

Insulators and Semiconductors

Mankind is a catalyzing enzyme for the transition from a carbon-based intelligence to a silicon-based intelligence.

Gérard Bricogne, French crystallographer

Although the ideal gas of fermions is a useful model for the behavior of metals, real systems deviate from this model in a very important way: real energy spectra contain gaps in which no energy levels occur. When the Fermi energy lies in one of those gaps, the behavior of the electrons is qualitatively different from that described in the previous chapter.¹ Such materials do not have the high electrical conductivity of metals. They are insulators or semiconductors, and their properties are the subject of the current chapter.

The origin of the strange band gaps that arise in the electron density of states in real materials is that the regular arrangement of the nuclei in a crystal lattice creates a periodic potential. This affects the energy levels of the electrons and gives rise to gaps in the energy spectrum. The mechanism by which the gaps arise can be understood from two perspectives. The first initially assumes that the atoms are very far apart, so that the electrons cannot easily jump from one atom to the next. Since an electron is tightly bound to a nucleus, this is known as the ‘tight binding approximation’. The second way of looking at it starts from the opposite extreme, where the electrons act almost like an ideal gas, but are presumed to be subject to a very weak periodic potential. This is known as the ‘nearly free electron approximation’. In both cases, gaps in the energy spectrum occur due to the periodicity of the lattice.

In the following sections we will explore both ways of seeing how gaps arise in the energy spectrum in a crystal. After that we will investigate the consequences of energy gaps, which will lead us to the properties of insulators and semiconductors.

30.1 Tight-Binding Approximation

Begin by considering a single, isolated atom. Since we know that electrons in isolated atoms have discrete energy levels, the energy spectrum of a single atom will look qualitatively like the diagram on the far left of Fig. 30.1.

¹ Be cautious in comparing our use of the Fermi energy with that of other books. We have defined the Fermi energy to be the $T \rightarrow 0$ limit of the chemical potential, as in eq. (29.6). This is often referred to in other books as the ‘Fermi level’.

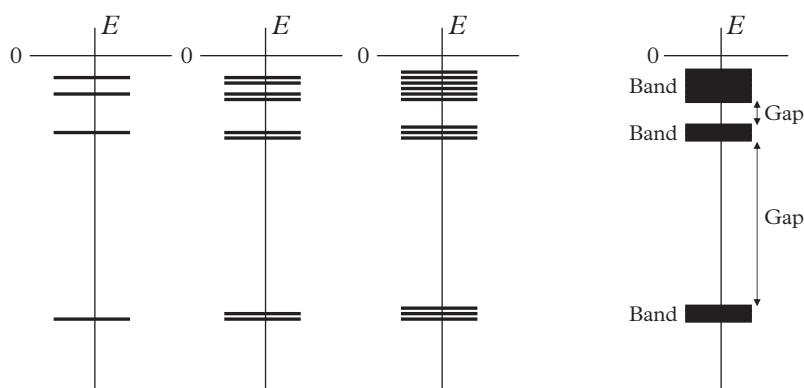


Fig. 30.1 Schematic plot of the energy levels in a crystal. The diagram on the far left shows the energy spectrum of an isolated atom. The second diagram shows the effect of bringing two atoms close together, and the third diagram corresponds to three neighboring atoms. The diagram on the far right shows the band structure that arises when the lattice constant is of the order of nanometers, as in a real crystal. The 10^{20} or more energy levels make each energy band a quasi-continuum.

If we bring a second atom to within a few tenths of a nanometer of the first, the wave functions will overlap, and the energy levels will be split according to quantum perturbation theory, as shown in the second diagram in Fig. 30.1.

If there are three atoms close together, each energy level will split into three levels, as shown in the third diagram in Fig. 30.1.

Now assume that we have a crystal; that is, a regular array of atoms in which the neighboring atoms are separated by a few tenths of a nanometer. If the crystal contains 10^{20} atoms, the energy levels will be split into 10^{20} closely spaced energy levels, which will form a quasi-continuum. Such a quasi-continuum of energy levels is called a ‘band’. This is shown schematically in the diagram on the far right of Fig. 30.1.

In the schematic diagram in Fig. 30.1, the spreading of the energy levels has been large enough to merge the top two single-atom energy levels into a single band. However, the gap between the bottom two two single-atom energy levels remains, despite the spreading of the energy levels into bands. These gaps, or ‘band gaps’, represent forbidden energies, since there are no single-particle wave functions with energies in the gaps.

This qualitative description of how gaps arise in the single-particle excitation spectrum can be made quantitative by applying perturbation theory to the weak overlap of the wave functions of nearly isolated atoms. Since the electrons are tightly bound to individual atoms when the overlap is very weak, this is known as the ‘tight-binding approximation’. We will not go into the details of the theory here, since our purpose is only to understand qualitatively how gaps in the excitation spectrum arise.

Before going on to discuss the opposite extreme, in which electrons experience only a weak periodic potential due to the crystal lattice, we need some mathematics. The following section introduces Bloch’s theorem for a particle in a periodic potential (Felix

Bloch (1905–1983), Swiss physicist, Nobel Prize 1952). Once armed with this theorem, we will discuss the effects of the crystal lattice on the single-electron energy spectrum.

