Electrons and Energy Band Structures in Crystals

5

5.1 Introduction

In Chap. 4, we introduced quantum mechanics as the proper alternative to classical mechanics to describe physical phenomena, especially when the dimensions of the systems considered approach the atomic scale. The concepts we learned will now be applied to describe the physical properties of electrons in a crystal. During this process, we will make use of the simple quantum mechanical systems which were mathematically treated in the previous chapter. This will lead us to the description of a very important concept in solid-state physics, namely, that of the "energy band structures."

5.2 Electrons in a Crystal

So far, we have discussed the energy spectrum of an electron in an atom, and more generally in a one-dimensional potential well. Modeling an electron in a solid is much more complicated because it experiences the combined electrostatic potential of all lattice ions and all other electrons. Nevertheless, the total potential acting on the electrons in a solid shares the symmetry of the lattice and thus reflects the periodicity of the lattice in the case of a crystal. This simplifies the mathematical treatment of the problem and allows us to understand how the energy spectrum, wavefunctions, and other dynamic characteristics (e.g., mass) of electrons in a solid are modified from the free particle case.

5.2.1 Bloch Theorem

The Bloch theorem provides a powerful mathematical simplification for the wavefunctions of particles evolving in a periodic potential. The solutions of the

M. Razeghi, *Fundamentals of Solid State Engineering*,
https://doi.org/10.1007/978-3-319-75708-7_5

Schrödinger equation in such a potential are not pure plane waves as they were in the case of a free particle (Eq. (4.33)) but are waves which are modulated by a function having the periodicity of the potential or lattice. Such functions are then called Bloch wavefunctions and can be expressed as:

$$\Psi\left(\vec{k}, \vec{r}\right) = \exp\left(i\,\vec{k}\,.\,\vec{r}\right).u\left(\vec{k}, \vec{r}\right) \tag{5.1}$$

where $\vec{k}$ is the wavenumber vector (in three dimensions) or wavevector of the particle, $\vec{r}$ its position, and $u\left(\vec{k}, \vec{r}\right)$ a space-dependent amplitude function which reflects the periodicity of the lattice:

$$u\left(\vec{k}, \vec{r} + \vec{R}\right) = u\left(\vec{k}, \vec{r}\right) \tag{5.2}$$

The expression in Eq. (5.1) means that the Bloch wavefunction is a plane wave, given by the exponential term in Eq. (5.1), which is modulated by a function which has the periodicity of the crystal lattice. An illustration of this is shown in Fig. 5.1 in the one-dimensional case.

Combining Eqs. (5.1) and (5.2) leads us to the form:

$$\Psi\left(\vec{k}, \vec{r} + \vec{R}\right) = \exp\left(i\,\vec{k}\,.\,\vec{R}\right)\Psi\left(\vec{k}, \vec{r}\right) \tag{5.3}$$

for any lattice vector $\vec{R}$. In a one-dimensional case, d being the period of the potential or lattice, this can be written as:

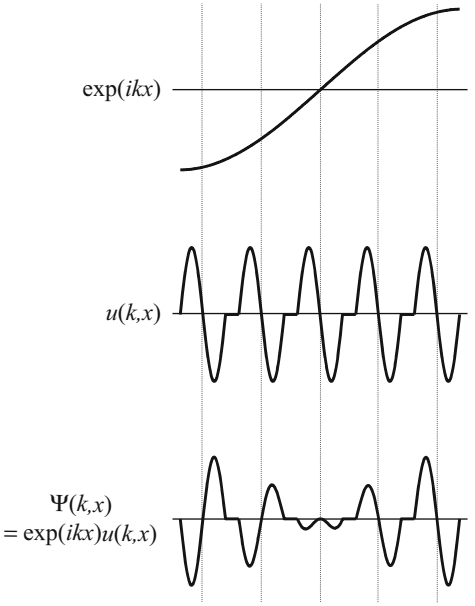


Fig. 5.1 One-dimensional illustration of a Bloch wavefunction (bottom) as a plane wave (top) modulated by a periodic function which has the period of the lattice (middle)

$$\Psi(k, x + d) = \exp(ikd)\Psi(k, x) \tag{5.4}$$

This shows that the wavefunction is the same for two values of k which differ by integral multiples of $\frac{2\pi}{d}$. We can therefore restrict the range of allowed values of k to the interval $-\frac{\pi}{d} < k \leq \frac{\pi}{d}$.

Another important limit of the Bloch theorem is for non-infinite crystals. In this case, it is common to use the periodic boundary conditions for the Bloch wavefunction, i.e., the wavefunction is the same at opposite extremities of the crystal. Assuming a linear periodic chain of N atoms (period d), the periodic boundary condition can be written as:

$$\Psi(k, x) = \Psi(k, x + Nd) = \exp(ikNd)\Psi(k, x) \tag{5.5}$$

which means that:

$$\exp(ikNd) = 1 \tag{5.6}$$

or:

$$k = \frac{2\pi n}{Nd} \tag{5.7}$$

where n is an integer. Since we restricted the range of k between $-\frac{\pi}{d}$ and $\frac{\pi}{d}$, n can only take integer values between $-\frac{N}{2}$ and $\frac{N}{2}$. There are thus only N distinct values for n and thus k.

5.2.2 One-Dimensional Kronig-Penney Model

In addition to the Bloch theorem, which simplified the wavefunction of a particle, there is a further simplification of the periodic potential, which is often used and is referred to as the Kronig-Penney model. We will continue with the one-dimensional formalism started in the previous section. In the Kronig-Penney model, the crystal is assumed to be infinite. In this model, the real crystal potential experienced by an electron is shown in Fig. 5.2a and is approximated by the one depicted in Fig. 5.2b.

The solution of the Kronig-Penney model partially utilizes the results from the finite potential well problem discussed in Sect. 4.4.4, and the same notations have therefore been used in Fig. 5.2b. The mathematical analysis will first be conducted locally, in the region $-b < x < a$, where the potential can be approximated by Eq. (4.48) except that there is a new limit for the variable x.

The wavefunction solution of the Schrödinger equation thus has two distinct components, $\Psi_1(x)$ and $\Psi_2(x)$, in different regions of space which must satisfy:

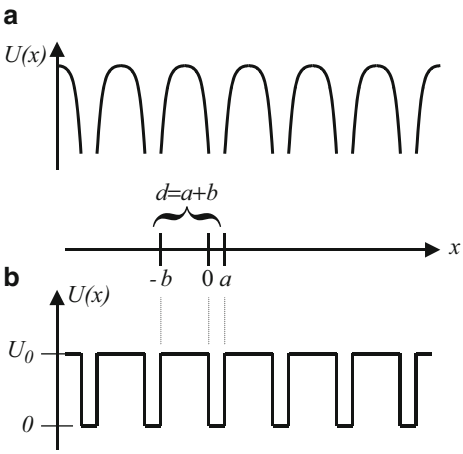


Fig. 5.2 (**a**) Real crystal potential experienced by electrons in a crystal and (**b**) simplified crystal potential used in the Kronig-Penney model

$$\begin{cases} \dfrac{d^2\Psi_1(x)}{dx^2} + \alpha^2\Psi_1(x) = 0 \quad \text{for} \quad -b < x < 0 \\ \dfrac{d^2\Psi_2(x)}{dx^2} + \beta^2\Psi_2(x) = 0 \quad \text{for} \quad 0 < x < a \end{cases} \tag{5.8}$$

by defining:

$$\begin{cases} \alpha = \begin{cases} i\alpha_-, \text{with}\, \alpha_- = \sqrt{\dfrac{2m(U_0 - E)}{\hbar^2}} & \text{when} \;\; 0 < E < U_0 \\ \alpha_+, \text{with}\, \alpha_+ = \sqrt{\dfrac{2m(E - U_0)}{\hbar^2}} & \text{when} \;\; U_0 < E \end{cases} \\ \beta = \sqrt{\dfrac{2mE}{\hbar^2}} \end{cases} \tag{5.9}$$

The general solution to Eq. (5.8) can be expressed as:

$$\begin{cases} \Psi_1(x) = A_1 \sin(\alpha x) + B_1 \cos(\alpha x) \\ \Psi_2(x) = A_2 \sin(\beta x) + B_2 \cos(\beta x) \end{cases} \tag{5.10}$$

with the understanding that $\sin(\alpha x)$ and $\cos(\alpha x)$ become $-i\sinh(\alpha_- x)$ and $\cosh(\alpha_- x)$, respectively, when the quantity $\alpha = i\alpha$ is imaginary.

The boundary conditions imply the continuity of $\Psi(x)$ and its first derivative $\frac{d\Psi(x)}{dx}$ at point $x = 0$ and include the periodicity condition of the wavefunction expressed through the Bloch theorem in Eq. (5.4) between points $x = a$ and $x = -b$:

$$\begin{cases} \Psi_1(0) = \Psi_2(0) \\ \dfrac{d\Psi_1}{dx}(0) = \dfrac{d\Psi_2}{dx}(0) \end{cases} \qquad \begin{cases} e^{ik(a+b)}\Psi_1(-b) = \Psi_2(a) \\ e^{ik(a+b)}\dfrac{d\Psi_1}{dx}(-b) = \dfrac{d\Psi_2}{dx}(a) \end{cases} \tag{5.11}$$

Utilizing Eq. (5.10), we obtain:

$$\begin{cases} B_1 = B_2 \\ \alpha A_1 = \beta A_2 \\ e^{ik(a+b)}[-A_1 \sin(\alpha b) + B_1 \cos(\alpha b)] = A_2 \sin(\beta a) + B_2 \cos(\beta a) \\ e^{ik(a+b)}[\alpha A_1 \cos(\alpha b) + \alpha B_1 \sin(\alpha b)] = \beta A_2 \cos(\beta a) - \beta B_2 \sin(\beta a) \end{cases} \tag{5.12}$$

which can be simplified by expressing A_2 and B_2 in terms of A_1 and B_1:

$$\begin{cases} A_1\left[e^{ik(a+b)} \sin(\alpha b) + \dfrac{\alpha}{\beta}\sin(\beta a)\right] + B_1\left[\cos(\beta a) - e^{ik(a+b)} \cos(\alpha b)\right] = 0 \\ A_1\left[\alpha e^{ik(a+b)} \cos(\alpha b) - \alpha \cos(\beta a)\right] + B_1\left[\beta \sin(\beta a) + \alpha e^{ik(a+b)} \sin(\alpha b)\right] = 0 \end{cases} \tag{5.13}$$

This system of two equations with two unknowns has a nonzero solution (i.e., A_1 and B_1 not both zero) if the determinant of the system is zero (for more details on the mathematics, the reader is referred to any introductory book on linear algebra). This means that the product of the first bracket in the top equation by the second bracket in the bottom equation minus the product of the second bracket in the top equation by the first bracket in the bottom equation is zero:

$$\begin{aligned} &\left[e^{ik(a+b)} \sin(\alpha b) + \frac{\alpha}{\beta}\sin(\beta a)\right]\left[\beta \sin(\beta a) + \alpha e^{ik(a+b)} \sin(\alpha b)\right] \\ &-\left[\cos(\beta a) - e^{ik(a+b)} \cos(\alpha b)\right]\left[\alpha e^{ik(a+b)} \cos(\alpha b) - \alpha \cos(\beta a)\right] = 0 \end{aligned} \tag{5.14}$$

or after simplification:

$$\cos k(a+b) = -\frac{\alpha^2 + \beta^2}{2\alpha\beta} \sin(\alpha b) \sin(\beta a) + \cos(\alpha b) \cos(\beta a) \tag{5.15}$$

Using the same constants as in Eq. (4.57), we can rewrite Eq. (5.9) as:

$$\begin{cases} \alpha = \begin{cases} i\alpha_-, \text{with}\,\alpha_- = \alpha_0\sqrt{1-\zeta} \;\; \text{when}\, 0 < E < U_0 \\ \alpha_+, \text{with}\,\alpha_+ = \alpha_0\sqrt{\zeta-1} \;\; \text{when}\, U_0 < E \end{cases} \\ \beta = \alpha_0\sqrt{\zeta} \end{cases} \tag{5.16}$$

Therefore, Eq. (5.15) can be simplified into:

$$\begin{cases} \cos k(a+b) = \dfrac{1-2\zeta}{2\sqrt{\zeta(1-\zeta)}} \sin\left(\alpha_0 a\sqrt{\zeta}\right)\sinh\left(\alpha_0 b\sqrt{1-\zeta}\right) \\ \qquad\qquad + \cos\left(\alpha_0 a\sqrt{\zeta}\right)\cosh\left(\alpha_0 b\sqrt{1-\zeta}\right) \\ \text{for } 0 < \zeta < 1 \\ \cos k(a+b) = \dfrac{1-2\zeta}{2\sqrt{\zeta(\zeta-1)}} \sin\left(\alpha_0 a\sqrt{\zeta}\right) \sin\left(\alpha_0 b\sqrt{\zeta-1}\right) \\ \qquad\qquad + \cos\left(\alpha_0 a\sqrt{\zeta}\right) \cos\left(\alpha_0 b\sqrt{\zeta-1}\right) \\ \text{for } 1 < \zeta \end{cases} \tag{5.17}$$

In these equations, the only variable in the right-hand side functions is the energy E, while the only variable in the left-hand side is the wavenumber k. Similar to the finite potential well case, a solution in ζ of Eq. (5.17) allows us to determine the values of the energy as well as the wavefunctions (after normalization).

5.2.3 Energy Bands

In the Kronig-Penney model, the crystal is assumed to be infinite. Therefore, the periodic boundary condition of the Bloch wavefunction is unnecessary, and the wavenumber k can take a continuous range of values and is real (i.e., not complex). Equation (5.17) is most easily solved graphically. The shape of the right-hand side function of Eq. (5.17), which we will call $f(\zeta)$, can be visualized in Fig. 5.3.

Because of the cosine on the LHS of Eq. (5.17), only values of $f(\zeta)$ that are between -1 and $+1$ lead to allowed (real) values for k. The areas where this occurs are shaded in Fig. 5.3. Because k is determined through a cosine function, two opposite values of k are possible for the same value for $f(\zeta)$. In these shaded areas, there is a continuous range of values for ζ (or E), corresponding to allowed energy bands. Some values of ζ, however, occur in non-shaded areas in Fig. 5.3 and are thus "forbidden," meaning that there is no possible state corresponding to these values of energy. Such regions are called regions of forbidden energy, or energy gaps. An illustration of these energy bands is given in Fig. 5.4.

