Chapter 6
Electron and Phonon Transport

In the preceding chapter on solid properties, we relied on the Drude–Sommerfeld model, which assumes that electrons are completely free and the Fermi surface is spherical and isotropic in all directions of the wavevector. While the concepts of electronic band structures and phonon dispersion in real solids were often mentioned, we have deliberately avoided any details. It is hoped that the free-electron model will help readers gain an intuitive picture of electrons without a deep knowledge of solid-state physics. Note that the free-electron model described in Sect. 5.1.3 is applicable only for metals, usually good conductors, and cannot be applied to semiconductors. The Sommerfeld theory, albeit successful in quantitatively describing electronic transport for certain metals, does not touch on the fundamental mechanisms of electron scattering and the shape of the Fermi surface. The free-electron model also fails to explain certain phenomena including thermoelectricity. The Hall effect and magnetoresistance, to be discussed in the following section, provide further evidence of the inadequacy of the free-electron model.

This chapter introduces electronic band theory after a brief discussion of electronic structures in atoms, binding in crystals, and crystal lattices. The phonon dispersion relations are presented subsequently and explained in terms of different branches of acoustic and optical phonons. Subsequently, the electron and phonon scattering mechanisms are outlined. The next section addresses electronic emission and tunneling phenomena, including photoelectric effect, thermionic emission, field emission, as well as electron tunneling through a potential. A significant portion of this chapter is then devoted to semiconductor materials and devices, with an emphasis on optoelectronic applications such as solar cells, thermophotovoltaic systems, light-emitting diodes (LEDs), and semiconductor lasers including quantum well lasers.

Z. M. Zhang, *Nano/Microscale Heat Transfer*, Mechanical Engineering Series,
https://doi.org/10.1007/978-3-030-45039-7_6

6.1 The Hall Effect

When a conductor carrying electric current is placed in a magnetic field perpendicular to the current flow, there is a Lorentz force acting on the conductor according to $\mathbf{F} = \sum q\mathbf{u}_\mathrm{d} \times \mathbf{B} = l\mathbf{I} \times \mathbf{B}$, where q is the charge of each carrier, $\mathbf{u}_\mathrm{d}$ is the drift velocity of the carrier, $\mathbf{B}$ is the magnetic induction, $\mathbf{I}$ is the current in the conductor, and l is the length of the conductor. This principle was used in the electromagnetic motor invented by Michael Faraday in 1821. Because electric current is always defined in the direction of the applied electric field $\mathbf{E} = -\nabla V$, the force acting on the conductor is independent of the nature of the carriers (electrons or holes). Microscopically, however, there is a subtle difference that can be distinguished by the experiment first performed by Edwin Hall in 1878 when he was a graduate student at Johns Hopkins University. As shown in Fig. 6.1, an electric current passes through a metal foil in the x-direction, while the electrons are drifted opposite to the x-direction. When a uniform magnetic field $\mathbf{B}$ is applied in the z-direction, the electrons are subjected to a force toward the negative y-direction. Gradually, an electric field is built up across the foil as manifested by a nonzero voltage V_H, which is called the Hall voltage. The electric potential in the y-direction eventually balances the magnetic force such that the electrons drift in the x-direction only. This effect is called the *Hall effect*. By setting the y-component of the Lorentz force $\mathbf{F} = q(\mathbf{E} + \mathbf{u}_\mathrm{d} \times \mathbf{B})$ to zero, one obtains

$$V_\mathrm{H} = \frac{IB}{nqd} \tag{6.1}$$

where n is the number density of the carrier and d is the thickness of the conductor [1, 2]. The *Hall coefficient* is defined as follows:

$$\eta_\mathrm{H} = \frac{V_\mathrm{H} d}{IB} = \frac{1}{nq} \tag{6.2}$$

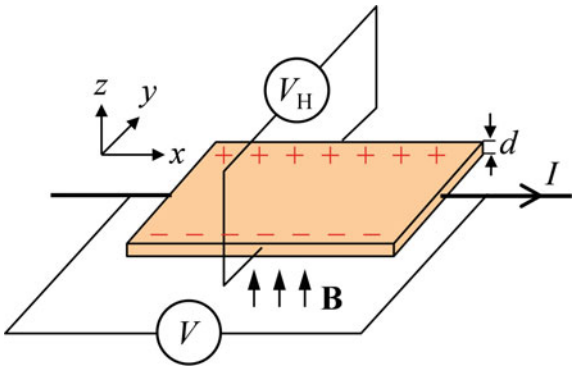

Fig. 6.1 Illustration of the Hall effect experiment

The *Hall resistance* can be defined as $R_H = V_H/I = B\eta_H/d$, and its inverse is called the *Hall conductance*. Similarly, the *Hall resistivity* is given by $r_H = B\eta_H = E_y/J_x$, where E_y is the electric field in the y-direction and J_x is the current density. For metals, $q = -e$ and $n = n_e$, the number density of free electrons, and one would expect a negative Hall resistance.

Example 6.1 Find the Hall coefficient and the Hall voltage for a copper foil of 2×2 cm^2 area, with a thickness of 10 μm. Given the electrical current $I = 0.5$ A and the magnetic induction $B = 1.0$ T (tesla) $= 1.0$ Wb/m^2, what is the voltage drop along the current flow direction?

Solution Based on the previous chapter, the number density of electrons in copper is $n_e = 8.45 \times 10^{28}$ m^{-3}. From Eq. (6.2), we obtain $\eta_H = -7.4 \times 10^{-5}$ cm^3/C, and from Eq. (6.1) we find $V_H = -3.7$ μV, which is a very small voltage but can be measured accurately. Using the resistivity of copper $r_e = 1.7 \times 10^{-8}$ Ω m, we see that $V = 850$ μV, which is much larger than the Hall voltage. The Hall coefficient is much larger for semiconductors because of their usually much lower carrier densities.

Before the discovery of the Hall effect, many people, including James Clerk Maxwell, believed that the force acted only on the conductor but not on the current carriers [3]. Measurement of the Hall coefficient allows the determination of the sign of the charge carriers as well as the carrier concentration. This is important especially for semiconductor materials. The Hall coefficient is positive for p-type semiconductors, but negative for n-type semiconductors. In reality, the Hall coefficient depends also on the applied magnetic field although such a dependence cannot be predicted by the Drude free-electron model. For some common metals like Al, Be, Cd, In, W, and Zn, the Hall coefficient can even become positive. Therefore, the Hall effect cannot be fully accounted by the free-electron model. It is necessary to understand the electronic band structures.

Magnetoresistance is the change in resistance of a material under an applied magnetic field. The magnetoresistance may be transverse, when the applied magnetic field is perpendicular to the current flow, and longitudinal, when the applied magnetic field is parallel to the current flow. In the free-electron theory, resistance is expected to be independent of the strength of the applied transverse magnetic field. In reality, most materials exhibit transverse magnetoresistance that depends on the magnetic field strength. In the late 1980s, researchers observed a giant magnetoresistive (GMR) effect, also called giant magnetoresistance, with extremely thin films of ferromagnetic and metallic layers. The GMR effect has been applied to read heads for magnetic hard disk drives [4] and the discovery is recognized by the 2007 Nobel Prize in Physics.

Klaus von Klitzing and coworkers in 1980 measured the Hall voltage of a 2D electron gas using a metal-oxide-semiconductor field-effect transistor (MOSFET), at very low temperatures ($T \approx 1.5$ K) with a high magnetic field ($B > 15$ T), at the Grenoble High Magnetic Field Laboratory in France [5]. They found that the Hall conductance is quantized and increases with the applied magnetic field by steps in

a staircase sequence. The Hall conductance is a multiple of a fundamental constant, $1/R_K$, where

$$R_K = h/e^2 = 25812.807449 \pm 0.000086\,\Omega \tag{6.3}$$

is called the von Klitzing constant. Note that e^2/h is proportional to the fine-structure constant, which is related to the strength of light–matter interaction in quantum electrodynamics. For this work, von Klitzing was awarded the Nobel Prize in Physics in 1985. The remarkable precision and gauge invariance of quantized conductance allowed the definition of a resistance standard used worldwide since 1990 [6]. As discussed in Chap. 5, quantized conductance has also been observed between nanocontacts and nanostructures with an increment of $2/R_K$. The discovery of the fractional quantum Hall effect in 1982, on the other hand, rendered three physicists (Robert Laughlin, Horst Störmer, and Daniel Tsui) the 1998 Nobel Prize. This has led to a breakthrough in our fundamental understanding of the physical world. For example, in a 2D system, electrons may switch between Fermi–Dirac statistics and Bose–Einstein statistics, continuously [7]. Strohm et al. [8] reported phonon Hall effect by applying a magnetic field perpendicular to the heat flow in a paramagnetic dielectric material at low temperatures. A transverse temperature difference was measured, which reverses sign when the magnetic field is inversed. Researchers have also observed magnon Hall effect and photonic spin Hall effect.

6.2 General Classifications of Solids

There are several ways to classify solids. Based on their electrical conductivities, solids may be classified as insulators, semiconductors, or conductors. They may exist in different forms, such as amorphous or crystalline phases, depending on how the atoms in the solids are arranged. A general introduction is given in this section considering chemical bonds and electrical properties of solids. Let us first take a look at the electron configuration in atoms because it is directly related to physical and chemical properties.

6.2.1 Electrons in Atoms

The periodic table of elements is arranged sequentially according to atomic number, which is determined by the number of protons inside the nucleus and equal to the number of electrons orbiting the nucleus, since an atom itself is charge neutral. The electrons occupy different quantum states, which are fully described by the Schrödinger wave equation as discussed in Chap. 3. By solving the wave equation in spherical coordinates [9, 10], the number of quantum states can be determined and identified using indices n, l, and m. The first or principal quantum number $n =$

1, 2, 3, 4, … corresponds to different shells, denoted as *K*, *L*, *M*, *N*, *O*,… In each shell, there are *n* subshells defined by the orbital number $l = 0, 1, 2, \ldots, (n-1)$. The corresponding symbols are *s*, *p*, *d*, *f*, *g*, *h*, and so forth. For each *l*, the magnetic quantum number $m = 0, \pm 1, \pm 2, \ldots, \pm l$, which gives a total of $2l + 1$ orbits (or *orbitals* since electrons do not follow an exact path that can be described by classical mechanics) for each subshell. Hence, there are a total of n^2 orbitals in the *n*th shell. When spin degeneracy is considered, the total allowable quantum states are $2n^2$ in the *n*th shell. In other words, there are 2, 8, 18, and 32 quantum states in the first (*K*), second (*L*), third (*M*), and fourth (*N*) shells, respectively. On the other hand, there are $2(2l + 1)$ quantum states in the *l*th ($l < n$) subshell. For example, the *s*, *p*, *d*, or *f* subshell contains correspondingly 2, 6, 10, or 14 quantum states. According to Pauli's exclusion principle, each quantum state can have no more than one electron, i.e., at most only two electrons (one with $+\frac{1}{2}$ and the other with $-\frac{1}{2}$ spin) can share the same orbital.

According to the Aufbau principle, electrons will fill the lowest energy states first. The electron configuration of an atom is expressed by the numbers in each subshell. For example, we can write for aluminum and calcium, respectively,

$$^{13}\text{Al: } 1s^2 2s^2 2p^6 3s^2 3p^1 \text{ and } ^{20}\text{Ca: } 1s^2 2s^2 2p^6 3s^2 3p^6 4s^2$$

Note that the 4*s* orbitals are filled before the 3*d* orbitals because the associated energy level of a 3*d* orbital is higher than that of a 4*s* orbital. However, the electron configuration for ^{29}Cu is

$$1s^2 2s^2 2p^6 3s^2 3p^6 4s^1 3d^{10} \text{ rather than } 1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^9$$

This is due to the fact that a half-filled or filled *d*-subshell is more stable than the *s* shell of the next level [10]. Similarly, the outermost shells for chromium (^{24}Cr) are $4s^1 3d^5$ not $4s^2 3d^4$, and those for gold (^{79}Au) are $6s^1 4f^{14} 5d^{10}$ not $6s^2 4f^{14} 5d^9$. The properties of an element depend largely on the filled state of the outermost orbitals. Alkali metals, such as ^{3}Li, ^{11}Na, and ^{19}K, have one electron in the outermost orbital and can easily lose it, especially when interacting with halogens whose outermost orbitals can be filled by adding only one electron each. The result is the formation of chemically stable compounds such as NaCl and CsF. The outermost electrons are called *valence electrons*. The $4s^1$ electron in copper is largely responsible for its high electrical conductivity because it can leave the atom relatively easily. When the outermost orbitals are completely filled, as in noble gases like He and Ne, the atoms are very stable and reluctant to react with others. Noble gases are also called inert gases since they are monatomic gases at ambient conditions. At the atmospheric pressure, helium must be cooled to 4.2 K for it to condense into liquid. The general sequence of electron configuration in order of increasing energy is schematically given in the following:

$$1s^2 \mid 2s^2\ 2p^6 \mid 3s^2\ 3p^6 \mid 4s^2 3d^{10} 4p^6 \mid 5s^2 4d^{10} 5p^6 \mid 6s^2 4f^{14} 5d^{10} 6p^6 \mid 7s^2 5f^{14} 6d^{10}$$

For convenience, each dashed line indicates the electron configuration of an inert gas listed underneath that line. Each noble gas contains a completely filled p subshell (with the exception of He which has a filled K shell) before the next s subshell. In atomic physics, *ionization energy* is the energy required to separate an electron from the atomic nucleus. The ionization energy varies periodically according to the atomic number: Alkali metals have the lowest ionization energy because of the single electron in the outermost s-orbitals. On the other hand, inert gases have the highest ionization energy. Helium is the most stable element with an ionization energy of 24.6 eV. The ionization energy of lithium is only 5.4 eV. For a hydrogen atom, the ionization energy is 13.6 eV as discussed in Chap. 3.

6.2.2 Insulators, Conductors, and Semiconductors

The picture of free-electron gas depicted in Chap. 5 is an oversimplified version in which the electron energies are limited to a nearly continuous band from the zero energy level up to the Fermi energy or Fermi level. Only those near the Fermi surface contribute to electronic transport properties. Electrons in a single atom are in various discrete energy levels, which are well predicted by quantum mechanics. In real solids, atoms are arranged in close proximity; hence, electrons interact strongly with one another as well as with the crystal lattices, resulting in complex wavefunctions as manifested by their band structures. There exist a large number of *allowable bands* that may be occupied by electrons. Between two consecutive allowable bands, there exists a *forbidden band* that cannot be occupied by any electron. Electrons occupy broad bands with allowable energy states up to the Fermi level. The distinction between insulators and metals can be understood by looking at the electronic states near the Fermi surface as illustrated in Fig. 6.2. A brief qualitative description is given here, whereas more detailed theories are deferred to subsequent sections.

For insulators, the highest occupied band is completely filled as shown in Fig. 6.2a. This is called a *valence band* due to the contribution of valence electrons. The next higher band is a *conduction band* which is completely empty. There exists a large energy gap between the valance band and the conduction band, usually between 5 and 15 eV. Examples are $E_g \approx 8$ eV for fused silica (SiO_2) and $E_g \approx 14$ eV for LiF. The Fermi level lies in the middle of the forbidden band. Because the valence band is completely filled, electrons are not free to move around (i.e., change from one quantum state to another) under the influence of an electric field. An electrical insulator is also called a dielectric. Pure crystalline dielectrics are transparent to visible light because their valence electrons cannot be excited unless the incoming radiation frequency is high enough that the photon energy exceeds the bandgap energy. Note that a photon energy of $h\nu = 2$ eV corresponds to a visible wavelength

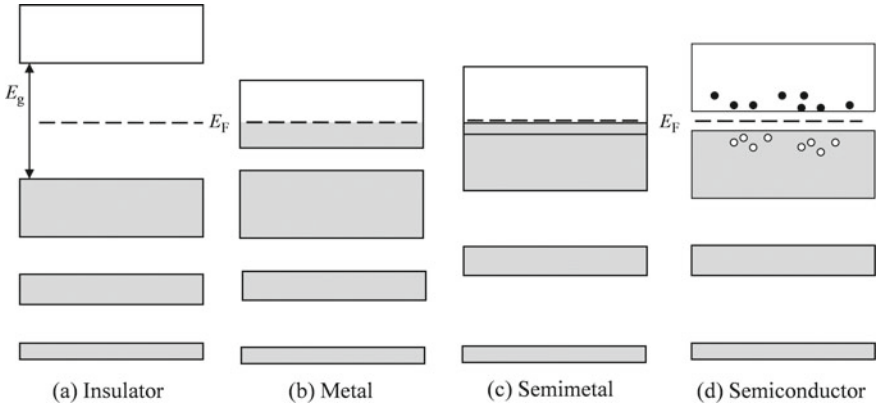

Fig. 6.2 Schematic of the energy band for different materials, where E_g is the bandgap energy and E_F is the Fermi energy. **a** An insulator has a completely filled valance band and a completely empty conduction band, with a wide bandgap between the two. **b** A metal has a partially filled conduction band and the Fermi level lies in this band. **c** A semimetal, also called a metal, has a conduction band that overlaps the filled valence band. **d** A semiconductor is like an insulator but with a much smaller bandgap and may conduct electricity at elevated temperatures due to thermally excited electrons and holes. Doping or impurities in a semiconductor can result in a large electrical conductance

$\lambda = 620\,\text{nm}$, and that of 10 eV corresponds to $\lambda = 124\,\text{nm}$, which lies in the deep ultraviolet. On the other hand, lattice vibrations or phonons in dielectric materials often yield absorption of radiation in the mid infrared.

A metal has a partially filled conduction band, which is the highest occupied band, as shown in Fig. 6.2b. The Fermi level lies inside this allowable band. For some metals like Bi and Sn, the conduction band overlaps the valence band as illustrated in Fig. 6.2c. These metals are sometimes called semimetals since their electrical conductivities are not as high as the alkali or noble metals. Because the energy states within the conduction band are continuous, the uppermost electrons in the partially filled conduction band or the top of the valence band can be excited to a higher unoccupied energy level by an arbitrarily applied field. Over 80% of the elements in the periodic table are metals (or semimetals). All group Ia (alkali, excluding hydrogen), group IIa (alkaline earth), group IIIa (except boron), and transition (all b groups from columns 3 to 12 of the periodic table) elements are metals. The interaction between electromagnetic radiation and a material is much like applying an electric field to the material, except that the frequency of the applied field is very high. Note that the frequency of red light at $\lambda = 632\,\text{nm}$ is $\nu = c/\lambda = 475\,\text{THz}$. Because of their relatively free electrons, metals interact with electromagnetic radiation strongly. This is manifested by the strong absorption by thin metallic films and the high reflection from polished bulk metals. The strong interaction of metals with microwaves can easily be demonstrated by placing a piece of aluminum foil in a microwave oven and then observing the noises and sparkles as the oven is turned on. At shorter wavelengths in the visible spectrum and in the ultraviolet, additional absorption mechanisms emerge that may be better explained by the particle nature of light.

Semiconductors have band structures similar to those of insulators, except that the energy bandgap E_g is much narrower, i.e., on the order of 1 eV. For example, diamond has a bandgap of 5.5 eV and is usually classified as an insulator, whereas silicon has a bandgap of 1.1 eV at room temperature and is a semiconductor. Some semiconductors can have a relatively large bandgap and hence are called *wideband semiconductors*. Examples are the III–V semiconductor GaN (3.4 eV) and the II–VI semiconductors CdS (2.4 eV) and ZnS (3.7 eV). Diamond may be considered as a wideband semiconductor because of its crystal structure similar to those of Si and Ge. Pure or *intrinsic* semiconductors are insulators at low temperatures. At higher temperatures, as illustrated in Fig. 6.2d, some electrons (dots) can be *thermally excited* from the valence band to the conduction band, leaving holes (circles) in the valence band. Subsequently, electrical current may flow through, although with a large resistance as compared to metals. Bandgap absorption is essential for the interaction of semiconductors with optical radiation. When the photon energy exceeds the energy gap, strong absorption occurs. This is why a silicon wafer looks dark and is opaque to visible light.

By doping the semiconductor with impurities, the charge distribution can be significantly changed, while, at the same time, the bandgap and the Fermi level are slightly modified. The semiconductor becomes *extrinsic*, meaning that the number of electrons is no longer the same as that of holes. A group V element, such as phosphorous with five valance electrons, may substitute a small fraction of silicon atoms. The extra valence electrons can be thermally excited to the conduction band via ionization of the impurities. The phosphorus atom is said to be a *donor*, and the doped semiconductor becomes *n*-type since majority of its carriers are electrons. The electron concentration can be significantly increased to enhance the electrical conductivity. From the band structure point of view, the donated electrons form a filled impurity band right below the conduction band. The difference in energy between the conduction band and the impurity band is called *ionization energy*, which is on the order of 0.05 eV. The ionization energy of a semiconductor has a different meaning from the ionization energy required to separate an electron from the atomic nucleus discussed earlier. Likewise, when impurities from a group III element such as boron with three valance electrons are introduced, additional holes are created such that the silicon semiconductor becomes a *p*-type semiconductor because of the additional positive charge carriers. The boron atoms are called *acceptors*, which form an empty impurity band right above the valence band [11]. The energy difference between these two bands is also called the ionization energy. Doping can strongly affect the infrared properties of semiconductors because of free-carrier absorption. Furthermore, impurities and defects tend to increase phonon scattering and reduce thermal conductivity since thermal transport in semiconductors is mainly by lattice vibration.

6.2.3 Atomic Binding in Solids

Two or more atoms can combine to form a molecule, mainly through the electrons in the outermost orbitals (i.e., valence electrons), since the electrons in the inner shells remain tightly bonded to their nuclei. The wavefunctions of the valance electrons are significantly modified as compared with those of the individual atoms. There are five major kinds of chemical bonds: the ionic, covalent, molecular, and hydrogen bonds for insulators and the metallic bond for conductors. Solids with identical chemical composition can have different stable forms or phases, which exhibit distinct differences in their appearances as well as electrical, mechanical, and thermal properties. A notable example is carbon, which may exist in the form of diamond, graphite, carbon black (amorphous carbon), or the fullerene family. A crystal contains periodic and densely packed atoms or lattices, whereas an amorphous solid does not have well-organized lattice structures. The atoms in an amorphous solid are disordered and irregular, like those in a liquid, except that they are firmly bonded together. Therefore, a crystal is usually denser and harder than the amorphous phase of the same composition. A crystal usually exhibits distinct facets along the crystalline planes and has a sharp transition between solid and liquid at a fixed melting point. An amorphous solid does not have clear facets when broken. When heated up, an amorphous solid is first softened and then gradually it melts over a wide temperature range. An example is quartz versus fused silica (glass), both made of SiO_2. For a given composition, the thermal conductivity is usually much higher in the crystal form because of lattice vibrations.

Alkali metals and alkaline earth metals have one and two valance electrons, respectively, that are loosely bonded. A metal atom can lose its outermost electrons to become a positive ion. On the other hand, the elements in groups VIIa and VIa tend to gain additional electrons to fill the outermost orbitals and become negative ions. The positive and negative ions attract each other by electrostatic force and form an *ionic bond*, which is quite strong. *Ionic crystals*, such as NaCl, CsCl, KBr, CaF_2, and MgO, are hard and usually have high melting points (above 1000 K). They are insulators because the ions cannot move around freely and are transparent in the visible spectrum because of the large bandgap. Nevertheless, some of these crystals are soluble and can be dissolved in water. The solution becomes conductive because of the ions. The positive and negative ions form an electrical dipole and can absorb infrared radiation through lattice vibrations. These solids belong to the group of *polar materials*, in terms of polarizability. Note that the elements in groups Ib (noble metals) and IIb (Zn, Cd, and Hg) resemble those in groups Ia and IIa because of the outermost *s*-orbital electrons. The difference is that groups Ib and IIb also have filled *d*-subshells. Therefore, II–VI semiconductors such as ZnSe and CdTe are largely ionic bonded.

The main contribution to the binding energy is the electrostatic or Madelung energy [2]. The long-range electrostatic force between two ions with charges q_1 and q_2 is proportional to $q_1 q_2 / r^2$, where r is the separation distance measured from the center of the ion cores. Depending on the sign of the charges, either attractive

or repulsive force may occur. The ions arrange themselves in a way that gives the strongest attractive interaction, which is balanced by the short-range repulsive force between atoms. The contribution of the Coulomb attraction to the total energy of the system is roughly proportional to $-1/r$. As atoms are brought very close to each other, the charge distributions or the electron orbitals begin to overlap with each other. Pauli's exclusion principle requires some of the electrons move to higher quantum states, resulting in an increased total energy of the system. Associated with the increased energy is a repulsive force between the atoms. The magnitude of this repulsive force varies with $1/r^{m+1}$ (where m is between 6 and 10 for alkali halides with NaCl structure), and thus is negligible at large distances but increases rapidly when the distance is less than 0.5 nm [1]. The repulsive force contributes to the energy of the system by $1/r^m$. There exists a minimum energy or equilibrium position of the system when all the repulsive and attractive forces balance each other. Readers are reminded about the similar discussion in Sect. 4.2.4 on the intermolecular force and potential; see Eq. (4.51) and Fig. 4.8. Understanding the binding energy or the interatomic potential is very important for atomic scale simulations, e.g., those using molecular dynamics.