30.2 Bloch's Theorem

For simplicity, we will again only treat the one-dimensional case. The three-dimensional case is not really more difficult, but the necessary vector notation makes it look more complicated than it is.

Consider a single particle in a periodic potential, $V(x)$, so that the Hamiltonian is given by

$$H = \frac{p^2}{2m} + V(x). \quad (30.1)$$

Denoting the period by a , we have

$$V(x+a) = V(x). \quad (30.2)$$

Define translation operators, T_n , for any integer n , such that

$$T_n H(p, x) = H(p, x+na) = H(p, x). \quad (30.3)$$

Translation operators have the property that they can be combined. For an arbitrary function $f(x)$,

$$T_n T_m f(x) = T_n f(x+ma) = f(x+ma+na) = T_{m+n} f(x), \quad (30.4)$$

or

$$T_n T_m = T_{m+n}. \quad (30.5)$$

The translation operators also commute with each other

$$[T_n, T_m] = T_n T_m - T_m T_n = 0, \quad (30.6)$$

and every T_n commutes with the Hamiltonian.

$$[T_n, H] = T_n H - H T_n = 0 \quad (30.7)$$

Therefore, we can define wave functions that are simultaneously eigenfunctions of all the translation operators and the Hamiltonian.

Denote a single-particle eigenfunction by $\phi(x)$, and the eigenvalue of T_n by C_n .

$$T_n \phi(x) = \phi(x + na) = C_n \phi(x) \quad (30.8)$$

From eq. (30.5), we can see that

$$T_n T_m \phi(x) = C_n C_m \phi(x) = T_{m+n} \phi(x) = C_{n+m} \phi(x) \quad (30.9)$$

or

$$C_n C_m = C_{n+m}. \quad (30.10)$$

We can solve eq. (30.10) for C_n in terms of a constant k ,

$$C_n = \exp(ikna), \quad (30.11)$$

where k is a real number. It is easy to confirm that this is a solution by inserting eq. (30.11) into eq. (30.10). The reason that k must be real is that $\phi(x)$ – and therefore C_n – must be bounded for very large (positive or negative) values of x .

The next step is to note that the eigenfunction can be written as

$$\phi_k(x) = \exp(ikx) u_k(x), \quad (30.12)$$

where $u_k(x)$ is periodic.

$$u_k(x + a) = u_k(x) \quad (30.13)$$

The eigenfunctions $\phi_k(x)$ also have an important periodicity with respect to the wave number k . If we define a vector

$$K = \frac{2\pi}{a}, \quad (30.14)$$

and consider the wave function

$$\phi_{k+K}(x) = \exp[i(k+K)x] u_{k+K}(x), \quad (30.15)$$

then

$$\begin{aligned} \phi_k(x) &= \exp(ikx) u_k(x) \\ &= \exp[i(k+K)x] \exp(-iKx) u_k(x) \\ &= \exp[i(k+K)x] u_{k+K}(x) \\ &= \phi_{k+K}(x). \end{aligned} \quad (30.16)$$

It is easy to confirm that $u_{k+K}(x) = \exp(-iKx)u_k(x)$ is a periodic function of x with period a .

By extending this argument to $\phi_{k+qK}(x)$, where q is any integer, we see that ϵ_k must be a periodic function of k with period $K = 2\pi/a$.

$$\epsilon_k = \epsilon_{k+K} \quad (30.17)$$

The following section uses Bloch's theorem to calculate the single-particle excitation spectrum for nearly free electrons. The periodicity of ϵ_k will be crucial to the discussion.

30.3 Nearly-Free Electrons

In this section we will look at the consequences of a weak periodic potential for a gas of non-interacting electrons. We will begin by assuming that the potential is exactly zero, and then investigate the effects of turning on a periodic potential, such as that due to a regular crystal lattice.

The discussion will consist of three stages:

1. Reminder of the energy spectrum of a non-interacting particle in a box.
2. The effects of Bloch's theorem on the representation of the energy spectrum of non-interacting particles in the limit that the amplitude of the periodic potential vanishes.
3. The effects of turning on a periodic potential with a small amplitude.

30.3.1 Free Electrons

We begin with the energy levels for free electrons, which are given in eq. (27.12) of Chapter 27.

$$\epsilon_k = \frac{\hbar^2}{2m} k^2 \quad (30.18)$$

Here

$$k = \frac{n\pi}{L}, \quad (30.19)$$

and n is a positive integer.

Fig. 30.2 shows a plot of ϵ_k as a function of the wave number k .

30.3.2 The 'Empty' Lattice

The next step is to look at the consequences of the periodicity of the lattice on the energy spectrum, while keeping the amplitude of the periodic potential equal to zero. This might

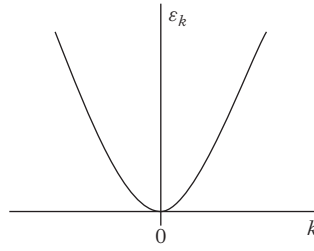


Fig. 30.2 Free electron energy spectrum.

seem rather strange, since a potential of amplitude zero cannot change the energy of a state. Nevertheless, the *representation* of the energy can change.

The representation of the energy spectrum in a periodic potential differs from that shown in Fig. 30.2 because of the periodicity in k -space imposed by eq. (30.17). This representation is known as the ‘empty lattice’, because even though we are assuming periodicity, there is nothing there to give rise to a non-zero potential.