Furthermore, as we can see from Fig. 5.3, for every given value of k between $-\frac{\pi}{a+b}$ and $\frac{\pi}{a+b}$, several values of ζ (thus E) are possible. An actual plot of the E–k relationship is given in Fig. 5.5 and is called the energy spectrum, the band diagram, or band structure. This type of diagram is very important in determining the

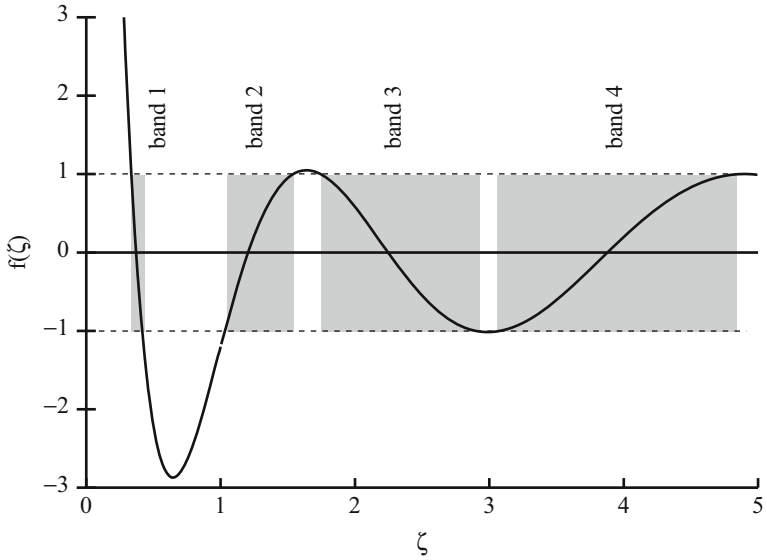


Fig. 5.3 Plot of the right-hand side of Eq. (5.17), showing the graphical determination of the E–k relationship. There exists a solution to Eq. (5.17) only when the right-hand side of the equation is between -1 and $+1$, which correspond to the shaded areas

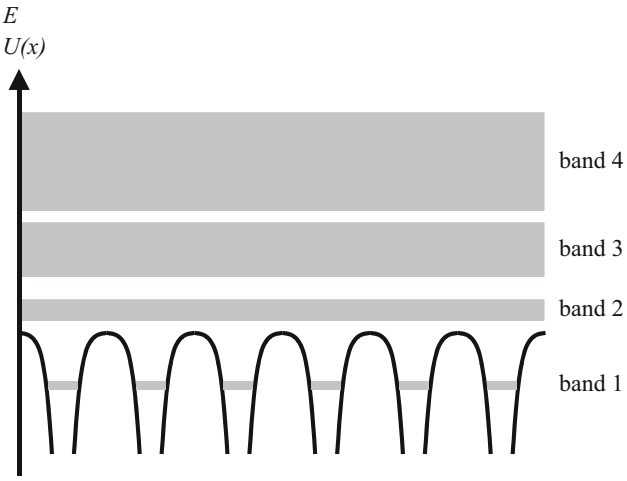


Fig. 5.4 Illustration of the concept of energy bands in the crystal

properties of an electron in a crystal. A noteworthy feature, which is true for real crystals and which can easily be seen in this diagram, is that the slope of the energy band, i.e. $\frac{dE}{dk}$, is equal to zero at the center ($k = 0$) and extremities ($k = \pm\frac{\pi}{a+b}$). This diagram, in which the value of k is restricted in the interval between $-\frac{\pi}{a+b}$ and $\frac{\pi}{a+b}$, is

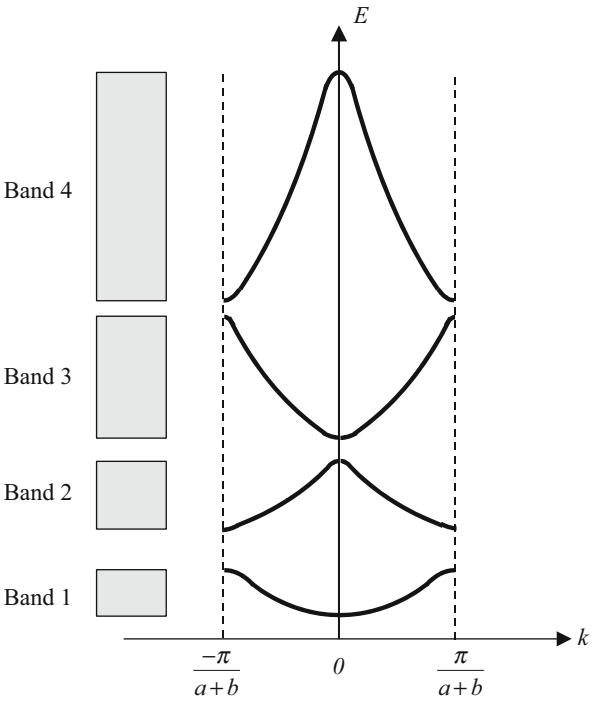


Fig. 5.5 One-dimensional E–k relationship in the reduced-zone representation in the Kronig-Penney model

often referred to as the reduced-zone representation of the energy versus k dispersion relation, as opposed to the extended-zone representation which we will now briefly discuss.

Because the energy is a periodic function of k, the reduced-zone scheme is the right way to think about the band structure of the system. All the information about the allowed energy bands is contained in the first Brillouin zone. Going outside the Brillouin zone simply repeats the same information; it does not add anything new to our knowledge. In the extended-zone representation, one can lift the previous restriction on the k-values and instead of being restricted to the values in the interval $-\frac{\pi}{a+b}$ and $\frac{\pi}{a+b}$, k is allowed to have any (larger) values. This however does not change the wavefunction because of the Bloch theorem: the k-values outside the first Brillouin zone can be reduced to ones inside the first Brillouin zone by "subtracting" a reciprocal; lattice vector $\vec{K}$. One can if one wishes unfold the band diagram into the diagram shown in Fig. 5.6, but the larger values of k can be reduced to equivalent values of k inside the first zone. Unlike for free particles, in a crystal subject to Bloch's theorem the higher values of k do not signify a higher value of momentum. Indeed, values of momentum differing from each other exactly by a reciprocal lattice vector are indistinguishable. This does not mean that $\vec{k}$ has nothing to do with momentum, it is related to the particle momentum, but it is defined and conserved only up to a reciprocal lattice vector: If one adds a reciprocal lattice vector to $\vec{k}$, the

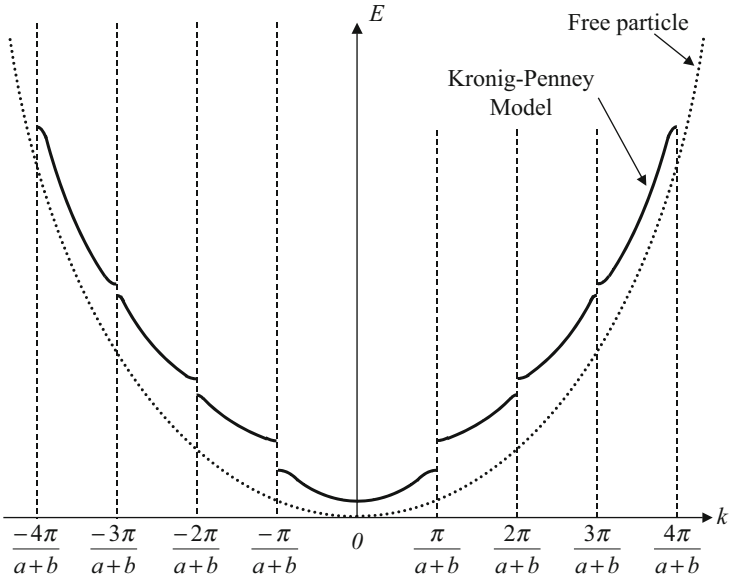


Fig. 5.6 One-dimensional E–k relationship in the extended-zone representation in the Kronig-Penney model. The parabolic relation for the free particle is shown in dotted lines for comparison. The deviation from a parabolic shape occurs mainly at the Brillouin zone boundaries

energy in the same band remains the same. The expression $\hbar k$, which corresponded to the particle momentum in the free particle case ($\langle p \rangle = \hbar k$), is now referred to as the quasi-momentum of the electron or the crystal momentum because it includes the interaction of the electron with the crystal. This explains why one can add integral multiples of $\frac{2\pi}{a+b}$ to the wavenumber without changing the band structure of the crystal, while this would be meaningless if it was a particle momentum. The reason why this quasi-momentum is not absolutely conserved in a lattice, and only conserved up to a reciprocal lattice vector, is ultimately connected to the fact that the Hamiltonian in a lattice is not translationally invariant over any arbitrary displacement as it would be in a space with no external forces, but it is only invariant when displaced by a lattice vector.

5.2.4 Nearly Free Electron Approximation

The Kronig-Penney model discussed previously is not the only method to determine the band structure in crystals, but it is the simplest and leads to a complete analytic solution. Many other methods have been developed which can be methodologically divided into two groups: one that uses the nearly free electron method and the other the tight-binding method (to be discussed below). Nevertheless, they all lead to similar results as they are merely different descriptions of the same phenomena. Here we have approximately described the band structure using the Kronig-Penney

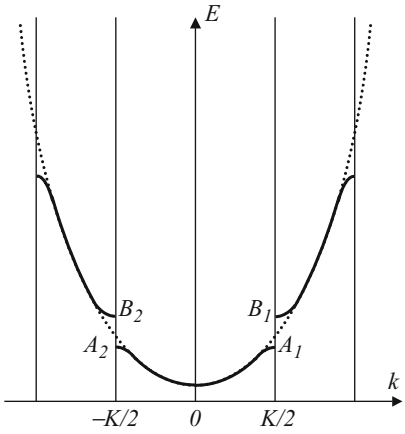


Fig. 5.7 Electron energy in a lattice (solid curve) and energy spectrum of free electrons (dashed curve). The deviation from the parabolic shape occurs at the Brillouin zone boundaries

model. In this subsection, we will briefly discuss the principle of the nearly free approximation (see Appendix A.7 for the pseudopotential approach).

This method is based on the assumption that the periodic potential introduces a small perturbation to the free electron state, i.e., a perturbation term is added to the potential energy in the Schrödinger equation, wavefunctions, and energy of the free particle to reflect this effect. Although these perturbations are small, the mathematical computation results in significant changes in the energy spectrum of a free electron. The reason is that the periodic potential scatters the electrons, and only the constructive interference of the waves survives and can propagate in the lattice as a Bloch function. The resulting band diagram in the extended-zone representation is depicted in Fig. 5.7 (solid line) and compared with that of a free electron (dashed lines).

The discontinuous curve results from the "reflections" that the electron waves with momenta of $\pm\hbar K/2$ experience at atomic lattice planes, where K is a reciprocal lattice vector (see Chap. 3 for reciprocal lattice). In the simple cubic lattice, $|K| = \frac{2\pi}{d}$ where d is the lattice constant. These locations correspond to the boundaries of the Brillouin zones defined in the previous subsection.

The energy difference between branches at points A_1 and B_1 (A_2 and B_2) is the energy gap that appears as a result of the periodic potential in the lattice. The value of the energy gap depends on the amplitude of the periodic potential. When the periodic potential reduces to be zero, the energy gaps close, and the spectrum becomes that of a free particle as shown in Fig. 4.5.

The band diagram can also be plotted in the reduced-zone representation where the energy spectrum is reduced to the smallest first Brillouin zone of range $\left[-\frac{K}{2}, +\frac{K}{2}\right]$ as shown in Fig. 5.8.

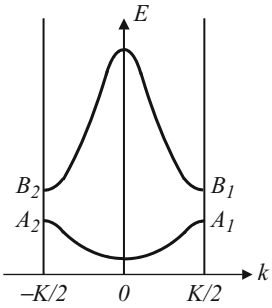


Fig. 5.8 Electron energy in the reduced-zone scheme

5.2.5 Tight-Binding Approximation

The other method commonly used to determine the band structure in a crystal, the tight-binding approximation, employs atomic wavefunctions as the basis set for the construction of the real wavefunction of an electron.

When initially isolated atoms with discrete electron energy levels are brought together and arranged in a lattice with small interatomic distances (typically ≈ 3–6 Å), the potential of each atom will be distorted due to the influence of other atoms. At the same time, the wavefunctions of electrons from different atoms will overlap, i.e., the probability of the presence of electrons from different atoms will be nonzero in the same position in space. These result in a nonzero probability for an electron to escape from one atom to the nearest neighbor. This causes a broadening of the initially discrete energy spectrum and creates energy bands of finite width instead. In other words, an electron does not live at a certain atomic energy level for an infinite time but travels from site to site which is equivalent to the movement of electrons in an energy band. Expressed mathematically, the Bloch superposition of localized orbitals gives us the tight-binding wavefunction:

$$\Psi_{\vec{k}}(\vec{r}) = \sum_{j,n} \beta_j \Phi_j(\vec{r} - \vec{R}_n) \exp(i\,\vec{k}\cdot\vec{R}_n) \tag{5.18}$$

where β_j are the admixture coefficients of the jth orbital and $\Phi_j(\vec{r} - \vec{R}_n)$ is the jth orbital itself on the atom located at $\vec{R}_n$, respectively. Substituting Eq. (5.18) into the time-independent Schrödinger equation allows us to calculate the energy bands. One does this to a good approximation by noting that the atomic problem (kinetic energy plus the potential of a given atom) is solved by the given orbital function, and the energy is known, i.e., using:

$$\left\{-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r} - \vec{R}_i)\right\}\Phi_\alpha(\vec{r} - \vec{R}_i) = E_\alpha \Phi_\alpha(\vec{r} - \vec{R}_i)$$

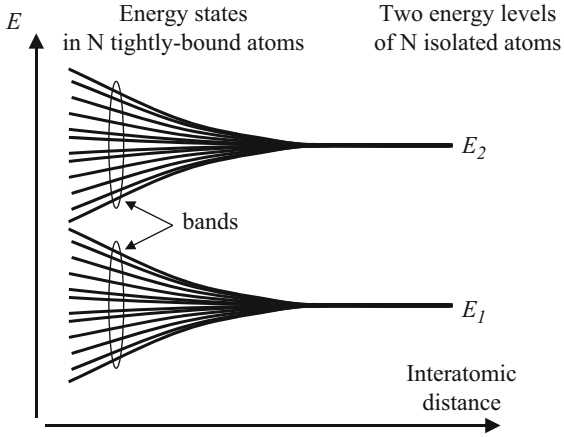


Fig. 5.9 Broadening of the atomic energy levels in a solid. When the atoms are isolated, they all have the same allowed discrete energy levels (e.g., E_1 and E_2). When the interatomic distance decreases, the atoms interact with one another and the allowed energy levels split: some increase while some others decrease

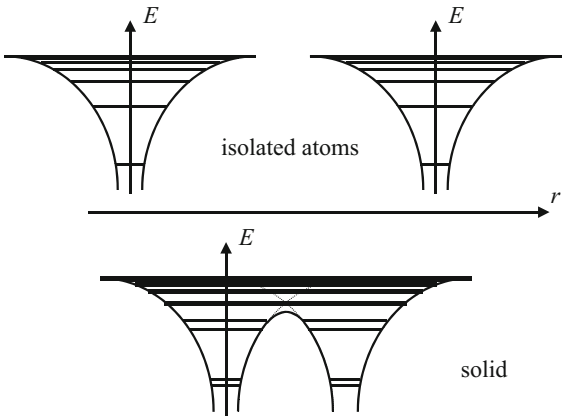


Fig. 5.10 Change in energy spectrum from single atoms to a solid. Each of the discrete energy levels in two isolated atoms split into two separate energy levels when the atoms are bound in a solid

where E_α is the energy of the atomic level and then multiplying both sides of the Schrödinger equation with a complex conjugate orbital state and then assuming the orthogonality of the orbitals centered on different sites. Normally, it is sufficient to keep only the nearest neighbor overlap terms $t_{l+1,l} = \int d\,\vec{r}\;\Phi^*\left(\vec{r}-\vec{R}_{l+1}\right)V\left(\vec{r}-\vec{R}_l\right)\Phi\left(\vec{r}-\vec{R}_l\right)$. This quantity is the so-called two-center integral, and this simplification makes the tight-binding method a good starting point for an approximate band structure calculation.

For the outer valence electrons which are usually of interest to us, the overlapping of wavefunctions is large, so the width of the energy band reaches several eV, i.e., is of the order of and even exceeds the spacing between the successive energy levels of an isolated atom. For electrons of the inner atomic shell, the level broadening is smaller, so the energy levels remain essentially sharp. The level broadening, which can be estimated to be zt where z is the number of nearest neighbors, and we take t_{ij}~t, is illustrated in Figs. 5.9 and 5.10.

Bringing atoms together and modifying their energy levels is the methodology of the "tight-binding approximation" because we start from tightly bound electrons in the atoms. This is in contrast with the previous nearly free electron approximation approach where we began with the free electron model and progressed by adding a periodic potential as a perturbation. With the tight-binding model, one arrives to a qualitatively similar band picture as that obtained from the nearly free electron model.

5.2.6 Dynamics of Electrons in a Crystal

The dynamics of electrons in a crystal can now be analyzed by considering an electron as a wavepacket. We will continue with the one-dimensional formalism of previous subsections.

Assuming that a wavepacket is centered on a frequency ω and a wavenumber k, the electron can be considered to be moving at a velocity v_g, called group velocity, which characterizes the speed of propagation of the energy that it transports. This velocity is defined by classical wave theory to be:

$$v_g = \frac{d\omega}{dk} \tag{5.19}$$

In quantum mechanics, this would correspond to the velocity of the electron. From the wave-particle duality, the frequency of the wave is related to the energy of the particle by $E = \hbar\omega$ and Eq. (5.19) thus becomes:

$$v_g = \frac{1}{\hbar}\frac{dE}{dk} \tag{5.20}$$

When an external force F acts on the wavepacket or electron so that a mechanical work is induced, it changes the energy E by the amount:

$$dE = Fdx = Fv_g dt \tag{5.21}$$

where dx is the distance over which the force is exerted during the interval of time dt. The force F can then be successively expressed as:

$$F = \frac{1}{v_g}\frac{dE}{dt} = \frac{1}{v_g}\frac{dE}{dk}\frac{dk}{dt} \tag{5.22}$$

or:

$$F = \hbar\frac{dk}{dt} = \frac{d(\hbar k)}{dt} \tag{5.23}$$

after using Eq. (5.20). On the other hand, differentiating Eq. (5.20) with respect to time leads to:

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

or:

$$\frac{dv_g}{dt} = \frac{1}{\hbar^2}\frac{d^2E}{dk^2}\frac{d(\hbar k)}{dt} \tag{5.24}$$

Eliminating $\frac{d(\hbar k)}{dt}$ in Eqs. (5.23) and (5.24), we find:

$$F = \left(\frac{1}{\frac{1}{\hbar^2}\frac{d^2E}{dk^2}}\right)\frac{dv_g}{dt} \tag{5.25}$$

This expression resembles Newton's law of motion when rewritten as:

$$F = m^*\frac{dv_g}{dt} \tag{5.26}$$

where we have defined m^* as:

$$m^* = \frac{\hbar^2}{{}^{d^2E}\!/_{dk^2}} \tag{5.27}$$

m^* is called the electron effective mass and has a very significant meaning in solid-state physics. Equation (5.26) shows that, in quantum mechanics, when external forces are exerted on the electron, the classical laws of dynamics can still be used if the mass is changed in the mathematical expressions for the effective mass of the electron.