Covalent bonds are formed between nonmetallic elements when the electrons in the outermost orbitals are shared by more than one atom. Covalent binding is important for gaseous molecules like Cl_2, N_2, and CO_2. When the atoms are brought close enough, the electron orbitals overlap, allowing them to share one or more electrons. Covalent interactions result in attractive forces, and the binding of atoms is associated with a reduced total energy. *Covalent crystals* consist of an infinite network of atoms joined together by covalent bonds. Examples are diamond, silicon, SiC, and quartz (SiO_2). The whole crystal is better viewed as a large molecule or supermolecule. In diamond structure, each atom is bonded to four neighboring atoms, which form a tetrahedron. In a SiC crystal, each silicon atom is bonded to four carbon atoms and vice versa. In a SiO_2 crystal, while each silicon atom is bonded to four oxygen atoms at tetrahedral angles, each oxygen atom is bonded only to two silicon atoms. Covalent solids are usually very hard with a high melting point and thermal conductivity. The melting points of quartz and silicon are 1920 K and 1690 K, respectively. Diamond has the highest melting point (3820 K) among all known materials. At room temperature, the thermal conductivity of diamond is 2300 W/m K, which is the highest of all known bulk materials. Pure diamond and intrinsic silicon do not absorb radiation at frequencies lower than that of the corresponding bandgap energy. Because of its wide bandgap, diamond is clear in the visible region and transparent throughout the whole infrared and microwave regions.

Some solids have both ionic and covalent characteristics. Examples are the III–V semiconductors such as GaN, GaAs, and InSb. II–VI materials such as ZnO and CdS have a large proportion (30%) of covalent bond characteristics. Even SiC has some ionic bond characteristics because of the dipoles formed due to different attractive forces by different atoms. Therefore, SiC is also a polar material that can absorb and emit infrared radiation through lattice vibrations.

Inert gases can be solidified at very low temperatures via *molecular bonds*. At atmospheric pressure, argon becomes liquid at temperatures between 84 and 87 K.

At temperatures below 84 K, it crystallizes into a dense solid, called a *molecular crystal*. Van der Waals' force caused by induced dipole moments between atoms is responsible for the attraction and binding of atoms. The van der Waals weak interaction gives a long-range potential that is proportional to $-1/r^6$, as discussed in Sect. 4.3.1. The repulsive potential for inert gas is proportional to $1/r^{12}$. Molecular bonds are also important for many organic molecules.

Hydrogen has only one electron per atom and can form a covalent bond with another to form H_2 molecule. When interacting with other atoms, a hydrogen atom may be attracted to form a *hydrogen bond*. The hydrogen bond and the resulting electrostatic attraction are important for H_2O molecules, with many striking physical properties in its vapor, water, and ice phases. Hydrogen bonds and molecular bonds are essential to organic molecules and polymers.

Metallic bonds are responsible for the binding energy in metals. Pure metals can form densely packed periodic lattices or crystals. Metals often exist in *polycrystalline* form in which small grains of crystals are joined together randomly, or in *alloy* form in which the atoms are arranged irregularly like an amorphous insulator. Unlike in a covalent crystal where atoms share a few electrons, in a metallic crystal, some valence electrons leave the ion cores completely and are shared by all the ions in the crystal. This is consistent with the picture of free-electron gas and describes well the behavior of alkali metals. Transition metals, including the noble metals, contain electrons in the d subshell. The metallic bonds are supplemented by covalent and molecular bonds. Due to the relatively free-electron gas, metals have high thermal and electrical conductivities. Metallic crystals are also more flexible than nonmetallic crystals, which are usually brittle. The melting points of metals vary significantly. Examples are Hg (234 K), Ga (303 K), Au (1338 K), and W (3695 K). As mentioned in previous chapters, the physical properties would change significantly as the structure is reduced down to hundreds, tens, or even a few atomic layers in one, two, or three dimensions. Examples are carbon nanotubes, silicon nanowires, ZnO nanobelts, and CdSe-ZnS quantum dots. In order to further understand the properties of solids, let us examine the crystal structures more closely in the following section.

6.3 Crystal Structures

A crystal is constructed by the continuous repetition in space of an identical structural unit. Geometrically speaking, a crystal is a 3D periodic array, or network, of lattices. All lattice points are identical to one another. For a crystal made of only one type of element, each lattice point may be treated as a single atom or ion. However, this is not necessary as will be illustrated later. In general, each lattice point represents a set of atoms, ions, or molecules, located in its neighborhood. This set of atoms, ions, or molecules is called a *basis*. A *unit cell* of a crystal structure contains both the lattice and the basis, and can be repeated by translations to cover the whole crystal.

It has long been hypothesized that crystalline materials must have some periodicity in their microstructures. In 1913, W. L. Bragg and his father W. H. Bragg used

x-rays to provide microscopic evidence of the existence of periodic lattice structures. This was a giant step because the distances between atoms are on the order of 0.1 nm. X-ray crystallography provided a powerful tool for the determination of the microscopic structure of solids. The Braggs received the Nobel Prize in Physics in 1915, when Lawrence Bragg was only 25 years old. It was not until 1983 that atomic images were obtained in real space using a scanning tunneling microscope (STM) as discussed in Chap. 1. The physical properties of crystalline solids are largely determined by the arrangement of atoms in a unit cell, in addition to the chemical bonds between atoms. It is of great importance to know the structure of a crystal first in order to understand its electrical, thermal, mechanical, and optical properties.

6.3.1 The Bravais Lattices

In three dimensions, crystal lattices can be grouped into 14 different types as required by translational symmetry. These are called Bravais lattices, named after French physicist Auguste Bravais (1811–1863), who showed that there are only 14 unique Bravais lattices from the point of view of symmetry. Bravais lattices are then categorized into seven crystal systems, resulting in seven types of *conventional unit cells*, namely, cubic, tetragonal, orthorhombic, hexagonal, rhombohedral, monoclinic, and triclinic, as illustrated in Fig. 6.3.

There are three cubic lattices: the simple cubic with lattice points only on its apexes, the body-centered cubic (bcc) with one additional lattice at the center, and the face-centered cubic (fcc) with one additional lattice at each face, as shown in Figs. 6.3a1, a2, and a3. To illustrate the difference between bcc and fcc lattices clearly, Fig. 6.4 displays the top views of these two structures with the same a, which is called the *lattice constant*. Some practical examples will be given soon. If one looks at Fig. 6.4b along the diagonals, the face-centered structure becomes body-centered. However, the lattice constant would become $a/\sqrt{2}$ along the lateral directions but remains a in the vertical direction. Such a structure is a special case of the tetragonal, because one side is not the same as the other two. There are two tetragonal Bravais lattices, the simple and the body-centered, because a face-centered tetragonal lattice can simply be rotated by 45° to become a body-centered one. A tetragonal lattice can be thought of as a cubic lattice stretched in one direction.

In the orthorhombic lattices shown in Fig. 6.3c, the three lattice constants, a, b, and c, are not equal to each other. Besides the simple, body-centered, and face-centered orthorhombic lattices, there exists a base-centered lattice structure, in which two additional lattices are placed at the center of the top and bottom faces. An orthorhombic lattice can be thought of as a corresponding tetragonal lattice stretched along one side of its square. To produce the additional two, one can simply rotate the tetragonal by 45° and then stretch it.

A hexagonal lattice contains equal triangular or honeycomb-layered structures (see Fig. 6.3d). The next three types of Bravais lattices have inclined faces (see Fig. 6.3e–g). The rhombohedral (or trigonal) has equal sides, whereas the triclinic

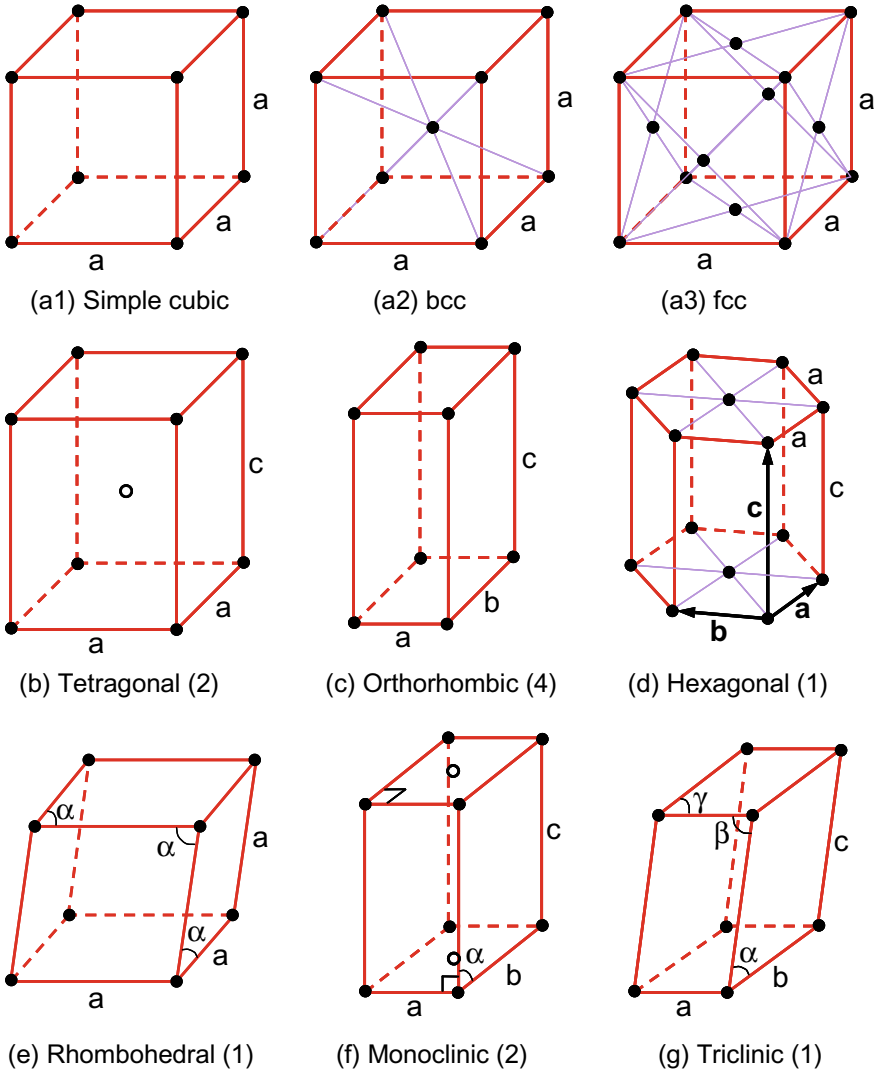

Fig. 6.3 The seven crystal systems with a total of 14 Bravais lattices, where each point is called a lattice point. The number in parentheses refers to the number of Bravais lattices in the crystal system. **a** Three types of the cubic: simple cubic, body-centered cubic (bcc), and face-centered cubic (fcc). **b** Tetragonal: either simple or body-centered as represented by the empty circle at the center. **c** Orthorhombic: simple, body-centered, face-centered, or base-centered. d Hexagonal. **e** Rhombohedral (also trigonal). **f** Monoclinic: simple or base-centered as represented by the empty circles on the opposite faces. **g** Triclinic

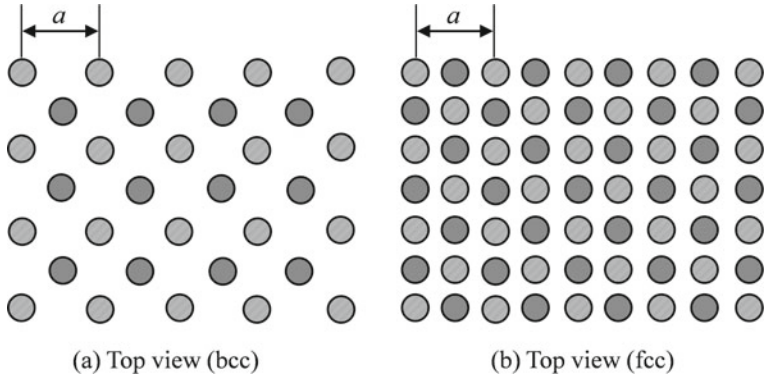

Fig. 6.4 Top views of **a** body-centered cubic and **b** face-centered cubic Bravais lattices. The two different filling patterns (hatched and shaded) represent lattice points on alternative layers as in Fig. 6.3a2, a3

has three different sides and angles. Both contain six parallelogram faces. The monoclinic, on the other hand, has four rectangular faces and two parallelogram faces.

Example 6.2 Copper is an fcc lattice. Estimate the lattice constant and the distance between nearest copper atoms (or ion cores to be exact) from the density and the molecular weight of copper.

Solution From Table 5.1, we have for Cu that $\rho = 8930\,\text{kg/m}^3$ and $M = 63.5\,\text{kg/kmol}$. The number density of Cu atoms is $n = \rho N_\text{A}/M = 8.47 \times 10^{28}\,\text{m}^{-3}$. If the structure were simple cubic, we would easily find that $a = n^{-1/3} = 0.228\,\text{nm}$, which would also be the closest distance between atoms. For an fcc lattice, there are eight corner points and six face points. If each lattice point is made of one atom, each corner point contains one-octant of an atom and each face point contains half of an atom inside the cube. Therefore, there are four atoms inside each fcc unit cell. The number of unit cells per unit volume becomes $n/4$ and the calculated lattice constant is $a = 0.361\,\text{nm}$ for Cu. The closest distance between atoms is $a/\sqrt{2} = 0.256\,\text{nm}$. If we assume that all the atoms are rigid spheres that are closely packed (touching one another), then we can calculate the packing density or the fraction of occupied space. Assume that the diameter of an atom is d. For a simple cubic lattice, $a = d$ and there is only one atom per lattice. Hence, $f = (1/6)\pi d^3/a^3 = 0.52$. For an fcc lattice, $a = d\sqrt{2}$ and $f = 4(1/6)\pi d^3/a^3 = 0.74$. What is the packing density for a bcc lattice then?

Some solids with bcc or fcc lattices are listed in Table 6.1, along with others that form a hexagonal close-packed (hcp) lattice. An hcp lattice can be considered as two Bravais hexagonal lattices that are interlocked at $c/2$. Each lattice point is surrounded by, at equal distances, 12 neighboring points: 3 above, 3 below, and 6 at the same height. Imagine that atoms are rigid spheres with a diameter d; it can

Table 6.1 Crystal structures and lattice constants of common elements [1, 2]. Room temperature values unless otherwise indicated. Note that 1 Å = 0.1 nm

fcc		bcc		hcp		
Element	a (Å)	Element	a (Å)	Element	a (Å)	c (Å)
Ar (4.2 K)	5.26	Ba	5.02	H (4 K)	3.75	6.12
Ag	4.09	Cr	2.88	Be	2.27	3.59
Al	4.05	Cs (78 K)	6.05	Cd	2.98	5.62
Au	4.08	Fe	2.87	Er	3.56	5.59
Ca	5.58	K (5 K)	5.23	Gd	3.64	5.78
Ce	5.16	Li (78 K)	3.49	Mg	3.21	5.21
Cu	3.61	Mo	3.15	Ti	2.95	4.69
Pb	4.95	Na (5 K)	4.23	Tl	3.46	5.53
Pd	3.89	Nb	3.30	Y	3.65	5.73
Pt	3.92	V	3.03	Zn	2.66	4.95
Yb	5.49	W	3.16	Zr	3.23	5.15

be shown that $a = d$ and $c = d\sqrt{8/3}$ for an hcp lattice. Each sphere is in contact with 12 others. It can be seen from Table 6.1 that these hcp crystals follow the ratio $c/a = \sqrt{8/3} \approx 1.633$ within $\pm 16\%$.

6.3.2 Primitive Vectors and the Primitive Unit Cell

A set of *primitive vectors* can be defined for Bravais lattices $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ so that the vector between any two lattice points can be expressed by the *lattice translation vector* (or operator)

$$\mathbf{R} = m\mathbf{a} + n\mathbf{b} + l\mathbf{c} \tag{6.4}$$

where m, n, and l are integers. For a simple cubic lattice, we can simply assign $\mathbf{a} = a\hat{\mathbf{x}}$, $\mathbf{b} = a\hat{\mathbf{y}}$, $\mathbf{c} = a\hat{\mathbf{z}}$, as can be seen from Fig. 6.3*a*1. However, the assignment of primitive vectors is not unique. The parallelepiped formed by the three vectors is called a *primitive unit cell*, whose volume $V_{\text{uc}} = \mathbf{a} \times \mathbf{b} \cdot \mathbf{c}$ remains the same no matter how the primitive vectors are chosen. Taking the bcc lattice as an example, we may choose the primitive vectors as either

$$\mathbf{a} = a\hat{\mathbf{x}}, \mathbf{b} = a\hat{\mathbf{y}}, \mathbf{c} = 0.5a(\hat{\mathbf{x}} + \hat{y} + \hat{z}) \tag{6.5a}$$

or

$$\mathbf{a} = 0.5a(\mathbf{-\hat{x} + \hat{y} + \hat{z}}), \mathbf{b} = 0.5a(\hat{\mathbf{x}} - \hat{y} + \hat{z}), \mathbf{c} = 0.5a(\hat{\mathbf{x}} + \hat{y} - \hat{z}) \tag{6.5b}$$

From Eq. (6.5b), we see that $\mathbf{a}+\mathbf{b}+\mathbf{c}$ points to the center point and $\mathbf{a}+\mathbf{b}=a\hat{\mathbf{z}}$. Either way, we end up with $V_{\text{uc}}=0.5a^3$, suggesting that the Bravais bcc lattice is not a primitive cell. In fact, only the simple Bravais lattices are primitive unit cells. Of course, there are other ways of choosing the primitive vectors. For a Bravais fcc lattice, we can write

$$\mathbf{a}=0.5a(\hat{\mathbf{y}}+\hat{z}),\ \mathbf{b}=0.5a(\hat{\mathbf{x}}+\hat{z}),\ \mathbf{c}=0.5a(\hat{\mathbf{x}}+\hat{y}) \tag{6.6}$$

Each vector conveniently ends at the three face-centered points. The total volume of the primitive cell becomes $V_{\text{uc}}=0.25a^3$, as expected. Each lattice point is associated with a basis of atoms whose locations relative to the lattice point can be specified by $\mathbf{r}_j=x_j\mathbf{a}+y_j\mathbf{b}+z_j\mathbf{c}$ with $0\le x_j, y_j, z_j\le 1$ for the jth atom.

Another way of choosing a unit cell is to follow the two steps: (1) Draw lines to connect a given lattice point to all nearby lattice points. (2) At the midpoint and normal to these lines draw new lines or planes. The smallest volume enclosed in this way is called the *Wigner–Seitz primitive cell*, as illustrated in Fig. 6.5. The Wigner–Seitz cell for a 2D lattice becomes a hexagon whose opposite sides are parallel, and that for an fcc lattice is a rhombic dodecahedron. The longer diagonal of each rhombic face is $\sqrt{2}$ times that of the shorter diagonal. There are six apexes where four surfaces meet and eight apexes where three surfaces meet. The distance between opposite axes joined by four faces is exactly the Bravais lattice constant a. The axes, x, y, and z, pass through these six apexes as well as the center. Each Wigner–Seitz cell contains only one lattice point, and it has been proven to be a primitive cell.

It is convenient to describe the orientation of the crystal plane by the Miller indices, which are three integers h, k, and l, without common factors, and denoted by (hkl). These numbers give a vector $h\mathbf{a}+k\mathbf{b}+l\mathbf{c}$ that is perpendicular to the plane. For example, if **a**, **b,** and **c** are along the x-, y-, and z-axes, respectively, the six surfaces of the cubic unit cell are represented by (001), (00$\bar{1}$), (010), (0$\bar{1}$0), (100), and ($\bar{1}$00),

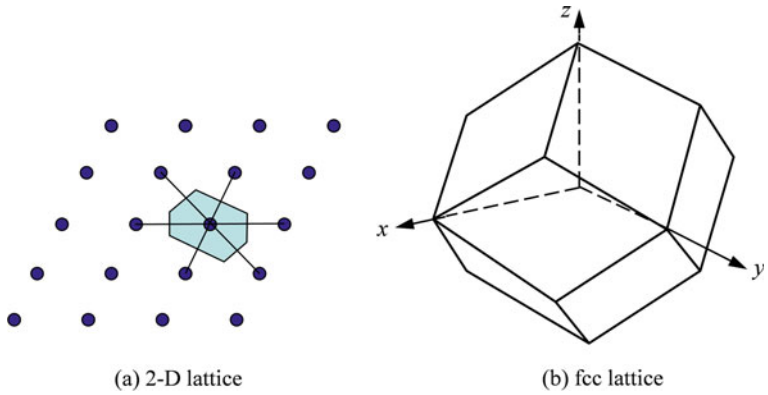

Fig. 6.5 The Wigner–Seitz cells: **a** For a 2D lattice as shown by the shaded region and **b** for an fcc lattice as shown by the rhombic dodecahedron

where negative sign is denoted by a bar on top of the number. The whole set of surfaces can be denoted by {100} due to symmetry. Most commercial semiconductor wafers are (100) oriented and some (111). The way to find the smallest h, k, l of any specified crystal facet is first to extend the plane so that it intersects the axes formed by the lattice vectors. Find the intercepts on each axis in terms of multiples of the unit cell vector, e.g., $(2, 4, -6)$; the numbers must be integers for any specified crystal plane. Take the reciprocals of these numbers, which are $(\frac{1}{2}, \frac{1}{4}, -\frac{1}{6})$. Multiply them by the least common multiple, which is 12 in this example. Then put into the Miller indices $(6, 3, \bar{2})$. All parallel planes are characterized by the same set of Miller indices.

Example 6.3 Find all angles between the (100), (111), and (311) surfaces in a cubic lattice.

Solution For two vectors **a** and **b**, $\mathbf{a} \cdot \mathbf{b} = ab\cos\alpha = x_a x_b + y_a y_b + z_a z_b$. Thus, the angle between (100) and (111) planes is $\alpha = \cos^{-1}\left(1/\sqrt{3}\right) = 54.7°$; that between (100) and (311) planes is $\alpha = \cos^{-1}\left(3/\sqrt{11}\right) = 25.2°$; and that between (111) and (311) planes is $\alpha = \cos^{-1}\left(\frac{3+1+1}{\sqrt{11\times 3}}\right) = 29.5°$.

6.3.3 Basis Made of Two or More Atoms

With respect to the primitive vector and basis, a bcc lattice can be thought of as a simple cubic with a basis made of two atoms, one at $(0, 0, 0)$ and the other at $a(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Each of the eight lattice points contains the same basis by translation, according to Eq. (6.4), and the unit vectors along the three orthogonal sides of the cubic. The simple cubic lattice having a basis of two atoms, however, breaks some of the symmetry of the Bravais cubic lattice and is called a non-Bravais lattice. Lattices with a basis consisting of more than one atom have important practical applications as discussed in the following. The cesium chloride structure is made of two types of elements, each forming a simple Bravais lattice, as shown in Fig. 6.6a. The two Bravais lattices can be thought of as being placed in identical positions first, and then one is moved by $a\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ so that the point at the origin is translated to the center of the other. It is not a body-centered cubic lattice. Rather, the crystal structure can be viewed as a simple cubic with a base of two ions, Cs at $(0, 0, 0)$ and Cl at $a\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$. The sodium chloride structure is more common. In this case, it can be considered as two fcc lattices made of different ions. The two fcc lattices are then translated exactly the same way as in the CsCl structure. The resulting structure is shown in Fig. 6.6b, where each ion is surrounded by six ions of the other type. The lattice constants of some common crystals are listed in Table 6.2. It can be seen that most ionic crystals form NaCl or CsCl structures.