The wave functions in the empty lattice representation can be written in the form given in eq. (30.12). The empty-lattice energy, ϵ_k , is periodic in k , as indicated in eq. (30.17). The empty-lattice energy spectrum is plotted in Fig. 30.3.

The degeneracies seen in Fig. 30.3 all come from the symmetry that $\epsilon_k = \epsilon_{k+K}$. The value of the empty-lattice representation is that it highlights those points at which the splitting due to a periodic lattice removes the degeneracies.

Fig. 30.3 shows how the periodicity indicated by eq. (30.17) affects the energy spectrum. The parabolic spectrum shown in Fig. 30.2 is repeated at intervals of $K = 2\pi/a$

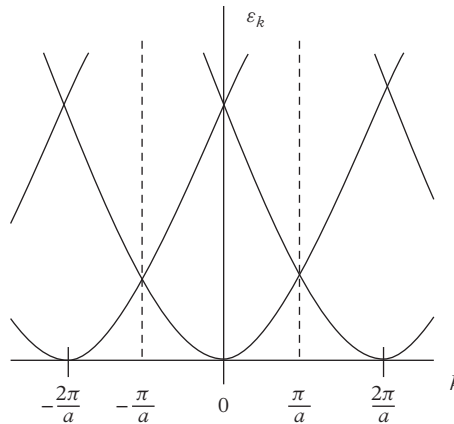


Fig. 30.3 Empty lattice energy spectrum.

along the entire k -axis. This repetition has the consequence that all information about the energy spectrum is contained in any region of length $K = 2\pi/a$ along the axis. It is conventional to choose the region from $-\pi/a$ to π/a . This region is called the Brillouin Zone, after Léon Brillouin. In Fig. 30.3, the limits of the one-dimensional Brillouin Zone are indicated by two vertical dotted lines.

In more than one dimension, a more general definition is used for the Brillouin Zone. The periodicity in \vec{k} -space is defined by a set of three vectors, $\{\vec{K}_1, \vec{K}_2, \vec{K}_3\}$. The set of all vectors of the form $n_1\vec{K}_1 + n_2\vec{K}_2 + n_3\vec{K}_3$, where the n_j s are integers, forms a ‘reciprocal lattice’. The Brillouin Zone is then defined as including those points closer to the origin in \vec{k} -space than to any other point in the reciprocal lattice. For the one-dimensional case, the reciprocal lattice consists of the points, $2n\pi/a$, where n is an integer. The points between $-\pi/a$ and π/a then satisfy this definition of the Brillouin Zone.

30.3.3 Non-Zero Periodic Potential

Now we are going to discuss the effects of making the amplitude of the periodic potential non-zero.

If we consider Fig. 30.3, we see that there are crossings of the energy levels at $k = n\pi/a$, where n is an integer. These crossings represent degeneracies, where two states have the same energy (in the original representation, they are traveling waves going in opposite directions). A non-zero periodic potential splits the degeneracy at these crossings, according to the usual result of quantum perturbation theory. After the splitting, the energy spectrum looks like Fig. 30.4.

After the splitting of the crossing degeneracies, gaps in the energy spectrum open up. There are now ranges of energies for which there are no solutions to the wave-function equation.

30.4 Energy Bands and Energy Gaps

The consequences of both the tight-binding and the nearly-free electron analyses are qualitatively the same. In certain ranges of energies, there exist single-particle energy states. These energy ranges are called ‘bands’. In other energy ranges, no such states are found. These energy ranges are called ‘gaps’.

Both the tight-binding and the nearly-free electron analysis can be used as starting points for quantitative calculations of the energy spectrum and the density of states. However, that would take us beyond the scope of this book.

The existence of energy gaps is clearly an important feature of the density of states. As we will see in the next section, when the Fermi energy falls within a band gap, the

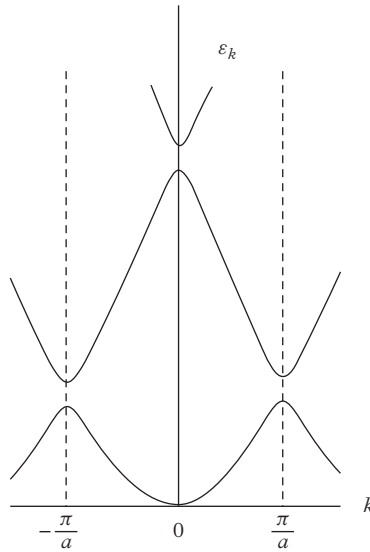


Fig. 30.4 Band structure with energy gaps occurring at $k = \frac{\pi}{a}$, 0 , and $-\frac{\pi}{a}$.

behavior of the system differs dramatically from what we found for the free-electron model.

30.5 Where is the Fermi Energy?

When the energy spectrum has band gaps, the behavior of the material will depend strongly on where the Fermi energy is – specifically, on whether the Fermi energy lies within a band or within a gap between bands. This, in turn, depends on whether the bands near the Fermi energy are empty, completely filled, or partially filled at zero temperature. If there is a partially filled band at $T = 0$, the Fermi energy will lie at the boundary between the filled and empty states. If there are only completely filled and completely empty bands at $T = 0$, the Fermi energy will lie halfway between the highest filled band and the lowest empty band.