Unlike the classical definition of mass, the effective mass is not a constant but depends on the band structure of the electron. The effective mass expresses a relationship between the band structure found in previous subsections and the dynamics of an electron in a solid. This shows us how important it is to determine the band structure in the first place and that an electron in a solid is very unlike an electron in vacuum.

For example, in the case of a free electron, the energy spectrum is parabolic (Eq. (4.35)):

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

where m is the mass of the electron. Using Eq. (5.27), the effective mass can be found to be $m^* = m$, which means that the effective mass of a free electron is equal to its classically defined mass.

However, when the energy spectrum is not parabolic with respect to the wavenumber k anymore, as, for example, depicted in Fig. 5.7, the effective mass differs from the classical mass. We thus see that the presence of a periodic potential

results in a value of effective mass different from the classical mass. The effective mass reflects the inverse of the curvature of the energy bands in k-space (i.e., $\frac{d^2E}{dk^2}$). Where the bands have a high curvature, m^* is small, while for bands with a small curvature (i.e., almost flat bands) m^* is large.

It is also worth noticing that since $\frac{d^2E}{dk^2}$ can be negative, m^* can also be negative, although it is not interpreted so, as we will see later by considering holes (Sect. 5.3.3). A negative effective mass means that the acceleration of the electron is in the direction opposite to the external force exerted on it, as shown in Eq. (5.26). This phenomenon is possible because of the wave-particle duality: an electron has wave-like properties and can therefore be reflected from the lattice planes when its wavevector satisfies the Bragg condition. Experimentally, if the momentum given to an electron from an external force is less than the momentum in the opposite direction given from the lattice (reflection), a negative electron effective mass will be observed.

Finally, it should also be noted that experiments conducted to measure the mass of an electron only lead to an estimate of its effective mass, or at least "components" of it.

Example

Q Assuming that the energy dispersion of a band in a semiconductor can be expressed as $E = Ak^2$, where $A = 84.67$ Å^2·eV, calculate the electron effective mass in this band, in units of free electron rest mass m_0.

A We make use of the formula: $m^* = \frac{1}{\frac{1}{\hbar^2}\frac{d^2E}{dk^2}} = \frac{1}{\frac{1}{\hbar^2}\frac{d^2(Ak^2)}{dk^2}} = \frac{\hbar^2}{2A}$. In units of free electron mass, we get:

$$\begin{aligned}\frac{m^*}{m_0} &= \frac{\hbar^2}{2Am_0} \\ &= \frac{\left(1.05458 \times 10^{-34}\right)^2}{2 \times \left(84.67 \times 10^{-20} \times 1.60218 \times 10^{-19}\right)\left(0.91095 \times 10^{-30}\right)} \\ &= 0.045\end{aligned}$$

5.2.7 Fermi Energy

We have seen so far that the electron energy spectrum in a solid consists of bands. These bands correspond to the allowed electron energy states. Since there are many electrons in a solid, it is not enough to know the energy spectrum for a single electron, but the distribution of electrons in these bands must also be known to understand the physical properties of a solid. Similar to the way the electrons fill the atomic orbitals with lower energies first (Chap. 1), the electrons in a crystal fill the lower energy bands first while satisfying the Pauli exclusion principle.

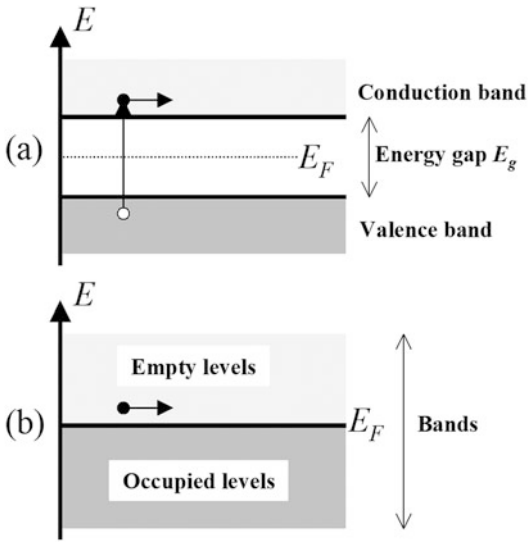


Fig. 5.11 Bands in (**a**) semiconductors and (**b**) metals. In most semiconductors, E_F is in the bandgap. In semiconductors, there is an energy region that does not contain allowed energy levels, and the Fermi energy is located in it. In metals, the Fermi energy is located inside an allowed energy band

Let us consider a solid where there are m energy levels and n electrons, at equilibrium. Usually these numbers are extremely large, and the number m of allowed energy levels (taking into account the spin degeneracy) in a solid is much larger than the number n of electrons ($m \gg n$): for instance, an iron metal with a volume of 1 cm^3 will have approximately 10^{22} atoms and 10^{24} electrons. At equilibrium, when no electron is in an excited state (e.g., at the absolute zero temperature, 0 K), the lowest n energy levels will be occupied by electrons, and the next remaining m-n energy levels remain empty.

If the highest occupied state is inside a band, the energy of this state is called the Fermi level and is denoted by E_F. That band is therefore only partially filled. This situation usually occurs for metals and is depicted in Fig. 5.11b. In the case of semiconductors, at $T = 0$ K, all bands are either full or empty. The Fermi level thus lies between the highest energy fully filled band (called valence band) and the lowest energy empty band (called conduction band), as shown in Fig. 5.11a. The energy gap between the valence band and the conduction band is called the bandgap and is denoted E_g.

The location of the Fermi level relative to the allowed energy bands is crucial in determining the electrical properties of a solid. Metals have a partially filled free electron band, since the Fermi level lies inside this band, which makes metals good electrical conductors because an applied electric field can push electrons easily into empty closely lying higher energy levels and in this way make them move in space and contribute to electrical conduction. By contrast, at 0 K, most semiconductors have completely filled or completely empty electron bands, which means that the Fermi energy lies inside a forbidden energy gap, and consequently the electric field cannot displace them from where they are in energy and therefore also not in space. Intrinsic semiconductors are poor electrical conductors at low temperatures. They

only conduct when carriers are thermally excited across the bandgap. The same can be said about insulators. Insulators differ from semiconductors in that their energy gap is much larger than k_bT, where $k_b(k_b = 1.38066 \times 10^{-23}\ \text{J}\cdot\text{K}^{-1} = 0.08625\ \text{meV}\cdot\text{K}^{-1})$ is the Boltzmann constant and T is the temperature in degrees K.

5.2.8 Electron Distribution Function

When the temperature is above the absolute zero, at thermal equilibrium, the electrons do not simply fill the lowest energy states first. We need to consider what is called the Fermi-Dirac statistics which gives the distribution of probability of an electron to have an energy E at temperature T:

$$f_e(E) = \frac{1}{\exp\left(\frac{E-E_F}{k_bT}\right) + 1} \tag{5.28}$$

where E_F is the Fermi energy and k_b is the Boltzmann constant. This distribution is called the Fermi-Dirac distribution and is plotted in Fig. 5.12 for various values of temperatures. This distribution function is obtained from statistical physics. In this description, the interaction between electrons is neglected, which is why we often talk of an electron gas.

In fact, a more general formulation of the Fermi-Dirac statistics involves a chemical potential μ instead of the Fermi energy E_F. This chemical potential depends on the temperature and any applied electrical potential. But in most cases of semiconductors, the difference between μ and E_F is very small at the temperatures usually considered.

At $T = 0$ K, the Fermi-Dirac distribution in Eq. (5.28) is equal to unity for $E < E_F$ and zero for $E > E_F$. This means that all the electrons in the crystal have their energy below E_F. At a temperature $T > 0$ K, the transition from unity to zero is less sharp. Nevertheless, for all temperatures, $f_e(E) = ½$ when $E = E_F$.

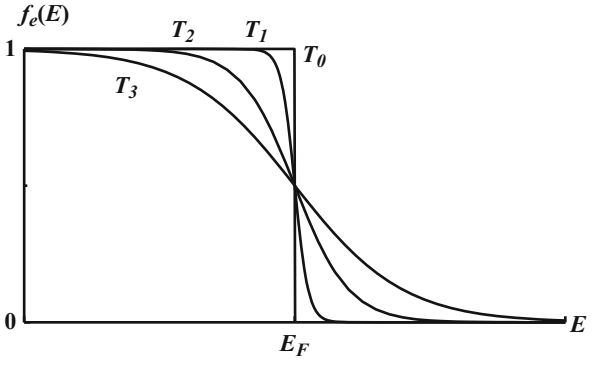


Fig. 5.12 Fermi-Dirac distribution function at different temperatures: $T_3 > T_2 > T_1, T_0 = 0$ K. At the absolute zero temperature, the probability of an electron to have an energy below the Fermi energy E_F is equal to 1, whereas its probability to have a higher energy is zero

To determine the Fermi energy, we must first introduce the concept of density of states. So far, we have somewhat indexed energy states individually, each having a certain energy. It is often more convenient to index these states according to their energy and determine the number of states which have the same energy.

5.3 Density of States (3D)

The concept of density of electronic states, or simply density of states, corresponds to the number of allowed electron energy states (taking into account spin degeneracy) per unit energy interval around an energy E. Most properties of crystals and especially semiconductors, including their optical, thermodynamic and transport properties, are determined by their density of states. In addition, one of the main motivations for considering low-dimensional quantum structures is the ability to engineer their density of states. In this section, we will present the calculation of the density of states in a bulk three-dimensional crystal, which will serve as the basis for that of low-dimensional quantum structures.

An ideal crystal has a periodic structure, which means that it has to be infinite since a surface would violate its periodicity. However, real crystals have a finite volume. We saw in Sect. 5.2 that one way to reconcile these two apparently paradoxical features in crystals was to exclude surfaces from consideration by using periodic boundary conditions (Born-von Karman). This allows us to just consider a sample of finite volume which is periodically repeated in all three orthogonal directions. A very important consequence of this was the quantization of the wavenumber k of the electron states in a crystal, as expressed through Eq. (5.7).

The analysis in Sect. 5.2 was primarily conducted in one spatial dimension (x) for the sake of simplicity. Here, it will be more appropriate to consider all three dimensions, i.e., to use $\vec{r} = (x, y, z)$.

5.3.1 Direct Calculation

Let us assume that the shape of the crystal is a rectangular parallelepiped of linear dimensions L_x, L_y, and L_z and volume $V = L_x L_y L_z$. The periodic boundary conditions, similar to Eq. (5.5), require the electron quantum states to be the same at opposite surfaces of the sample:

$$\Psi(x + L_x, y, z) = \Psi\big(x, y + L_y, z\big) = \Psi(x, y, z + L_z) = \Psi(x, y, z) \tag{5.29}$$

Using the Bloch theorem, these conditions mean that:

$$\exp\left(ik_x L_x\right) = \exp\left(ik_y L_y\right) = \exp\left(ik_z L_z\right) = 1 \tag{5.30}$$

or:

$$\begin{cases} k_x = \dfrac{2\pi}{L_1} n_x \\ k_y = \dfrac{2\pi}{L_2} n_y \\ k_z = \dfrac{2\pi}{L_3} n_z \end{cases} \tag{5.31}$$

where $n_x, n_y, n_z = 0, \quad \pm 1 \ldots$ are integers, while k_x, k_y, and k_z are the wavenumbers in the three orthogonal directions. These are in fact the coordinates of the electron wavenumber vector or wavevector $\vec{k} = (k_x, k_y, k_z)$. Therefore, the main result of the periodic boundary conditions is that the wavevector $\vec{k}$ of an electron in a crystal is not a continuous variable but is discrete. Equation (5.31) actually defines a lattice for the wavevector $\vec{k}$, and the space in which this lattice exists is in fact the k-space or reciprocal space.

The volume of the smallest unit cell in this lattice is then $\frac{(2\pi)^3}{L_x L_y L_z} = \frac{(2\pi)^3}{V}$. From Chap. 3, we know that there is exactly one lattice point in each such volume, which means that the density of allowed $\vec{k}$ is uniform and equal to $\frac{V}{(2\pi)^3}$ in k-space. Moreover, from Chap. 3, we recall that the wavevector $\vec{k}$ was used to index electron wavefunctions and therefore allowed electron states. The density of electron states per unit k-space volume is therefore equal to:

$$g(\vec{k}) = 2\frac{V}{(2\pi)^3} \tag{5.32}$$

where the extra factor of 2 arises from the spin degeneracy of electrons.

Example

Q Calculate the density of states in k-space for a cubic crystal with a side of only 1 mm. Is the density of state in k-space too low?

A The density of states in k-space is given by: $g(\vec{k}) = 2\frac{V}{(2\pi)^3} = 2 \times \frac{1\text{mm}^3}{8\pi^3} = 8.063 \times 10^{-3}$ mm^3. This number may look small, but if we compare with the volume of the first Brillouin zone, we will find that this density of states is actually very high. For example, for a face-centered cubic lattice with a lattice constant of $a = 5.65325$ Å (e.g., GaAs), the volume of its first Brillouin zone in k-space is given by: $V_k = 32\left(\frac{\pi}{a}\right)^3 = 5.492$ Å^{-3}. Therefore, the total number of possible states in this first Brillouin zone is:

$$N = V_k g(\vec{k}) = \left(5.492 \times 10^{21}\text{mm}^{-3}\right)\left(8.063 \times 10^{-3}\text{mm}^3\right) \approx 4.43 \times 10^{19}$$

The density of states $g(E)$ as defined earlier is therefore related to its counterpart in k-space, $g(\vec{k})$, by:

$$g(E)dE = g(\vec{k})d\ \vec{k} \tag{5.33}$$

where dE and $d\ \vec{k}$ are unit interval of energy and the unit volume in k-space, respectively. In order to obtain $g(E)$, one must first know the $E(\vec{k})$ relationship, which is equivalent to the E–k relationship in one dimension and which gives the number of wavevectors $\vec{k}$ associated with a given energy E. This is a critical step because the differences in the density of states of a bulk semiconductor crystal, a quantum well, a quantum wire, and a quantum dot arise from it.

For a bulk semiconductor crystal, the electron density of states is calculated near the bottom of the conduction band because this is where the electrons which give rise to the most important physical properties are located. Furthermore, we choose the origin of the energy at the bottom of this band, i.e., $E_C = 0$. Extrapolating from the results of Sect. 5.2, the shape of the $E(\vec{k})$ relationship near the bottom of the conduction band can generally be considered parabolic:

$$E(\vec{k}) = \frac{\hbar^2 k^2}{2m^*} \tag{5.34}$$

where k is the norm or length of the wavevector $\vec{k}$, and m^* is the electron effective mass as defined in Sect. 5.2.6. Using this expression, we can express successively:

$$dE = \frac{\hbar^2}{2m^*}(2k)dk \tag{5.35}$$

When considering orthogonal coordinates, the unit volume in k-space is defined given by:

$$d\ \vec{k} = dk_x dk_y dk_z \tag{5.36}$$

which is equal, when using spherical coordinates, to:

$$d\ \vec{k} = d\left(\frac{4\pi}{3}k^3\right) = 4\pi k^2 dk \tag{5.37}$$

Therefore, by replacing into Eq. (5.35), we get:

$$dE = \frac{\hbar^2}{2m^*}\left(\frac{1}{2\pi k}\right)d\ \vec{k} \tag{5.38}$$

Using Eq. (5.34) to express k in terms of E, and replacing into Eq. (5.38):

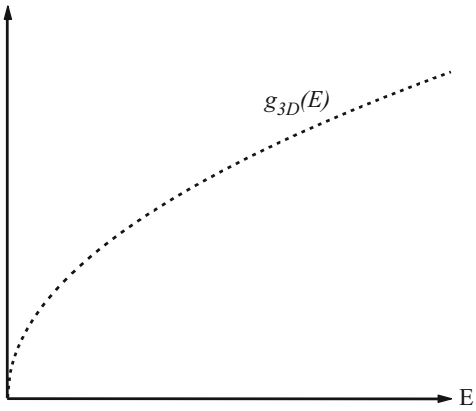


Fig. 5.13 Energy dependence of density of states for a three-dimensional semiconductor conduction band. The density of states follows a parabolic relationship

$$\begin{aligned} dE &= \frac{\hbar^2}{2m^*}\left(\frac{1}{2\pi}\sqrt{\frac{\hbar^2}{2m^*E}}\right)d\vec{k} \\ &= \frac{1}{2\pi}\left(\frac{\hbar^2}{2m^*}\right)^{3/2}\frac{1}{\sqrt{E}}d\vec{k} \end{aligned} \tag{5.39}$$

Now, by replacing into Eq. (5.33), we obtain successively:

$$g(E) = 2\pi\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\sqrt{E}g\left(\vec{k}\right)$$

Finally, using Eq. (5.32), we get:

$$g_{3D}(E) = \frac{V}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\sqrt{E} \tag{5.40}$$

where a "3D" subscript has been added to indicate that this density of states corresponding to the conduction band of a bulk three-dimensional semiconductor crystal. This density of states is shown in Fig. 5.13.