The crystal structures of diamond and zincblende semiconductors are also derivatives of the cubic structure. The zincblende structure is formed from two fcc lattices

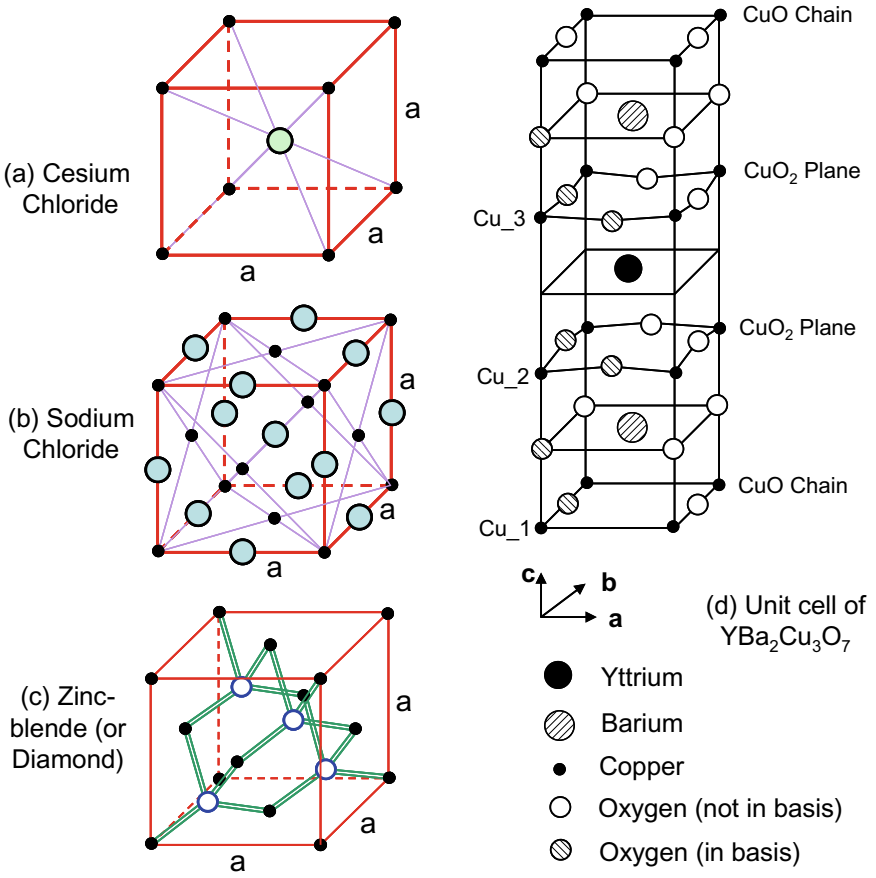

Fig. 6.6 Crystalline structures. **a** Cesium chloride; **b** Sodium chloride. **c** Zincblende, which becomes a diamond structure when the atoms in the empty circles are the same as the filled ones. **d** $YBa_2Cu_3O_7$ superconductor whose lattice constants are approximately $a = 0.38$, $b = 0.39$, and $c = 1.17$ nm

with different types of atoms, displaced along the body diagonal by one-quarter the length of the diagonal. Specifically, the basis is made of one atom at $(0, 0, 0)$ and the other atom at $a\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$, as shown in Fig. 6.6c. A total of four atoms are moved completely inside the cube, and each atom has a covalent bond with each of the four adjacent atoms, which together form a tetrahedron. Examples of zincblende structure are GaAs, SiC, and so forth. A diamond structure can be viewed as a special case of a zincblende structure for which there is only one type of element, such as C, Si, or Ge. The outermost subshell of Si is $3s^2 3p^2$, and the s subshell is filled. By promoting an s-electron to a p-orbital to form sp^3 hybrid orbitals, four covalent bonds can be formed. This is also true for C and Ge. In essence, the diamond lattice can be thought of as an fcc lattice with a basis containing two identical atoms: one is on

Table 6.2 Crystal properties of some compounds and semiconductors at room temperature [1, 2]. For semiconductors, the bandgap energy is indicated, and "i" in parentheses denotes an indirect bandgap

Compound		Semiconductors		
Composition	a (Å)	Composition	a (Å)	E_g (eV)
Sodium chloride structure		Diamond structure		
LiF	4.02	C	3.57	5.47 (i)
LiCl	5.13	Si	5.43	1.11 (i)
NaBr	5.97	Ge	5.66	0.66 (i)
NaCl	5.64			
KBr	6.60	Zincblende structure		
KCl	6.29	BN	3.62	7.5 (i)
CsF	6.01	CdS	5.82	2.42
AgCl	5.55	CdSe	6.05	1.70
AgBr	5.77	CdTe	6.48	1.56
MgO	4.21	GaAs	5.65	1.42
MgS	5.20	GaN (w)	5.45	3.36
CaO	4.81	GaP	5.45	2.26 (i)
CaS	5.69	GaSb	6.43	0.72
CaSe	5.91	HgTe	6.04	<0
BaTe	6.99	InAs	5.87	0.36
		InP	6.48	1.35
Cesium chloride structure		InSb	4.35	0.17
CsCl	4.12	SiC	4.63	2.36
CsBr	4.29	ZnO	5.41	3.35
CsI	4.57	ZnS	5.67	3.68
TlBr	3.97	ZnSe	6.09	2.58

the corner and the other on the body diagonal at a distance of one-quarter diagonal. Table 6.2 also presents commonly used diamond and zincblende semiconductors with associated lattice constants and bandgap energies. Notice that GaN crystal is wurtzite in its stable form with a hexagonal symmetry. This is also the case for AlN and InN, which are not shown in the table. The III-nitride materials have a wide-band, and thus are important for UV-blue-green LEDs and lasers. On the other hand, ZnS, ZnO, CdS, and CdSe can also be wurtzite. HgTe is a semimetal with a negative bandgap and can be mixed with the wideband semiconductor CdTe to form the ternary compound of $Hg_{1-x}Cd_xTe$, which can be used as infrared detectors, namely, mercury–cadmium–telluride (MCT) detectors.

Yttrium–barium–copper oxide ($YBa_2Cu_3O_7$) is a high-temperature superconductor, which becomes superconducting at temperatures below 91 K [12]. It belongs to the cuprate-perovskite family and is a ceramic material when one oxygen atom

is removed from the unit cell to form $YBa_2Cu_3O_6$ [13]. The crystal structure of $YBa_2Cu_3O_7$ is a simple orthorhombic lattice, whose basis contains 13 atoms, as shown in Fig. 6.6d. The structure is very close to a tetragonal one since $a \approx b$. The properties of $YBa_2Cu_3O_7$ are highly anisotropic in the c-axis direction. Superconductivity is found in the a–b plane, which is presumed due to the CuO_2 planes above and below the yttrium atom. Other examples of Bravais lattices include As, Sb, and Bi with rhombohedral lattices; In and Sn with tetragonal lattices; and Ga, Cl, Br, and S (rhombic) with orthorhombic lattices [1].

Graphite is a form of carbon made of layered structures as shown in Fig. 6.7. When separated from others, each individual layer or sheet is called a graphene. In the graphite structure, each carbon atom is covalently bonded to three others in the plane and loosely bonded between planes. There are relatively free electrons, and hence graphite is a conductor along the plane. The layer of graphite has honeycomb shape, and at first sight, it may be difficult to link it with the arrays of triangles in the hexagonal lattice. It becomes more obvious, however, if a basis is chosen to contain two atoms so that a hexagon with all diagonals can be seen by the dashed lines in Fig. 6.7a. In this way, graphite can be considered as a hexagonal Bravais lattice. The primitive unit cell of graphite is a rhombic prism (with six surfaces) formed using three layers, as illustrated by the dashed lines in Fig. 6.7b. Each unit cell contains a total of four carbon atoms.

The structure of carbon nanotubes (CNTs) can be understood based on the graphene structure and the chiral vector,

$$\mathbf{C}_\mathrm{h} = m\mathbf{a}_1 + n\mathbf{a}_2 \tag{6.7}$$

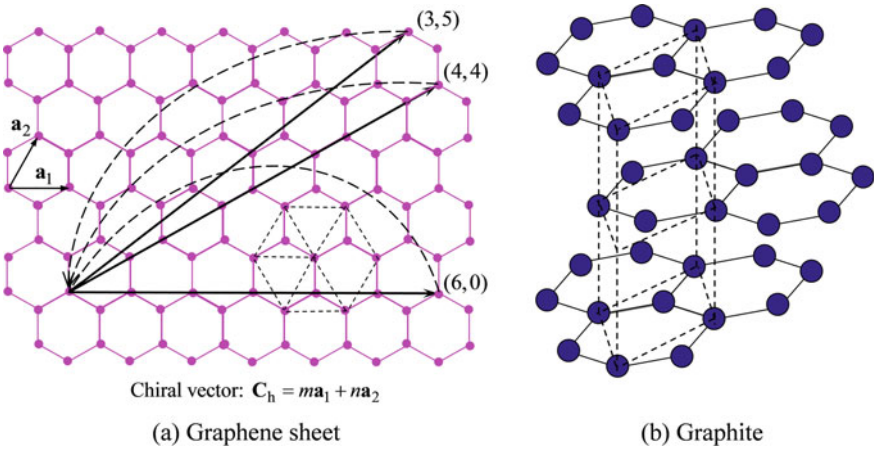

Fig. 6.7 Crystal structures of **a** graphene layer and **b** graphite. Carbon nanotubes can be viewed as rolling a graphene sheet in a direction perpendicular to the chiral vector

Different CNTs are based on rolling in the chiral vector so that the axis is perpendicular to the chiral vector and the magnitude of the chiral vector becomes the perimeter of the tube. The diameter of the tube becomes

$$d_t = \frac{C_h}{\pi} = \frac{a_{C-C}}{\pi}\sqrt{3(m^2 + mn + n^2)} \tag{6.8}$$

where $a_{C-C} = 0.1421$ nm is the nearest distance between the carbon atoms in graphene [14]. Notice that the chiral vector has a magnitude $a = a_{C-C}\sqrt{3} = 0.246$ nm. In calculating the cross-sectional surface area of a single-walled nanotube (SWNT), one could use a as the wall thickness and obtain

$$A_c = \pi d_t a = 3(a_{C-C})^2\sqrt{(m^2 + mn + n^2)} \tag{6.9}$$

Take the (20, 20) SWNT as an example, we have $d_t = 2.7$ nm and $A_c = 2.1\,\text{nm}^2$. Some researchers suggested using a layer thickness equal to the separation of graphite as 0.335 nm, which gives $A_c = \pi d_t * 0.335 = 2.9\,\text{nm}^2$. Note that $A_c = \pi D^2/4 = 5.8\,\text{nm}^2$.

6.4 Electronic Band Structures

The behavior of electrons in solid is complicated because the solution of wave functions involves a rather complicated many-body problem. Electrons in solids can be thought of as in a periodic potential due to the periodic arrays of atoms. Electronic band structures are functions that describe the electron states in the energy versus wavevector space. Let us first look at the reciprocal lattice in three dimensions.

6.4.1 Reciprocal Lattices and the First Brillouin Zone

The reciprocal lattice of a crystal structure is defined in the **k**-space (wavevector space). Since a crystal is a periodic array of lattices in real space, the reciprocal lattice can be obtained by performing a spatial Fourier transform of the crystal. For a simple orthorhombic lattice with the primitive vectors $\mathbf{a} = a\hat{\mathbf{x}}$, $\mathbf{b} = b\hat{\mathbf{y}}$, and $\mathbf{c} = c\hat{\mathbf{z}}$, the reciprocal lattice can be defined by the three vectors $\mathrm{A} = \frac{2\pi}{a}\hat{\mathbf{x}}$, $\mathbf{B} = \frac{2\pi}{b}\hat{\mathbf{y}}$, and $\mathbf{C} = \frac{2\pi}{c}\hat{\mathbf{z}}$, which define another orthorhombic. The product of the volumes of the unit lattice and the reciprocal lattice is $8\pi^3$. Some of this aspect was discussed in Chap. 5. In general, the reciprocal primitive vectors can be generated by

$$\mathrm{A} = 2\pi\frac{\mathbf{b}\times\mathbf{c}}{\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})};\ \mathbf{B} = 2\pi\frac{\mathbf{c}\times\mathbf{a}}{\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})};\ \mathbf{C} = 2\pi\frac{\mathbf{a}\times\mathbf{b}}{\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})} \tag{6.10}$$

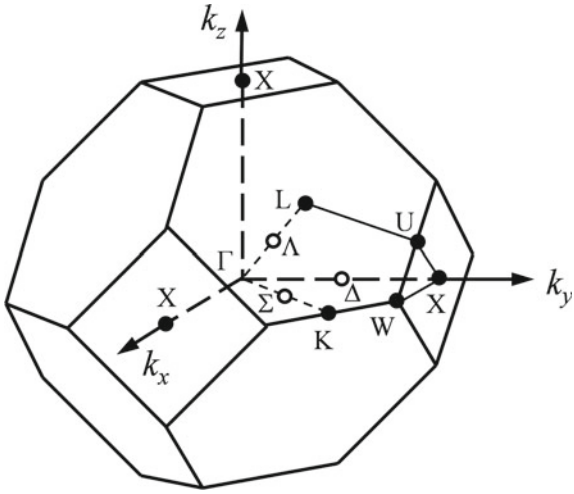

Fig. 6.8 The first Brillouin zone of a face-centered cubic structure. The shape is a truncated octahedron with eight hexagons and six squares. This is also the Wigner–Seitz cell for a bcc lattice, whose first Brillouin zone has the same shape as the Wigner–Seitz cell for an fcc lattice shown in Fig. 6.5b

In solid-state physics, a Brillouin zone is defined as a Wigner–Seitz cell in the reciprocal lattice and the smallest of which is called the first Brillouin zone. The definition of the Brillouin zone gives a vivid geometric interpretation of the Bragg diffraction condition and thus is of importance in the study of electron and phonon states in crystals, as well as their interactions with electromagnetic waves. Figure 6.8 shows the first Brillouin zone of a face-centered cubic lattice. The directions k_x, k_y, and k_z are called the [100], [010], and [001] directions, respectively. The center of the Brillouin zone is called the Γ-point, and the intersection of the three axes with the zone edge is called the X-point. The *body diagonal*, or the [111] direction, meets the zone edge at the L-point. Other points of interest such as K, W, and U at the zone edges and Δ, Λ, and Σ, located halfway between the zone center and an edge, can also be defined.

6.4.2 Bloch's Theorem

The total potential in a crystal includes the core–core, electron–electron, and electron–core Coulomb interactions. For solving electron wave functions subjected to such a potential, one would have to deal with a many-body problem, which turned out to be very difficult in mathematics. However, this problem can be simplified using the so-called *nearly free-electron model*, in which each electron moves in the average field created by the other electrons and ions. This is also called the one-electron model. The Hamiltonian operator H for the one-electron model is given as

$$H = \frac{p_e^2}{2m_e} + U(\mathbf{r}) \tag{6.11}$$

where p_e and m_e are the momentum and the mass of the electron, respectively, and $U(\mathbf{r})$ is a periodic potential function resulted from both the electron–electron and electron–core interactions. The one-electron Schrödinger equation is (see Sect. 3.5.1) given as follows:

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2 + U(\mathbf{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{6.12}$$

where E is the electron energy and $\psi(\mathbf{r})$ is the electron wave function. The periodicity of the lattice structure yields the boundary condition,

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R}) \tag{6.13}$$

where $\mathbf{R}$ is the vector between two lattice points, called the periodic potential, which ensures that $U(\mathbf{r})$ can be expanded as a Fourier series in terms of the reciprocal lattice vector $\mathbf{G}$ as follows:

$$U(\mathbf{r}) = \sum_{\mathbf{G}} U_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \tag{6.14}$$

The reciprocal lattice vector can be expressed as $\mathbf{G} = l_1\mathbf{A} + l_2\mathbf{B} + l_3\mathbf{C}$, where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are primitive vectors of the reciprocal lattic e as given in Eq. (6.10), and the integers l_1, l_2, and l_3 are indices. In Eq. (6.14), $U'_{\mathbf{G}}s$ are complex Fourier expansion coefficients for a given set of l_1, l_2, and l_3.

According to the Bloch theorem, the wave function of an electron in a periodic potential must have the form:

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) \tag{6.15}$$

where $u_{\mathbf{k}}(\mathbf{r})$ is a periodic function with the periodicity of the lattice, similar to Eq. (6.13), and thus $\psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi(\mathbf{r})$. The wave function $\psi(\mathbf{r})$ can also be expressed as a Fourier series summed over all values of the permitted wavevector such that

$$\psi(\mathbf{r}) = \sum_{\mathbf{k}} C_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{6.16}$$

The summation is over all wavevectors $\mathbf{k}$'s. From Eq. (6.16), we have

$$\nabla^2\psi(\mathbf{r}) = \sum_{\mathbf{k}} C_k (ik)^2 e^{i\mathbf{k}\cdot\mathbf{r}} = -\sum_{\mathbf{k}} k^2 C_k e^{i\mathbf{k}\cdot\mathbf{r}} \tag{6.17}$$

The combination of Eqs. (6.14) and (6.16) gives

$$U(\mathbf{r})\psi(\mathbf{r}) = \sum_{\mathbf{k}} \sum_{\mathbf{G}} U_{\mathbf{G}} C_{\mathbf{k}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{6.18}$$

Using Eqs. (6.16) through (6.18), we can rewrite the Schrödinger equation as follows:

$$\sum_{\mathbf{k}} \frac{\hbar^2 k^2}{2m_e} C_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} + \sum_{\mathbf{k}} \sum_{\mathbf{G}} U_{\mathbf{G}} C_{\mathbf{k}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} = \sum_{\mathbf{k}} E C_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{6.19}$$

The coefficients of each Fourier component must be equal on both sides of the equation. Thus,

$$\left(\frac{\hbar^2 k^2}{2m_e} - E\right) C_{\mathbf{k}} + \sum_{\mathbf{G}} U_{\mathbf{G}} C_{\mathbf{k}-\mathbf{G}} = 0 \tag{6.20}$$

where $C_{\mathbf{k}-\mathbf{G}}$ is the coefficient for the term with $\mathbf{k} - \mathbf{G}$ in the exponent, i.e., $\exp[i(\mathbf{k} - \mathbf{G}) \cdot \mathbf{r}]$ in Eq. (6.16). Equation (6.20) is paramount in the electronic band theory of crystals, and it is, therefore, called *the central equation* [2]. When $U(\mathbf{r}) \equiv 0$, Eq. (6.20) reduces to $E_k^0 = \hbar^2 k^2/(2m_e)$ by noting that $\hbar k = p_e$ for free electrons, as used in the Sommerfeld theory. Under the influence of a periodic potential, the relationship becomes more complex because it is a set of linear equations for infinite numbers of coefficients. Because the equation is homogeneous, the determinant of the characteristic matrix must be zero. In some cases, the terms can be significantly reduced to yield simple solutions with insightful physics.

Consider the 1D case when the Fourier components are relatively small compared with the kinetic energy of electrons at the zone boundary. This is the weak-potential assumption. At the first Brillouin zone boundaries, we have

$$k = G/2 = \pi/a \tag{6.21}$$

Because there are only two values of k and G, Eq. (6.20) reduces to the following two equations due to symmetry:

$$(E_\mu^0 - E)C_\mu + UC_{-\mu} = 0 \tag{6.22a}$$

and

$$(E_{-\mu}^0 - E)C_{-\mu} + UC_\mu = 0 \tag{6.22b}$$

where $\mu = \frac{1}{2}G$ is introduced merely for the convenience of notation. These equations have solutions only when the determinant becomes zero, i.e.,

$$\begin{vmatrix} E_\mu^0 - E & U \\ U & E_{-\mu}^0 - E \end{vmatrix} = 0 \tag{6.23}$$

Because $E^0_\mu = E^0_{-\mu} = \hbar^2\mu^2/(2m_\mathrm{e})$, the two roots are then obtained as

$$E = E^0_\mu \pm U = \frac{(\hbar\pi/a)^2}{2m_\mathrm{e}} \pm U \tag{6.24}$$

The two solutions at the zone edge, i.e., $k = \pi/a$, are actually on two $E(k)$ curves. When k is near the zone edge, we can express the central equation, Eq. (6.20), as the following two equations [1, 2]:

$$(E^0_k - E)C_k + UC_{k-G} = 0 \tag{6.25a}$$

and

$$(E^0_{k-G} - E)C_{k-G} + UC_k = 0 \tag{6.25b}$$

By setting its determinant to zero, we obtain

$$E(k) = \tfrac{1}{2}(E^0_k + E^0_{k-G}) \pm \left[\tfrac{1}{4}(E^0_k - E^0_{k-G})^2 + U^2\right]^{1/2} \tag{6.26}$$

which gives two branches near the zone edge, as shown in Fig. 6.9. A bandgap of $2U$ is formed at the first Brillouin zone edge. The corresponding wave functions at the zone edge are

$$\psi_{1,2}(x) = \frac{1}{\sqrt{2L}}\left(\mathrm{e}^{\mathrm{i}\pi x/a} \pm \mathrm{e}^{-\mathrm{i}\pi x/a}\right) \tag{6.27a}$$

where L is the length of the crystal. This forms two standing waves:

$$\psi_1(x) = \sqrt{2/L}\cos(\pi x/a) \quad \text{and} \quad \psi_2(x) = \mathrm{i}\sqrt{2/L}\sin(\pi x/a) \tag{6.27b}$$

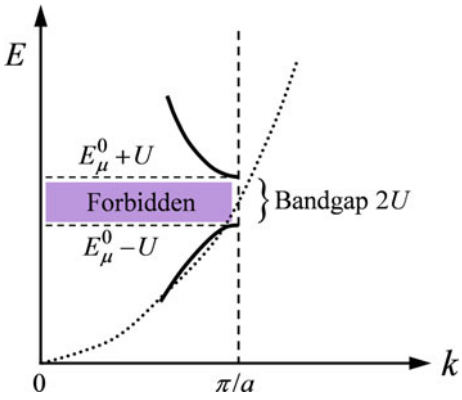

Fig. 6.9 Illustration of the energy bands, where the solid curves are calculated from Eq. (6.26). The lower and upper bands correspond to the choice of the minus and plus signs, respectively. When $k = G/2 = \pi/a$, the two bands are separated by a gap of magnitude $2U$. The dashed line, on the other hand, represents the free-electron behavior according to $E \propto k^2$

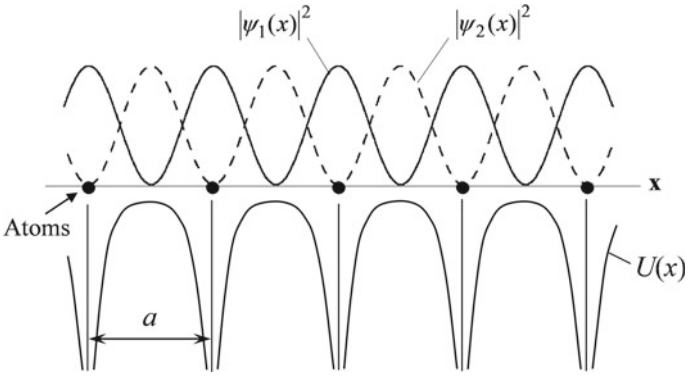

Fig. 6.10 The upper part of the figure plots the probability density $|\psi|^2$ in a 1D weak potential at the edge of the first Brillouin zone; the lower part of the figure illustrates the actual potential $U(x)$ of electrons

As shown in Fig. 6.10, the lower energy state $E_\mu^0 - U$ corresponds to ψ_1 with a probability density $|\psi_1|^2$ peaked at core sites. The probability density function describes electrons that are piled up close to the core site. The upper energy state $E_\mu^0 + U$ corresponds to ψ_2 with a probability density $|\psi_2|^2$ that distributes electrons between the cores. The energy difference between these two states is the origin of formation of the gap at the Brillouin zone edge. On the other hand, away from the zone edge, the electron wave functions can be expressed as

$$\psi(x) \approx L^{-1/2} \mathrm{e}^{\pm \mathrm{i}kx} \tag{6.28}$$

which are propagating waves that characterize the wavelike behavior of free electrons [2, 15].

When all the Brillouin zones and their associated Fourier components are included, the result is a set of curves, as those shown in Fig. 6.11a. An easier way to show this is to use the Kronig–Penney model, first formulated in 1931, in which the potential is assumed to be a square-well array [2, 9]. The details are left as an exercise (see Problem 6.12). The allowable bands are illustrated by the solid curves in Fig. 6.11. If the electrons were completely free, then $E(k) = E_k^0 = \hbar^2 k^2/(2m_\mathrm{e})$ would be a parabola, as illustrated by the dashed curve in Fig. 6.11a, without any bandgap. It is useful to plot all the energy levels in the first Brillouin zone. This can be done by folding the branches in Fig. 6.11a, which is known as the *extended-zone scheme*, using the reciprocal lattice vector. The result is shown in Fig. 6.11b, which is called the *reduced-zone scheme* for the representation of the electronic bands.

