field required to achieve a field-emission current density of 100 mA cm^2 in the absence of field enhancement ($\beta = 1$) and with a field enhancement of $\beta = 800$ (typical value for a CNT emitter).

4.21 **Nordheim–Fowler field emission in an FED** Table 4.10 shows the results of I–V measurements on a Motorola FED microemitter. By a suitable plot show that the I–V follows the Nordheim–Fowler emission characteristics. Can you estimate Φ?

Table 4.10 Tests on a Motorola FED micro field emitter

V_G	40.0	42	44	46	48	50	52	53.8	56.2	58.2	60.4
I_{emission}	0.40	2.14	9.40	20.4	34.1	61	93.8	142.5	202	279	367

4.22 **Lattice waves and heat capacity**

a. Consider an aluminum sample. The nearest separation $2R$ (2 × atomic radius) between the Al–Al atoms in the crystal is 0.286 nm. Taking a to be $2R$, and given the sound velocity in Al as 5100 m s^{-1}, calculate the force constant β in Equation 4.66. Use the group velocity v_g from the actual dispersion relation, Equation 4.55, to calculate the "sound velocity" at wavelengths of $\Lambda = 1$mm, 1 μm, and 1 nm. What is your conclusion?

b. Aluminum has a Debye temperature of 394 K. Calculate its specific heat at 30 °C (Darwin, Australia) and at −30 °C (January, Resolute Nunavut, Canada).

c. Calculate the specific heat capacity of a germanium crystal at 25 °C and compare it with the experimental value in Table 2.5.

4.23 **Specific heat capacity of GaAs and InSb**

a. The Debye temperature T_D of GaAs is 344 K. Calculate its specific heat capacity at 300 K and at 30 °C.

b. For InSb, $T_D = 203$ K. Calculate the room temperature specific heat capacity of InSb and compare it with the value expected from the Dulong–Petit rule ($T > T_D$).

4.24 **Thermal conductivity**

a. Given that silicon has a Young's modulus of about 110 GPa and a density of 2.3 g cm^{-3}, calculate the mean free path of phonons in Si at room temperature.

b. Diamond has the same crystal structure as Si but has a very large thermal conductivity, about 1000 W m^{-1} K^{-1} at room temperature. Given that diamond has a specific heat capacity c_v of 0.50 J K^{-1} g^{-1}, Young's modulus Y of 830 GPa, and density ρ of 0.35 g cm^{-3}, calculate the mean free path of phonons in diamond.

c. GaAs has a thermal conductivity of 200 W m^{-1} K^{-1} at 100 K and 80 W m^{-1} K^{-1} at 200 K. Calculate its thermal conductivity at 25 °C and compare with the experimental value of 44 W m^{-1} K^{-1}. (Hint: Take $\kappa \propto T^{-n}$ in the temperature region of interest; see Figure 4.48.)

4.25 Overlapping bands Consider Cu and Ni with their density of states as schematically sketched in Figure 4.61. Both have overlapping $3d$ and $4s$ bands, but the $3d$ band is very narrow compared to the $4s$ band. In the case of Cu the band is full, whereas in Ni, it is only partially filled.

a. In Cu, do the electrons in the $3d$ band contribute to electrical conduction? Explain.

b. In Ni, do electrons in both bands contribute to conduction? Explain.

c. Do electrons have the same effective mass in the two bands? Explain.

d. Can an electron in the $4s$ band with energy around E_F become scattered into the $3d$ band as a result of a scattering process? Consider both metals.

e. Scattering of electrons from the $4s$ band to the $3d$ band and vice versa can be viewed as an additional scattering process. How would you expect the resistivity of Ni to compare with that of Cu, even though Ni has two valence electrons and nearly the same density as Cu? In which case would you expect a stronger temperature dependence for the resistivity?