To understand why the behavior of a material depends strongly on where the Fermi energy is, first recall that when a single-particle state is filled, no other electron can enter that state. If all single-particle states at a given energy are filled, nothing can happen at that energy. Neither can anything happen at an energy for which all states are empty. It is only when electrons can change states without changing their energy significantly that they can move from one area to another within the system.

The explicit dependence of the electrical properties of systems on their band structures gives rise to the differences between metals, insulators, and semiconductors, which will be discussed in the following sections.

30.6 Fermi Energy in a Band (Metals)

If the Fermi energy falls within a band, it requires only a tiny amount of thermal energy to raise an electron into an empty energy level with a slightly higher energy. The electron is then able to move to other states with the same energy in different parts of the system. Since electrons are charged, their motion is electric current. Systems with Fermi energies that fall within a band are metals.

In Chapter 29 we found from the Sommerfeld expansion that the specific heat at low temperatures was given by eq. (29.46),

$$C_V = \frac{\pi^2}{3} D_F(\epsilon_F) k_B^2 T. \quad (30.20)$$

For free electrons, this gave the particular result given in eq. (29.53).

If the Fermi energy lies in one of the energy bands, eq. (30.20) still holds. The only change from the free electron result is quantitative, due to the change in the density of single-electron states at the Fermi energy. For some important conductors, such as copper, silver, and gold, the free-electron specific heat is within 40% of the measured value of the specific heat at low temperatures. For other metals, the magnitude of the change can be quite large—up to an order of magnitude larger than the free-electron specific heat. The details of band structure that give rise to these differences can be quite interesting, but go beyond the scope of this book.

If the Fermi energy falls in an energy gap, the behavior of the system is qualitatively different from that of a metal, and we find insulators and semiconductors. Their properties are dramatically different from those of metals—and from each other! We will consider each case separately in the following sections, but first we will discuss the generic equations that govern both insulators and intrinsic semiconductors (defined below).

30.7 Fermi Energy in a Gap

When there is an energy gap between the highest filled single-particle state and the lowest empty one, the Fermi energy must be exactly in the middle of the gap. This is a very general principle, which was already implicit in the calculations for the cases of discrete sets of energy levels that were treated in Chapter 29. To see that this is so, introduce a specific model for which the density of states is given by

$$D(\epsilon) = \begin{cases} A(\epsilon - \epsilon_C)^a & \epsilon > \epsilon_C \\ 0 & \epsilon_C > \epsilon > \epsilon_V \\ B(\epsilon_V - \epsilon)^b & \epsilon_V > \epsilon, \end{cases} \quad (30.21)$$

where A , B , a , and b are positive constants. The usual values of the exponents are $a = b = 1/2$ —the demonstration of which is left as an exercise for the reader—but for now we will keep the notation general.

The subscripts C and V in (30.21) refer to the ‘conduction’ (upper) and ‘valence’ (lower) bands. Conduction will occur only in so far as electrons are thermally excited from the valence band into the conduction band.

Since we are interested in very low temperatures when calculating the Fermi energy, states with energies much greater than ϵ_C are nearly empty, while those with energies much less than ϵ_V are nearly full. Therefore, the form of the density of states away from the gap is often unimportant.

It might seem extremely improbable that the filled single-particle state with the highest energy (at $T = 0$) is exactly at the top of a band, while the lowest empty state is at the bottom of the next higher band. Actually, it is quite common. The reason is most easily seen from the tight-binding picture. A band is formed from the splitting of atomic energy levels. If these energy levels are filled, then the bands states will also be filled. If the energies in a band do not overlap the energies of any other band, it will be completely filled, while the higher bands will be empty.

As usual, we will use eq. (27.61), which gives the number of particles N as a function of μ , to find the chemical potential

$$N = \int_0^{\infty} D(\epsilon) (\exp[\beta(\epsilon - \mu)] + 1)^{-1} d\epsilon. \quad (30.22)$$

Splitting the integral into two parts—one for each band—gives

$$N = \int_0^{\epsilon_V} D(\epsilon) (\exp[\beta(\epsilon - \mu)] + 1)^{-1} d\epsilon + \int_{\epsilon_C}^{\infty} D(\epsilon) (\exp[\beta(\epsilon - \mu)] + 1)^{-1} d\epsilon. \quad (30.23)$$

By our assumptions, the lower band is exactly filled at zero temperature,

$$N = \int_0^{\epsilon_V} D(\epsilon) d\epsilon. \quad (30.24)$$

Subtracting eq.(30.24) from eq.(30.23), we can eliminate N from the equation for the chemical potential

$$0 = \int_{\epsilon_C}^{\infty} D(\epsilon) (\exp[\beta(\epsilon - \mu)] + 1)^{-1} d\epsilon + \int_0^{\epsilon_V} D(\epsilon) \left[(\exp[\beta(\epsilon - \mu)] + 1)^{-1} - 1 \right] d\epsilon. \quad (30.25)$$