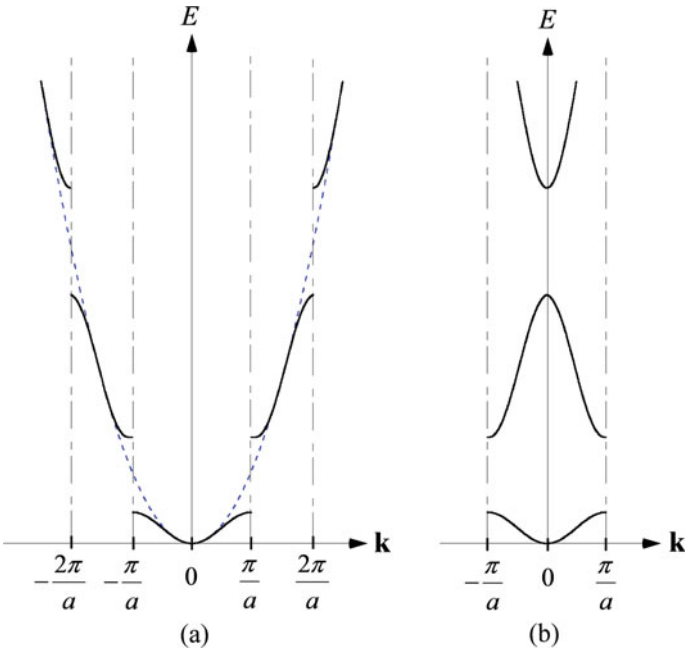

Fig. 6.11 Representation of the electronic band structure. **a** The extended-zone scheme. **b** The reduced-zone scheme

6.4.3 Band Structures of Metals and Semiconductors

The nearly free-electron model described in the previous section assumes a weak potential and cannot predict the behavior of electrons in the inner orbitals or near the nuclei. A simple way to calculate the electronic structure of inner electrons, such as those in the *d* subshells, is the *tight-binding method*, which assumes that the potential is so large that electrons can hardly move out of the ion core. Due to the complicated 3D structure and the multiple numbers of outermost electrons in each atom, the actual electronic band structures are rather complicated. More advanced methods include the augmented plane wave (APW) method, the Korringa–Kohn–Rostoker (KKR) Green function method, and the pseudopotential method. More details can be found from Ashcroft and Mermin [1], Kittel [2], and Omar [15], and references therein.

It can be shown that the number of orbitals in a band in the first Brillouin zone is the same as the number of unit cells in the crystal, *N*. According to the Pauli exclusion principle, the number of electrons that can occupy a band is 2*N*. For copper, the outermost electron configuration is $4s^1 3d^{10}$. The *s*- and *d*-subshell electrons result in six bands (with some overlap), as can be seen from Fig. 6.12, along the direction according to the first Brillouin zone depicted in Fig. 6.8 [16–18]. The *d* bands are from 2 to 5.5 eV below the Fermi level and are completely filled. The *s* band, illustrated by the thicker line segments, is interrupted by the *d* bands. The *s* band is only half filled

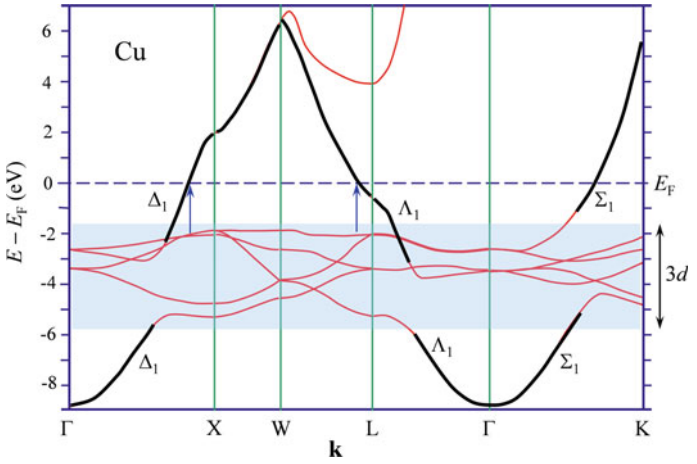

Fig. 6.12 Calculated energy band structure of copper, *adapted from Refs.* [16–18] *with permission*

and half empty. For alkali metals, there is only one valence electron and the *s* band is continuous. Electrons in the *s* band can be easily excited from below the Fermi level to above the Fermi level within the same band. This explains why copper is a conductor. When radiation is incident on a copper surface, because of the relatively high frequency, free electrons have an inductive characteristic and tend to reflect the radiation. The absorption of photons will cause the electrons in the *s* band to reach a higher level within the same band. If the phonon energy exceeds 2 eV, transition from the top *d* band to the *s* band right above the Fermi level is possible, as indicated by the two arrows in Fig. 6.12. The *interband transitions* result in strong absorption as well as a reduction in reflection of copper at wavelengths shorter than about 0.6 μm. Pure copper has a red-brown color because it does not reflect blue and violet colors. Gold has a similar interband transition that absorbs short-wavelength visible light. On the other hand, for silver, the interband transition occurs at a much shorter wavelength. Thus, silver can reflect light in the whole visible spectrum.

The Fermi surface is anisotropic and not spherical for real crystals. For alkali metals with bcc lattices, such as Na and K, the Fermi surface is nearly spherical lying inside the first Brillouin zone [1]. The Fermi surface of Al is close to the free-electron surface for an fcc lattice with three conduction electrons per atom. For noble metals, due to the effect of *d* bands, the Fermi surface is characterized by a sphere that bulges out in the eight <111> directions.

The electronic band structures of Si and GaAs in the first Brillouin zone are shown in Fig. 6.13, along reciprocal lattice directions [19–21]. Si and GaAs are chosen here because these two types of semiconductors have distinct energy gap features that can represent a wide range of semiconductor materials. Degeneracy causes additional subbands within the conduction and valence bands. *Intraband transitions* refer to the excitation or relaxation of electrons between subbands. For intrinsic semiconductors, the Fermi level lies right in the middle between the bottom of the conduction band

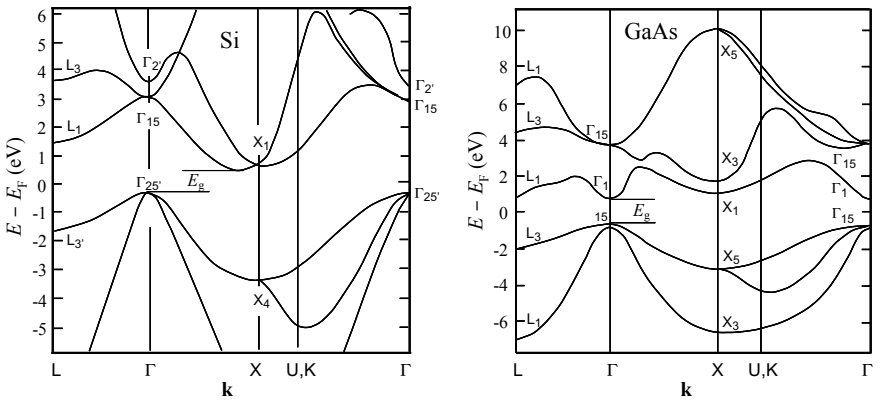

Fig. 6.13 Calculated energy band structure of silicon (left) and gallium arsenide (right), *adapted from Refs.* [19–21] *with permission*

and the top of the valence band. The valence bands are formed by the bonded valence electrons, and they are completely filled at low temperatures. The electrons in the conduction band are dissociated from the atom and hence become free charges. The bandgap energy, or energy gap, E_g is the difference between the energies at the top of the valence band (E_V) and the bottom of the conduction band (E_C). The values of E_g for some semiconductors are included in Table 6.2. For Si, as shown in Fig. 6.13a, the bottom of the conduction band and the top of the valence band do not occur at the same k. This type of semiconductor is called an *indirect gap* semiconductor. For a *direct gap* semiconductor, such as GaAs, the bottom of the conduction band and the top of the valence band occur at the same value of k at the Γ-point, as shown in Fig. 6.13b. The mechanism for electron transition between the valence band and the conduction band in a direct gap semiconductor is different from that in an indirect gap semiconductor. Additional discussion about radiation absorption processes will be given in Chap. 8.

At absolute zero temperature, there are no electrons in the conduction band and the valence band is completely filled. When the temperature increases or there exist optical excitations, electrons in the valence band can transit to the conduction band, leaving behind some vacancies in the valence band. The vacancies left in the valence band are called *holes*, which carry opposite charge as electrons. Usually the electrons are found almost exclusively in levels near the conduction band minima, while the holes are found in the neighborhood of the valence band maxima. Therefore, the energy versus wavevector relations for the carriers can generally be approximated by quadratic forms in the neighborhood of such extrema, i.e.,

$$E_e(k) = E_C + \frac{\hbar^2 k^2}{2m_e^*} \quad \text{and} \quad E_h(k) = E_V - \frac{\hbar^2 k^2}{2m_h^*} \tag{6.29}$$

where subscript e and h are for electrons and holes, respectively, E_C is the energy at the bottom of the conduction band and E_V is the energy at the top of the valence band. In the 1D case, the effective mass m^* for electrons and holes is defined as

$$\frac{1}{m_e^*} = \frac{1}{\hbar^2}\frac{d^2E_e}{dk^2} \quad \text{and} \quad \frac{1}{m_h^*} = -\frac{1}{\hbar^2}\frac{d^2E_h}{dk^2} \tag{6.30}$$

where the negative sign is assigned to make the effective mass of the hole positive at the top of the valance band. Effective mass is defined based on the quantum mechanical description of the group velocity and the acceleration of charge carriers, respectively, as

$$v_g = \frac{1}{\hbar}\frac{\partial E}{\partial k} \quad \text{and} \quad a = \frac{dv_g}{dt} = \frac{1}{\hbar}\frac{\partial^2 E}{\partial k^2}\frac{dk}{dt} = \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k^2}F \tag{6.31}$$

where $F = \frac{dp_e}{dt} = \hbar\frac{dk}{dt}$ is the force exerted on the charge carrier due to an electric field. In 3D case, the effective mass depends also on the direction and is a 3×3 tensor [15]. Note that the above definition of effective mass is for parabolic bands only according to Eq. (6.29) and hence does not apply to 2D solids such as graphene, to be discussed next.

6.4.4 Electronic Properties of Graphene

As a layered 2D material with carbon atoms arranged in a honeycomb lattice, as shown in Fig. 6.7a, graphene has unique electronic, mechanical, thermal, and optical properties. Due to its large carrier mobility and electrical conductivity, along with the feasibility of controlling the carrier density by a gate voltage, graphene is a promising material for the next generation of transistors and 2D flexible nanoelectronics [22, 23]. Graphene can be synthesized chemically (e.g., by chemical vapor deposition on a metal surface) or isolated using mechanical or liquid-phase exfoliation from graphite [24]. As discussed in Chap. 5, the thermal conductivity of graphene can be as high as or even higher than that of diamond [25]. Graphene and related materials also hold promise for energy conversion and storage [26]. In addition, graphene exhibits unique optical and infrared properties [27] for optoelectronics and photonics applications (to be discussed in Chap. 9). Knowledge of the electronic structure of carbon and its related materials is critical for understanding the unique electronic and other properties of graphene.

A carbon atom has six electrons configured as ^{6}C: $1s^2 2s^2 2p^2$. There are four electrons in the second shell, two in the s-orbital and two in p-orbitals. However, this is merely the ground-state configuration without excitation. Note that the electron cloud for the s-orbital is isotropic or spherical shaped. Each of the three p-orbitals shapes like a dumbbell (or the number 8), identified as p_x, p_y, and p_z with the direction along the orbital axis. In methane (CH_4), a carbon atom is bonded to four hydrogen atoms,

forming a tetrahedral molecular geometry (referring to Fig. 3.9b). The hydrogen atoms are at the vertices of the regular tetrahedron while the carbon atom is at the centroid. All four bonds between the C and H are equally spaced with equal strength measured by the bond energy. The underlying mechanism can be explained by orbital hybridization, which may be explained in two steps. Firstly, one electron is promoted from the $2s$ orbital to the $2p$ orbital due to excitation, such that each of the orbital in the outmost shell of carbon $(s,\ p_x,\ p_y,$ and $p_z)$ is occupied by one electron; secondly, the four electron orbitals combine and rearrange themselves so that each contains 25% of s, p_x, p_y, and p_z components. The hybrid orbitals look like asymmetric dumbbell, and each points out in one of the tetrahedral directions to form a chemical bond with an H atom. This is called the sp^3 hybridization and each hybrid orbital possesses 25% s-orbital and 75% p-orbital characteristics. The orbital hybridization theory was originally developed in 1931 by Linus Carl Pauling, who received the Nobel Prize in Chemistry in 1954. The sp^3 hybridization is also responsible to the diamond (or crystalline silicon) structure where each carbon (or silicon) atom is bonded to four other carbon (silicon) atoms.

In an ethylene (or ethene) molecule (C_2H_4), as shown in Fig. 6.14a, each carbon atom is bonded to two hydrogen atoms and the carbon atoms form a double bond with each other. In this case, the s-orbital electron is hybridized with the electrons in the p_x and p_y orbital to form a sp^2 hybridization (with 1/3 s-orbital component). The hybrid electrons form a sigma (σ) bond with each hydrogen ($1s$ electron) as well as between the carbon atoms. All six atoms lie in a plane and the angle between H–C–H bonds is close to 120°. Nevertheless, for each carbon atom, there is one lonely electron in the p_z orbital whose axis is perpendicular to the plane. The two lonely electrons form a pi (π) bond to share the orbital. In essence, the double bond between the carbon atoms contains a σ bond and a π bond with very different characteristics. Note that σ bonds are the strongest type of covalence bond. In a σ bond, the atomic orbitals overlap with each other in a head-on position, so that their orbitals are symmetrical with respect to rotation about the bond axis. On the other hand, in a π bond, the orbital axes are perpendicular to the bond axis between the two atoms. In acetylene (or ethyne) molecule (C_2H_2), all four atoms are aligned in the x-direction, as shown in Fig. 6.14b. The electrons in the s and p_x orbitals form sp hybridization, which is responsible for the strong σ bond between C and H as well as between C and C atoms. Furthermore, the electrons in the p_y (or p_z) orbital of each carbon atom form a π bond whose orbital axis is in the y (or z) direction. The triple bond between the carbon atoms consists of one σ bond and a pair of π bonds; the latter is much weaker than the former.

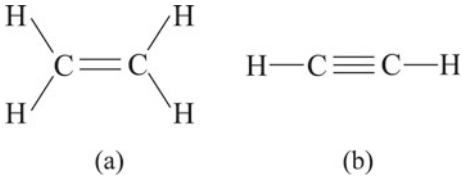

Fig. 6.14 The chemical structure of **a** ethylene and **b** acetylene

The sp^2 hybridization is also responsible to the carbon bonds in graphene, where each carbon is bonded to three neighborhood carbon atoms via σ bonds. These electrons are also called σ electrons. The unhybridized electron in the p_z orbital forms a π bond with another carbon atom's p_z orbital. In a way, the carbon–carbon bonding in graphene contains alternating single and double bonds, forming a conjugated system like benzene. In contrary to benzene, the π electrons in graphene are shared by the atoms and are highly mobile along the graphene sheet, like free electrons in metal. What is more, there are some unique properties of graphene that can only be explained quantum mechanically, with the help of the electron band structures.

Figure 6.15 shows the band structure of graphene calculated in the first Brillouin zone based on the density-functional theory (DFT), when the Fermi energy is set to zero [28, 29]. The bands below (or above) the Fermi level are completely filled (or empty). The filled bands, called π or σ bands, are associated with electrons in the π and σ bonds. The unfilled bands are associated with π^* or σ^* antibonding orbitals. It can be seen that the σ and σ^* bands are spaced far away from the Fermi level. The transition from $\pi \rightarrow \pi^*$ is responsible for nearly all electronic and optical properties of graphene, except with high-energy excitations (> 5 eV) such as irradiation by photons in the deep ultraviolet and x-ray regions [27]. The most striking feature of 2D graphene is the gapless feature of the bands at the K point (one of the six Dirac points), as indicated by the dashed box. Furthermore, the π and π^* bands are conical as shown in Fig. 6.16. For 3D semiconductor materials, there is a bandgap between the conduction band and the valence band and the band structures are parabolic in a 2D diagram or parabola in a 3D diagram.

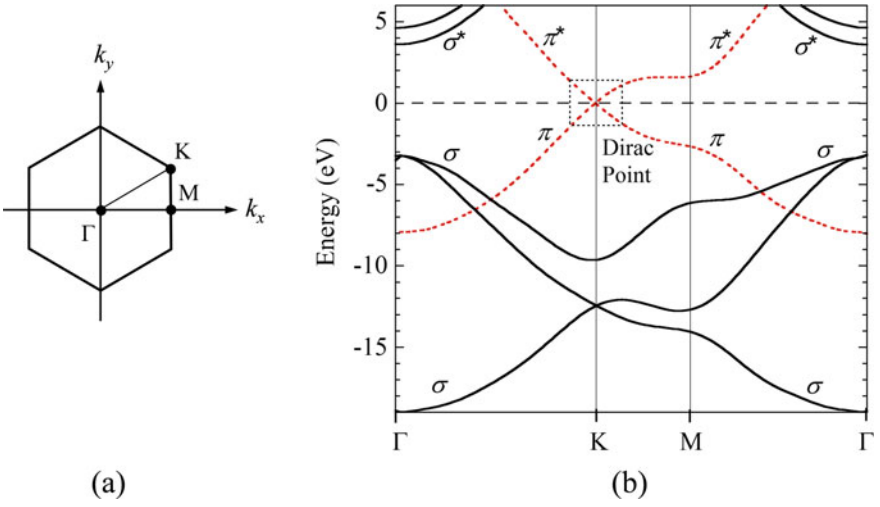

Fig. 6.15 **a** First Brillouin zone of graphene; **b** band structure of a graphene sheet, where the dashed curves are for π or $\pi*$ bands, and solid curves are for σ or $\sigma*$ bands. *Adapted from* [29] *with permission of American Physical Society*

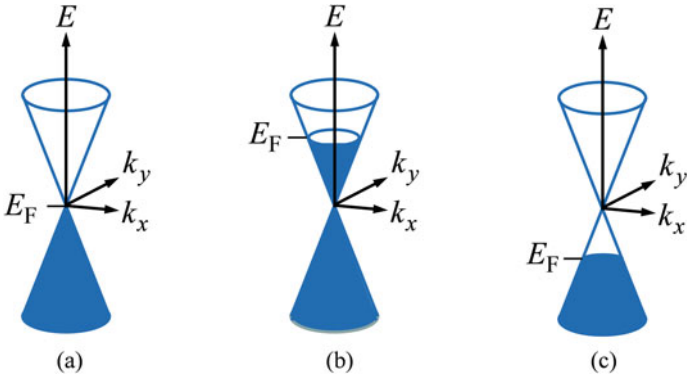

Fig. 6.16 Band shape near the Dirac point with different Fermi levels: **a** $E_F = 0$; **b** $E_F > 0$; **c** $E_F < 0$

According to the tight-binding model with some approximation [22], the bands can be described by

$$E = \pm\gamma_0\sqrt{3 + 2\cos(K_y a) + 4\cos(\sqrt{3}K_x a/2)\cos(K_y a/2)} \tag{6.32}$$

where $E = 0$ corresponds to the Dirac point, γ_0 is a constant about 2.8 eV, and $a \approx 0.246$ nm is the magnitude of the chiral vector defined in Fig. 6.7a.

Equation (6.32) predicts conical band structures at the six Dirac points when the energy is within about 1 eV from the apex. For example, at the K point, $K_x = 2\pi/(\sqrt{3}a)$ and $K_y = 2\pi/(3a)$. This can be schematically shown in Fig. 6.16 with different Fermi levels. It can be seen that

$$E = \pm v_F\hbar\sqrt{k_x^2 + k_y^2} = \pm v_F\hbar k \tag{6.33}$$

Here, $v_F \approx 1 \times 10^6$ m/s is the Fermi velocity, and the wavevector **k** at the K point is set to zero. In 1928, Paul Dirac derived a relativistic wave equation that modifies the Schrödinger equation and can be applied to massless particles. According to Eq. (3.122) as discussed in Sect. 3.7, $E^2 = m^2c^4 + p^2c^2$ for a particle with mass m and momentum p that travels at the speed c. Therefore, for a massless particle, E is the product of the momentum and speed. Note that in Eq. (6.33), $\hbar k$ is the momentum of the electron and v_F is essentially a constant. Therefore, it can be said that the electrons in pure graphene at low temperatures ($E_F = 0$) are massless Dirac fermions. Gating with a voltage across the graphene sheet allows the Fermi level to be changed, as shown in Fig. 6.16b, c, similar to chemically doping with a small amount of impurities. The cyclotron effective mass is given as [22, 30].

$$m_c = \hbar k_F/v_F = |E_F|/v_F^2 \tag{6.34}$$

which has been experimentally determined by measuring the cyclotron frequency $\omega_c = eB/m_c$ by applying a magnetic field B. Note that Eq. (6.34) resembles Einstein's equation, $E = mc^2$. Measurements under applied magnetic field have revealed another exotic behavior in graphene: the anomalous (half-integer) quantum Hall effect at room temperature [22, 30].

From the semi-classical Boltzmann theory, the electrical conductivity of 2D graphene (unit: S or Ω^{-1}) can be expressed in terms of the relaxation time (at the Fermi level) [31]:

$$\sigma_{2D} = \frac{e^2 v_F^2}{2} D(E_F)\tau \tag{6.35a}$$

where τ is the relaxation time and D is the density of states (DOS) per unit area. For a relatively pure graphene sheet, the DOS can be approximated as

$$D(E) = \frac{2|E|}{\pi \hbar^2 v_F^2} \tag{6.35b}$$

Substituting Eq. (6.35b) into Eq. (6.35a), we obtain

$$\sigma_{2D} = \frac{e^2 |E_F|}{\pi \hbar^2}\tau \tag{6.36}$$

Furthermore, the carrier concentration n (number per unit surface area) and the wavevector at the Fermi level are related by $k_F = \sqrt{\pi |n|}$ or $|E_F| = \hbar v_F \sqrt{\pi |n|}$, which can be used to estimate the carrier concentration. Equation (6.36) breaks down $E_F \to 0$ or $n \to 0$. Due to the gapless feature, there exists a universal minimum conductivity theoretically given as [22]

$$\sigma_{min} = \frac{4e^2}{\pi h} \approx 4.932 \times 10^{-5}\ \text{S} \tag{6.37}$$

which has been experimentally observed with somewhat higher values due to impurity, size, and other experimental factors [32].

Using the relation $\sigma_{2D} = e|n|\mu$, the electron (or hole) mobility in graphene can be calculated from

$$\mu = \frac{e v_F}{\hbar \sqrt{|n|\pi}}\tau \tag{6.38}$$

where the relaxation time τ is on the order of $10^{-14} - 10^{-12}$ s for impurity concentrations in the range $10^{11} - 10^{12}\text{cm}^{-2}$ [31]. Note that the carrier concentration is different from the impurity concentration and can be tuned either way by a gating voltage. The carrier mobility μ in graphene typically ranges from 2,000 to 20, 000 $\text{cm}^2/\text{V s}$ [32], but could be as high as 200, 000 $\text{cm}^2/\text{V s}$ in suspended graphene, even at room temperature [33].

The success in obtaining single or few layers of graphene and measuring their exotic properties have spurred growing interest in a variety of other 2D or quasi-2D materials, whose properties are dramatically distinctive from their 3D counterparts [34–38]. Examples are hexagonal boron nitride (hBN), transition metal dichalcogenides (TMD) such as MoS_2 and WSe_2, black phosphorus (or phosphorene), layered Bi_2Te_3 and GeSe, etc. The intensive theoretical and experimental investigations of these materials systems expand the current understanding of the electrical, mechanical, optical, and thermal properties of existing materials and may provide pathways for new technologies in novel electronic devices, energy systems, nanophotonics, and biomedical applications [26, 34–39].

6.5 Phonon Dispersion and Scattering

In the above discussion of electronic band structures, it is assumed that the cores of atoms are fixed. In a real crystal, however, the cores of atoms are vibrating about their equilibrium positions and the vibration of atoms has an important influence on energy storage and transport in crystals. Lattice vibration causes elastic waves to propagate in crystalline solids. Phonons are the energy quanta of lattice waves. For a given vibration frequency ω, the energy of a phonon $\hbar\omega$ is the smallest discrete value of energy. Thermal vibrations in crystals are thermally excited phonons, like the thermally excited photons in a blackbody cavity.