Note that, if the origin of the energies has not been chosen to be the bottom of the band (i.e., $E_c \neq 0$), then $\sqrt{E}$ would be replaced by $\sqrt{E - E_c}$.

Example

Q: Calculate the number of states from the bottom of the conduction band to 1 eV above it, for a 1 mm^3 GaAs crystal. Assume the electron effective mass is $m^* = 0.067m_0$ in GaAs.

A: The number of states from 0 to 1 eV above the bottom of the conduction band is obtained by integrating the three-dimensional density of states $g_{3D}(E)$:

$N = \int_0^{1\,eV} g_{3D}(E)dE$. Since the expression for $g_{3D}(E)$ is given by: $g_{3D}(E) = \frac{V}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\sqrt{E}$, we obtain:

$$
\begin{aligned}
N &= \int_0^{1\,eV} g_{3D}(E)dE = \frac{V}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\int_0^{1\,eV}\sqrt{E}dE \\
&= \frac{V}{3\pi^2}\left(\frac{2m^* \times 1eV}{\hbar^2}\right)^{3/2} \\
&= \frac{\left(10^{-3}\right)^3}{3\pi^2}\left(\frac{2\left(0.067^*0.91095 \times 10^{-30}\right) \times \left(1.60218 \times 10^{-19}\right)}{\left(1.05458 \times 10^{-34}\right)^2}\right)^{3/2} \\
&\approx 7.88 \times 10^{16}
\end{aligned}
$$

5.3.2 Other Approach

A more elegant approach, but more mathematically challenging way, to calculate the density of states is presented here. This method will prove easier when calculating the density of states of low-dimensional quantum structures. The density of states $g(E)$ as defined earlier can be conceptually written as the sum: $g(E) = 2 \times$ (number of states which have an energy $E\left(\vec{k}\right)$ equal to E) which can be mathematically expressed as:

$$
g(E) = 2\sum_{\vec{k}} \delta\left[E\left(\vec{k}\right) - E\right] \tag{5.41}
$$

where the summation is performed over all values of wavevector $\vec{k}$, since it is used to index the allowed electron states. $\delta(x)$ is a special even function, called the Dirac delta function, and is defined as:

$$
\begin{cases}
\delta(x) = 0 & \text{for } x \neq 0 \\
\int_{-\infty}^{+\infty} \delta(x)dx = 1
\end{cases} \tag{5.42}
$$

Some of the most important properties of the Dirac delta function include:

$$\begin{cases} \int_{-\infty}^{+\infty} \delta(x)Y(x)dx = Y(0) \\ \int_{-\infty}^{+\infty} \delta(x-x_0)Y(x)dx = Y(x_0) \end{cases} \tag{5.43}$$

In addition, in crystals of macroscopic sizes, the differences between nearest values of $\vec{k}$ are small, as they are proportional to $\frac{1}{L_x}, \frac{1}{L_y}$, or $\frac{1}{L_z}$. Therefore, in practice, the discrete variable $\vec{k}$ can be considered as quasi-continuous. For this reason, the summation of a function $\mathrm{Y}\left(\vec{k}\right)$ over all allowed states represented by a wavevector $\vec{k}$ in k-space can be replaced by an integration over a continuously variable $\vec{k}$ such that:

$$\sum_{\vec{k}} \mathrm{Y}\left(\vec{k}\right) \equiv \frac{V}{(2\pi)^3} \iiint_{\vec{k}} \mathrm{Y}\left(\vec{k}\right) d\,\vec{k} = \frac{V}{(2\pi)^3} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{Y}\left(k_x, k_y, k_z\right) dk_x dk_y dk_z \tag{5.44}$$

The factor $\frac{V}{(2\pi)^3}$ is the volume occupied by a reciprocal lattice point in k-space. Eq. (5.41) can therefore be rewritten into:

$$g(E) = \frac{1}{4\pi^3} \iiint_{\vec{k}} \delta\left[E\left(\vec{k}\right) - E\right] d\,\vec{k} \tag{5.45}$$

Now, we need to use the expression of $d\left[E\left(\vec{k}\right)\right]$ as a function of $d\,\vec{k}$ found in Eq. (5.39):

$$d\left[E\left(\vec{k}\right)\right] = \frac{1}{2\pi}\left(\frac{\hbar^2}{2m^*}\right)^{3/2} \frac{1}{\sqrt{E\left(\vec{k}\right)}} d\,\vec{k} \tag{5.46}$$

Equation (5.45) therefore becomes:

$$g(E) = \frac{2\pi V}{4\pi^3}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} \int_0^{\infty} \delta\left[E\left(\vec{k}\right) - E\right] \sqrt{E\left(\vec{k}\right)} d\left[E\left(\vec{k}\right)\right] \tag{5.47}$$

and after the change of variable $E\left(\vec{k}\right) \rightarrow x$:

$$g(E) = \frac{V}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\int_0^\infty \delta(x-E)\sqrt{x}dx \tag{5.48}$$

Using Eq. (5.43), and because $E > 0$:

$$g(E) = \frac{V}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\sqrt{E} \tag{5.49}$$

which is the same expression as Eq. (5.40) for $g_{3D}(E)$.

Therefore, the knowledge of the *Fermi-Dirac distribution*, which gives us the probability of the presence of an electron with energy E, and the *density of states*, which tells how many electrons are allowed with an energy E, together permit the determination of the distribution of electrons in the energy bands. The total number of electrons in the solid, n_{total}, is therefore obtained by summing the product of the Fermi-Dirac distribution and the density of states over all values of energy:

$$n_{\text{total}} = \int_0^\infty g(E)f_e(E)dE \tag{5.50}$$

Because E_F is embedded into the function $f_e(E)$, this equation shows us how the Fermi energy can be calculated.

One important parameter for semiconductor devices is the concentration or density of electrons n in the conduction band. The following discussion provides a simplified overview of the formalism commonly used for this parameter and illustrates well the use of the Fermi-Dirac distribution. A more detailed analysis will be provided in Chap. 7 in which we will discuss the equilibrium electronic properties of semiconductors. Here, the density of electrons n, with effective mass m_c, in the conduction band is given by:

$$n = \frac{1}{V}\int_{E_C}^\infty g(E)f_e(E)dE \tag{5.51}$$

where the integration starts from E_C which is the energy at the bottom of the conduction band. In a bulk semiconductor, the density of states $g(E)$ in the conduction band is, as derived above, given by:

$$g(E) = \frac{V}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}(E-E_C)^{1/2} \tag{5.52}$$

Combining this expression with Eq. (5.28), the density of electrons becomes:

$$n = \frac{1}{2\pi^2}\left(\frac{2m_c}{\hbar^2}\right)^{3/2}\int_{E_C}^{\infty}(E-E_C)^{1/2}\frac{1}{\exp\left(\frac{E-E_F}{k_bT}\right)+1}dE \tag{5.53}$$

or:

$$n = N_c F_{\frac{1}{2}}\left(\frac{E_F - E_C}{k_bT}\right) \tag{5.54}$$

where:

$$N_c = 2\left(\frac{2\pi k_b T m_e}{h^2}\right)^{3/2} \tag{5.55}$$

is the effective density of states in the conduction band, and:

$$F_{\frac{1}{2}}(x) = \frac{2}{\sqrt{\pi}}\int_0^{\infty}\frac{y^{1/2}}{1+\exp(y-x)}dy \tag{5.56}$$

is the Fermi-Dirac integral. A more detailed discussion on the effective density of states and the Fermi-Dirac integral will be given in Chap. 7.

5.3.3 Electrons and Holes

We have seen that when the curvature of the *E–k* energy spectrum is positive, such as near point O in the bottom band in Fig. 5.8, the electron effective mass is positive.

However, when the curvature is negative, such as near point A_1 in this same band, the effective mass of the electron as calculated in Sect. 5.2.6 would be negative. In this case, it is more convenient to introduce the concept of holes. A hole can be viewed as an allowed energy state that is non-occupied by an electron in an almost filled band. Figures 5.14a, b are equivalent descriptions of the same physical phenomenon. In Fig. 5.14a, we are showing the energy states occupied by electrons. In Fig. 5.14b, we are showing the energy states in the valence band which are occupied by holes, i.e., vacated by electrons.

Electrons can move in such a band only through an electron filling this non-occupied state and thus leaving a new non-occupied state behind. By doing so, it is as if the vacated space or hole had also moved, but in the *opposite direction*, which means that the effective mass of the hole is therefore opposite that of the electron that would be at that same position, in other words, the effective mass of the hole is positive near point A_1 in Fig. 5.8 and is computed as:

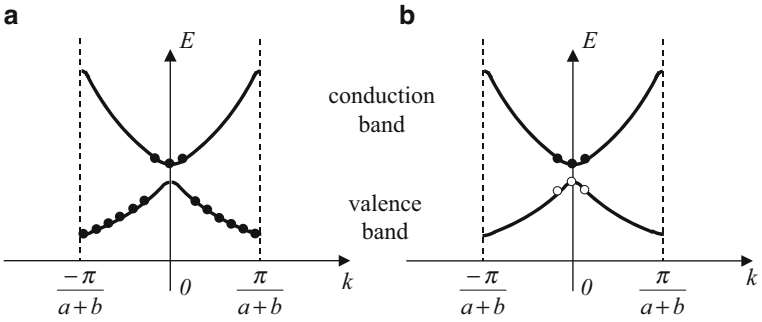


Fig. 5.14 Electron energy states in the reduced-zone scheme. In (**a**), the solid circles show the states occupied by electrons. In (**b**), the closed circles show the states in the conduction band which are occupied by electrons, and the open circles the states in the valence band occupied by holes

$$m^* = -\frac{\hbar^2}{d^2E/dk^2} \tag{5.57}$$

A hole can be viewed as a positively charged particle (energy state vacated by an electron). Holes participate in the electrical charge transfer (electrical current) and energy transfer (thermal conductivity).

Let us consider the concept of holes in more details. The probability of the state k to be occupied by an electron is $f_e(k)$. The probability of the state not to be occupied is the probability to find hole in the state k and can be written as:

$$f_h(k) = 1 - f_e(k) \tag{5.58}$$

The electrical current from the electrons in the band is:

$$j = -2q \sum_k f_e(k) v_k \tag{5.59}$$

where ν_k is the electron velocity at state k, q is the electron charge ($q > 0$) and the summation is performed over all states with wavenumber k in the first Brillouin zone. This can be rewritten as:

$$\begin{aligned} j &= -2q \sum_k f_e(k) v_k = -2q \sum_k [1 - f_h(k)] v_k \\ &= -2q \sum_k v_k + 2q \sum_k f_h(k) v_k \end{aligned} \tag{5.60}$$

We can now use the fact that the electron energy spectrum is always symmetrical, i.e., $E(k) = E(-k)$; hence $v_k = -v_{-k}$ from Eq. (5.20), and the sum of velocities over the entire first Brillouin zone is zero. The first sum in Eq. (5.60) is thus equal to zero and we obtain:

$$j = +2q \sum_k f_{\mathrm{h}}(k) v_{\mathrm{k}} \tag{5.61}$$

Therefore, the electrical current in a band incompletely filled with electrons moving at speed v_{k} is equivalent to the current of positively charged holes moving at speed v_{k}. We thus see that in a band incompletely filled with electrons, the electrical current can be represented by flow of positively charged particles-holes.

5.4 Band Structures in Real Semiconductors

In three-dimensional crystals with three-dimensional reciprocal lattices, the use of a reduced-zone representation is no longer merely a convenience. It is essential; otherwise, the representation of the electronic states becomes too complex. How then can we display the band structure information from a three-dimensional crystal, which needs of course four dimensions (E, k_x, k_y, and k_z) to describe it? The answer is to make representations of certain important symmetry directions in the three-dimensional Brillouin zone as one-dimensional E versus k plots. Only by doing so can we get all the important information onto a two-dimensional page. Therefore, when looking at an E–k diagram, one is looking at different sections cut out of the k-space. In addition, to simplify the diagram, we consider that k varies continuously. Indeed, the difference between two values of k is $\Delta k = \frac{2\pi}{Na}$, where the lattice parameter a is around several angstroms and the order of magnitude of N is 10^8. And the length of the side of the Brillouin zone is $\frac{2\pi}{a} \approx 6.28 * 10^{10}\ \mathrm{m}^{-1} \gg \frac{2\pi}{Na} \approx 6.28 * 10^{10}\ \mathrm{m}^{-1}$. As a result, at the scale of the reciprocal lattice, the wavenumber can be considered to vary continuously.

5.4.1 First Brillouin Zone of an fcc Lattice

The first Brillouin zone of an fcc lattice is shown in Fig. 5.15. Certain symmetry points of the Brillouin zone are marked. Roman letters are mostly used for symmetry points and Greek letters for symmetry directions, specifically the Γ, X, W, K, and L points and the directions Δ, Λ, and Σ. The following is a summary of the standard symbols and their locations in k-space, with a the side of the conventional cubic unit cell:

$$\Gamma \quad \frac{2\pi}{a}(0,0,0)$$

$$\mathrm{X} \quad \frac{2\pi}{a}(0,0,1)$$

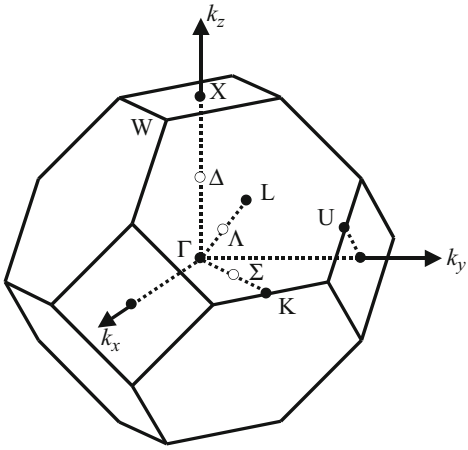


Fig. 5.15 First Brillouin zone of an fcc lattice

$$\text{W} \quad \frac{2\pi}{a}(½, 0, 1)$$

$$\text{K} \quad \frac{2\pi}{a}(¾, ¾, 0)$$

Note that there may be several equivalent positions for each of these points. For example, there are six equivalent X symmetry points, located at coordinates $\frac{2\pi}{a}$ (0,0,±1), $\frac{2\pi}{a}$(0, ±1,0), and $\frac{2\pi}{a}$(±1,0,0).

Using Miller indices, the symmetry directions can be denoted as:

$$\Delta: \quad \Gamma \rightarrow \text{X (parallel to } < 100 >)$$

$$\Lambda: \quad \Gamma \rightarrow \text{L (parallel to } < 111 >)$$

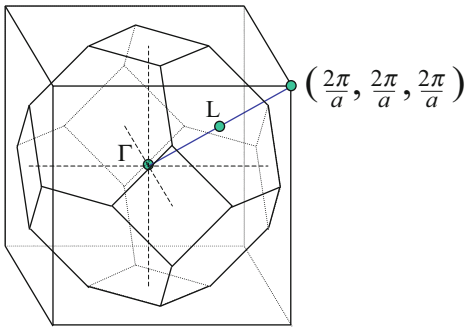
$$\Sigma: \quad \Gamma \rightarrow \text{K (parallel to } < 110 >).$$

These notations come from the crystal group theory where they are used to label the symmetry operation groups at those particular high-symmetry points and directions. For example, Γ is the symmetry group at the zone center ($\vec{k} = (0, 0, 0)$) and is isomorphic to the lattice point group.

Example

Q: Determine the coordinates of the L point in the first Brillouin zone of a face-centered cubic lattice.