Using the identity from Section 29.3, we can rewrite the resulting equation in a form that shows the equality between the number of particles taken out of the lower band and the number of particles added to the upper band

$$\int_{\epsilon_C}^{\infty} D(\epsilon) (\exp[\beta(\epsilon - \mu)] + 1)^{-1} d\epsilon = \int_0^{\epsilon_V} D(\epsilon) (\exp[-\beta(\epsilon - \mu)] + 1)^{-1} d\epsilon. \quad (30.26)$$

Anticipating that $\mu \approx \epsilon_F = (\epsilon_C + \epsilon_V)/2$ at low temperatures, we see that the exponentials in the denominators of both integrands are large. This allows us to approximate eq. (30.26) as

$$A \int_{\epsilon_C}^{\infty} (\epsilon - \epsilon_C)^a \exp[-\beta(\epsilon - \mu)] d\epsilon \approx B \int_{-\infty}^{\epsilon_V} (\epsilon_V - \epsilon)^b \exp[\beta(\epsilon - \mu)] d\epsilon, \quad (30.27)$$

where we have inserted the approximate form for the density of states in eq. (30.21). We have also changed the lower limit of the integral on the right to minus infinity, since the exponential in the integrand makes the lower limit unimportant at low temperatures.

We do not need to evaluate the integrals in eq. (30.27) explicitly at this point, but we do need to extract the β -dependence. We do this by defining new variables for the integral on the left of eq. (30.27),

$$x = \beta(\epsilon - \epsilon_C), \quad (30.28)$$

and for the integral on the right,

$$y = -\beta(\epsilon - \epsilon_V). \quad (30.29)$$

Eq. (30.27) then becomes

$$A \exp(\beta(\mu - \epsilon_C)) \beta^{-1-a} \int_0^{\infty} x^a \exp[-x] dx = B \exp(\beta(\epsilon_V - \mu)) \beta^{-1-b} \int_0^{\infty} y^b \exp[-y] dy. \quad (30.30)$$

Now bring the exponentials with μ to the left side of the equation,

$$\exp(\beta(2\mu - \epsilon_C - \epsilon_V)) = \left(\frac{B \int_0^{\infty} y^b \exp[-y] dy}{A \int_0^{\infty} x^a \exp[-x] dx} \right) (k_B T)^{b-a}. \quad (30.31)$$

Denoting the dimensionless constant X by

$$X = \frac{B \int_0^{\infty} y^b \exp[-y] dy}{A \int_0^{\infty} x^a \exp[-x] dx}, \quad (30.32)$$

we find

$$\mu = \frac{1}{2}(\epsilon_C + \epsilon_V) + \frac{k_B T}{2}(b - a) \ln(k_B T) + \frac{k_B T}{2} \ln X. \quad (30.33)$$

Clearly, the Fermi energy is given by

$$\epsilon_F = \lim_{T \rightarrow 0} \mu = \frac{1}{2}(\epsilon_C + \epsilon_V), \quad (30.34)$$

for any values of the constants A, B, a , and b .

As mentioned at the beginning of this section, the most usual values of the exponents are $a = b = 1/2$. For this case, eq. (30.33) simplifies.

$$\mu = \frac{1}{2}(\epsilon_C + \epsilon_V) + \frac{k_B T}{2} \ln \left(\frac{B}{A} \right) \quad (30.35)$$

The calculation in this section also gives us an explicit estimate for the number of electrons in partially filled energy levels that can contribute to an electric current. The number of electrons that are thermally excited to the upper band is given by the first term on the right of eq. (30.23),

$$\begin{aligned} N_e &= \int_{\epsilon_C}^{\infty} D(\epsilon) (\exp[\beta(\epsilon - \mu)] + 1)^{-1} d\epsilon \\ &\approx \int_{\epsilon_C}^{\infty} D(\epsilon) \exp[-\beta(\epsilon - (\epsilon_C + \epsilon_V)/2)] d\epsilon \\ &\approx \exp(-\beta(\epsilon_C - \epsilon_V)/2) \int_{\epsilon_C}^{\infty} D(\epsilon) \exp[-\beta(\epsilon - \epsilon_C)] d\epsilon \\ &\approx \exp(-\beta(\epsilon_C - \epsilon_V)/2) A \int_{\epsilon_C}^{\infty} (\epsilon - \epsilon_C)^{1/2} \exp[-\beta(\epsilon - \epsilon_C)] d\epsilon \\ &= \exp(-\beta(\epsilon_C - \epsilon_V)/2) A \int_0^{\infty} (x/\beta)^{1/2} \exp[-x] \beta^{-1} dx \\ &= \exp(-\beta(\epsilon_C - \epsilon_V)/2) A (k_B T)^{3/2} \int_0^{\infty} x^{1/2} \exp[-x] dx \\ &= \frac{A\sqrt{\pi}}{2} (k_B T)^{3/2} \exp(-\beta(\epsilon_C - \epsilon_V)/2). \end{aligned} \quad (30.36)$$

The factor $\exp(-\beta(\epsilon_C - \epsilon_V)/2)$ in the last line of eq. (30.36) shows that the number of electrons in the upper band will be exponentially small when the band gap $\epsilon_{gap} = \epsilon_C - \epsilon_V$ is much greater than $k_B T$.