6.5.1 The 1D Diatomic Chain

Phonon dispersion describes the relationship between the vibration frequency and the phonon wavevector. A simple example is given first for a diatomic chain of linear spring–mass arrays, as shown in Fig. 6.17. It is assumed that the spring constant K is the same between the nearest-neighbor atoms. The spring is a conceptual representation of the combined attractive and repulsive forces, which can be assumed linear if the displacement is sufficiently small. Anharmonic vibrations may become

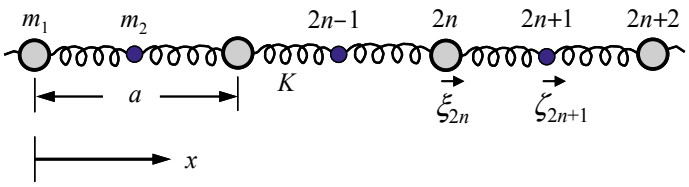

Fig. 6.17 A chain of two atoms with different masses m_1 and m_2 linked by springs of the same spring constant K, where ξ and ζ denote the displacements of individual atoms from their equilibrium positions

significant at high temperatures. Another assumption of the nearest-neighbor model is that the forces on an atom come from the nearest neighbors only [40]. The equation of motion of the atoms can be written as follows:

$$m_1 \frac{d^2 \xi_{2n}}{dt^2} = K(\zeta_{2n+1} + \zeta_{2n-1} - 2\xi_{2n}) \tag{6.39a}$$

and

$$m_2 \frac{d^2 \zeta_{2n+1}}{dt^2} = K(\xi_{2n+2} + \xi_{2n} - 2\zeta_{2n+1}) \tag{6.39b}$$

where ξ_{2n} is the displacement of the atom with mass m_1 indexed by an even number and ζ_{2n+1} is the displacement of the atom with mass m_2 indexed by an odd number [41, 42]. To solve these equations, we substitute the general solutions $\xi_{2n} = A_1 \exp[i(nka - \omega t)]$ and $\zeta_{2n+1} = A_2 \exp[i(n + 1/2)ka - i\omega t]$ into Eq. (6.39a). After some manipulations, we can obtain

$$(2K - m_1\omega^2)A_1 - 2K\cos(ka/2)A_2 = 0 \tag{6.40a}$$

$$(2K - m_2\omega^2)A_2 - 2K\cos(ka/2)A_1 = 0 \tag{6.40b}$$

The determinant of Eq. (6.40a) must be zero, that is,

$$m_1 m_2 \omega^4 - 2K(m_1 + m_2)\omega^2 + 4K^2[1 - \cos^2(ka/2)] = 0 \tag{6.41}$$

The two roots for ω^2 can be expressed as

$$\omega_{1,2}^2 = K\left(\frac{1}{m_1} + \frac{1}{m_2}\right) \pm K\left[\left(\frac{1}{m_1} + \frac{1}{m_2}\right)^2 - \frac{4\sin^2(ka/2)}{m_1 m_2}\right]^{1/2} \tag{6.42}$$

The resulting $\omega - k$ curves are the dispersion relations, as shown in Fig. 6.18. Two branches are formed when $m_1 \neq m_2$. The upper branch that corresponds to the plus sign is called the *optical phonon branch*, or simply *optical branch*, because it is important for infrared activities in ionic solids. The lower branch that corresponds to the minus sign is called the *acoustic branch*. At very low frequencies, the atoms in the unit cell move in phase with each other. Such a behavior is characteristic for a sound wave.

It can be seen that the dispersion curves vary periodically with k. The results outside the first Brillouin zone merely reproduce lattice dynamics that can be fully described by the dispersion curves in the first Brillouin zone. Due to the periodicity of the solution in terms of k, we may treat a value of k outside the first Brillouin zone by subtracting an appropriate integer times the reciprocal lattice constant $2\pi/a$ to give a value of k within the limits of the first Brillouin zone. Given that $|k| \leq \pi/a$,

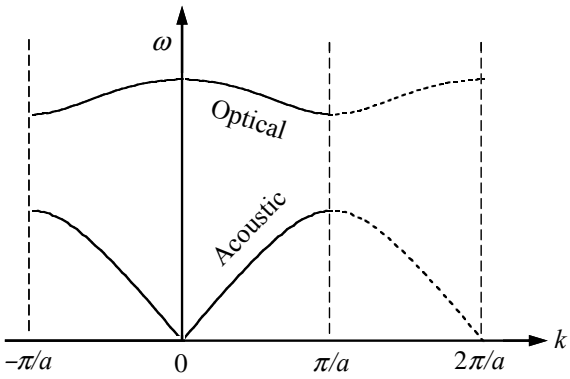

Fig. 6.18 Phonon dispersion of the linear diatomic chain, calculated by the nearest-neighbor model. The first Brillouin zone is between $-\pi/a$ and π/a

the phonon wavelength is specified by

$$\lambda = \frac{2\pi}{k}, 2a \leq \lambda < \infty \tag{6.43}$$

This makes perfect sense as the wavelength should not be smaller than the lattice constants, as explained in previous chapter; see Fig. 5.3. For solids with small dimensions, there is also a limit of the maximum wavelength $2L$. For $k << \pi/a$, the acoustic branch gives $\omega \propto k$, which is a linear dispersion relation. At $k = \pi/a$, $\omega = \sqrt{2K/m_1}$ and $\sqrt{2K/m_2}$, and the two branches are separated when $m_1 \neq m_2$. In this case, it should be noticed that the group velocity $v_g \equiv \mathrm{d}\omega/\mathrm{d}k = 0$. Only standing waves exist. If $m_1 = m_2$, then the upper and lower branches will be continuous at $k = \pi/a$ and the slope is not zero. However, the lattice constant needs to be modified to $a/2$ in Fig. 6.17, and thus the range of the first Brillouin zone is between $-2\pi/a$ and $2\pi/a$. If the upper branch is not folded at $k = \pi/a$, it will connect smoothly with the lower branch and extend to $2\pi/a$. Detailed discussion can be found from Ref. [41].

6.5.2 *Dispersion Relations for Real Crystals*

The above discussion can be extended to 3D systems, in which lattice vibrations allow both transverse and longitudinal modes. For the case of two atoms per primitive cell, there are one longitudinal and two transverse branches for both acoustic and optical vibration modes. The phonon dispersion relations for silicon and silicon carbide are shown in Fig. 6.19 [43–46]. Experimental determination of the phonon dispersion curves was made with neutron scattering [43, 44] for Si and Raman scattering for SiC [46]. Because $m_1 = m_2$ for Si, the longitudinal optical (LO) and longitudinal acoustic (LA) branches meet at the zone edge and thus the group velocity is not equal to zero there. For SiC, on the other hand, the two roots in Eq. (6.42) are different

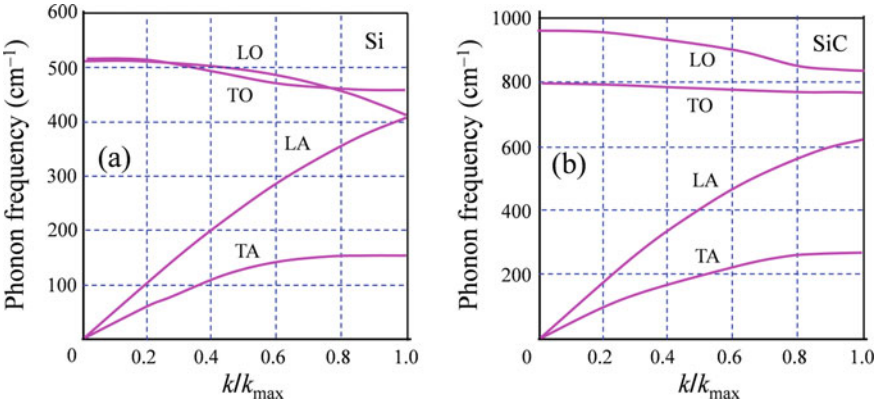

Fig. 6.19 Optical and acoustical branches of phonon dispersion. **a** Si [101] direction. *Adapted with permission from Refs.* [43–45]. **b** SiC. *Adapted with permission from Ref.* [46]

because $m_1 \neq m_2$. There exists a frequency gap between the LO and LA branches at the zone edge. The frequency gap is forbidden for propagating waves, i.e., no phonons can propagate at frequencies within the gap, similar to the bandgap for electrons. The group velocities of LO and LA phonon modes are zero at the zone edge; this can be seen by the flat dispersion curves. One should not worry about the negative or positive sign of the group velocity as it is merely a result of folding the dispersion curves. The group velocity is always in the direction of energy transfer. It should be mentioned that the speed of sound and the phonon propagation speed refer to the group velocity, not to the phase velocity.

According to the wave–particle duality, a phonon with energy $\hbar\omega$ should also have an associated momentum, given by

$$\mathbf{p} = \hbar\mathbf{k} \tag{6.44}$$

where $\mathbf{k}$ is the wavevector of the phonon. There is a distinction between phonons and photons. Phonons do not carry any physical momentum because the physical momentum associated with lattice vibration is zero, except when all lattices are in phase. On the other hand, when interacting with other elementary particles, such as electrons or photons, the wavevector must follow the selection rule such that it looks as if a phonon has a real momentum given by Eq. (6.44). This momentum is often called the *crystal momentum* [1, 2].

The group velocity of phonons in the optical branches is usually small, and subsequently optical phonons contribute little to the thermal conduction in solids. On the other hand, optical phonons can interact or scatter with acoustic phonons, especially at elevated temperatures, to reduce the thermal conductivity [41]. Although LA phonons have higher group velocities than TA phonons, one must also consider the frequency distribution of phonons since phonons obey Bose–Einstein statistics; see Eq. (5.77) and discussions in Chap. 5. At low temperatures, the TA phonon mode

is the dominant contributor to both the thermal conductivity and the specific heat of insulators and semiconductors. As the temperature goes up, LA phonons also play a significant role. While optical phonons contribute little to the heat conduction, they contribute about half of the heat capacity above room temperature. This is because group velocity does not enter the equation for specific heat; refer to Eq. (5.30). In general, if there are q atoms in the primitive cell or basis, there will be one longitudinal and two transverse acoustic branches, and $q-1$ longitudinal and $2(q-1)$ transverse optical branches. However, degeneracy of the transverse branches may occur due to symmetry [40–42]. An example of complex materials is the family of zeolites, which are hydrated aluminosilicate minerals that exhibit nanoporous crystalline structures. Zeolites have important applications as filters, catalysts, solar collector, and adsorption refrigeration. Greenstein et al. [47] studied the thermal properties of MFI zeolite films considering phonon dispersion. MFI is a special type of zeolite that has ordered channel directions and an average pore size of 0.6 nm. The calculation of specific heat and thermal conductivity involved summation over 864 polarizations (phonon branches) over all wavevectors in the first Brillouin zone. The modeling results were in reasonable agreement with experiments [47].

In recent years, lots of studies have been done on the phonon transport in graphene and other 2D materials, as well as the interface between 2D materials and substrates [25, 36, 48]. For example, recent measurement and simulation have shown that a thermal conductivity in monolayer of hBN can be as high as 750 W/m K [37].

Recently, researchers have demonstrated experimentally very high thermal conductivity (900–1300 W/m K) in boron arsenide (BAs) crystals at room temperature; the result verified the previous predictions by first-principles simulations in 2013 and 2017; see Ref. [49, 50] and works cited therein. Cubic BAs has since replaced cubic BN (around 740 W/m K) to become the bulk material with the second highest (next to diamond) thermal conductivity at room temperature [49].

Another important aspect of phonon transport is scattering. The mean free path of phonons is often small compared with the size of crystals. For nanostructures, on the contrary, the mean free path can be larger than the characteristic length, resulting in boundary scattering. Some qualitative discussions have been given in the previous chapter. A summary of the characteristics of phonon and photon is given in Table 6.3. In most situations, phonons are treated as particles, especially in dealing with interactions among phonons themselves as well as with electrons, photons, and defects. For long-wavelength phonons, lattice vibration can also be described by a sound wave or an acoustic wave of three polarizations. To analyze the acoustic wave behavior, the crystal is viewed as a continuous medium because the individual vibration of atoms is not of interest. A brief discussion of the microscopic conservation (or selection rules) during scattering events involving phonons and/or electrons is presented next.

Table 6.3 Comparison of the characteristics of phonon and photon

Phonon	Photon
Bose–Einstein statistics	Bose–Einstein statistics
Massless	Massless
Energy: $\varepsilon = h\nu$	Energy: $\varepsilon = h\nu$
Phase speed: $v_p = \omega/k$	Phase speed: $v_p = \lambda\nu$
Mechanical vibration (existence in solids and some liquids, such as liquid helium)	Electromagnetic waves (existence in any medium as well as in vacuum)
Both transverse and longitudinal	Transverse only
Crystal momentum: $\mathbf{p} = \hbar\mathbf{k}$	Physical momentum: $\mathbf{p} = \hbar\mathbf{k}$
Frequency: less than $\approx$ 50 THz	Frequency: no limit
Group velocity: $<\approx 2 \times 10^4$ m/s	Group velocity: order of 10^8 m/s
Mean free path: $\approx$ 10 to 100 nm (except at very low temperatures and in nanotubes)	Mean free path: no limit (largely dependent on the medium)

6.5.3 *Scattering Mechanisms*

Phonon scattering governs the thermal transport properties of dielectric and semiconductor materials. Proper modeling of phonon scattering is important for the application of the Boltzmann transport equation (BTE) or Peierls–Boltzmann equation, considering the frequency-dependent scattering rate. The anharmonic nature of the interatomic potential offers a coupling mechanism for phonon–phonon interactions, which was not included in the linear oscillator model. The phonon–phonon scattering is inelastic because the phonon frequency before the scattering event is different from that after the event. The energy conservation requires the scattering to involve at least three phonons. A three-phonon process is mostly common since the probability is usually much larger than the values for processes involving four or more phonons. In a three-phonon process, either two phonons interact to form a third one or one phonon breaks into two others. The phonon energy and crystal momentum are conserved as given by [1, 2]

$$\hbar\omega_1 + \hbar\omega_2 = \hbar\omega_3 \quad \text{or} \quad \hbar\omega_1 = \hbar\omega_2 + \hbar\omega_3 \tag{6.45}$$

$$\hbar\mathbf{k}_1 + \hbar\mathbf{k}_2 = \hbar\mathbf{k}_3 \quad \text{or} \quad \hbar\mathbf{k}_1 = \hbar\mathbf{k}_2 + \hbar\mathbf{k}_3 \tag{6.46}$$

In Eqs. (6.45) and (6.46), the left-hand side terms are for phonon(s) before scattering and the right-hand side terms are for phonon(s) after scattering. The processes just described are called *normal (or N) processes*, in which the wavevectors of phonons are inside the first Brillouin zone. Since both the energy and the momentum are conserved, *N*-processes do not alter the direction of energy flow. Hence, *N*-processes make no contribution to the thermal resistance and do not affect the thermal conductivity.

Scattering is also permitted when two phonons interact to form a third one, whose wavevector is outside the Brillouin zone. This can be understood by the equivalence of phonons with the same energy but with different wavevectors $\mathbf{k}'$ and $\mathbf{k}$ that follow the relationship:

$$\mathbf{k}' = \mathbf{k} + \mathbf{G} \tag{6.47}$$

where $\mathbf{G}$ is a reciprocal lattice vector. The reverse process is also possible with the assistance of $\mathbf{G}$ so that one phonon is annihilated to create two others. The momentum relations given in Eq. (6.46) need to be modified as follows after dropping $\hbar$ in all terms:

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3 + \mathbf{G} \quad \text{or} \quad \mathbf{k}_1 + \mathbf{G} = \mathbf{k}_2 + \mathbf{k}_3 \tag{6.48}$$

These equations, combined with the energy conservation described by Eq. (6.45), describe *the umklapp (or U) processes*. The net momentum is not conserved in the *U*-processes, which introduce thermal resistance and thus reduce the thermal conductivity. Figure 6.20 schematically shows the relationship between the wavevectors for an *N*-process and a *U*-process. An *N*-process can be viewed as the general case of a *U*-process when $\mathbf{G} = 0$.

Above room temperature, *U*-processes dominate and the thermal conductivity decreases linearly as temperature increases. This is because the scattering rate $\gamma = 1/\tau$ between acoustic phonons due to *U*-processes can be described by [41]

$$\gamma_{\mathrm{U}} = \left(A\omega + B\omega^2\right)T \tag{6.49}$$

where A and B are positive constants. When the temperature is reduced, the *U*-process becomes weaker because of the shift in phonon distribution function toward longer wavelengths. Scattering of phonons by defects becomes important. As shown in Fig. 5.13, as the temperature is decreased below room temperature, the thermal conductivity increases to a maximum and then decreases due to the reduction in

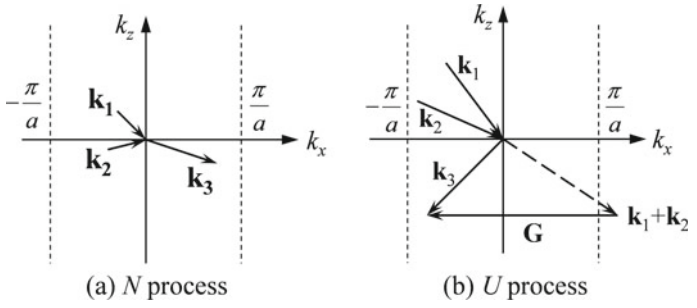

Fig. 6.20 Schematic illustrations of phonon–phonon scattering processes: **a** the *N*-process and **b** the *U*-process

the specific heat. Four-phonon processes are also possible. Four-phonon scattering includes the annihilation of two phonons to create two others, the annihilation of one phonon to create three others, and the annihilation of three phonons to create another. The calculation of the probability of scattering is more involved [41]. Ecsedy and Klemens [51] estimated the scattering rate due to four-phonon processes to be

$$\gamma_{\text{Four}} \propto \omega^2 T^2 \tag{6.50}$$

Their simplified model also suggests that the probability of four-phonon processes in the temperature range from 300 to 1000 K is negligibly small compared with the three-phonon *U*-processes [41, 51]. Recent first-principles calculations by Feng et al. [50] have shown that the four-phonon process cannot be neglected even for common materials like silicon and diamond above room temperature. At 1000 K, the inclusion of four-phonon scattering could reduce the predicted κ by 30%, resulting in excellent agreement with the experimental value. Furthermore, the predicted κ value for single-crystal BAs with and without considering four-phonon processes is 2200 W/m K and 1400 W/m K, respectively [50]. Later, the value around 1000 W/m K was experimentally demonstrated by several groups in 2018 [49]. According to the study by Feng et al. [50], the four-phonon scattering rate should scale with

$$\gamma_{\text{Four}} \propto \omega^4 T^2 \tag{6.51}$$

In addition to phonon–phonon interactions, phonons may also interact with defects (such as impurities, vacancies, or dislocations) and boundaries. These scattering processes can also influence the mean free path of phonons. Scattering of phonons by defects is elastic since the phonon energy remains the same. At temperatures near and above the Debye temperature, phonon–phonon interactions are dominant. As the temperature drops, the dominant wavelengths of phonons become comparable to the size of defects; therefore, scattering of phonons by defects becomes important. The scattering rate for phonon-defect scattering is independent of temperature but dependent on the phonon wavelength. This can be modeled using the Rayleigh scattering theory for small particles such that the scattering rate due to defects is inversely proportional to the fourth power of the phonon wavelength λ, viz.,

$$\gamma_{\text{ph-d}} \propto \lambda^{-4} \quad \text{or} \quad \omega^4 \tag{6.52}$$

When the bulk mean free path is comparable or greater than the characteristic dimension, such as the thickness of the film or the diameter of the wire, scattering of phonons by boundaries becomes important. Boundary scattering is important for nanostructure materials and at low temperatures when the phonon mean free path is large, as extensively discussed in the previous chapter.

In metals and semiconductors, electronic transport becomes important. The scattering of charge carriers controls the electric conduction in solids and dominates the thermal conduction in metals. Carrier–carrier inelastic scattering is negligible except for highly conductive materials, such as a high-temperature superconductor. Since

lattice vibrations are enhanced with increasing temperature, electron–phonon scattering usually dominates the scattering process at high temperatures, while at low temperatures, lattice vibrations are weak and defect scattering becomes important. The vibration of lattice ions causes deviations from the perfect periodic lattice and distorts the carrier wave function. This is more easily visualized as the scattering of electrons by phonons. Both the acoustic branch and the optical branch can scatter electrons. Usually, the energy of acoustic phonons can be neglected compared with the electron energy. Therefore, scattering by acoustic phonons is essentially elastic. Scattering by optical phonons is inelastic because the exchange of energy between the carriers and the phonons can be significant. This process facilitates the energy transfer between electrons and phonons, which is associated with Joule heating. For materials with two different atoms per primitive cell, the asymmetric charge distribution in the chemical bond forms a dipole. Scattering by optical phonons in these materials is called *polar scattering*, which can effectively scatter electrons or holes. The energy and momentum conservations for carrier–phonon scattering can be written as

$$E_{\mathrm{f}} = E_{\mathrm{i}} \pm \hbar\omega_{\mathrm{phonon}} \tag{6.53a}$$

and

$$\mathbf{k}_{\mathrm{f}} + \mathbf{G} = \mathbf{k}_{\mathrm{i}} \pm \mathbf{k}_{\mathrm{phonon}} \tag{6.53b}$$

where subscripts i and f indicate the initial and final states of the carrier, the minus sign corresponds to phonon emission, and the plus sign corresponds to phonon absorption. The momentum of an electron is similar to that of a phonon and is also referred to as the *crystal momentum*. If $\mathbf{G}$ is set to zero, the process is an *N*-process; otherwise, it is a *U*-process as in phonon–phonon scattering. In semiconductors at low temperatures, only *N*-processes are energized. In metals and semiconductors, the electron–phonon scattering rate typically ranges from 10^{12} to 10^{14} Hz at room temperature. Near or above the Debye temperature, the specific heat is almost a constant and the number of phonons increases linearly with temperature. Hence, the electron–phonon scattering rate is proportional to temperature in metals, resulting in nearly temperature-independent thermal conductivity, while the electrical resistance is proportional to temperature.

An electron or hole in a periodic lattice does not really collide with ions. The transport of free carriers can be viewed as the propagation of a wave in a periodic potential created by the ions. In addition to lattice vibrations, defects or impurities may break the periodicity of the potential or alter its amplitude. Kinetic theory gives the defect scattering rate as

$$\gamma_{\mathrm{e-d}} = n_{\mathrm{d}}\sigma_{\mathrm{d}}v_{\mathrm{e}} \tag{6.54}$$

where n_{d} and σ_d are the defect number density and scattering cross section, respectively, and v_{d} is the average carrier velocity. The scattering cross-section is an effective

area related to the scattering probability and not the actual geometric cross-sectional area. For metals, the electron velocity is the Fermi velocity v_F, which is on the order of 10^6 m/s. For semiconductors, the random velocity of electrons or holes can be calculated by

$$v_{th} = (3k_B T/m^*)^{1/2} \tag{6.55}$$

which is called the *thermal velocity* and is on the order of 10^5 m/s at room temperature.

In semiconductors, the interband transition requires the conservation of both energy and momentum. This can occur by electronic transitions when interacting with the incident radiation. For indirect gap semiconductors, however, the photon itself cannot provide a large enough change in momentum. Therefore, a phonon is either emitted or absorbed for momentum conservation. The energy and momentum conservation equations are, respectively,

$$E_f - E_i = \hbar\omega_{photon} \pm \hbar\omega_{phonon} \tag{6.56a}$$

and

$$\mathbf{k}_f - \mathbf{k}_i = \mathbf{k}_{photon} \pm \mathbf{k}_{phonon} \tag{6.56b}$$

where the plus and minus signs correspond to phonon absorption and emission, respectively. This kind of transition is called the indirect interband transition. For a direct interband transition, there is no need to emit or absorb a phonon and, thus, the last term in both Eqs. (6.56a) and (6.56b) should be dropped out. The interaction of photons with solids will be left to Chap. 8 (Sect. 8.4) for a more detailed discussion about the absorption and emission processes.