A: The first Brillouin zone of a face-centered cubic lattice with side a is body-centered cubic with a side equal to $\frac{4\pi}{a}$ in the k_x, k_y, and k_z directions, as shown in the figure below. Let us take the Γ point at the center of the first Brillouin zone. The L point is exactly at the bisection point of Γ and the lattice point at ($\frac{2\pi}{a}$,$\frac{2\pi}{a}$,$\frac{2\pi}{a}$). Its coordinates are thus: ($\frac{\pi}{a}$,$\frac{\pi}{a}$,$\frac{\pi}{a}$).



5.4.2 First Brillouin Zone of a bcc Lattice

Similarly, the first Brillouin zone of a bcc lattice can be described in terms of its principal symmetry directions as it is shown in Fig. 5.16.

The symmetry points are conventionally represented as Γ, H, P, and N, and the symmetry directions as Δ, Λ, D, Σ, and G. The various symmetry points are:

$$\Gamma \quad \frac{2\pi}{a}(0,0,0)$$

$$\mathrm{H} \quad \frac{2\pi}{a}(0,0,1)$$

$$\mathrm{P} \quad \frac{2\pi}{a}(½,½,½)$$

$$\mathrm{N} \quad \frac{2\pi}{a}(½,½,0).$$

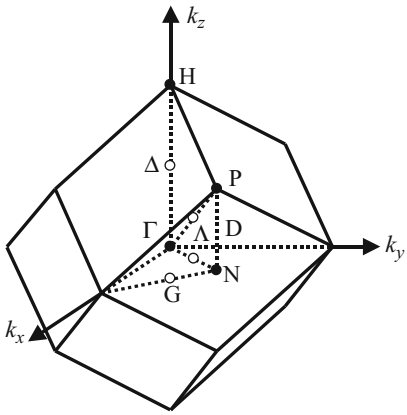


Fig. 5.16 First Brillouin zone of a bcc lattice

Using Miller indices for the directions:

$$\Delta: \quad \Gamma \rightarrow \mathrm{H}\ (\text{parallel to } < 100 >)$$

$$\Lambda: \quad \Gamma \rightarrow \mathrm{P}\ (\text{parallel to } < 111 >)$$

$$\mathrm{D}: \quad \mathrm{N} \rightarrow \mathrm{P}\ (\text{parallel to } < 100 >)$$

$$\Sigma: \quad \Gamma \rightarrow \mathrm{N}\ (\text{parallel to } < 110 >)$$

$$\mathrm{G}: \quad \mathrm{N} \rightarrow \mathrm{H}\ (\text{parallel to } < 1\bar{1}0 >).$$

5.4.3 First Brillouin Zones of a Few Semiconductors

As discussed in Chap. 3, many semiconductors have the diamond or zinc blende lattice structures. In these cases, the extrema in the E–k relations occur at the zone center or lie, for example, along the high-symmetry Δ (or <100>) and Λ (or <111>) directions. The important physical properties involving electrons in a crystal can thus be derived from plots of the allowed energy E versus the magnitude of k along these high-symmetry directions.

Figure 5.17 depicts the E–k diagrams characterizing the band structures in Ge (Fig. 5.17a), Si (Fig. 5.17b), and GaAs (Fig. 5.17c). The lines shown here represent bands in the semiconductor. The three lower sets of lines correspond to the valence band, while the upper bands correspond to the conduction bands. Note that the energy scale in these diagrams is referenced to the energy at the top of the valence band, E_V is the maximum valence band energy, E_C the minimum conduction-band energy, and$E_g = E_C$–E_V the bandgap. This is only a conventional choice, and the origin of energy can be chosen elsewhere.

The plots in Fig. 5.17 are two-direction composite diagrams. The <111> direction is toward point L, and the <100> direction is toward point X. Because of crystal symmetry, the $-\vec{k}$ portions of the diagrams are just the mirror images of the corresponding $+\vec{k}$ portions. It is therefore standard practice to delete the negative portions of the diagrams. The left-hand portions($\Gamma \rightarrow$ L) of the diagrams are shorter than the right-hand portions ($\Gamma \rightarrow$ X) as expected from the geometry of Brillouin zone.

Valence Band

In all cases, the valence band maximum occurs at the zone center, at $k = 0$. The valence band in each of the materials is actually composed of three subbands. Two of the bands are degenerate (have the same energy) at $k = 0$, while the third band is split from the other two. In Si, the upper two bands are almost indistinguishable in Fig. 5.17b and the maximum of the third band is only 0.044 eV below E_V at $k = 0$.

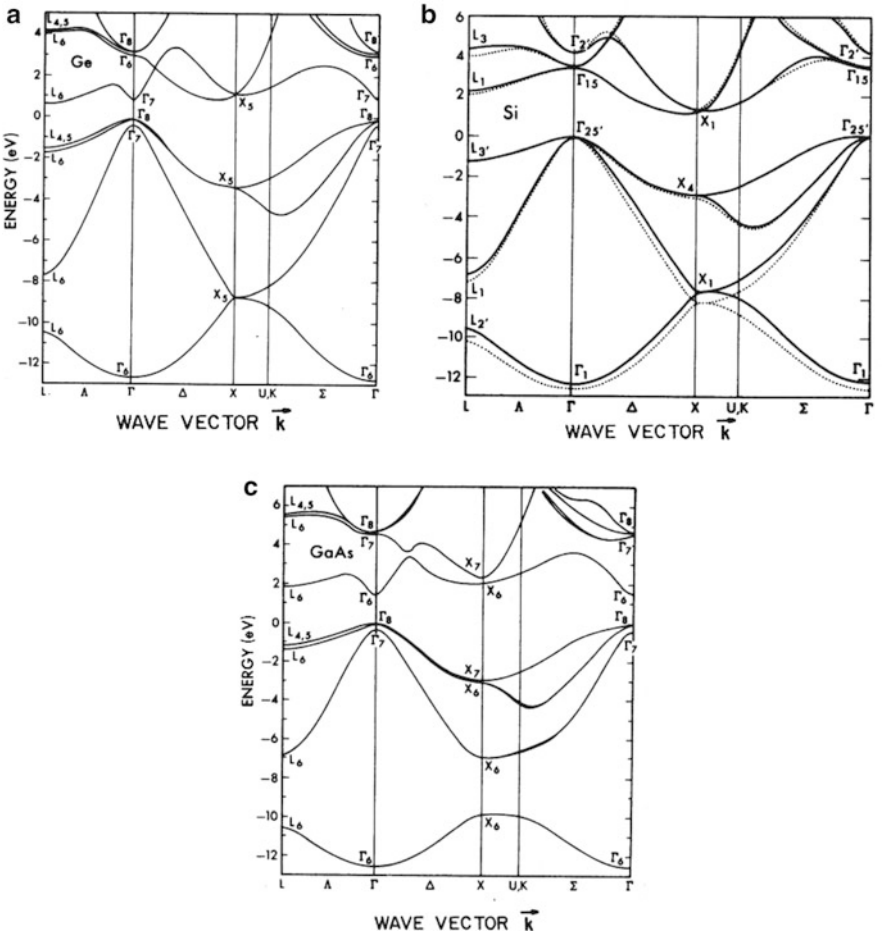


Fig. 5.17 *E–k* diagram of a few semiconductor crystals: (**a**) Ge, (**b**) Si, and (**c**) GaAs. The structures of the conduction and valence bands are plotted. The origin of the energy is chosen to be at the top of the valence band. (Reprinted figure with permission from Chelikowsky and Cohen (1976). Copyright 1976 by the American Physical Society)

The degenerate band with the smaller curvature about $k = 0$ is called the heavy-hole band, and the other with larger curvature is called the light-hole band. The band maximizing at a slightly reduced energy is called the spin-orbit split-off band (see the Kane effective mass method in Sect. 5.6).

Conduction Band

There are a number of subbands in each of the conduction bands shown in Fig. 5.17. These subbands exhibit several local minima at various positions in the Brillouin zone. However – and this is very significant – the position of the conduction band

absolute minimum in *k*-space, which is the lowest minimum among all these subbands and which is where the electrons tend to accumulate, varies from material to material.

In Ge the conduction band (absolute) minimum occurs right at point *L*, the zone boundary along the Λ or <111> direction in Fig. 5.17a. Actually, there are *eight equivalent conduction band minima* since there are eight equivalent <111> directions. However, each minimum is equally shared with the neighboring zone, and there is therefore only a *fourfold degeneracy* or a *multiplicity of four*. The other local minima in the conduction band occurring at higher energies are less populated and are therefore less important.

The Si conduction band absolute minimum occurs at $k \approx 0.8(2\pi/a)$ from the zone center along the Δ or <100> direction. The sixfold symmetry of the <100> directions gives rise to *six equivalent conduction band minima* within the Brillouin zone. The other local minima in the Si conduction band occur at considerably higher energies and are typically not important as they would only have a negligible electron population unless some very strong force could activate carriers to these higher extrema or if the temperature is much higher.

Among the materials considered in Fig. 5.17, GaAs is unique in that the conduction band minimum occurs at the zone center directly over the valence band maximum. Moreover, the L-valley minimum at the zone boundary along the <111> directions lies only 0.29 eV above the absolute conduction band minimum at Γ. Even in thermal equilibrium at room temperature, the L-valley contains a non-negligible electron population. The transfer of electrons from the Γ-valley to the L-valley can, for example, happen at high electric fields when electrons are heated up to high velocity. The transfer keeps the high energy but gives them a high effective mass which slows them down in space. When they slow down, they force the new electrons coming in to slow down too, until they, the transferred valley charge has exited. This results in a self-oscillating current state and is an essential feature for some device operations such as in charge-transferred electron devices (e.g., Gunn diodes, etc.).

Having discussed the properties of the conduction and valence bands separately, we must point out that the relative positions of the band extreme points in *k*-space are in itself an important material property. When the conduction band minimum and the valence band maximum occur at the same value of *k*, the material is said to be direct-gap type. Conversely, when the conduction band minimum and the valence band maximum occur at different values of *k*, the material is called indirect-gap type.

Of the three semiconductors considered, GaAs is an example of a direct-gap material, while Ge and Si are indirect-gap materials. The direct or indirect nature of a semiconductor has a very significant effect on the properties exhibited by the material, particularly its optical properties. The direct nature of GaAs, for example, makes it ideally suited for use in semiconductor lasers and infrared light-emitting diodes.

5.5 Two-Dimensional Semiconductors and Transition Metal Dichalcogenides "TMDC"

5.5.1 Examples: Graphene (G) and TMDC

One of the great discoveries of recent times is the exfoliation of the material which has been named graphene (G) [Geim, Castro, Avouris]. It was fabricated initially by exfoliation from graphite and consists ideally of a single carbon sheet. The good news is that this monolayer of carbon is strong enough to survive experimental manipulation and temperature, even in suspension, and can therefore be used in making devices. Graphene turned out to be so interesting that the discoverers Novoselov and Geim were awarded the Noble prize. The literature on graphene is now huge, and we will here only focus on a few noteworthy aspects which have enriched semiconductor science. The interested reader is strongly urged to consult the vast literature.

The structure of G is hexagonal two-dimensional carbon and shown in Fig. 5.18.

5.5.2 Graphene Band Structure: Nearest Neighbor Tight Binding

The simplest and most popular way of deriving the graphene band structure is to use the tight-binding method described in our *Appendix 2 in Chap. 2*. Using A and B to denote the two types of atoms which form the hexagonal lattice (see Fig. 5.18), we can assign a valence orbital to every carbon atom and allow these orbitals to couple to generate the graphene energy bands. Consider the t-b (tight-binding) Hamiltonian for this lattice and drop the Coulomb interaction between the electrons, steps which can be justified later. Using second quantization, and the creation $c_{i\sigma}^{+}$ and annihilation $c_{i\sigma}$ operators for electrons at atomic orbitals at a given site "i" [Da Sarma et al, Castro et al] [see Chap. 16 electron phonon interaction for second quantization]:

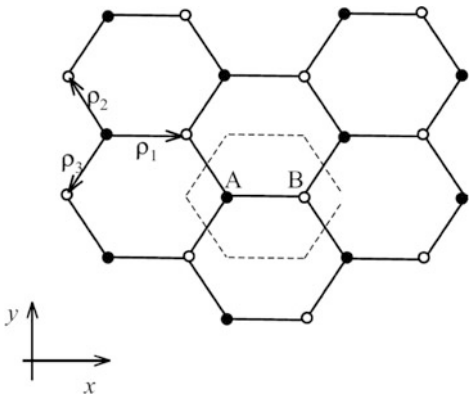


Fig. 5.18 Hexagonal lattice of a 2D material

$$H_e = \sum_{i,\sigma} \varepsilon_{i\sigma} c_{i\sigma}^{+} c_{i\sigma} + \sum_{i \neq j,\sigma} t_{ij} c_{i\sigma}^{+} c_{j\sigma} \tag{5.62}$$

$\varepsilon_{i\sigma}$ are the atomic orbital energies with spin index σ, and "t" is the tight-binding coupling matrix element linking two neighboring orbital orbitals.

Now let us set up the Heisenberg equation of motion [see FSSE Chap. 4] for the amplitude of the wavefunction at atomic sites or corresponding operator with atom of type "a," for example, E = energy:

$$(E - \varepsilon_{i\sigma}) c_{i\sigma}^{a} = \sum_{j} t_{ij} c_{j\sigma}^{b} \tag{5.63}$$

The sum j goes over the n.n and we note that the neighboring atoms a, b are *not equivalent by translational symmetry*, though apparently physically completely equivalent, so that the Bloch periodicity argument cannot be used straight away. In order to recover an *equivalent atom* and solve the equations by symmetry, we have to go one step further and set up a similar relation for the b-sites as we did for the a-sites with Eq. 5.63, thus we have:

$$(E - \varepsilon_{i\sigma}) c_{i\sigma}^{b} = \sum_{j} t_{ij} c_{j\sigma}^{a} \tag{5.64}$$

Now, we substitute Eq. (5.64) in Eq. (5.63) and relate two equivalent **a** or **b** atoms at distance **R** by the Bloch's phase factor exp(**ik.R**) in the usual way we can solve the problem. The unusual linear dispersion at the points in **k** space **K and K′** as shown in Fig. 5.19 now called Dirac points, where the gap is zero, is due to the fact that the two sets of lattice points "a and b" are completely equivalent apart from the fact that they are mirror symmetrical not translationally symmetrical. This topological restriction, analogous to the restriction of having to move below light speed at all times, then gives rise to space splitting and a new pseudo spin-like quantum number. One can see that this notion can be generalized to an infinity of topologies, which could have this, and indeed far more complex chiral symmetries. Energies can be degenerate, but one level can be "hole like" and the other "particle like."