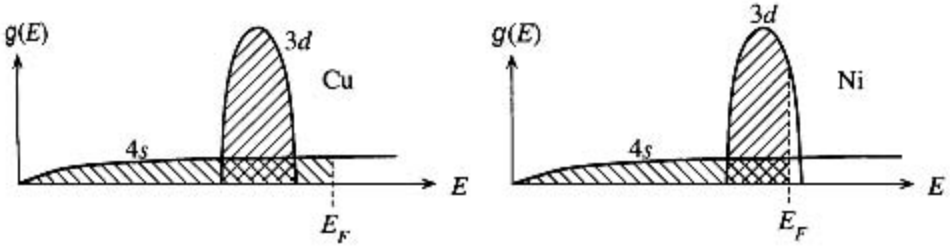
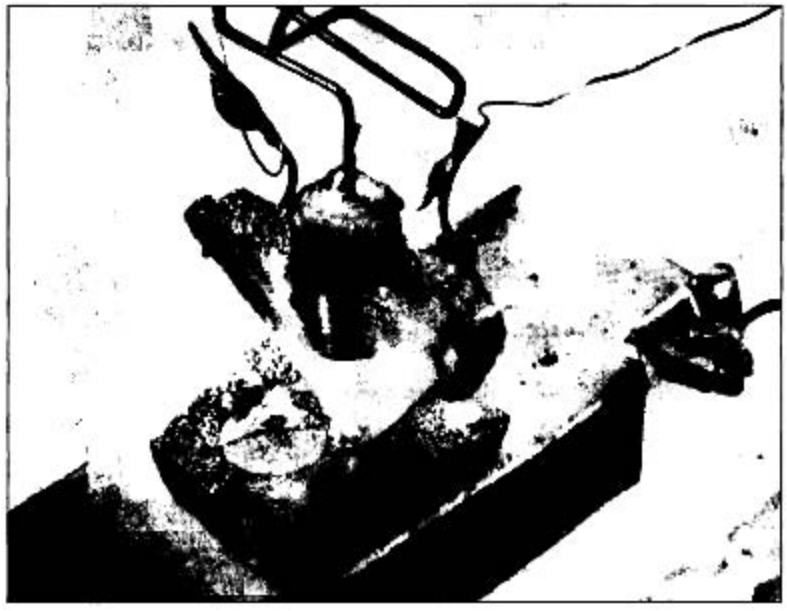


Figure 4.61 Density of states and electron filling in Cu and Ni.

***4.26 Overlapping bands at E_F and higher resistivity** Figure 4.61 shows the density of states for Cu (or Ag) and Ni (or Pd). The d band in Cu is filled, and only electrons at E_F in the s band make a contribution to the conductivity. In Ni, on the other hand, there are electrons at E_F both in the s and d bands. The d band is narrow compared with the s band, and the electron's effective mass in this d band is large; for simplicity, we will assume m_e^* is "infinite" in this band. Consequently, the d-band electrons cannot be accelerated by the field (infinite m_e^*), have a negligible drift mobility, and make no contribution to the conductivity. Electrons in the s band can become scattered by phonons into the d band, and hence become relatively immobile until they are scattered back into the s band when they can drift again. Consider Ni and one particular conduction electron at E_F starting in the s band. Sketch schematically the magnitude of the velocity gained $|v_x - u_x|$ from the field E_x as a function of time for 10 scattering events; v_x and u_x are the instantaneous and initial velocities, and $|v_x - u_x|$ increases linearly with time, as the electron accelerates in the s band and then drops to zero upon scattering. If τ_{ss} is the mean time for s to s-band scattering, τ_{sd} is for s-band to d-band scattering, τ_{ds} is for d-band to s-band scattering, assume the following sequence of 10 events in your sketch: τ_{ss}, τ_{ss}, τ_{sd}, τ_{ds}, τ_{ss}, τ_{sd}, τ_{ds}, τ_{ss}, τ_{sd}, τ_{ds}. What would a similar sketch look like for Cu? Suppose that we wish to apply Equation 4.27. What does $g(E_F)$ and τ represent? What is the most important factor that makes Ni more resistive than Cu? Consider Matthiessen's rule. (Note: There are also electron spin related effects on the resistivity of Ni, but for simplicity these have been neglected.)

4.27 Grüneisen's law Al and Cu both have metallic bonding and the same crystal structure. Assuming that the Grüneisen's parameter γ for Al is the same as that for Cu, $\gamma = 0.23$, estimate the linear expansion coefficient λ of Al, given that its bulk modulus $K = 75$ GPa, $c_v = 900\ \text{J K}^{-1}\ \text{kg}^{-1}$, and $\rho = 2.7\ \text{g cm}^{-3}$. Compare your estimate with the experimental value of $23.5 \times 10^{-6}\ \text{K}^{-1}$.



First point-contact transistor invented at Bell Labs.
| SOURCE: Courtesy of Bell Labs.

The three inventors of the transistor: William Shockley (seated), John Bardeen (left), and Walter Brattain (right) in 1948; the three inventors shared the Nobel prize in 1956.
| SOURCE: Courtesy of Bell Labs.