For an insulator, the band gap, ϵ_{gap} , between the highest filled ('valence') and lowest empty ('conduction') band is large; that is, greater than roughly $3eV$. The reason for

the choice of this (somewhat arbitrary) number is that a band gap of 3 eV implies that at room temperature ($T = 300\text{ K}$), $\beta\epsilon_{\text{gap}}/2 \approx 58$, so that $\exp(-\beta\epsilon_{\text{gap}}/2) \approx 6.5 \times 10^{-26}$. This factor in eq. (30.36) is so small that essentially no electrons are found in the conduction band of an insulator with a band gap of 3 eV or greater. Without conduction electrons, no current can flow, and the material is an insulator.

30.8 Intrinsic Semiconductors

A rough definition of a pure (or intrinsic) semiconductor is a material with a band structure like an insulator, but with a smaller band gap. Examples are given in Table 30.1.

For an intrinsic semiconductor like germanium (Ge), the factor of $\exp(-\beta\epsilon_{\text{gap}}/2)$ in eq. (30.36) is about $\exp(-\beta\epsilon_{\text{gap}}/2) \approx \exp(-13) \approx 2 \times 10^{-6}$ at room temperature (300 K). While this value is small, there are still a substantial number of electrons that are thermally excited. The conductivity of pure germanium is much smaller than that of the good insulator, but it is not negligible.

30.9 Extrinsic Semiconductors

The main reason why semiconductors are so interesting—and extraordinarily useful in modern electronics—is that their electrical properties are very strongly influenced by the presence of relatively small concentrations of impurities. Indeed, that property provides a better definition of a semiconductor than having a small band gap.

Since the word ‘impurities’ tends to have negative connotations that would not be appropriate in this context, impurities in semiconductors are generally called ‘dopants’, and impure semiconductors are called ‘extrinsic’ semiconductors. There are two basic classes of dopant: donors and acceptors. The distinction is based on whether the particular dopant has more (donor) or fewer (acceptor) electrons than the intrinsic semiconductor. In the following two subsections, the two kinds of dopant are discussed.

Table 30.1 *Band gaps of semiconductors. The third column expresses the band gap in units of $k_B T$ at room temperature.*

Material	ϵ_{gap} (eV)	$\epsilon_{\text{gap}}/k_B T$
Ge	0.67	26
Si	1.14	43
GaAs	1.43	55
GaSb	0.7	27
InAs	0.36	24
InN	0.7	27

To have a concrete example, we will use silicon (Si) in both cases as the intrinsic semiconductor that is being doped.

The presence of dopants is important because they introduce new energy levels into the band gap. These energy levels have the effect of moving the Fermi energy from the middle of the gap, where it would be for an intrinsic semiconductor, to either just under the conduction band (donor) or just above the valence band (acceptor).

30.9.1 Donor Impurities

Phosphorus (P) is an example of a donor impurity for silicon (Si). The phosphorus ion, P^+ , goes into the lattice in place of a Si atom, keeping its net positive charge. The remaining electron from the P atom goes into the conduction band. A semiconductor that contains donor dopants is called an ' n -type' semiconductor because the dopants provide extra carriers with *negative* charge.

Since there is a Coulomb attraction between the net positive charge on the P^+ ion in the lattice and the donated electron in the conduction band, the combination looks very much like a hydrogen atom. The main differences are that the effective mass of electrons in the conduction band differs from that of a free electron, and there is a screening effect due to the dielectric constant of Si . The combination of these two effects results in a bound state with an energy of about -0.044 eV , relative to the bottom of the conduction band. This donor state is close to the bottom of the conduction band, because the binding energy is much smaller than the band gap of 1.14 eV . The extra electron is in the new donor-state energy level, which lies in the band gap, about 0.044 eV below the bottom of the conduction band.

The new energy level is the highest occupied level in the system. The lowest unoccupied level at the bottom of the conduction band, which is only about 0.044 eV higher. As a result, the Fermi energy is moved from the center of the gap to about 0.022 eV below the conduction band. Since the binding energy is only about half of $k_B T$ at room temperature for the example of P impurities in Si , even without calculations we can see that many more electrons will be excited into the conduction band with dopants than without.

30.9.2 Acceptor Impurities

Boron (B) is an example of an acceptor impurity for Si . The B atom goes into the lattice in place of a Si atom and binds an electron from the valence band, creating a B^- ion. Since an electron was removed from the valence band, the band is no longer full. The state with the missing electron is known as a 'hole'. For reasons that go beyond the scope of this book, the missing electron acts just like a *positive* mobile charge. A semiconductor that contains acceptor dopants is called a ' p '-type semiconductor, because the dopants provide extra carriers with *positive* charge.

In analogy to the situation for donors, the mobile positive hole in the valence band is attracted to the negative B^- ion fixed in the lattice. Again, this creates a new localized state with an energy in the band gap. However, for ' p '-type semiconductors the state is

located just above the top of the valence band (about 0.010 eV above for B in Si). The equations for calculation of the energy are essentially the same as those for the bound state in ‘ n ’-type semiconductors. The resultant Fermi energy is located about 0.005 eV above the top of the valence band.

30.10 Semiconductor Statistics

Although electrons are fermions, so that we might expect the equations for Fermi statistics to apply to the impurity states, the true situation is rather more interesting. The occupation of the donor and acceptor states is determined by what we might call ‘semiconductor statistics’, which we have not seen before in this book.