Another way to derive the band structure is to use the real spatial wavefunction ψ and then the Wallace expectation value and optimization process [P R Wallace Phys Rev. 71, 622, (1947)]:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathrm{A}} \exp(i\mathbf{k} \cdot \mathbf{R}_{\mathrm{A}}) X(\mathbf{r} - \mathbf{R}_{\mathrm{A}}) + \lambda \sum_{\mathrm{B}} \exp(i\mathbf{k} \cdot \mathbf{R}_{\mathrm{B}}) X(\mathbf{r} - \mathbf{R}_{\mathrm{B}}), \tag{5.65}$$

r is the particle coordinate, $\mathbf{R_A}$ and $\mathbf{R_B}$ are position of the two types of atoms (see Fig. 4.7), and **k** is the Bloch wavevector:

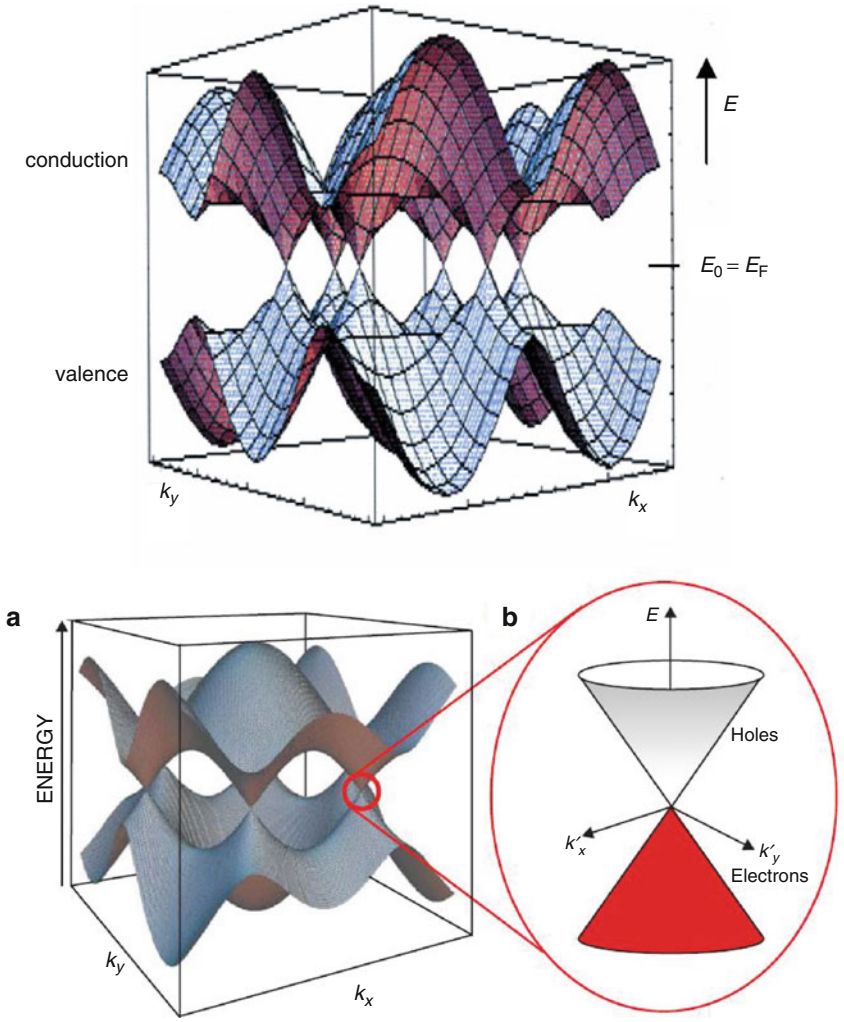


Fig. 5.19 *From Phaedon Avouris.* "Graphene electronic and photonic properties and devices" Nanoletters vol 10, p. 4285, (2010)

$$t_i = \int X^*(\mathbf{r} - \mathbf{R}_A) H X(\mathbf{r} - \mathbf{R}_{B,i}) d\mathbf{r},$$
$$E_{\mathbf{k}} = E_0 \pm \left| \sum_i t_i \exp(-i\mathbf{k} \cdot \boldsymbol{\rho}_i) \right|, \tag{5.66}$$

$\lambda = 1$ or -1

Band structure of graphene plotted in 3D to exhibit the zero gap or Dirac points K,K′.

Analytic tight-binding band structure is:

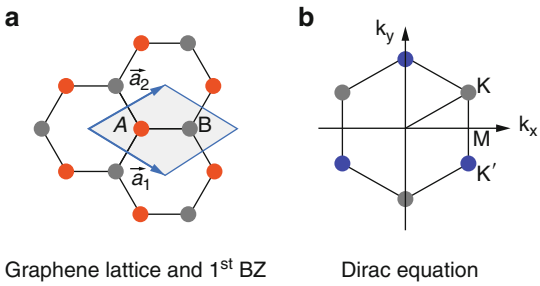


Fig. 5.20 Near the K K's points, shown above, now called Dirac points, the dispersion (energy momentum relation) is linear implying a zero effective mass

$$E(k) = \pm\gamma_0\sqrt{1 + 4\cos\left(\frac{3}{2}k_x a\right)\cos\left(\frac{\sqrt{3}}{2}k_y a\right) + 4\cos^2\left(\frac{\sqrt{3}}{2}k_y a\right)}, \tag{5.67}$$

where γ_0 is the banding energy t (atom to atom overlap) (Fig. 5.20).

The honeycomb structure can be thought of as a triangular lattice with a basis of two atoms per unit cell with 2D lattice vectors $\mathbf{A_0} = (a/2)(3, \sqrt{3})$ and $\mathbf{B_0} = (a/2)(3, -\sqrt{3})$, where $a = 0.142$ nm is the carbon-carbon distance, and $\mathbf{K} = (2\pi/(3a), 2\pi/(3\sqrt{3}\,a))$ and $\mathbf{K}' = (2\pi/(3a), -2\pi/(3\sqrt{3})\,a)$ as the inequivalent corners of the BZ and are called "Dirac points." The Dirac points play a role similar to the role of Γ points in direct bandgap semiconductors.

Relative to the Dirac point, the dispersion is:

$$E_{\pm}(q) = \pm\hbar v_F q + O(q/k)^2 \tag{5.68}$$

The dispersion depends on the Fermi velocity v_F. In tight binding v_F can be expressed in terms of the nearest neighbor hopping integral t so that:

$$\begin{aligned} &\hbar v_F = \frac{3ta}{2} \\ &a = 0.14 \text{ nm}, \quad t = 2.5 \text{ eV}, \quad v_F = 10^8 \text{ cm/s} \end{aligned} \tag{5.69}$$

The linear dispersion is like the dispersion of light or photons with:

$$E = c\hbar q \tag{5.70}$$

where c is the velocity of light. But there are here two sublattices A, B in the structure of G which allows us to write the Hamiltonian on the two sides of the bandgap as a relativistic Dirac-like Hamiltonian:

$$H = v_F \sigma \bullet \hbar q \tag{5.71}$$

where σ is a spinor-like wavefunction, v_F is the Fermi velocity of G, and $\mathbf{q}$ is the wavevector of the electron. Linear dispersion can be thought of as zero effective mass. The spinor nature of the wavefunction is not a consequence of electron spin as in the Dirac equation, but rather from the fact that there are two atoms per unit cell A,

B, and the electron can be thought of as jumping between the components A, B which is then analogous to having a pseudo spin coordinate (in the Dirac equation, the electron can be thought of as hopping into its antiparticle and back again on the time scale short enough that it is allowed by Heisenberg uncertainty principle). Whereas in the latter case, the energy to be overcome is the energy to create a particle-hole pair ($2mc^2$) in empty space, the great new effect here is that the bandgap is zero, so that the particle and hole can be created with zero energy cost. This exciting property implies that one can think of the electron as moving not into empty space but into the "graphene vacuum" of virtual electron-hole pairs and thus oscillating back and forth into the hole component of these virtual pairs created around as the particle moves. This is very much like a photon which moves by alternatively going from electric field (A) to magnetic field excitation (B) . The analogy with relativistic quantum mechanics now also follows by noting that in relativity, the energy of a particle is given by (p-momentum):

$$E = \left[(mc)^2 + (pc)^2\right]^{1/2} \tag{5.72}$$

so that the linear dispersion follows in the limit of zero mass.

Remember from Chap. 4 that in the Dirac equation, *the spin does not disappear in the nonrelativistic limit* but remains a fundamental property of quantum particles in a four-dimensional quantum space time. In other words, even when the mean particle velocity is slow, the short-time (light speed) visits into the antiparticle space which is the origin of the spin, are always allowed by Heisenberg's uncertainty principle. Now one can understand why the zero gap nature of semiconductors, graphene being an example, can be so exciting. The visits into hyperspace are now zero energy visits into the valence holes and back (A, B sublattices) and giving rise to a new pseudo spin quantum number. If we now generate a gap, then this can change the quantum dynamics and properties drastically. This is in principle relatively straightforward to produce by external means (gate, multilayer, doping, etc.). Finally, we note that whereas parabolic electrons have constant density of state in 2D, graphene electrons will have linearly increasing density of states with energy (Fig. 5.21).

The short section here does not do justice to the enormously interesting field. The reader is advised to consult the excellent reviews in the literature and in particular see, for example, the excellent review by Phaedon Avouris "Graphene Electronic and Photonic Properties and Devices" Nanoletters vol 10, p4285, (2010).

5.5.3 Two-Dimensional Metal-Dichalcogenide TMDC: Electronic Structures

Introduction

After Graphene, researchers tried to find new types of graphene like 2D layered materials in order may be to discover some of the exciting photonic like electron band structures. Various groups around the world have found out how to make

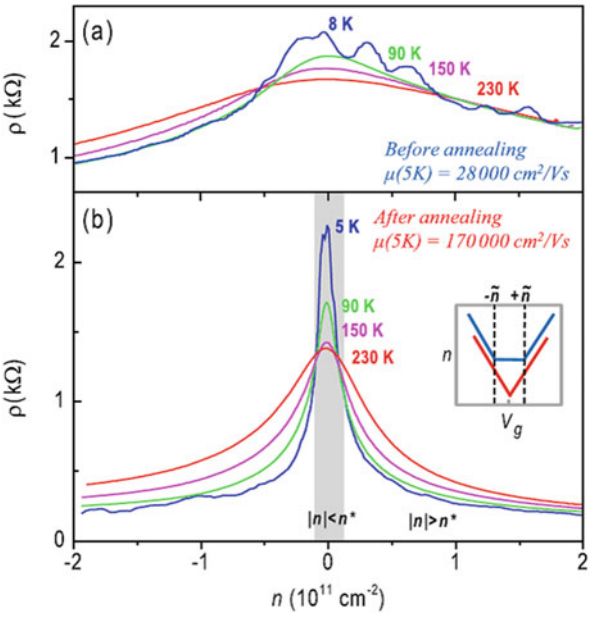


Fig. 5.21 From "Temperature-Dependent Transport in Suspended Graphene" *Bolotin et al.* (2008). "Temperature dependence of resistance of suspended graphene device before and after current annealing. Inset sketch of gate voltage dependence of the carrier density in clean and charge inhomogeneous graphene"

freestanding barrier layers like h-BN, and then finally more recently, they discovered how to exfoliate 2D layers of the transition metal dichalcogenides TMDs [Manish Chewola] . Artificial multilayer fabrication technology is now a very active and popular field of science and technology. Notable discoveries are the ultrathin film transparent high ratio field-effect transistors TFT which exhibit a high degree of plasticity and are promising for tattoo electronics and many other highly commercial applications. Examples of TMD are given in Table 5.1 (Fig. 5.22).

Making nanosheets: description

The TMD sheets have so far been made in several ways described in the work of Chhowalla et al.

1. Scotch tape exfoliation
2. Liquid exfoliation with selected surfactant with right surface energy penetrating the layers and dissolving them
3. Chemical vapor deposition CVD (Figs. 5.23 and 5.24)

5.5.4 Example: Fabrication of Flexible Transistors

Ten atomic thick high-mobility transparent TFTs with ambipolar device characteristics fabricated on both conventional silicon platform and on a flexible substrate have been demonstrated by Saptarshi Das et al. *Nano Letters Vol.14,*

Table 5.1 Electronic character of different layered TMDs[25]

Group	M	X	Properties
4	Ti, Hf, Zr	S, Se, Te	Semiconducting ($E_g = 0.2 \sim 2$ eV). Diamagnetic
5	V, Nb, Ta	S, Se, Te	Narrow band metals ($\rho \sim 10^{-4}$ Ω.cm) or semimetals. Superconducting. Charge density wave (CDW). Paramagnetic, antiferromagnetic, or diamagnetic
6	Mo, W	S, Se, Te	Sulfides and selenides are semiconducting ($E_g \sim 1$ eV). Tellurides are semimetallic ($\rho \sim 10^{-3}$ Ω cm). Diamagnetic
7	Tc, Re	S, Se, Te	Small-gap semiconductors. Diamagnetic
10	Pd, Pt	S, Se, Te	Sulfides and selenides are semiconducting ($E_g = 0.4$ eV) and diamagnetic. Tellurides are metallic and paramagnetic. $PdTe_2$ is superconducting

ρ, in-plane electrical resistivity

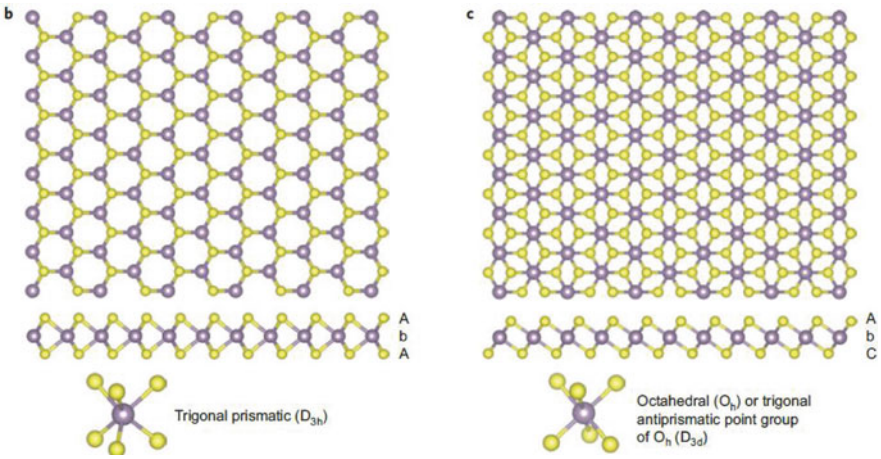


Fig. 5.22 Structure of monolayered TMD. About 40 different layered TMD compounds exist. The transition metals and three chalcogen elements predominantly crystallize in those layered structures. From Chhowalla et al. (2013) [9]

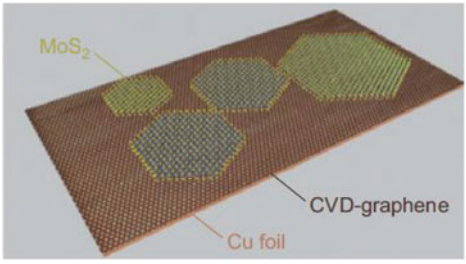


Fig. 5.23 Illustration of the quality achievable for heterostructures MoS_2 on graphene

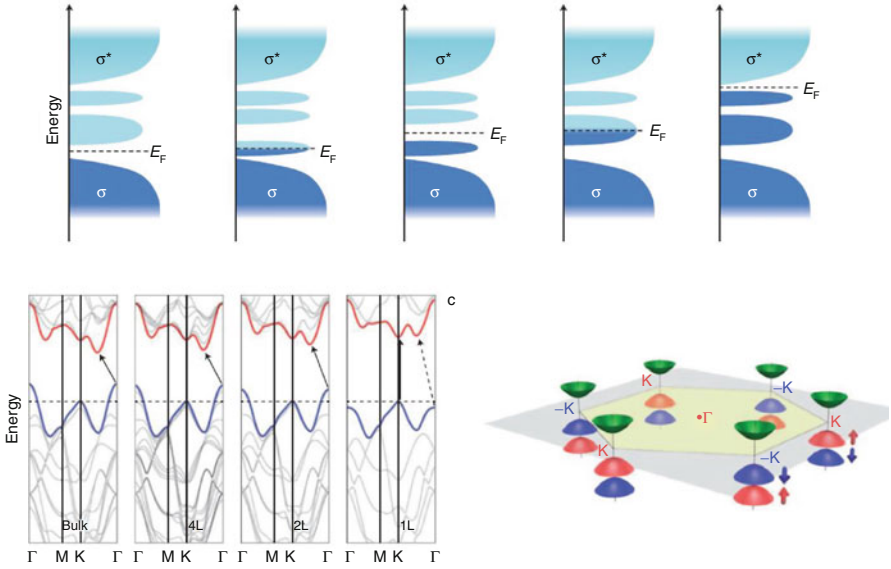


Fig. 5.24 *Qualitative schematic illustration* showing the progressive filling of d-orbitals that are located within the bandgap of bonding and antibonding in groups 4,5,6,7, and 10 TMDs. The D_{3h} and D_{3d} refer to the point groups associated with the trigonal prismatic and the octahedral coordination of the transition metal oxides. (From Chhowalla et al. (2013))

p. 2861, (2014). Monolayer graphene was used as gate electrode and 3–4 atomic layers thick h-BN was used as the gate dielectric, and finally bilayers of WSe_2 were used as the semiconducting channel material for the TFT. The active device stack was found to be 88% transparent over the entire visible spectrum. On to off ratios of 10^7 were observed in all the two-dimensional TFTs.

5.5.5 Summary: Discussion

The atomically thin 2D nanosheets of TMD derived from layered materials exhibit excellent electronic properties, exceptional mechanical flexibility, and partial optical transparency. Beyond the typical semiconducting properties, the various 2D layered materials can also exhibit superconductivity ($NbSe_2$) magnetic ($CrSe_2$) insulating (InN) and thermoelectric (Bi_2Te_3). The 2D sheets can be grown on top of each other to build superlattices with van der Waals bonded layers with exciting new prospects; see the excellent review by Xidong Duan et al.

Xidong Duan et al. review Chem Soc Rev. Vol. 44 p. 8859 (2015): "Two-dimensional transition metal dichalcogenides as atomically thin semiconductors: opportunities and challenges."

Previously in this chapter, we investigated the new "wonder material" called graphene. We described its band structure and explained why this two-dimensional

perfect semimetal with high electron mobility is expected to, and indeed shows, new physical properties which can have serious technical applications. There is by now a massive literature on this subject; indeed, both G and TMD sheets and the interested reader are encouraged to consult some of this material.