In addition to the absorption and the emission, photons may be scattered by phonons, causing a nonlinear effect. There exists inelastic scattering when photons are scattered by phonons, resulting in x-ray scattering, neutron scattering, Raman scattering, and Brillouin scattering. In Raman scattering, the creation (emission) and annihilation (absorption) of a phonon cause a shift in the frequency of the radiation, namely, the Stokes and anti-Stokes shifts, as shown in Fig. 6.21. The energy conservation equations are

$$\hbar\omega_s = \hbar\omega_i - \hbar\omega_{ph}, \text{ for a Stokes shift} \tag{6.57a}$$

and

$$\hbar\omega_s = \hbar\omega_i + \hbar\omega_{ph}, \text{ for an anti-Stokes shift} \tag{6.57b}$$

where subscripts i, s, and ph are for incident photon, scattered photon, and phonon, respectively. Because the interaction involved two photons and one phonon, the

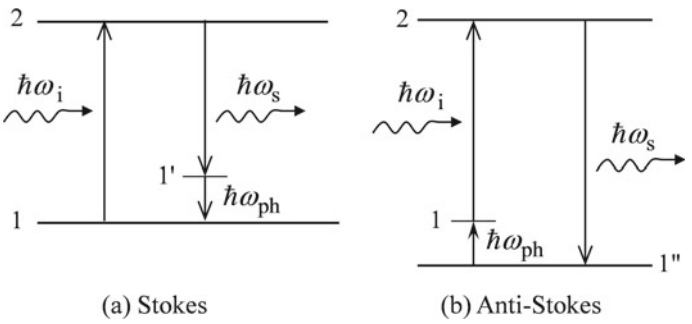

Fig. 6.21 Illustration of Raman scattering and **a** the Stokes and **b** anti-Stokes processes

momentum of the phonon is restricted to small values. The Raman effect, or the Raman scattering, was named after Indian physicist C. V. Raman (1888–1970), who won the Nobel Prize in Physics in 1930 for the discovery. The intensity of the anti-Stokes shift is usually much weaker than that of the Stokes shift. In certain cases, however, the phonons generated by the Stokes process can subsequently participate in the anti-Stokes process, causing a strong excitation to the anti-Stokes component. It is interesting to note that the anti-Stokes component actually pumps energy out from the material, resulting in a radiative cooling effect.

Note that the resulting photon can interact with the phonon again, creating a cascade process that emits m phonons. The photon energy is reduced by m times the energy per phonon. The probability decreases as the order increases. Raman spectroscopy has become a major analytical instrument for the study of solids. High-intensity lasers, high-resolution spectrometers, and sensitive detectors such as photomultiplier tubes (PMTs) are often employed to measure narrow Raman lines. The Raman intensity and intensity ratio depend upon temperature, as illustrated in Fig. 6.22. The ratio of the Raman intensities can be expressed by

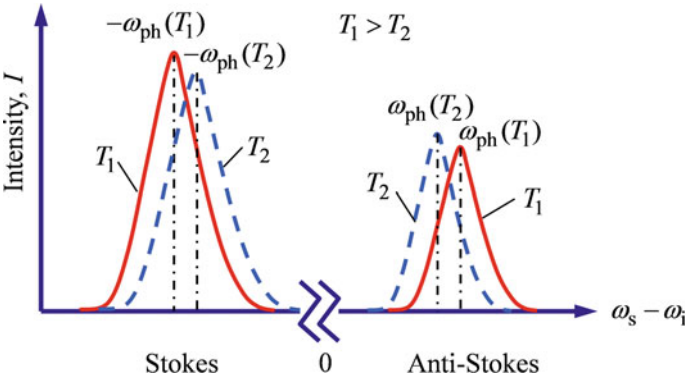

Fig. 6.22 Raman intensity for the Stokes and anti-Stokes scattering at two different temperatures

$$\frac{I_{\text{anti-Stokes}}}{I_{\text{Stokes}}} = \left(\frac{\omega_\text{i} - \omega_\text{ph}}{\omega_\text{i} + \omega_\text{ph}}\right)^2 \exp\left(-\frac{\hbar\omega_\text{ph}}{k_\text{B}T}\right) \tag{6.58}$$

which can be used for surface temperature measurements in microelectronics and microcantilever heaters [52, 53].

Example 6.4 Neutron scattering by phonons is important for measuring the dispersion relations. Express the energy conservation and the momentum conservation during the neutron–phonon scattering in terms of the wavevector and the mass of the neutron, and the wavevector and the frequency of the phonon. Assume the process involves one phonon only.

Solution A neutron has a mass $m_\text{ne} = 1.673 \times 10^{-27}$ kg that is 1834 times that of an electron. Based on the wave–particle duality, the kinetic energy of a neutron can be expressed as $E_\text{ne} = \frac{p^2}{2m_\text{ne}} = \frac{\hbar^2 k^2}{2m_\text{ne}}$; thus, the energy conservation becomes

$$\frac{\hbar^2 k_\text{s}^2}{2m_\text{ne}} = \frac{\hbar^2 k_\text{i}^2}{2m_\text{ne}} \pm \hbar\omega_\text{ph} \tag{6.59}$$

where k_i and k_s are the magnitude of wavevector of the incident and scattered neutrons. The wavevector selection rule gives

$$\mathbf{k}_\text{s} + \mathbf{G} = \mathbf{k}_\text{i} \pm \mathbf{k}_\text{ph} \tag{6.60}$$

These relations characterize the inelastic scattering of neutrons by phonons. The plus and minus signs refer to the process that absorbs or releases phonons, respectively.

6.5.4 Phononics and Coherent Phonons

Phonons are quantized lattice vibration waves. The wavelength of phonons that are important for thermal transport falls in the region from 1 to 10 nm at room temperature and can be longer at cryogenic temperatures. While phonons are typically treated as particles, their wave nature may become important in 1D superlattices and 2D or 3D periodic structures due to interference effects, especially at low temperatures. These effects may enable tailoring the thermal transport properties of semiconductors and insulators, resulting in phonon engineering that belongs to a research branch called *phononics*. Phonon engineering may allow us to control heat transfer in unprecedented manner, such as thermal diode, thermal transistors, thermal memory, topological phonon hall effect, heat cloaking, thermal lens for heat focusing, to name a few [54–56]. The development of graphene and 2D van der Waals materials increases the possibilities and broadens the applications of phonon engineering [25, 57]. Furthermore, with femtosecond to picosecond excitations, phonons may

exhibit coherent nature. Resonance lattice vibrations or coherent phonons have been experimentally observed not only in bulk solids but also in superlattices [58–62].

In a solid crystal, valence electrons are confined and their properties are manifested by the band structure. The electron wave functions can be analyzed based on the Bloch theory. Given the periodicity of typical solids from 0.1 to 1 nm, the crystalline structure also affects the phonon dynamics, resulting in phonon dispersion. Since phonons can extend to longer wavelengths than electrons do, periodic structures with a period of 1–200 nm can affect the phonon band structure (dispersion relations), forming *phononic crystals* that can modify the DOS and group velocity of phonons, thereby reducing the thermal conductivity due to the coherence effect [54, 63]. One of the promising applications is in thermoelectricity where the confinement reduced the thermal conductivity dramatically with little effect on the Seebeck coefficient or electrical conductivity. This is desired to improve the figure of merit of thermoelectric devices as discussed in Chap. 5. It is worth mentioning here that *photonic crystals* with a period from 100 nm to 10 μm can exhibit unique optical properties in the visible and infrared regions, as to be discussed in Chap. 9.

Luckyanova et al. [64] demonstrated coherent phonon transport on thermal conductivity using GaAs/AlAs superlattices fabricated by metal-organic chemical vapor deposition. Each period contains 12 nm GaAs and 12 nm AlAs films, as confirmed by transmission electron microscopic images. The cross-plane thermal conductivities of five samples whose periods are 1, 3, 5, 7, and 9 were measured using a time-domain thermoreflectance (TDTR) technique at temperatures from 30 to 300 K. For periodic layers in the incoherent regime, interface scattering often dominates the thermal resistance, which can be modeled as a combination of the internal and interface resistances. The resulting cross-plane thermal resistance increases with the number of layers, and subsequently the effective thermal conductivity of superlattices with the same period is independent of the number of layers. When interface resistance is small and the phonon mean free path in the bulk material is much longer than the total thickness of the superlattice, the thermal resistance is dominated by boundary resistance at the front and back of the superlattice. This situation resembles ballistic phonon transport in a homogeneous film, when the thermal conductivity decreases linearly as the thickness is reduced as already discussed in Chap. 5. Luckyanova et al. [64] observed a linear dependence of the thermal conductivity at temperatures below 150 K, and thereby demonstrated coherent phonon contributions to thermal conductivity. As the temperature increases, due to phonon scattering at interfaces, the linearity breaks down. The phonon mean free path (Λ) distribution for an infinitely extended superlattice was evaluated using first-principles calculation. It was found that phonons whose mean free path exceed 216 nm (the total thickness of the superlattices used in the experiments) contribute to the thermal conductivity about 87% at 100 K and 71% even at 300 K. Due to the lack of internal scattering and interface scattering, these low-frequency phonons can be described by the dispersion relations based on the superlattice band structures. The results unambiguously demonstrate coherent phonon transport in superlattices even at room temperature [64].

In the incoherence limit, the thermal conductivity of a superlattice decreases as the interface density increases or period decreases in superlattices made of the same

materials with similar interfaces. If the interfaces are made very smooth to avoid scattering, when the period is smaller than the phonon coherence length (such that coherent phonons can play a significant role), the thermal conductivity increases as the period is reduced due to the modification of the phonon dispersion and DOS. The combined effects give rise to a minimum when the thermal conductivity is plotted against the period of the superlattice or the interface density. Such a minimum thermal conductivity separates the particle regime and wave regime of heat conduction by phonons. The predicted minimum thermal conductivity was experimentally demonstrated by Ravichandran et al. [65] using perovskite superlattices made of either $SrTiO_3/CaTiO_3$ or $SrTiO_3/BaTiO_3$ at room temperature. They used molecular epitaxy deposition to grow (001) oxides with atomically sharp interfaces. The layer thickness varies from a single unit cell to 40 nm for $SrTiO_3/CaTiO_3$. For this kind of superlattices with a 50:50 volume ratio between the two materials and a total thickness of 200 nm, a minimum cross-plane thermal conductivity was observed when the period is between 2 and 3 nm. The valley gets deeper when the temperature is lowered from 307 to 142 K and then to 84 K, with a gradual shift toward large period as the temperature is lowered [65]. The experimental results are consistent with theoretical predictions of the crossover between the wave–particle behaviors and the value of the minimum thermal conductivity. The fact that the thermal conductivity can be manipulated by changing the superlattice period may have extensive applications in thermoelectric devices.

Phononic nanomesh structure refers to periodically perforated thin membranes or films, which form a 2D phononic crystal [54–57]. Significant reduction of the in-plane thermal conductivity at room temperature has been observed in Si nanomeshes. Yu et al. [66] fabricated nanomeshes on 22-nm-thick doped epitaxial Si film with a period of a few tens of nanometers and positioned the nanomesh structure between two suspended membranes (one for heating and one for sensing) to measure the thermal conductance at steady state. They reported a nanomesh thermal conductivity near 1.9 W/m K at temperatures from 150 to 280 K, which then decreases to approximately 1.4 W/m K at 100 K. Hopkins et al. [67] fabricated relatively large nanomesh structures with 500-nm-thick single crystalline Si. The period was varied from 500 to 800 nm while the diameter of the holes is 300 or 400 nm. The thermal conductivity of suspended nanomesh structure was measured with a TDTR setup to be as low as 6.8 W/m K, more than an order of magnitude lower than the value of 148 W/m K obtained for bulk Si. While low-frequency coherent or partially coherent phonons with a longer mean free path may play a role in the thermal transport of nanomeshes [68], later theoretical studies [69, 70] and recent experiments [71] suggest that coherent phonons may not play a significant role due to the relatively large surface roughness of microfabricated structures (1–3 nm) as compared with the phonon wavelength at room temperature. Classical phonon scattering including backscattering and native oxide layers may be largely responsible for the thermal conductivity reduction at near room temperature [57, 70, 71].

Nevertheless, in the sub-kelvin temperature range where the dominant thermal phonons are in the gigahertz frequency region, coherent phonon effects on thermal conduction have been observed in suspended silicon-nitride nanomeshes with a

period of 970 nm and 2425 nm, respectively [72]. At such low temperatures, the dominant phonon wavelength is in the micrometer range, much larger than the surface roughness. In this regime, a strong modification of the phonon band structure can affect the thermal transport as demonstrated experimentally [72]. Maire et al. [73] fabricated ordered and disordered 1D and 2D hole arrays with a nominal period of 300 nm on a 145 nm single crystalline Si layer. The Si nanobeam or membrane with a nominal hole diameter of 150 nm was patterned on a silicon-on-insulator wafer and etched into suspended structures. The roughness of the surface was measured with an atomic force microscope (AFM) to be 0.5 nm, while the roughness of the inner surface of the hole was estimated to be 2.5 nm from scanning electron microscopic images. Measurements with a micro-TDTR setup suggested that the thermal decay rates for ordered and disordered samples remain the same at temperatures higher than about 10 K. Below 10 K, especially at 4 K, disorder can significantly increase the decay rate [73]. Coherent phonons with frequencies less than 200 GHz are expected to play a significant role at 4 K, resulting in a decreased decay rate (or thermal conductivity) in the ordered structure. Their results are consistent with coherent acoustic phonon spectroscopy performed with femtosecond pump-probe thermoreflectance experiments [74].

6.6 Atomistic Simulation of Lattice Thermal Properties

In Chap. 5, we extensively discussed the relaxation time approximation in modeling thermal conductivity based on BTE, where the Debye model is often assumed in predicting the phonon DOS. While the results can predict the temperature dependence of thermal conductivity as well as classical size effects, the accuracy is rather limited. Furthermore, the parameters that enter the model are typically obtained by fitting the measurement data. In Sect. 6.5.3, we gave a brief introduction of phonon dispersion and scattering mechanisms. In the last 20 years, especially since the publication of the first edition of this book in 2007, more and more researchers have applied atomistic simulations rooted in quantum mechanics to predict lattice thermal properties for semiconductors, superlattices, and low-dimensional materials (including polymer chains) [75–81]. Classical molecular dynamics (MD) simulations have also been extensively used to predict the thermal conductivity of solid materials based on empirical potentials (refer to Sect. 4.3) using either the linear response theory (i.e., the Green–Kubo relation) or the direct method (i.e., nonequilibrium approach) [82]. Phonon properties can be extracted by analyzing the wave-packet dynamics [82, 83]. Furthermore, the phonon scattering rate or lifetime can also be determined from MD simulations by analyzing the normal modes [83–85]. The interatomic potentials can also be obtained from the first-principles calculations and used in MD simulations [78, 80]. This section overviews the ab initio simulation methodology and how it can be applied to obtain phonon properties and calculate lattice thermal conductivity. The equilibrium and nonequilibrium MD methods will also be introduced.

6.6.1 Interatomic Force Constants (IFCs)

Nowadays, most first-principles calculations for materials properties are based on the density-functional theory (DFT) [86, 87], which has been referred to or cited by over one hundred-thousand papers per year in the last decade according to Google Scholar. While the term "first principles" can have different meanings in different fields, it generally refers to a basic set of postulates or physical laws that do not rely on any other assumptions or fitting parameters that are frequently used in empirical modeling. Therefore, solving the Schrödinger equation to obtain the electronic structure of a crystal under suitable approximations without using parameters obtained from fitting the experimental data is a first-principles approach. Some of the methods were mentioned in Sect. 6.4.3, but they are limited to a few atoms in the unit cell. Rather than solving the many-body Schrödinger equation to obtain the wave functions, DFT treats it as a vibrational problem to obtain the electron densities by minimizing the energy functional using the method of Lagrange multipliers as described in Appendix B.2. The details are rather complicated [86, 87]; thus, only a conceptual description is given next.

The many-electron wave function can be expressed as a product of the single-electron wavefunctions as

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N) = \psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2)\ldots\psi_N(\mathbf{r}_N) \tag{6.61}$$

where N is the number of electrons in the system. Each of the wave function satisfies the Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2 + V_{\text{ext}}(\mathbf{r}) + \Phi_{\text{e}}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}) \tag{6.62}$$

where ψ_i and ε_i are, respectively, the wave function and orbital energy of the ith electron, V_{ext} is the potential due to all nuclei that is determined by the structure and elemental composition of the system, and Φ_e is the potential due to the existence of other electrons (the Coulomb potential). Most solution methods start with a suitable potential V_{ext} to obtain the wave function $\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N)$. From there, all observables can be obtained, and among them, the electron density can be expressed as

$$n(\mathbf{r}) = N\int d\mathbf{r}_2 \int d\mathbf{r}_3 \ldots \int d\mathbf{r}_N \Psi^*\Psi \tag{6.63}$$

Hohenberg and Kohn [86] proved that all ground-state properties can be expressed as functionals of the electron density $n(\mathbf{r})$ and the total energy is such a functional that attains its minimum for the correct ground-state density. In essence, DFT is a variational approach in which the electron density is obtained first and then used to obtain the many-electron wave function, along with the potential. All other observables can be obtained consequently. Compared with traditional first-principles methods, the

computational cost of DFT is relatively low. This advantage becomes more obvious especially with more and more complex systems.

The local-density approximation (LDA) could be used for a slowly varying density so that the ground-state energy can be obtained in terms of single-particle equations for the interacting system [86]. The Kohn–Sham equations provide a means to obtain the electron density by solving N noninteracting one-electron Schrödinger's equations [86, 87]. Before 1990, DFT was not considered as an accurate method in band structure computations [21]. Since then, some approximations have been greatly refined to better model the exchange and correlation potentials [88]. The 1998 Nobel Prize in Chemistry was bestowed on Walter Kohn for his establishment of DFT and John A. Pople for his development of computational methods in quantum chemistry including the implementation of DFT [89]. More recently, DFT has been widely used to predict the chemical, electronic, structural, lattice dynamics, and even magnetic properties of materials from the atomic scale. A number of software packages are available for solid-state simulations based on DFT [88, 90]. It should be noted that the term ab initio (from the beginning) is often used with the same meaning as from first principles, especially in computational chemistry based on quantum mechanics.

DFT can be used to predict phonon properties by using the theory of lattice dynamics [88]. The Born–Oppenheimer approximation assumes that the motion of atomic nuclei can be treated separately from that of electrons. This is the so-called adiabatic approximation, which allows us to decouple the vibrational degrees of freedom from the electronic degrees of freedom [91]. A similar Hamiltonian for a crystal can be written in terms of the interatomic force constants (IFCs) and then used in lattice dynamics to obtain phonon properties.

The density-functional perturbation theory (DFPT) is a linear response theory that aims at obtaining IFCs through a small perturbation of the system from its equilibrium [91]. Suppose $\hat{H}$ is a Hamiltonian and write the Schrödinger equation of the ith particle as follows:

$$\hat{H}\psi_i = \varepsilon_i \psi_i \tag{6.64}$$

Then by applying a small perturbation (parameter ξ), we can write the perturbed Hamiltonian, wave function, and energy of the particle as [92]

$$\hat{H} = \hat{H}^{(0)} + \xi \hat{H}^{(1)} + \xi^2 \hat{H}^{(2)} + \cdots \tag{6.65a}$$

$$\psi_i = \psi_i^{(0)} + \xi \psi_i^{(1)} + \xi^2 \psi_i^{(2)} + \cdots \tag{6.65b}$$

and

$$\varepsilon_i = \varepsilon_i^{(0)} + \xi \varepsilon_i^{(1)} + \xi^2 \varepsilon_i^{(2)} + \cdots \tag{6.65c}$$

By substituting Eq. (6.65a) into Eq. (6.64) and equalizing terms with the same order on both sides, we obtain a series of equations based on the order of perturbation.

The first equation is

$$\hat{H}^{(0)}\psi_i^{(0)} = \varepsilon_i^{(0)}\psi_i^{(0)} \tag{6.66}$$

Apparently, this is for the unperturbed system or the system at equilibrium. The first-order perturbation gives

$$\hat{H}^{(0)}\psi_i^{(1)} + \hat{H}^{(1)}\psi_i^{(0)} = \varepsilon_i^{(0)}\psi_i^{(1)} + \varepsilon_i^{(1)}\psi_i^{(0)} \tag{6.67}$$

and the second-order perturbation gives

$$\hat{H}^{(2)}\psi_i^{(0)} + \hat{H}^{(1)}\psi_i^{(1)} + \hat{H}^{(0)}\psi_i^{(2)} = \varepsilon_i^{(2)}\psi_i^{(0)} + \varepsilon_i^{(1)}\psi_i^{(1)} + \varepsilon_i^{(0)}\psi_i^{(2)} \tag{6.68}$$

The third-order and higher order perturbation relations can also be applied when necessary. It should be noted that

$$\varepsilon_i^{(n)} = \frac{1}{n!}\frac{\mathrm{d}^n\varepsilon_i}{\mathrm{d}\xi^n}\bigg|_{\xi=0} \quad \text{and} \quad \psi_i^{(n)} = \frac{1}{n!}\frac{\mathrm{d}^n\psi_i}{\mathrm{d}\xi^n}\bigg|_{\xi=0} \tag{6.69}$$

The Hellmann and Feynman theorem gives $\varepsilon_i^{(1)}$ in terms of $\psi_i^{(0)}$ and $\hat{H}^{(1)}$, which are supposed to be known. The second-order energy derivative and higher order derivatives can also be obtained [92]. The $(2n+1)$ theorem states that if one knows the wave functions up to order of n, i.e., $\psi_i^{(n)}$, one can deduce the energy derivative up to the order $(2n+1)$, i.e., $\varepsilon_i^{(2n+1)}$. The actual multivariable perturbation problem is very complicated and readers are referred to [91, 92] and references therein.

Let us assume that, based on the elementary composition of the system, the unit cell parameters and equilibrium atomic positions can be obtained from DFT. If the atoms are allowed to have small displacement around their equilibrium position, then the potential energy of the system U can be expressed as a Taylor expansion [80, 81]:

$$U = U_0 + \sum_i\sum_\alpha \Pi_i^\alpha \zeta_i^\alpha + \frac{1}{2!}\sum_{i,j}\sum_{\alpha,\beta}\Phi_{ij}^{\alpha\beta}\zeta_i^\alpha\zeta_j^\beta + \frac{1}{3!}\sum_{i,j,k}\sum_{\alpha,\beta,\gamma}\Psi_{ijk}^{\alpha\beta\gamma}\zeta_i^\alpha\zeta_j^\beta\zeta_k^\gamma + O(\zeta^4) \tag{6.70}$$

Here, $i, j, k \ldots = 1, 2, \ldots N$ are the atom indices, and $\alpha, \beta, \gamma \ldots = 1, 2,$ or 3 are the coordinate indices (x, y, or z in the Cartesian coordinates system). The IFCs are given as follows:

$$\Pi_i^\alpha = \frac{\partial U}{\partial \zeta_i^\alpha} = -F_i^\alpha \tag{6.71}$$

where F_i^α is the force component on the ith atom and it is zero at equilibrium. Thus, the first summation on the right-hand side of Eq. (6.70) disappears. The second-order

derivatives are called the harmonic force constants,

$$\Phi_{ij}^{\alpha\beta} = \frac{\partial^2 U}{\partial \zeta_i^\alpha \partial \zeta_j^\beta} \tag{6.72}$$

This will become clear in the next section. The third-order derivatives are known as the cubic force constants, which are related to the anharmonic processes due to three-phonon interactions:

$$\Psi_{ijk}^{\alpha\beta\gamma} = \frac{\partial^3 U}{\partial \zeta_i^\alpha \partial \zeta_j^\beta \partial \zeta_k^\gamma} \tag{6.73}$$

The next anharmonic constants are related to four-phonon processes [50]. Once the IFCs are obtained using DFPT [78, 91, 92], lattice dynamics can be used to obtain phonon dispersion relations, as well as the wavevectors and wave functions of the normal modes (phonon modes) from harmonic lattice dynamics. The higher order terms have little impact on the phonon dispersion, which determines the group velocity and DOS of phonons. However, the anharmonic terms are related and can be used to obtain phonon–phonon scattering rates, which are temperature dependent. Besides DFPT, other methods, such as the real-space small displacement method and frozen-phonon method, have also been used to extract the IFCs for various materials [78, 93–96]. When the phonon properties are fully determined, the data can be combined with the Peierls–Boltzmann equation. The solution allows the determination of the thermal conductivity from first principles as discussed in the following two sections [75, 76, 79].

6.6.2 *Lattice Dynamics and Fermi's Golden Rule*

The phonon properties can be obtained once the IFCs are calculated. The harmonic force constants allow the determination of phonon dispersion and normal modes according to lattice dynamics theory. The anharmonic force constants can be used to obtain the scattering rate for three-phonon or four-phonon interactions using Fermi's golden rule. This section provides a brief coverage of lattice dynamics and the golden rule.