The reason for the difference between semiconductor statistics and Fermi statistics arises from two features of the impurity energy levels:

1. The impurity levels in doped semiconductors are two-fold degenerate because there are both spin-up and spin-down states.
2. The two states in a given level cannot both be occupied, because of a strong Coulomb repulsion.

These features mean that the probabilities for occupying the two degenerate states that make up the impurity energy level cannot be regarded as independent: the possibility of the energy level containing two electrons is excluded. Because of the lack of independence of the two impurity states, the derivation of semiconductor statistics given in the next subsection calculates the occupation of the entire energy level.

30.10.1 Derivation of Semiconductor Statistics

To derive semiconductor statistics—that is, the occupation number for an impurity energy level—we return to eq. (27.31) in Chapter 27,

$$\mathcal{Z} = \prod_{\alpha} \sum_{n_{\alpha}} \exp[-\beta(\epsilon_{\alpha} - \mu)n_{\alpha}]. \quad (30.37)$$

For an donor energy level with energy ϵ_d , the sum over n_{ϵ_d} must only contain terms corresponding to (1) both states empty, (2) only the spin-up state occupied, and (3) only the spin-down state occupied. Double occupancy of the donor energy level is excluded due to the Coulomb repulsion energy,

$$\begin{aligned} \sum_{n_{\epsilon_d}} \exp[-\beta(\epsilon_d - \mu)n_{\epsilon_d}] &= \exp[0] + \exp[-\beta(\epsilon_d - \mu)] + \exp[-\beta(\epsilon_d - \mu)] \\ &= 1 + 2\exp[-\beta(\epsilon_d - \mu)]. \end{aligned} \quad (30.38)$$

Inserting eq. (30.38) in eq. (27.44) for the average occupation number, we find

$$\begin{aligned}
 \langle n_d \rangle &= \frac{\sum_{n_\alpha} n_\alpha \exp[-\beta(\epsilon_\alpha - \mu)n_\alpha]}{\sum_{n_\alpha} \exp[-\beta(\epsilon_\alpha - \mu)n_\alpha]} \\
 &= \frac{2 \exp[-\beta(\epsilon_d - \mu)]}{1 + 2 \exp[-\beta(\epsilon_d - \mu)]} \\
 &= \left[\frac{1}{2} \exp[\beta(\epsilon_d - \mu)] + 1 \right]^{-1}.
 \end{aligned} \tag{30.39}$$

The extra factor of $1/2$ in the denominator of the last line in eq. (30.39) distinguishes semiconductor statistics from fermion statistics. Naturally, the single-particle states in semiconductor electron bands still obey Fermi–Dirac statistics.

30.10.2 Simplified Illustration of Semiconductor Statistics

As an example of the consequences of semiconductor statistics, we will first consider a simplified model in which there are N_d donor impurities with a corresponding energy level ϵ_d , which lies slightly below the bottom of the conduction band (ϵ_C). We will assume that the valence band is filled, so that we can ignore it. To simplify the mathematics and the interpretation of the results, the conduction band will be approximated by assigning all N_C states the same energy, ϵ_C —not because it is realistic (it is not), but to show the effects of semiconductor statistics. A more realistic model will be presented in the next subsection.

As was the case for problems with either Fermi–Dirac or Bose–Einstein statistics, the first task is to find the chemical potential. The equation governing the chemical potential is, as usual, the equation for the total number of particles. Since each donor impurity contributes a single electron, the total number of electrons is equal to N_d , the number of donor levels,

$$N = N_d = N_d \left[\frac{1}{2} \exp[\beta(\epsilon_d - \mu)] + 1 \right]^{-1} + N_C [\exp[\beta(\epsilon_C - \mu)] + 1]^{-1}. \tag{30.40}$$

Note that the electrons in the conduction band still obey Fermi–Dirac statistics.

We might have imagined that the Fermi energy coincides with the energy of the impurity level, ϵ_d , because of the two-fold degeneracy and single occupancy at $T = 0$. However, due to the exclusion of double occupancy of the impurity level, the Fermi energy turns out to be half way between the impurity-state energy and the bottom of the conduction band, $\epsilon_F = (\epsilon_C + \epsilon_d)/2$. We will assume that this is true in the following derivation of the Fermi energy, and confirm it by the results of the calculation.

To solve eq. (30.40), we can use a variation on the identity in eq. (29.7).

$$N_d [2 \exp[-\beta(\epsilon_d - \mu)] + 1]^{-1} = N_C [\exp[\beta(\epsilon_C - \mu)] + 1]^{-1}. \tag{30.41}$$

At low temperatures, since we expect $\mu \approx (\epsilon_d + \epsilon_C)/2$, we have both $\beta(\mu - \epsilon_d) \gg 1$ and $\beta(\epsilon_C - \mu) \gg 1$. This means that the exponentials dominate in both denominators in eq. (30.41), and the neglect of the '+1' in both denominators is a good approximation,

$$\frac{N_d}{2} \exp[\beta(\epsilon_d - \mu)] \approx N_C \exp[-\beta(\epsilon_C - \mu)]. \quad (30.42)$$

We can solve this equation for the chemical potential.

$$\mu \approx \frac{\epsilon_C + \epsilon_d}{2} + \frac{1}{2} k_B T \ln \left(\frac{N_d}{2N_C} \right) \quad (30.43)$$