5.6 Band Structures in Metals

Although this chapter was primarily devoted to the band structures of semiconductors, which is of great importance in solid-state devices, it would not be complete without a few words on the band structures of metals. Figures 5.25 and 5.26 are examples of electron band structures of two such metals, aluminum and copper.

As mentioned earlier in this chapter, very different behaviors can be seen between the band structures of metals and semiconductors. First of all, there is no forbidden energy region (bandgap) in metals. All the energy range drawn in these diagrams is allowed in metals, which is the most critical difference between metals and semiconductors. Even at a temperature of zero K, a metal has a band which is partially filled with electrons and its Fermi level thus lies within this band. There is no such distinction as valence and conduction bands as encountered in a semiconductor.

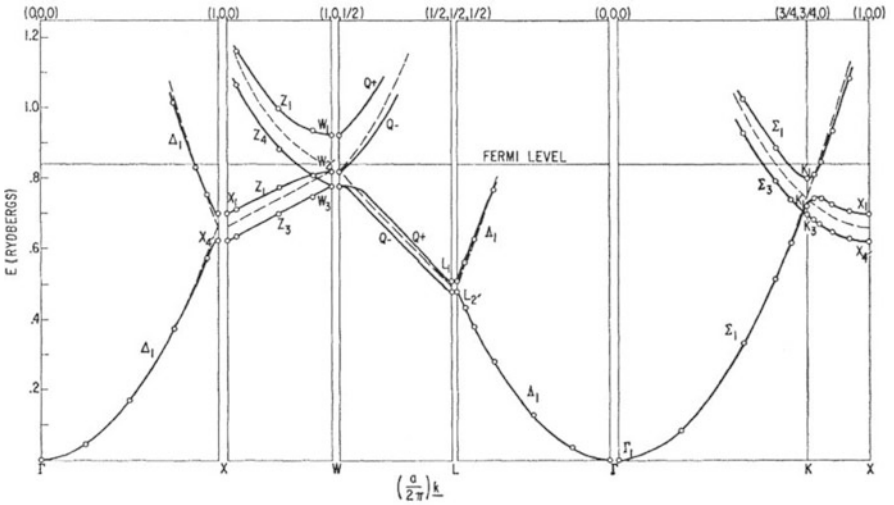


Fig. 5.25 Electron band structure diagram of aluminum. The energy is expressed in units of Rydberg. The dashed lines show the energy bands for a free electron (Reprinted figure with permission from Segall B The Physical Review, vol. 124, p. 1801, Fig. 3, Copyright 1961 by the American Physical Society)

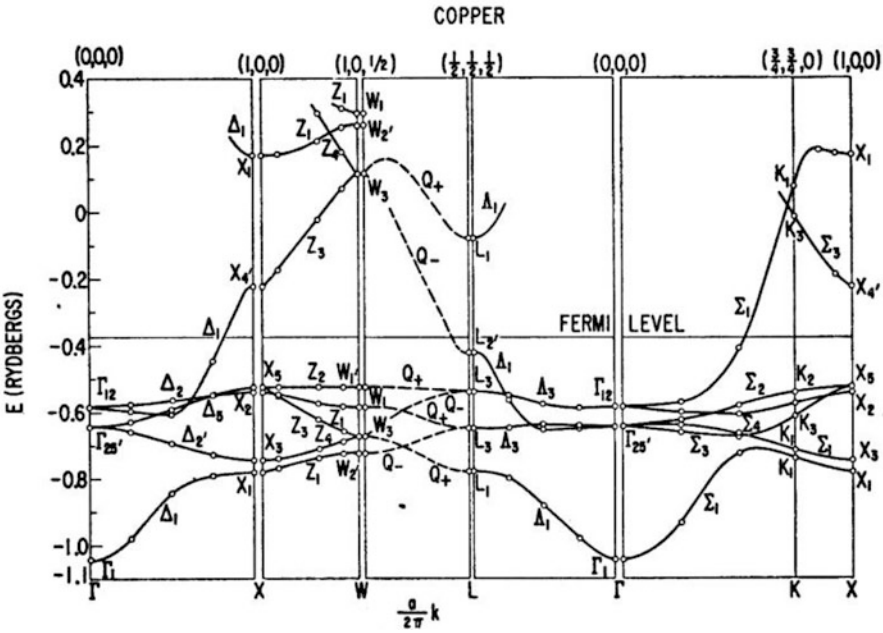


Fig. 5.26 Electron band structure diagram of copper. The energy is expressed in units of Rydberg. There are a few narrow bands located just below the Fermi energy, corresponding to the 4d orbitals in copper (Reprinted figure with permission from Segall B The Physical Review, vol. 125, p. 113, Fig. 5, Copyright 1962 by the American Physical Society)

The band structures in the Γ→X, Γ→K, and Γ→L directions are nearly parabolic and are therefore similar to the free electron case. Electrons in aluminum thus behave almost like free electrons.

The dashed lines in Figs. 5.25 and 5.26 are the *E*–*k* relation for a free electron. One can see that the band structure in aluminum is very close to that of free electrons. The energy spectrum of copper has less resemblance to the free electron *E*–*k* parabolic relation. The major difference between copper and aluminum is the presence of a number of narrow bands below E_F in copper. These narrow bands are attributed to the 4*d*-orbitals of copper atoms. The presence of these *d*-orbital-originated bands is a common feature of most transition metals (such as iron and nickel) and noble metals (such as copper, gold, and silver). These provide a degree of screening effect for electrons. The absence or presence of these *d*-band electrons is also at the origin of the gray and red color appearance of aluminum and copper, respectively. Indeed, when there is a *d*-band, as in copper, not all the photons reaching the metal surface are reflected, but those photons with sufficient energy can be absorbed by the *d*-electrons (see Chapter 0). As a result of this "deficiency" of photons with certain energies, the copper appears red. A similar explanation is valid for the yellow color of gold.

There are always many nearly free electrons in metals that contribute to the electrical and thermal conduction. On the contrary, semiconductors do not have many free electrons when they are intrinsic (i.e., without impurities), and carriers must be provided by a process called doping. The controllability of the doping level in semiconductors is one of the most important reasons why semiconductors are useful in making electronic and optoelectronic devices and will be discussed later in this textbook.

5.7 The Kane Effective Mass Method

In the chapter on band structure, we made the observation, and indeed used this later also throughout the book, that the band dispersion in the majority of semiconductors near $\boldsymbol{k} = 0$ could be approximated as a parabola in $\boldsymbol{k}$ but with an effective mass which is determined by rigorous band structure computation. One finds in practice that the scheme works very well and that the true effective masses can be very different from the free electron masses. From the "exact results" shown in this chapter, one cannot easily understand why the effective mass behaves in the way it does, and one cannot see how it would correlate with the other features of the material, such as its band gap, for example. Also it would be nice to have a scheme which could predict the effective mass, was versatile, and could be applied to confined and multilayer structures as well. Some years ago, Evan O. Kane discovered that it was possible, with rather simple mathematical methods, to shed light on this question. He worked out a scheme with which it is possible to obtain a good approximation to the effective mass near the $\boldsymbol{k} = \mathbf{0}$ points in semiconductors, and a correlation between the effective mass and the band gap.

Kane's method is a brilliant example on how one "piece of information," normally obtained by experiment, can be used to derive another piece of information using the logical structure of a theory. The Kane argument goes as follows.

Consider the full Hamiltonian and Schrödinger equation (SE) of the electron in the periodic potential $V(\vec{r})$ of the lattice. Now assume that the wavefunction is a Bloch wave and must mathematically have the structure:

$$\psi_{n\vec{k}}(\vec{r}) = u_{n\vec{k}}(\vec{r})e^{i\vec{k}.\vec{r}} \tag{5.62}$$

with energy $E_n(\vec{k})$ we know that this must be true, so we substitute it in the SE

$$\left[\frac{p^2}{2m_0} + V(\vec{r})\right]\Psi_{n\vec{k}}(\vec{r}) = E_{n\vec{k}}\Psi_{n\vec{k}}(\vec{r}) \tag{5.63}$$

differentiate, collect the terms, and find

$$\left[\frac{p^2}{2m_0}+\frac{\hbar^2}{m_0}\vec{k}\cdot\vec{p}+V(\vec{r})\right]u_{\vec{nk}}(\vec{r})=\left[E_{\vec{nk}}-\frac{\hbar^2}{2m_0}k^2\right]u_{\vec{nk}}(\vec{r}) \tag{5.64}$$

This is now an equation for the unknown modulating part of the wavefunction $u_{\vec{nk}}(\vec{r})$. The known part has been incorporated and has given an energy shift and a new term in the Hamiltonian. We can rewrite Eq. (5.64) as:

$$\left[H_0+\frac{\hbar^2}{m_0}\vec{k}\cdot\vec{p}\right]u_{\vec{nk}}(\vec{r})=\left[E_{\vec{nk}}-\frac{\hbar^2}{2m_0}k^2\right]u_{\vec{nk}}(\vec{r}) \tag{5.65}$$

and taking the limit $\boldsymbol{k}=\mathbf{0}$, we have the eigenvalue equation:

$$H_0u_n(\vec{r})=[E_n(0)]u_n(\vec{r}) \tag{5.66}$$

for the $\boldsymbol{k}=\mathbf{0}$ envelope. So now one can ask what is the gain in all this, since we are back at the usual Schrödinger equation for the band? There are two observations to be made: the wavefunctions $u_n(\vec{r})$ only have band indices n, there are as many of them as we have energy bands in the semiconductors. In particular, there are valence band functions and conduction band functions. There is a finite energy difference between each band. We could use these functions even though we do not know them, as basis functions, and expand the $\boldsymbol{k}$-dependent term of the Hamiltonian Eq. (5.65) as a perturbation near $\boldsymbol{k}=0$. In this way, we derive the additional $\mathbf{k}$-dependence of the energy and the $\boldsymbol{k}$-dependence of the core wavefunction $u_{\vec{nk}}(\vec{r})$. In this way, we also automatically get an expression for the effective mass in terms of the matrix elements of these basis functions and the energy difference. Thus applying second-order perturbation theory from Chap. 4 to the $\boldsymbol{k}$-dependent term in Eq. (5.65), we have for the energy and wavefunction:

$$E_n(\vec{k})=E_n(0)+\frac{\hbar^2k^2}{2m_0}+\frac{\hbar}{m_0}\vec{k}\cdot\vec{p}_{nn}+\frac{\hbar^2}{m_0^2}\sum_{n'\neq n}\frac{\left|\vec{k}\cdot\vec{p}_{nn'}\right|^2}{E_n(0)-E_{n'}(0)} \tag{5.67}$$

$$u_{\vec{nk}}(\vec{r})=u_{n0}(\vec{r})+\sum_{n'\neq n}\left[\frac{\hbar}{m_0}\frac{\vec{k}\cdot\vec{p}_{n'n}}{E_n(0)-E_{n'}(0)}\right]u_{n'0}(\vec{r}) \tag{5.68}$$

Remember that the complete wavefunction is of the form of Eq. (5.62). Now, we see that progress has indeed been made. When we look at Eq. (5.67), then indeed Eq. (5.67) with $E_{\vec{nk}}\sim E_n(0)+\frac{\hbar^2k^2}{2m^*}$ and the observation that by symmetry $\vec{p}_{nn}=0$ tells us that the effective mass near $\boldsymbol{k}=0$ is given by (i, j denote the x,y,z components)):

$$\left(\frac{1}{m^*}\right)_{ij} = \frac{1}{m_0}\delta_{ij} + \sum_{n' \neq n} \frac{p^i_{nn'} p^j_{n'n} + p^j_{nn'} p^i_{n'n}}{E_n(0) - E_{n'}(0)} \tag{5.69}$$

The inverse of the effective mass is a sum of the free electron mass and term which depends on the momentum matrix elements of the $\boldsymbol{k} = \boldsymbol{0}$ envelope but is also dependent on the energy difference between the bands. If for simplicity, we now consider just two bands, namely, the conduction and valence band, then to a good approximation, we see that the inverse effective mass scales as the inverse of the energy gap of the semiconductor. In other words, we have the result that semiconductors with smaller band gaps should have the lower effective mass. If this statement turns out to be generally true, then it helps to establish an important principle and correlation between band gap and effective mass (see the data in Appendix A4).

At this stage, the most important unknown is the momentum matrix element. The next step is therefore to establish empirically that the momentum matrix elements are not strongly dependent on the band gap and to include the other bands when necessary. Here one also uses the fact that the exact wavefunctions are *s*-like near the bottom of the conduction band and *p*-like near the top of the valence band. This is known from first principle and tight-binding band structure theories. This interplay between theory and experiment then gives us useful and simple empirical rules and numbers for the above matrix element in Eq. (5.69). For example, one finds that the Kane parameter $E_{\mathrm{P}} = \frac{2m_0}{\hbar^2}P^2$ where $P = \frac{\hbar}{m_0}p^z_{cv}$ is roughly 20–25 eV for most semiconductors of interest, where the subscripts c and v denote the conduction and valence band, respectively. The Kane method of expanding around the $\boldsymbol{k} = \boldsymbol{0}$ envelope states can be extended to treat also the spin-orbit interaction. The spin-orbit coupling is of the form:

$$V_{so} = \frac{\hbar}{4m_0^2c^2}\,\vec{\sigma}\cdot\left(\nabla V\left(\vec{r}\right)\times\vec{p}\right) \tag{5.70}$$

where σ is the electron spin operator and $V\left(\vec{r}\right)$ is the total potential experienced by the electrons. The spin-orbit interaction is a small but non-negligible effect in semiconductors. It is ideally treated using the Kane model because the energy shifts up to second order in perturbation theory and involved the same type of matrix elements of the momentum as before. Indeed, one can say that the Kane method provides a very natural way to treat the spin-orbit interaction. The method can be extended to also treat confined systems. The results can at the end be expressed as functions of E_{g} and P. The first of which, E_{g}, is known, and the second of which, P, can be estimated to good accuracy.

Kane theory tells us that the effective mass is related to the structure of the envelope momentum matrix elements. These as it happens do not change all that much from one system to another system. The band gap which also enters the formula, however, changes quite a lot. If for some reason, such as strain or

confinement, the band gap changes, even locally, then we can expect the effective mass also to change locally. The changes in P or wavefunction shapes are of lower order than the band gap changes, and this is why the Kane method is so useful. The Kane method is therefore a very practical way of handling strain effects in semiconductor interfaces. This happens when there is lattice mismatch forcing the top grown lattice to adopt the lattice parameters of the substrate. The mismatch can force the top layer bonds to be stretched or compressed. Compression or dilation affects both Kane parameters E_g and P locally. But the gap is more sensitive than P to first order. In quantum dots strain, one also has strain which can vary locally and give rise to local effective mass. The reader is referred to the book by L Chuang for a detailed treatment of the Kane model and its applications.