In a solid crystal, due to periodicity, the atom index is expressed in terms of double indices, that is, $i \to lb$, $j \to l'b'$, $k \to l''b''$, etc. Here, l is the index of the unit cell and b is the index of the atom inside the unit cell l. Note that the lattice translation vector of the lth unit cell can be represented by $\mathbf{R}_l$ according to Eq. (6.4), though l means all three indices here. Another vector $\mathbf{r}_b$ can be used to indicate the position of the atom with respect to the lattice point. With these notations, the Hamiltonian for the crystal with potential energy U becomes

$$\hat{H} \approx \sum_{lb} \frac{\hat{p}_{lb}^2}{2m_b} + \frac{1}{2!} \sum_{lb,l'b'} \sum_{\alpha,\beta} \Phi_{lb,l'b'}^{\alpha\beta} \zeta_{lb}^{\alpha} \zeta_{l'b'}^{\beta} + \frac{1}{3!} \sum_{lb,l'b',l''b''} \sum_{\alpha,\beta,\gamma} \Psi_{lb,l'b',l''b''}^{\alpha\beta\gamma} \zeta_{lb}^{\alpha} \zeta_{l'b'}^{\beta} \zeta_{l''b''}^{\gamma} \tag{6.74}$$

Note that $\hat{p}_{lb}$ is the momentum operator for the bth atom in the lth cell and m_b is the mass of the atom. The other parameters are defined in Eqs. (6.69–6.72). The fourth-order term is not given for simplicity. The equation of motion of atom b in unit cell l can be expressed as follows:

$$m_b \frac{\partial^2 \zeta_{lb}^{\alpha}}{\partial t^2} = -\sum_{l'b'} \sum_{\beta} \Phi_{lb,l'b'}^{\alpha\beta} \zeta_{l'b'}^{\beta} \tag{6.75}$$

The dynamic equation may be solved using a Fourier transform and expressed in terms of a series of Fourier components, each one is a plane wave with a wavevector $\mathbf{k}$ and angular frequency ω,

$$\zeta_{lb}^{\alpha} = \frac{1}{\sqrt{m_b}} \sum_{\mathbf{k}} \eta_b^{\alpha}(\mathbf{k}) \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{R}_l - \omega t)} \tag{6.76}$$

Note that $\mathbf{R}_l$ is the equilibrium position vector of the lth unit cell and the coefficient $\eta_b^{\alpha}(\mathbf{k})$ is independent of l. Substituting Eq. (6.76) into Eq. (6.75) gives

$$\omega^2 \eta_b^{\alpha}(\mathbf{k}) = \sum_{b',\beta} D_{bb'}^{\alpha\beta}(\mathbf{k}) \eta_{b'}^{\beta}(\mathbf{k}) \tag{6.77}$$

where the dynamic matrix is expressed as

$$D_{bb'}^{\alpha\beta}(\mathbf{k}) = \frac{1}{\sqrt{m_b m_{b'}}} \sum_{l'} \Phi_{0b,l'b'}^{\alpha\beta} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{R}_0 - \mathbf{R}_{l'})} \tag{6.78}$$

In writing Eq. (6.78), the translational invariance has been used [40, 41]. For periodic boundary conditions, the dynamic matrix is essentially a Fourier transform of the harmonic force constant matrix. The determinant of the following matrix must be zero, viz.,

$$\left| D_{bb'}^{\alpha\beta}(\mathbf{k}) - \omega^2 \delta_{\alpha\beta} \delta_{bb'} \right| = 0 \tag{6.79a}$$

which can be written in a matrix form,

$$\det\left[\mathbf{D}(\mathbf{k}) - \omega^2 \mathbf{I}\right] = 0 \tag{6.79b}$$

This allows the determination of phonon dispersion curves for all polarizations in the first Brillouin zone. The analysis of a 1D chain with two atoms was illustrated

in Sect. 6.5.1. If a unit cell has one atom, there will be three polarizations for each wavevector; if a unit cell has two atoms, there will be six polarizations (or branches), and so forth. For 2D and 3D crystals, the dispersion curves are rather complicated since they are different along different directions between the lattice points [45], similar to the electron band structures. Note that Fig. 6.19a only plots Si dispersion curves along the direction from the zone center Γ to point X in the first Brillouin zone, while Fig. 6.19b is for several SiC polytypes along the direction with a maximum magnitude of wavevector at the zone edge.

Equation (6.77) may be written in terms of the eigenvectors $\mathbf{e}_{b,\lambda}$ of the normal modes, where λ specifies a (phonon) mode, in the following [41]:

$$\omega^2 e^{\alpha}_{b,\lambda}(\mathbf{k}) = \sum_{b',\beta} D^{\alpha\beta}_{bb'}(\mathbf{k}) e^{\beta}_{b',\lambda}(\mathbf{k}) \tag{6.80}$$

For a system with N atoms, there exists $3N$ discrete eigenvectors that are orthogonal and normalized. These normal modes or Bloch modes represent $3N$ harmonic vibrational modes or phonons with specific wavevectors. The corresponding wavevectors and frequencies (eigenvalues) can be obtained from the dynamic matrix. All vibrational motion can be expressed as a superposition of the normal modes. As discussed previously, the phonon DOS, group velocity, and specific heat can be calculated using the dispersion relations. While the harmonic vibrations are for low temperatures, anharmonic vibrations do not affect the dispersion significantly unless the temperature is very high (e.g., close to the melting temperature). Anharmonic lattice dynamics can be used to predict the frequency shift, linewidth, and lifetime by including anharmonic perturbation or using the self-consistent phonon formulation [78, 97].

Fermi's golden rule, or simply the golden rule, was derived from the time-dependent perturbation theory in quantum physics to express the scattering rate or decay rate. The general form of the Fermi golden rule can be applied to nuclear decay, atomic electron transitions, interband and intraband transitions, electron scattering, phonon scattering, etc. For a system that undergoes a transition from an initial state described by the wavevector $\mathbf{k}$ to a final state denoted by $\mathbf{k}'$, the golden rule can be expressed as [42, 90]

$$\gamma_{\mathbf{k}\to\mathbf{k}'} = \sum_{\mathbf{k}'} \frac{2\pi}{\hbar} M^2_{\mathbf{k}\mathbf{k}'} \delta(E_{\mathbf{k}'} - E_{\mathbf{k}} \mp \hbar\omega) \tag{6.81}$$

where γ is the transition rate (probability of transition per unit time) and its inverse is called the lifetime or relaxation time, $M_{\mathbf{k}\mathbf{k}'}$ is the interaction matrix element for transition from state $\mathbf{k}$ to $\mathbf{k}'$, and the Dirac delta function signifies energy conservation. Note that momentum conservation is also necessary for the transition to occur as discussed in Sect. 6.5.3. As an example, for electron–phonon interaction, ω is the frequency of the emitted or absorbed phonon, referring to Eq. (6.45). The term $\hbar\omega$ can be dropped when electrons are scattered elastically such as in the case of

electron-defect scattering. Note that $M_{\mathbf{k}\mathbf{k}'}$ can be expressed in terms of a suitable Hamiltonian that represents the scattering potential (or strength) so that

$$M_{\mathbf{k}\mathbf{k}'} = \frac{1}{V}\int_V \Psi^*_{\mathbf{k}'}\hat{H}_{\text{scat}}\Psi_{\mathbf{k}}\mathrm{d}\mathbf{r} \tag{6.82}$$

The Fermi golden rule can be expressed in terms of phonon–phonon scattering involving three or four phonons [50, 90]. This is very important for modeling anharmonic scattering considering three-phonon and four-phonon processes. In the evaluation of the interaction matrix elements, the IFCs of the third order allow the determination of the three-phonon processes (both *N*-process and *U*-process) and fourth-order IFC allows the determination of four-phonon processes [50, 81]. Formulations for defect scattering and interfacial scattering can also be obtained based on Fermi's golden rule [78]. For heavily doped semiconductors, phonon–electron scattering may also play a role in reducing the lattice thermal conductivity [81, 98]. The detailed formulations and computational methodology can be found from the cited literature.

6.6.3 Evaluation of Thermal Conductivity

This section describes several methods used to calculate thermal conductivity of insulators and semiconductors. These methods can be used for both bulk and nanostructured materials if boundary scattering is included. They can also be used to study thermal transport in inhomogeneous media and thermal transport across interfaces, though the focus of this chapter is mostly on the intrinsic properties of solids.

If Fourier's law of heat conduction is extended to include anisotropy, we can write

$$q''_\beta = -\sum_\alpha \kappa_{\alpha\beta}\frac{\partial T}{\partial \alpha} \tag{6.83}$$

Here again, α or β specifies a coordinate axis. The Peierls–Boltzmann equation or phonon BTE can be solved using relaxation time approximation for each mode. The thermal conductivity can be obtained by comparing the BTE solution with Eq. (6.83) [75, 78]:

$$\kappa_{\alpha\beta} = \kappa_{\alpha\beta} = \sum_P \sum_K c(\mathbf{k})v_\alpha(\mathbf{k})v_\beta(\mathbf{k})\tau(\mathbf{k}) \tag{6.84}$$

Here, $c(\mathbf{k}) = h\omega(\mathbf{k})\partial f_{\text{BE}}/\partial T$, and v_α and v_β are the α and β components of the group velocity, respectively. These quantities can be obtained from the dispersion relations. The scattering rate $\tau(\mathbf{k})$ for each mode λ (in terms of the wavevector index K and polarization index P) is obtained from the golden rule as discussed

previously. Equation (6.84) is written in a way that is consistent with Eq. (5.71) but including anisotropy. The thermal conductivity tensor becomes a diagonal tensor $\bar{\bar{\kappa}} = \text{diag}(\kappa_{xx}, \kappa_{yy}, \kappa_{zz})$ in the principal coordinates and becomes $\bar{\bar{\kappa}} = \kappa \mathbf{I}$ for an isotropic medium. Since 2007, this method has been used to obtain the thermal conductivity of numerous semiconductor materials and oxide [78–81].

In the past 20 years, molecular dynamics simulations have been extensively used in modeling thermal transport and thermal properties of bulk and nanostructured materials. A brief introduction was given in Sect. 4.5.1. Molecular dynamics is advantageous for modeling crystals whose basis contains a large number of atoms, unstructured and disordered materials (such as amorphous solids and liquid phase), as well as soft matters (such as polymer and organic compounds) [76, 82–85, 99]. It is particularly suitable for modeling nanostructured materials such as nanotubes, nanowires, and fullerenes and related materials. Classical MD simulations are based on empirical or semi-empirical intermolecular or interatomic potentials. The number of atoms that can be simulated is rather limited by the earlier computational capabilities. The computational speed and capabilities have been significantly improved in the last two decades. Nowadays, advanced large-scale molecular dynamics simulation packages, e.g., LAMMPS [100], are accessible. Various functional forms obtained by comparison with experiments and from ab initio methods are available to model the interatomic potentials of different types of materials. The combination of these functions and parameters is often called a *force field* in molecular modeling. Another advantage of MD simulation of thermal transport is that it inherently includes anharmonicity and is particularly suitable at high temperatures. The wave characteristics of lattice vibration (phonon) cannot be easily observed since MD is a time-domain simulation. This problem has been addressed using various post-processing techniques such as Fourier analysis of the time variation of the locations and velocities of particles, wave-packet analysis, and mode decomposition [80, 82–84].

There are two MD simulation methods in studying thermal transport. One is called the direct method or nonequilibrium molecular dynamics (NEMD) and the other is often called the Green–Kubo method or equilibrium molecular dynamics (EMD). NEMD is more intuitive since it is based on 1D Fourier's law to determine the thermal conductivity:

$$k_x = -\frac{q_x''}{\partial T/\partial x} \tag{6.85}$$

In the direct simulation, either the temperatures at both ends of the structure (heat baths) are preset or the heat flux is preset. The simulation is then run for a sufficient duration of time using millions of time steps, each at subpicosecond timescale, to determine the unknown heat flux or the temperature gradient, respectively [85]. Periodic boundary conditions may be used to model bulk solids. Equation (6.85) is then used to calculate the thermal conductivity. In the MD simulation, the temperature is calculated according to [82]

$$T = \frac{1}{3Nk_{\mathrm{B}}}\left\langle \sum_{i}^{N} m_i v_i^2 \right\rangle \tag{6.86}$$

In the study where stress is considered, the pressure can be calculated by the virial equation [83],

$$P = \frac{Nk_{\mathrm{B}}T}{V} + \frac{1}{3V}\left\langle \sum_{i} \sum_{j>i} \mathbf{r}_{ij} \cdot \mathbf{F}_{ij} \right\rangle \tag{6.87}$$

In general, the temperature distribution is not linear due to temperature jumps near the two heat baths. However, the temperature profile can be approximated as linear in the middle section. The size effect and length effect on thermal conductivity of nanostructures can be easily modeled [76]. NEMD can also be used to study interface thermal resistance as well as non-Fourier heat conduction using the transient method [85]. These topics will be further discussed in Chap. 7.

In EMD, the Green–Kubo linear response theory is used. Statistical fluctuations always exist at temperatures exceeding absolute zero, causing random motion of atoms. If the sampling time is sufficiently small, the summation of the magnitude (or the square of the fluctuation term) depends on the temperature and the properties of the system. EMD is based on the Green–Kubo fluctuation-dissipation theorem and the thermal conductivity is computed from the *fluctuating heat current* as follows [82–84]:

$$k_{\alpha\beta} = \frac{1}{Vk_{\mathrm{B}}T^2}\int_0^{\infty} \langle J_\alpha(t) J_\beta(0)\rangle \mathrm{d}t \tag{6.88}$$

where V is the simulation volume, and J_α or J_β is the *heat current* component in the α or β direction, respectively, since the heat current is a vector. Note that $\alpha = \beta$ along the principle thermal conductivity coordinates. The operator $<>$ signifies ensemble average. The fluctuating heat current vector can be expressed as follows:

$$\mathbf{J} = \frac{\mathrm{d}}{\mathrm{d}t}\sum_i \varepsilon_i \mathbf{r}_i \tag{6.89a}$$

where ε_i is the total energy of the ith particle that includes both kinetic and potential energies. Note that the heat current defined here has a unit of heat flux multiplied by volume or [W m]. If only pairwise potentials are considered, we have

$$\mathbf{J} = \sum_i E_i \mathbf{v}_i + \frac{1}{2}\sum_{i,j} \left(\mathbf{F}_{ij} \cdot \mathbf{v}_i\right)\mathbf{r}_{ij} \tag{6.89b}$$

Three-body interaction terms can also be introduced and may be necessary to describe the force field, such as in the Stillinger–Weber potential commonly used in modeling diamond structure semiconductors [82].

By analyzing the EMD simulation in the reciprocal lattice space or *k*-space (sometimes called phonon space) and using the theory of lattice dynamics, phonon relaxation time and mean free paths can be extracted [80, 83, 84]. The phonon DOS can be obtained by a Fourier transform of the velocity autocorrelation function [80]. The spectral partition ratio and spatial energy distribution can be obtained using the DOS spectrum. These quantities are important to assess phonon confinement and localization especially in determination of the minimum thermal conductivity. The ensemble average of the heat current is called the heat current autocorrelation function, which is key to perform phonon space analysis in the first Brillouin zone. The analysis can provide information on the phonon spectral energy distribution and relaxation time. Anharmonic effects can also be examined [84]. The uncertainties associated with EMD calculations have also been systematically investigated [101].

First-principles calculations can be used to obtain harmonic and anharmonic force field and then applied to calculate the thermal conductivity [102, 103]. Recently, ab initio MD simulations have also been applied to model thermal conductivity [104–106].

6.7 Electron Emission and Tunneling

In all the discussions given so far, electrons are confined to the solid. Emission or discharge of electrons from a solid surface to vacuum or through a barrier (such as in a metal–insulator–metal multilayer structure) is possible, under the influence of an incident electromagnetic wave, an electric field, or a heating effect. Because of the importance of electron emission and tunneling to fundamental physics and device applications, the basic concepts are described in this section.

6.7.1 Photoelectric Effect

In 1887, Heinrich Hertz observed the *photoelectric effect* or *photoemission*. Shortly afterward, the phenomenon was experimentally studied by several others, including J. J. Thomson, who discovered electron as a subatomic particle. When radiation is incident on a metal plate, the electrons in the metal can be excited by absorbing the energy of the electromagnetic wave to escape the surface, as illustrated in Fig. 6.23a. The actual apparatus used for measuring the ejected photoelectrons was to use another electrode and measure the current flow via a closed circuit. This is similar to the arrangement shown in Fig. 6.23b for thermionic emission, but with photons incident on the left plate without heating up any of the plates. If the frequency of the incident radiation is not high enough, no electrical current can be measured no matter how

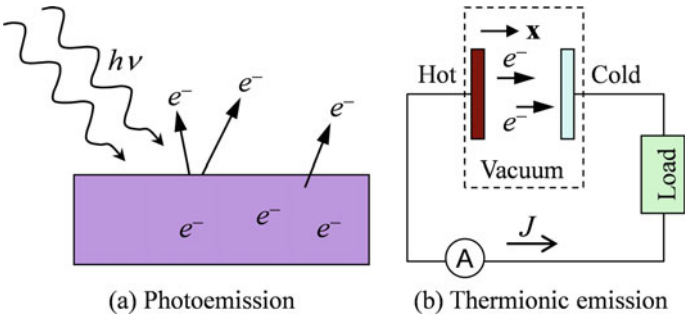

Fig. 6.23 Illustration of **a** the photoelectric effect and **b** the thermionic emission

intense the incident radiation is. Saying in other words, there appears to be a threshold frequency for photoemission to occur in a given material. The photoelectric effect was explained in 1905 by Albert Einstein with the concept of light quanta, postulated by Max Planck a few years earlier. Although Einstein also made seminal contributions to the theory of relativity and Brownian motion, he was awarded the Nobel Prize in Physics in 1921 mainly for his discovery of the law that governs the photoelectric effect.

From the Fermi–Dirac distribution function of free-electron gas, we can see that at low temperatures, electrons fill all energy levels up to the Fermi energy E_F. Note that we use E and E_F as the relative electron energy and, thus, they can be either positive or negative. Because of the binding of the electron with the rest of the solid, an additional energy, called the *work function* ψ, must be provided to the electron for it to escape from the solid. For Ag, Al, Au, Cu, Fe, Pb, and W, the work function ranges from 4 to 5 eV, which corresponds to a wavelength in the ultraviolet region from 250 to 300 nm. For Na, K, Cs, and Ca, the work function ranges from 2 to 3 eV, which falls in the visible spectrum. Because a photon can interact with only one electron at a time, the photon energy $h\nu$ must exceed the work function in order for the incident radiation to eject electrons from the surface. If $h\nu > \psi$, the photon energy may be absorbed by an electron right at the Fermi level. Subsequently, the electron will have a kinetic energy of

$$\tfrac{1}{2} m_e v_{e,\max}^2 = h\nu - \psi \tag{6.90}$$

after leaving the surface. If an electron is below the Fermi level, the kinetic energy of the ejected electron will be smaller than that given by Eq. (6.90). Therefore, Eq. (6.90) predicts the *maximum kinetic energy* of an electron for the prescribed photon frequency. A direct method for the determination of the work function is to measure the kinetic energy distribution of the photoelectrons, for a given frequency of the incident radiation.

One of the applications of photoemission is to measure the electron binding energy using the x-ray photoelectron spectroscopy (XPS), which is also called the electron spectroscopy for chemical analysis (ESCA). The basic principle for XPS is

$$E_{\mathrm{bd}} = h\nu - \tfrac{1}{2}m_{\mathrm{e}}v_{\mathrm{e}}^2 - \psi \tag{6.91}$$

where E_{bd} stands for the binding energy with respect to the Fermi energy. The high-energy photons from an x-ray source (200–2000 eV) can interact with the inner electrons and eject them out of the surface. The photoelectron intensity can be plotted as a function of the electron kinetic energy using an electron energy analyzer. The intensity peaks are associated with the binding energies of the particular atomic structures. Comparing with the recorded photoelectron spectra, XPS allows the determination of the chemical composition of the substance near the surface. Swedish physicist Kai Siegbahn shared the Nobel Prize in Physics in 1981 for his contribution leading to the practical application of XPS. Furthermore, ultraviolet photoemission spectroscopy (UPS) with photon energies ranging from 5 to 100 eV, often from a synchrotron radiation source, has been used to study the band structures of crystalline solids [107].

6.7.2 Thermionic Emission

The charge emission from hot bodies was independently discovered by British scientist Frederich Guthrie in 1873, with a heated iron ball, and Thomas Edison in 1880, while working on his incandescent bulbs. *Thermionic emission* was extensively studied in the early 1900s by Robert Millikan, Nobel Laureate in Physics in 1923; Owen Richardson, Nobel Laureate in Physics in 1928; and Irving Langmuir, Nobel Laureate in Chemistry in 1932, among others.

With the understanding of the work function as the threshold energy that an electron must gain to escape the solid, it becomes straightforward to explain the emission of electrons from a heated metal. We use metal here to illustrate thermionic emission because good conductors can be better approximated by the Sommerfeld theory. The distribution function of a free-electron gas has been extensively discussed in Chap. 5 (Sect. 5.1.3). At absolute zero temperature, all states below the Fermi level are filled by electrons and all states above the Fermi level are empty. Note that this picture is consistent with the electronic band theory. At elevated temperatures, the distribution function is modified as illustrated in Fig. 5.5. Some electrons will have energies above E_{F} (or μ_{F} as was used in Chap. 5). Because the distribution function becomes zero only when $E \to \infty$, a small fraction of electrons must occupy energy levels exceeding $E_{\mathrm{F}} + \psi$. We wish to quantitatively evaluate the current density or the charge flux from the hot plate to the cold plate, as illustrated in Fig. 6.23b. Let the electron flow be along the x-direction.

From Eq. (5.16), the number of electrons per unit volume between $\mathbf{v}$ and $\mathbf{v} + \mathrm{d}\mathbf{v}$ is

$$n(\mathbf{v})\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z = 2\left(\frac{m_{\mathrm{e}}}{h}\right)^3 \frac{\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z}{\mathrm{e}^{(E-E_{\mathrm{F}})/k_{\mathrm{B}}T} + 1} \tag{6.92}$$

where $E = \frac{1}{2}m_e(v_x^2 + v_y^2 + v_z^2)$ is the kinetic energy of an electron. The current density in the x-direction is given by

$$J_x = (1 - r') \iiint (-e) v_x n(\mathbf{v}) \mathrm{d}v_x \mathrm{d}v_y \mathrm{d}v_z \tag{6.94}$$

where r' is the electron reflection coefficient or the fraction of electron reflected by the receiver.

The integration is from $-\infty$ to ∞ in both the y- and z-directions. In order for an electron to escape in the x-direction, the following criterion must be satisfied:

$$v_x > v_{x,0} = \sqrt{2(E_F + \psi)/m_e} \tag{6.95}$$

This equation suggests that the integration is carried out only in the tail of the distribution function, where the x velocity is positive and the kinetic energy is sufficiently large, i.e., $E - E_F > \psi$, which is on the order of several electron volts. Note that $k_B T = 0.086$ eV at 1000 K and 0.026 eV at 300 K. When $\psi / k_B T = 4$, dropping the unity term in the denominator of Eq. (6.92) causes less than 2% error. The error becomes even smaller at a larger v_x so that its impact on the integration is negligibly small. For this reason, it appears safe to substitute the Fermi–Dirac distribution by the Maxwell–Boltzmann distribution, viz.,

$$J_x = -2e(1 - r')\left(\frac{m_e}{h}\right)^3 \mathrm{e}^{E_F/k_B T} \int_{v_{x,0}}^{\infty} v_x \exp\left(-\frac{m_e v_x^2}{2k_B T}\right) \mathrm{d}v_x$$
$$\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{m_e v_y^2}{2k_B T} - \frac{m_e v_z^2}{2k_B T}\right) \mathrm{d}v_y \mathrm{d}v_z \tag{6.95}$$

The result is the famous *Richardson–Dushman equation* for the current density [1, 2]:

$$J = A_{RD}(1 - r') T^2 \mathrm{e}^{-\psi/k_B T} \tag{6.96}$$

where $A_{RD} = 4\pi m_e e k_B^2 / h^3 = 1.202 \times 10^6$ A/m^2 K^2 is called the *Richardson constant*, and the direction of J is as shown in Fig. 6.23b. The heat transfer associated with the electron flow can be evaluated by considering the kinetic energy associated with each electron, $\frac{1}{2}m_e v^2 \approx \frac{1}{2}m_e v_x^2$, so that

$$q_x'' = (1 - r') \iiint v_x \left(\tfrac{1}{2}m_e v_x^2\right) f(\mathbf{v}) \mathrm{d}v_x \mathrm{d}v_y \mathrm{d}v_z = (\psi + E_F + k_B T) \frac{J_x}{e} \tag{6.97}$$

This equation suggests that the average energy of the "hot electron" is $\psi + E_F + k_B T$, as expected. Vacuum tubes operate based on the principle of thermionic emission. Vacuum tubes had wide applications in the mid-twentieth century in radio, TV, and computer systems, but have largely been replaced by transistors nowadays. Thermionic generators produce electricity without any moving parts and belong to the category of direct energy converters. Extensive discussion of the thermodynamics and efficiency of thermionic converters can be found from Hatsopoulos and Gyftopoulos [108].