Since the Fermi energy is the zero temperature limit of the chemical potential, we can immediately see from eq. (30.43) that

$$\epsilon_F = \frac{\epsilon_C + \epsilon_d}{2} \quad (30.44)$$

On the other hand, because the difference in energy between ϵ_C and ϵ_d can be of the order of $k_B T$, the second term in eq. (30.43) is not always negligible. There are usually far fewer impurities than conduction states, and we can easily have $N_d \approx 10^{-5} N_C$. The second term in eq. (30.43) would be roughly $-12 k_B T$. To see how significant this effect is, we need to calculate the occupancy of the donor levels:

$$\langle n_d \rangle = N_d \left[\frac{1}{2} \exp[\beta(\epsilon_d - \mu)] + 1 \right]^{-1} \quad (30.45)$$

Inserting eq. (30.43) for the chemical potential, we find

$$\langle n_d \rangle \approx N_d \left[\frac{1}{2} \exp \left[\beta \left(\epsilon_d - \frac{\epsilon_C + \epsilon_d}{2} \right) - \frac{1}{2} \ln \left(\frac{N_d}{2N_C} \right) \right] + 1 \right]^{-1} \quad (30.46)$$

or

$$\langle n_d \rangle \approx N_d \left[\frac{1}{2} \left(\frac{2N_C}{N_d} \right)^{1/2} \exp \left[-\beta \left(\frac{\epsilon_C - \epsilon_d}{2} \right) \right] + 1 \right]^{-1}. \quad (30.47)$$

There are two competing factors in the first term in the denominator of eq. (30.47). While the factor of $\exp[-\beta(\epsilon_C - \epsilon_d)/2]$ can be small for very low temperatures, for *As* in *Ge* at room temperature it has a value of about 0.6. On the other hand, if $N_d \approx 10^{-5} N_C$, the factor of $\sqrt{2N_C/N_d} \approx 450$, so that it would dominate, making the occupation of the donor state quite small. Under such circumstances, almost all of the donor levels would be empty, and the electrons would be in the conduction band.

Note that the emptying of the donor levels is an entropic effect. There are so many more states in the conduction band that the probability of an electron being in a donor level is very small, even though it is energetically favorable. This feature is not an artifact of the simple model we are using for illustration in this section; it is a real feature of semiconductor physics.

In the next section we will consider a (slightly) more realistic model that removes the unphysical assumption of an infinitely thin conduction band.

30.10.3 A More Realistic Model of a Semiconductor

Although our neglect of the valence band in the model used in the previous subsection is often justifiable, giving all states in the conduction band the same energy is certainly not. In this subsection, we will introduce a model that is realistic enough to bear comparison with experiment – although we will leave such a comparison to a more specialized book.

We will assume the donor levels all have energy ϵ_d , as in Subsection 30.10.2. However, we will model the conduction band by a density of states as in eq. (30.21). The total density of states can then be written as

$$D(\epsilon) = \begin{cases} A(\epsilon - \epsilon_C)^{1/2} & \epsilon \geq \epsilon_C \\ N_d \delta(\epsilon - \epsilon_d) & \epsilon < \epsilon_C. \end{cases} \quad (30.48)$$

Since only single-particle states near the bottom of the conduction band will be occupied, the upper edge of the band is taken to be infinity.

The equation for the total number of donor electrons, which gives us the chemical potential, is then

$$N_d = N_d \left[\frac{1}{2} \exp[\beta(\epsilon_d - \mu)] + 1 \right]^{-1} + \int_{\epsilon_C}^{\infty} A(\epsilon - \epsilon_C)^{1/2} \left[(\exp[\beta(\epsilon - \mu)] + 1)^{-1} - 1 \right] d\epsilon. \quad (30.49)$$

As in the previous section, we will again use a variation on the identity in eq. (29.7) to simplify the equation. We will also introduce the same dimensionless variable defined in eq. (30.28),

$$N_d [2 \exp[-\beta(\epsilon_d - \mu)] + 1]^{-1} = A \exp(\beta(\mu - \epsilon_C)) \beta^{-1-a} \int_0^{\infty} x^{1/2} \exp[-x] dx. \quad (30.50)$$

Since we can evaluate the integral, this becomes

$$N_d [2 \exp[-\beta(\epsilon_d - \mu)] + 1]^{-1} = \frac{A\sqrt{\pi}}{2} (k_B T)^{3/2} \exp(\beta(\mu - \epsilon_C)). \quad (30.51)$$

The Fermi energy again turns out to be half way between the energy level of the donor state and the bottom of the conduction band,

$$\epsilon_F = \frac{\epsilon_d + \epsilon_C}{2}. \quad (30.52)$$

The remaining behavior of this more realistic model is qualitatively the same as that of the simplified illustration in the previous subsection. The details will be left to the reader.

30.11 Semiconductor Physics

This brings us to the end of our introduction to insulators and semiconductors. The field of semiconductors, in particular, is extremely rich. We have covered the bare basics, but we have not even mentioned how semiconductors can be used to create lasers or the transistors that are the basis of all modern electronics. However, this introduction should be sufficient preparation to read a book on semiconductors without too much difficulty. The field is both fascinating and extraordinarily important in today's world.