5.7.1 The Effect of the Spin-Orbit Coupling

Let us now consider the effect of the spin-orbit coupling explicitly. Let us go back to Eq. (5.63) and include the spin-orbit interaction:

$$\left[\frac{p^2}{2m_0}+V(\vec{r})+\frac{\hbar}{4m_0^2c^2}\left[\vec{\nabla}\ V\times\ \vec{p}\right]\cdot\ \vec{\sigma}\right]\Psi_{n\vec{k}}(\vec{r})=E_{n\vec{k}}\Psi_{n\vec{k}}(\vec{r}) \tag{5.71}$$

Substituting the Bloch function then gives:

$$\left\{\frac{p^2}{2m_0}+\frac{\hbar}{m_0}\ \vec{k}\cdot\vec{p}\ +V(\vec{r})+\frac{\hbar}{4m_0^2c^2}\left[\vec{\nabla}\ V\times\ \vec{p}\right]\cdot\ \vec{\sigma}\right.$$
$$\left.+\frac{\hbar^2}{4m_0^2c^2}\left[\nabla V\times\ \vec{k}\ \right]\cdot\ \vec{\sigma}\ \right]u_{n\vec{k}}(\vec{r})\Bigg\}=\left[E_{n\vec{k}}-\frac{\hbar^2}{2m_0}k^2\right]u_{n\vec{k}}(\vec{r}) \tag{5.72}$$

Following the book by L. Chuang (see references list), we will define:

$$E^{'}=E-\frac{\hbar^2k^2}{2m_0} \tag{5.73}$$

The last term in Eq. (5.72) depends on the Bloch wavevector $\boldsymbol{k}$ and is much smaller than the term which involves the momentum operator. The reason is that the momentum around the nucleus is much larger than the band momentum $\boldsymbol{k}$ so that we can neglect the last term to obtain

$$\left\{\frac{p^2}{2m_0}+\frac{\hbar^2}{m_0}\ \vec{k}\ .\ \vec{p}\ +V(\vec{r})+\frac{\hbar}{4m_0^2c^2}\left[\vec{\nabla}\ V\times\ \vec{p}\right]\cdot\ \vec{\sigma}\ \right]\Big\}u_{n\vec{k}}(\vec{r})\Big\}=\left[E^{'}\right]u_{n\vec{k}}(\vec{r}) \tag{5.74}$$

In order to solve this equation in the $\boldsymbol{k}\cdot\boldsymbol{p}$ approach, we assume as before that the wavefunction can be written as a superposition of the $\boldsymbol{k}=0$ subbands:

$$u_{n\vec{k}} = \sum_{n'} a_{n'\vec{k}} u_{n'0}(r) \tag{5.75}$$

The $u_{n'0}(\vec{r})$ are chosen to roughly correspond to what one knows about the system from first principle band structure techniques, namely, that for the eigenstate near the conduction band edge, the wavefunctions have S symmetry. Those near the top of the valence band have p-symmetry so that

$$\text{Conduction bands} |S\uparrow\rangle, |S\downarrow\rangle$$

$$\text{Valence bands} |X\uparrow\rangle, |Y\uparrow\rangle, |Z\uparrow\rangle, |X\downarrow\rangle, |Y\downarrow\rangle, |Z\downarrow\rangle$$

Also these wavefunctions satisfy the conditions $H_0|S\uparrow\rangle = E_S|S\uparrow\rangle$ without a magnetic field, and the two spin states have the same energy, and $H_0|X\uparrow\rangle = E_p|X\uparrow\rangle$, $H_0|Y\uparrow\rangle = E_p|Y\uparrow\rangle$, and $H_0|Z\uparrow\rangle = E_p|Z\uparrow\rangle$. For practical purposes, it is convenient to choose the basis states for spin and angular momentum raising and lowering operators:

$$|S\uparrow\rangle, \left|\frac{1}{\sqrt{2}}(X - iY)\uparrow\right\rangle, |Z\downarrow\rangle, \left|-\frac{1}{\sqrt{2}}(X + iY)\uparrow\right\rangle \tag{5.76}$$

$$|S\downarrow\rangle, \left|-\frac{1}{\sqrt{2}}(X + iY)\downarrow\right\rangle, |Z\uparrow\rangle, \left|\frac{1}{\sqrt{2}}(X - iY)\downarrow\right\rangle$$

The valence band basis states can be selected from the eigenstates of angular momentum $\boldsymbol{L}$ in Chap.4. So a p-state corresponds to $l = 1$ and we have:

$$Y_{1,\pm 1} = \mp\frac{1}{\sqrt{2}}|X \pm iY\rangle \quad \text{and} \quad Y_{10} = |Z\rangle \tag{5.77}$$

Now we can generate the matrix representation of the Hamiltonian Eq. (5.74) using this basis set to find an 8×8 matrix which as a result of spin degeneracy reduces to a 4×4 matrix:

$$\begin{pmatrix} E_s & 0 & kP & 0 \\ 0 & E_p - \frac{\Delta}{3} & \frac{\sqrt{2}}{3}\Delta & 0 \\ kP & \frac{\sqrt{2}\Delta}{3} & E_p & 0 \\ 0 & 0 & 0 & E_p + \frac{\Delta}{3} \end{pmatrix} \tag{5.78}$$

where the Kane parameter is defined as:

$$\begin{aligned} P &= -i\frac{\hbar}{m_0}\langle S|p_z|Z\rangle \\ \Delta &= \frac{3\hbar i}{4m_0^2c^2}\left\{\langle X|\frac{\partial V}{\partial x}p_y - \frac{\partial V}{\partial y}p_x|Y\rangle\right\} \end{aligned} \tag{5.79}$$

Let us measure the eigenvalues of this system such that the conduction band E_s is at E_g and the top of the valence band is at 0. The solutions are the roots of the equation:

$$E'\left(E' - E_g\right)\left(E' + \Delta\right) - k^2P^2\left(E' + \frac{2}{3}\Delta\right) = 0 \tag{5.80}$$

which we can solve analytically if we expand to first order in k^2. The result, for the energy and effective mass in the conduction band, is:

$$E_c(k) = E_g + \frac{\hbar^2k^2}{2m_0} + \frac{k^2P^2}{3}\frac{\left(3E_g + 2\Delta\right)}{E_g\left(E_g + \Delta\right)} \tag{5.81}$$

$$\frac{1}{m_c^*} = \frac{1}{m_0} + \frac{2P^2}{3\hbar^2}\frac{\left(3E_g + 2\Delta\right)}{E_g\left(E_g + \Delta\right)} \tag{5.82}$$

For the heavy-hole valence states, we have:

$$E_{hh}(k) = \frac{\hbar^2k^2}{2m_0} \tag{5.83}$$

$$\frac{1}{m_{hh}^*} = \frac{1}{m_0} \tag{5.84}$$

For the light hole:

$$E_{lh}(k) = \frac{\hbar^2k^2}{2m_0} - \frac{2k^2P^2}{3}\frac{1}{E_g} \tag{5.85}$$

$$\frac{1}{m_{lh}^*} = \frac{1}{m_0} - \frac{4P^2}{3\hbar^2}\frac{1}{E_g} \tag{5.86}$$

For the spin orbit shifted band:

$$E_{so}(k) = -\Delta + \frac{\hbar^2k^2}{2m_0} - \frac{k^2P^2}{3}\frac{1}{\left(E_g + \Delta\right)} \tag{5.87}$$

$$\frac{1}{m_{so}^*} = \frac{1}{m_0} - \frac{2P^2}{3\hbar^2}\frac{1}{\left(E_g + \Delta\right)} \tag{5.88}$$

To zero order in k^2, the conduction band wavefunctions are unchanged at $|S\uparrow\rangle$ and $|S\downarrow\rangle$. The valence light-hole states have a spin-orbit shift even to this order. So we have for the heavy hole the two states:

$$\left|-\frac{1}{\sqrt{2}}(X+iY)\uparrow\right\rangle. = \left|\frac{3}{2},\frac{3}{2}\right\rangle. \tag{5.89}$$

$$\left|\frac{1}{\sqrt{2}}(X-iY)\downarrow\right\rangle. = \left|\frac{3}{2},-\frac{3}{2}\right\rangle. \tag{5.90}$$

and the light holes:

$$\left|\frac{1}{\sqrt{6}}(X-iY)\uparrow\right\rangle. + \sqrt{\frac{2}{3}}|Z\downarrow\rangle. = \left|\frac{3}{2},-\frac{1}{2}\right\rangle. \tag{5.91}$$

$$-\left|\frac{1}{\sqrt{6}}(X+iY)\downarrow\right\rangle. + \sqrt{\frac{2}{3}}|Z\uparrow\rangle. = \left|\frac{3}{2},\frac{1}{2}\right\rangle. \tag{5.92}$$

The $\boldsymbol{k}\cdot\boldsymbol{p}$ perturbation result can be generated as in Eqs. (5.67) and (5.68). The exact result can be obtained by solving for the eigenvalues $E_n^{'} = E_n - \dfrac{\hbar^2 k^2}{2m_0}$, as done above, then substituting back to solve the linear equations. The matrix is effectively 3×3 so that:

$$\begin{pmatrix} E_g - E_n' & 0 & kP \\ 0 & -\dfrac{2\Delta}{3} - E_n^{'} & \dfrac{\sqrt{2}}{3}\Delta \\ kP & \dfrac{\sqrt{2}\Delta}{3} & -\dfrac{\Delta}{3} - E_n^{'} \end{pmatrix} \begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix} = 0 \tag{5.93}$$

with

$$\left\{|a_n|^2 + |b_n|^2 + |c_n|^2\right\}^{1/2} = 1 \tag{5.94}$$

Note that the present approach neglects the remote band effects, and it does not reproduce the correct heavy-hole mass. In order to do that, one has to go further and consider the Luttinger-Kohn model which is similar in spirit but takes into account remote energy bands and will not be considered here.

The k-dependence of the wavefunctions has not been studied here. They can be of great interest in doped magnetic semiconductors and magnetic metals where in the presence of a finite spin polarization, they can give rise to the so-called anomalous Hall effect and spin Hall effect (see Jungwirth et al. 2006). But again here, one would go further and use the Luttinger-Kohn model which includes the remote band effects. The magnetism can then be treated as an effective uniform self-consistent Curie field which acts on the spin system (see Jungwirth et al. reference).

5.7.2 Summary

In this chapter, using simple quantum mechanical concepts and methods, we have described the energy states of electrons in a periodic potential. We have modeled the crystal using the Kronig-Penney model. Nearly free electron and the tight-binding approximations were briefly introduced. We familiarized the reader with the notion of band structure, band gap, Bloch wavefunction, effective mass, Fermi energy, and Fermi-Dirac distribution and holes. The band structures for the common semiconductors, including Si, Ge, and GaAs, have been illustrated after first describing the conventionally used high-symmetry points and orientations. The main features in these band structures have been outlined. The band structures of a few metals, including aluminum and copper, have also been presented, and the main features were described and compared to those of semiconductors. We have shown how one can evaluate the Bloch wavefunctions and effective masses of semiconductors near $k = 0$ using a scheme called the $k{\cdot}p$ method. The method is simple and very powerful. It was applied to derive the effective mass including the spin-orbit coupling.

Problems

1. *Equations of motion of an electron in the presence of an electric field.*

$$\text{Assuming a dispersion relation : } \varepsilon = \varepsilon_C + \frac{\hbar^2}{ma^2}[1 - \cos(ka)]$$

   (a) Calculate the velocity of the electron at $k = \pi/a$.
   (b) If the electric field E is applied in the $-x$ direction, derive the time dependence of k for an electron initially at $k = \pi/a$ and position $x = 0$.
   (c) Derive the time dependence of the electron velocity, $v(t)$, and the time dependence of the electron position, $x(t)$.
   (d) For $a = 5$ nm, $E = 104$ V·cm^{-1}, and $m = 0.2\ m_0$, what are the maximum and minimum values of x that the electron will reach?
   (e) What is the period of the oscillation?
   (f) For the parameter of part (e), derive an expression for the effective mass as a function of k. Sketch the function.

2. *The period of the Bloch oscillations.*

   Consider an electron that is subjected to an electric field. The electric field exerts a force $F = -qE$ on the electron. Assume that the electron is initially not in motion, i.e., $k = 0$. Upon application of the electric field, the k value of the electron increases from 0 to π/a. At this value of k, Bragg reflection occurs, and the electron assumes a k value of $-\pi/a$. Then, the electron is again accelerated to $k = \pi/a$. At this point, the electron again undergoes Bragg reflection, and the cycle starts from the beginning. The process described above is called the Bloch oscillation of the electron in an energy band of the solid-state crystal.

(a) Show that the period of the Bloch oscillation is given by $\tau = \frac{2\pi\hbar}{qEa}$, where a is the periodicity of a one-dimensional atomic chain.

(b) Calculate the period of the Bloch oscillations for $a = 4$ Å and $E = 1250$ V·cm^{-1}. Compare the period of the Bloch oscillations with a typical inelastic scattering times. What conclusions do you draw from the comparison? Are the Bragg reflections important scattering events for the movement of electrons in a crystal? Typical inelastic scattering times are 10^{-11} s for low fields and 10^{-13} s for high fields.

3. *Idealized electron dynamics.*

A single electron is placed at $k = 0$ in an otherwise empty band of a bcc solid. The energy versus k relation of the band is given by:

$$\varepsilon\left(\vec{k}\right) = -\alpha - 8\gamma \cos\left(\frac{k_x a}{2}\right).$$

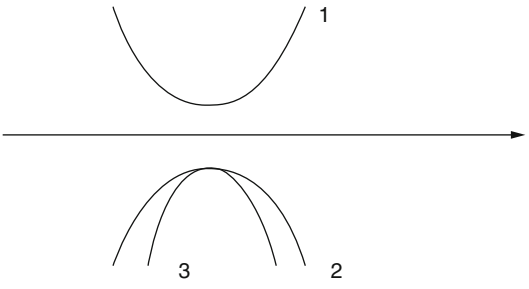
At $t = 0$, a uniform electric field E is applied in the x-axis direction. Describe the motion of the electron in k-space. Use a reduced-zone picture. Discuss the motion of the electron in real space assuming that the particle starts its journey at the origin at $t = 0$. Using the reduced-zone picture, describe the movement of the electron in k-space. Discuss the motion of the electron in real space assuming that the particle starts its movement at the origin at $t = 0$.

4. *Effective mass.*

For some materials, the band structure of the conduction band around $k = 0$ can be represented by $\varepsilon\left(\vec{k}\right) = \frac{\hbar^2}{2m} A\left(k_x^2 - \frac{a^2}{2\pi^2}k_x^4\right)$.

What is the effective mass of a free electron under these conditions?

On the figure, name the different bands and point out which one of the two in the lower band has the higher effective mass.



5. *Calculate the coordinates of the high-symmetry point U in Fig. 5.15.*

6. *Origin of electronic bands in materials.*

Explain how electronic energy bands arise in materials.

The periodic potential in a one-dimensional lattice of spacing a can be approximated by a square wave which has the value $U_0 = -2$ eV at each

atom and which changes to zero at a distance of $0.1a$ on either side of each atom. Describe how you would estimate the width of the first energy gap in the electron energy spectrum.

7. *Position of the Fermi level in intrinsic semiconductors.*
   Assume that the density of states is the same in the conduction band (N_C) and in the valence band (N_V). Then, the probability p that a state is filled at the conduction band edge (E_C) is equal to the probability p that a state is empty in the valence band edge (E_V). Where is the Fermi level located?
8. *Plot of the Fermi distribution function at two different temperatures.*
   Calculate the Fermi function at 6.5 eV if $E_F = 6.25$ eV and $T = 300$ K. Repeat for $T = 950$ K assuming that the Fermi energy does not change. Plot the energy dependence of the electron distribution function at $T = 300$ K and at $T = 950$ K assuming $E_F = 6.25$ eV.
9. *Numerical evaluation of the effective densities of states of Ge, Si, and GaAs.*
   Calculate the effective densities of states in the conduction and valence bands of germanium, silicon and gallium arsenide at 300 K. Note in analogy to Eq. (5.55) we have $N_V = 2\left(\frac{2\pi k_b T m_h}{h^2}\right)^{3/2}$
10. *Density of states of a piece of Si.*
    Calculate the number of states per unit energy in a 100 by 100 by 10 nm piece of silicon ($m^* = 1.08\ m_0$) 100 meV above the conduction band edge. Write the results in units of eV^{-1}.
11. *Number of conduction electrons in a Fermi sphere of known radius.*
    In a simple cubic quasi-free electron metal, the spherical Fermi surface just touches the first Brillouin zone. Calculate the number of conduction electrons per atom in this metal as a function of the Fermi-Dirac integral. Consider the energy at the bottom of the conduction band to be $E_C = 0$ eV.

References

Chhowalla M et al (2013) Electronics and optolectronics of two –dimensional transition metal dichalcogenides. Nat Chem 5:263
Das S et al (2014) Nano Lett 14:2861
Duan X et al (2015) Review Chem Soc Rev 44:8859
Chelikowsky JR, Cohen ML (1976) Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors. Phys Rev B 14:556–582
Jungwirth T, Sinova J, Masek J, Kucera J, Macdonald AH (2006) Theory of ferromagnetic (III,Mn) V semiconductors. Rev Mod Phys 78:809

Further Reading

Andre G (2011) Nobel lecture: random walk to graphene. Rev Mod Phys 83
Avouris P (2010) Graphene electronic and photonic properties and devices. Nano 10:4285
Bolotin K, Sikes K, Stormer HL, Kim P (2008) PRL 101:096802

Bonaccorso F, Sun Z, Hasan T, Ferrari A (2010) Graphene photonics and optoelectronics. Nat Photonics 4:611

Bonnacorso F et al (2015) Graphene related two-dimensional crystal and hybrid systems for energy conversion and storage. Science 347(6217):1246501–1246501.

Falkovsky L (2008) Optical properties of graphene and IV-VI semiconductors. The electronic properties of graphene. Physics Uspekhi 51:887

Jungwirth T, Sinova J, Masek J, Kucera J, Macdonald AH Theory of ferromagnetic (III,Mn)V semiconductors. Rev Mod Phys 78:809 2006

Neto CAH (2009) Rev Mod Phys 81. Jan–March 2009

Peres NM, Guinea F, Castro AH (2006) Electronic properties of disordered two- dimensional carbon. PRB 73:125411

Wallace PR (1947) Phys Rev 71:622