In some applications, a voltage can be applied between the electrodes. Furthermore, a semiconductor can be used to form a Schottky barrier between a metal and a semiconductor [11]. The applied voltage changes the potential distribution so that it gradually decreases inside the barrier. Furthermore, the work function can be significantly reduced. Assuming the transmission coefficient is unity, Eq. (6.96) can be modified to the following for the net charge transfer:

$$J_{\text{net}} = A^* T^2 e^{-\psi^*/k_B T} (e^{e\Delta V/k_B T} - 1) \tag{6.98}$$

where A^* should be calculated according to the effective mass, ψ^* is the effective work function, and ΔV is the applied voltage [11]. In deriving Eq. (6.98), we assumed that hot electrons from the cathode will go through the barrier through ballistic processes. This means that the electron mean free path must be larger than the thickness of the semiconductor film. Otherwise, the electron transport is governed by diffusion because of collisions with phonons or impurities. When diffusion occurs, the electron transport under the influence of a temperature difference is described by the thermoelectric effect, based on irreversible thermodynamics, as discussed in Chap. 5. When the barrier thickness is extremely small, another phenomenon called quantum tunneling may occur such that an electron whose energy is lower than the potential barrier has a chance to transmit through the barrier. Tunneling effect will be discussed in the next subsection. Mahan and coworkers pointed out that, for the thermionic phenomenon to be the dominant transport mechanism, the electron mean free path in the barrier must be greater than the thickness of the barrier [109]. Furthermore, the latter must exceed the characteristic length, below which tunneling becomes significant. Thermionic emission in semiconductor heterogeneous structures has been extensively studied in the last decades for both refrigeration and power generation [109–111]. The refrigeration process is a reversed thermionic power generation process. In *thermionic refrigeration*, the cold cathode emits electrons to the room-temperature anode as a result of the applied voltage. In order to achieve any cooling effect, energy that is carried through by the electrical current must be greater than that by heat conduction via lattice vibration from the hot electrode to the cold electrode. The nonequilibrium electron and phonon transport phenomena have also been investigated. In some cases, both thermionic and thermoelectric effects may show up [111, 112]. In other cases, thermionic and tunneling effects can work together or against each other [113, 114].

6.7.3 Field Emission and Electron Tunneling

From the above discussion, we have noticed that thermionic emission may be enhanced or even reversed (from a colder cathode to a hotter anode) by an applied electric field. Some thermal excitation is necessary for part of the electrons to occupy energy levels above the Fermi level by a finite amount, prescribed by the work function. This is commonly referred to as a *potential barrier* or a *potential hill*. An electron must acquire sufficient energy for it to surmount the barrier. When the field strength is very high, however, electrons at energy levels lower than the height of the barrier can tunnel through the potential hill. The word *tunneling* gives a vivid (but inaccurate) picture of the tunneling phenomenon as if a hole were drilled for the electrons without sufficient energy to pass through a potential hill, without climbing to its top first. This phenomenon of electron emission at high applied field is called *field emission*, which can occur at very low temperatures. The applied electric field can exceed several billion volts per meter. Because of the high field, field emission can occur only in ultrahigh vacuum (UHV); otherwise, ionization of the gas molecules would occur that can cause *discharge glow*. In essence, field emission is a form of *quantum tunneling*, which cannot be understood within the framework of classical mechanics. The electron motion is governed by Schrödinger's wave equation, and the transmission can be predicted by the probability of finding an electron on the other side of the potential hill, as illustrated in Fig. 6.24.

In 1928, Fowler and Nordheim [115] provided the first quantum mechanical derivation of the field emission current density J as follows:

$$J = C\left(\frac{\Delta V}{L}\right)^2 \exp\left(-\frac{\alpha \psi^{3/2}}{\Delta V/L}\right) \tag{6.99}$$

This is called the Fowler–Nordheim equation, in which $\Delta V/L$ is the electric field, C and α are two positive constants, and ψ is the work function defined previously.

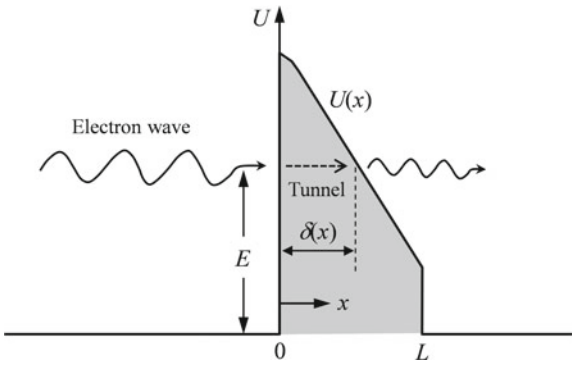

Fig. 6.24 Illustration of quantum tunneling through a potential barrier by an electron wave

The WKB approximation is commonly used to find the transmission probability τ' of tunneling. WKB (also KWB or BWK) stands for Wentzel, Kramers, and Brillouin, although a fourth person Jeffreys was also included in some literature—so, the abbreviation appeared as JWKB. The main assumption in the WKB approximation is that the potential $U(x)$ is a slow function of x [116]. In the region where the electron energy E is greater than $U(x)$, the wave function is of the form

$$\Psi(x,t) = A\exp(-\mathrm{i}\omega t)\exp\left[\pm\frac{\mathrm{i}}{\hbar}\sqrt{2m_\mathrm{e}(E-U)}\right] \tag{6.100}$$

where A is the amplitude of the electron wave. In the region where $E < U(x)$, the wave function is of the form

$$\Psi(x,t) = A\exp(-\mathrm{i}\omega t)\exp\left[\pm\frac{1}{\hbar}\sqrt{2m_\mathrm{e}(U-E)}\right] \tag{6.101}$$

The transmission probability or transmission coefficient can be approximated as

$$\tau'(E) = \exp\left[-\frac{2}{\hbar}\int_0^\delta \mathrm{d}x\sqrt{2m_\mathrm{e}(U-E)}\right] \tag{6.102}$$

where δ is the width of the potential at E [116].

Example 6.5 Consider a potential barrier in the region $0 \le x \le L$ whose potential is the highest but linear decreases with x according to $U(x) = \psi - \frac{x}{L}e\Delta V$ and $\delta(E) = \frac{\psi - E}{e\Delta V}L$. Find the transmission coefficient.

Solution For the triangular barrier shown in Fig. 6.24, we note that

$$\int_0^\delta \mathrm{d}x\sqrt{U-E} = \int_0^\delta \mathrm{d}x\sqrt{\psi - E - e\Delta V x/L} = \frac{(\psi-E)^{3/2}}{e\Delta V/L}\int_1^0 \sqrt{1-u}\mathrm{d}u = \frac{2(\psi-E)^{3/2}}{3e\Delta V/L}$$

Substituting this equation into Eq. (6.102), we obtain

$$\tau'(E) = \exp\left[-\frac{4(\psi-E)^{3/2}}{3\hbar\Delta V/L}\sqrt{2m_\mathrm{e}}\right] \tag{6.103}$$

When $E << \psi$, we see that $\tau' \approx \exp\left(-\frac{\alpha\psi^{3/2}}{\Delta V/L}\right)$, where $\alpha = \frac{4\sqrt{2m_\mathrm{e}}}{3\hbar}$. At elevated temperatures, however, we need to consider the energy distribution of electrons. Esaki and coworkers demonstrated that the resonant tunneling of electron waves may allow the transmission coefficient to approach unity in superlattice and double-barrier structures [117]. Electron tunneling is similar to photon tunneling of electromagnetic waves, to be discussed in Chap. 10.

The tunneling current density can be calculated by

$$J_{\mathrm{t}} = \int_{E_{\min}}^{E_{\max}} e\tau'(E)n(E)\mathrm{d}E \tag{6.104}$$

where E is the kinetic energy in the x-direction, $E_{\max}$ corresponds to the energy at the top of the potential barrier, $E_{\min}$ is a reference energy, and $n(E)\mathrm{d}E$ is the number of available electrons, with energy between E and $E + \mathrm{d}E$, per unit area per unit time, given as

$$n(E) = \frac{m_{\mathrm{e}} k_{\mathrm{B}} T}{2\pi^2 \hbar^3} \ln\left[1 + \exp\left(-\frac{E - E_{\mathrm{F}}}{k_{\mathrm{B}} T}\right)\right] \tag{6.105}$$

Some analytical expressions similar to Eq. (6.100) have been presented [118, 119] to approximate the integration of Eq. (6.104).

The energy transfer during field emission or electron tunneling can also be evaluated [113, 120, 121]. A salient difference between thermionic emission and field emission is that thermionic emission always gives out energy as the electrons are emitted and transfer the energy to the other side of the barrier. This is because the emitted electrons are in the high-energy tail of the distribution function, called hot electrons, with a much higher effective temperature than the equilibrium cathode temperature. On the other hand, field emission allows electrons with energies much lower than that corresponding to the equilibrium temperature to escape the surface. Since the replacement electrons have a higher average energy than the emitted electrons, a heating effect occurs that increases the cathode temperature. Depending on the geometry, temperature, transmission coefficient, and energy distribution, both heating and cooling of the cathode are made possible by field emission. This is known as the Nottingham effect originally published in 1941.

Some applications of quantum tunneling in semiconductors and superconductors were discussed in Chap. 1. One of the applications of electron tunneling was the invention of scanning tunneling microscope (STM). Xu et al. [120] developed a model for the energy exchange by the tunneling electrons and made a comparison with STM measurements. They considered the Nottingham effect on both electrodes, as well as resistive heating. At short distances, thermionic emission, field emission, and photon tunneling could occur simultaneously. Photon tunneling will be studied in Chap. 10. Fisher and Walker [121] analyzed the energy transport in nanoscale field emission processes by considering the geometry of the emission tip. Quantum size effect may play a role in modifying some of the critical parameters. Field emission by nanotubes has been proposed for nanoscale manufacturing and thermal writing [122]. Wong et al. [123] performed a detailed thermal analysis during electron beam heating and laser processing. Carbon nanotube field emission display (CNT-FED) has been demonstrated. While CNT-FEDs resemble the cathode-ray tubes (CRTs) in many ways, it can be made thin and flat with a much lower applied voltage [124].

6.8 Electrical Transport in Semiconductor Devices

Semiconductors are the most important materials for microelectronics, MEMS, and optoelectronics. Much of the discussions in Chap. 5 and the previous sections of this chapter are applicable to semiconductors, especially for the energy storage and transport by phonons. This section focuses on the basics of electrical transport and properties for some common semiconductor devices used in optoelectronics.

6.8.1 Number Density, Mobility, and the Hall Effect

The calculation of the number density of electrons and holes at any given temperature T is very important for the determination of the electrical, optical, and thermal properties of semiconductor materials and devices. The free-electron gas model can be modified to describe the electron and hole distributions and the transport in semiconductors. The Fermi–Dirac distribution function is applicable to electrons and holes according to

$$f_{\mathrm{e}}(E) = \frac{1}{\mathrm{e}^{(E-E_{\mathrm{F}})/k_{\mathrm{B}}T}+1} \quad \text{and} \quad f_{\mathrm{h}}(E) = \frac{1}{\mathrm{e}^{(E_{\mathrm{F}}-E)/k_{\mathrm{B}}T}+1} \tag{6.106}$$

Note that $f_{\mathrm{e}}(E) + f_{\mathrm{h}}(E) \equiv 1$. The number density of electrons or holes is given by

$$n_{\mathrm{e}} = \int_{E_{\mathrm{C}}}^{\infty} \frac{D_{\mathrm{e}}(E)\mathrm{d}E}{\mathrm{e}^{(E-E_{\mathrm{F}})/k_{\mathrm{B}}T}+1} \quad \text{and} \quad n_{\mathrm{h}} = \int_{-\infty}^{E_{\mathrm{V}}} \frac{D_{\mathrm{h}}(E)\mathrm{d}E}{\mathrm{e}^{(E_{\mathrm{F}}-E)/k_{\mathrm{B}}T}+1} \tag{6.107}$$

where $D_{\mathrm{e}}(E)$ and $D_{\mathrm{h}}(E)$ are the densities of states in the conduction and valence bands, respectively. With the approximated quadratic forms of the conduction and valence bands, Eq. (6.29), the densities of states can be written as

$$D_{\mathrm{e}}(E) = M_{\mathrm{C}} \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}E}\bigg|_{\mathrm{C}} = \frac{M_{\mathrm{C}}}{2\pi^2}\left(\frac{2m_{\mathrm{e}}^*}{\hbar^2}\right)^{3/2}(E-E_{\mathrm{C}})^{1/2} \tag{6.108}$$

and

$$D_{\mathrm{h}}(E) = \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}E}\bigg|_{\mathrm{V}} = \frac{1}{2\pi^2}\left(\frac{2m_{\mathrm{h}}^*}{\hbar^2}\right)^{3/2}(E_{\mathrm{V}}-E)^{1/2} \tag{6.109}$$

where M_{C} is the number of equivalent minima in the conduction band. Equations (6.108) and (6.109) are derived based on the parabolic shape near the bottom of the conduction band for electrons or the top of the valance band for holes.

The effective mass of electrons is a geometric average over the three major axes because the effective mass of silicon depends on the crystal direction. The effective mass of holes is an average of heavy holes and light holes because there exist different subbands [11]. At moderate temperatures, $E_C - E_F >> k_B T$ and $E_F - E_V >> k_B T$ are satisfied; subsequently, f_e and f_h can be approximated with the classical Maxwell–Boltzmann distribution:

$$f_e(E) \approx e^{(E_F - E)/k_B T} \quad \text{and} \quad f_h(E) \approx e^{(E - E_F)/k_B T} \tag{6.110}$$

We can carry out the integrations in Eq. (6.107) and thus obtain

$$n_e = N_C e^{-(E_C - E_F)/k_B T} \tag{6.111}$$

and

$$n_h = N_V e^{-(E_F - E_V)/k_B T} \tag{6.112}$$

where

$$N_C = 2M_C \left(\frac{m_e^* k_B T}{2\pi\hbar^2} \right)^{3/2} \quad \text{and} \quad N_V = 2 \left(\frac{m_h^* k_B T}{2\pi\hbar^2} \right)^{3/2} \tag{6.113}$$

are called the *effective density of states* in the conduction band and in the valance band, respectively. The combination of Eqs. (6.111) and (6.112) gives, in terms of $E_g = E_C - E_V$,

$$n_e n_h = N_{th}^2 = N_C N_V e^{-E_g/k_B T} \propto T^3 e^{-E_g/k_B T} \tag{6.114}$$

This expression does not involve the Fermi energy. Therefore, it holds for both intrinsic and doped semiconductors. The number density Nth can be viewed as thermally excited electron–hole pairs per unit volume. It is also referred to as the number density of intrinsic carriers because $n_e = n_h = N_{th}$, in an intrinsic semiconductor. It can be seen that the number densities increase with temperature so that the electrical conductivity of an intrinsic semiconductor increases with temperature. The Fermi energy for an intrinsic semiconductor can be obtained by setting $n_e = n_h$ in Eqs. (6.111) and (6.112), yielding

$$E_F = \frac{E_C + E_V}{2} + \frac{k_B T}{2} \ln\left(\frac{N_V}{N_C} \right) \approx \frac{E_C + E_V}{2} \tag{6.115}$$

The Fermi energy for an intrinsic semiconductor is expected to lie in the middle of the forbidden band or the bandgap. The requirement for the approximate distributions given in Eq. (6.110) to hold with less than 2% error is $E_g/k_B T > 8$, such that $\exp\left[-E_g/(2k_B T)\right] < 0.02$. For $E_g > 0.8$ eV, we have $T < 1150$ K. One should keep

in mind that E_g reduces as temperature increases. For silicon, $E_g \approx 1.11$ eV at 300 K and ≈ 0.91 eV at 900 K.

When impurities of either donors or acceptors or both are involved, the calculation of Fermi energy and number densities becomes more involved [11, 15]. Let N_D and N_A stand, respectively, for the number densities (i.e., doping concentrations) of donors (e.g., P and As) and acceptors (e.g., B and Ga). In brief, the energy level of donors E_D is usually lower but very close to E_C. As a result, the Fermi energy E_F goes up but is always below E_D. The difference $E_C - E_D$ is called the ionization energy of donors, which is required for the donors to become ionized. The ionization of donors increases the number of free electrons, and the semiconductor is said to be of n-type. For the semiconductor Si, the ionization energy for P is 45 meV and that of As is 54 meV. Likewise, the energy level of acceptors E_A is slightly above E_V, and $E_A - E_V$ is called the ionization energy of acceptors. The ionization of acceptors increases the number of holes, and the semiconductor is said to be of p-type. For the semiconductor Si, the ionization energy for B is 45 meV and that of Ga is 72 meV. Note that there are 5.0×10^{22} cm^{-3} (atoms per cubic centimeters) for silicon. For n-type silicon with an arsenic doping concentration of $N_D = 5.0 \times 10^{16}$ cm^{-3}, the impurities occupy one atomic site per million. Because of the change in Fermi energy, most of the impurities are ionized at room temperature when the doping concentration is less than 5.0×10^{17} cm^{-3}. For fully ionized impurities, the charge neutrality requires that

$$n_e + N_A = n_h + N_D \tag{6.116}$$

If the impurities are partially ionized, N_D and N_A in Eq. (6.116) should be replaced by the ionized donor and acceptor concentrations, respectively.

Example 6.6 For boron-doped Si with $N_A = 2.0 \times 10^{16}$ cm^{-3}, find N_{th}, n_e, and n_h at temperatures from 300 to 1000 K; compare your answers with the values for intrinsic silicon. Assume $m_e^* = 0.3m_e$ and $m_h^* = 0.6m_e$. Use $M_C = 6$ and $E_g(T) = 1.155 - 0.000473T^2/(T + 636)$ eV where T is in kelvin.

Solution This is a p-type semiconductor with $N_D = 0$, and from Eq. (6.116), we have $n_h = n_e + N_A$. Substituting it into Eq. (6.114), we have $n_e^2 + N_A n_e = N_{th}^2 = N_C N_V e^{-E_g/k_B T}$. The solution gives

$$n_e = \frac{1}{2}\left(\sqrt{N_A^2 + 4N_{th}^2} - N_A\right) \text{ and } n_h = \frac{1}{2}\left(\sqrt{N_A^2 + 4N_{th}^2} + N_A\right) \tag{6.117}$$

The calculated values of $N_C = 2.47 \times 10^{19}$ cm^{-3} and $N_V = 1.17 \times 10^{19}$ cm^{-3} at 300 K are somewhat lower than the recommended values of $N_C = 2.86 \times 10^{19}$ cm^{-3} and $N_V = 2.66 \times 10^{19}$ cm^{-3} [11]. The results are plotted in Fig. 6.25 for comparison. In the extrinsic region when $T < 700$ K, the majority carriers are holes, and $n_h \approx N_A$ depends little on temperature. In the intrinsic region when $T > 800$ K, $n_e \approx n_h \approx N_{th}$, due to thermal excitation. It should be mentioned that at very low temperatures,

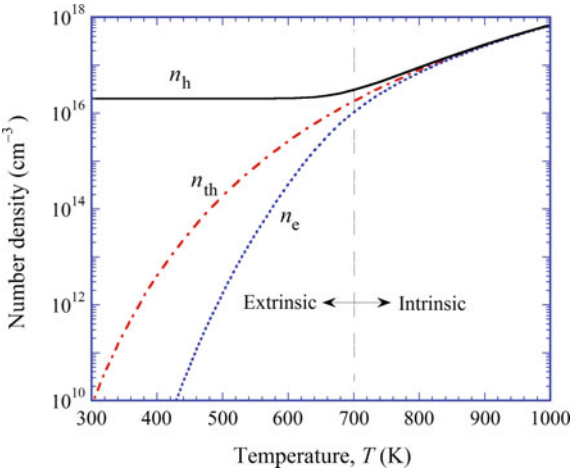

Fig. 6.25 Calculated number densities for Example 6.6

i.e., $T < 100\,\text{K}$, ionization is not very effective and $n_h << N_A$. Therefore, the low-temperature region is called freeze-out zone, which is not shown in the plot.

The Drude free-electron model predicts $\sigma = \tau n_e e^2/m_e$, as given in Eq. (5.52), which can be applied for both electrons and holes, using proper effective masses and relaxation times. In semiconductor physics, the electron or hole mobility is defined based on the effective mass:

$$\mu_e = \frac{e\tau_e}{m_e^*} \quad \text{and} \quad \mu_h = \frac{e\tau_h}{m_h^*} \tag{6.118}$$

The physical significance is that mobility is the drift velocity per unit applied field, i.e.,

$$\mathbf{u}_{d,e} = -\mu_e \mathbf{E} \quad \text{and} \quad \mathbf{u}_{d,h} = \mu_h \mathbf{E} \tag{6.119}$$

The electrical conductivity of a semiconductor is thus

$$\sigma = en_e\mu_e + en_h\mu_h \tag{6.120}$$

Depending on the impurity and temperature range, one term may be dominant, or both the terms may be comparable. It is crucial to understand the scattering mechanism in semiconductors. In metals, all the conducting electrons are near the Fermi surface, and their average energy cannot be described by the classical statistics because $\bar{\varepsilon} = \frac{1}{2}m_e\overline{v_e^2} \approx \frac{3}{5}\mu_F \neq \frac{3}{2}k_BT$ (see Example 5.2). For semiconductors, on the other hand, the Boltzmann distribution given in Eq. (6.110) suggests that $\bar{\varepsilon} = \frac{1}{2}m_e^*\overline{v_e^2} \approx \frac{3}{2}k_BT = \frac{1}{2}m_e^*v_{th}^2$, and the classical statistics is applicable to a large temperature range. Thermal velocity is the velocity of electrons or holes at the equilibrium temperature and was given in Eq. (6.55). At sufficiently high temperatures when phonon scattering dominates, the electron mean free path $\Lambda_e \propto T^{-1}$. Based

on the relation $\tau_e = \Lambda_e/\bar{v}_e$, we have

$$\mu_{ph} \propto T^{-3/2} \tag{6.121}$$

where μ_{ph} is the contribution of carrier–phonon scattering. Equation (6.121) describes intrinsic semiconductor without defects. The scattering by impurities results in a mobility given by

$$\mu_d \propto T^{3/2}/N_d \tag{6.122}$$

where N_d stands for the concentration of the ionized impurities. The combination gives the mobility for either electron or hole as follows:

$$\frac{1}{\mu} = \frac{1}{\mu_{ph}} + \frac{1}{\mu_d} \tag{6.123}$$

For intrinsic semiconductor, the electrical conductivity is very small and proportional to $\exp\left[-E_g/(2k_BT)\right]$ so that the electrical conductivity increases with temperature. For intermediately doped semiconductors, there exists a maximum value of the mobility below room temperature due to the opposite temperature dependence of μ_{ph} and μ_d. At that temperature, the electrical conductivity is maximum. As the temperature goes up beyond room temperature, the conductivity decreases due to the increased phonon scattering. When the semiconductor reaches the intrinsic region, the number density suddenly increases and the conductivity increases again with temperature.

The Hall effect is very useful in measuring the mobility of semiconductors. In the extrinsic region, the Hall effect allows measurement of the type and concentration of the carriers. The measurements are usually carried out with the van der Pauw method, which is a four-probe technique for determining the electrical resistance and the Hall coefficient. The data of electrical resistivity and number density allow the extraction of the mobility, based on the effective mass determined using cyclotron resonance technique.

When both the carriers are significant to the transport properties, the situation is rather interesting. Referring to Fig. 6.1, when current flows to the positive x-direction, we have $v_{e,x} < 0$ and $v_{h,x} > 0$. The magnetic force drives both the electrons and the holes toward the negative y-direction, such that $v_{e,y} < 0$ and $v_{h,y} < 0$ if $E_y = 0$. At steady state, a finite E_y, known as the Hall field, may exist. Since there is no net current flow in the y-direction, we must have

$$J_x = -ev_{e,x}n_e + ev_{h,x}n_h \tag{6.124}$$

$$J_y = -ev_{e,y}n_e + ev_{h,y}n_h = 0 \tag{6.125}$$

In general, both $v_{e,y}$ and $v_{h,y}$ are not zero. The Lorentz force $\mathbf{F} = q(\mathbf{E}+\mathbf{u}_d \times \mathbf{B})$ in the y-direction is related to the drift velocities for electrons or holes by

$$-eE_y + ev_{e,x}B = ev_{e,y}/\mu_e \tag{6.126}$$

and

$$eE_y - ev_{h,x}B = ev_{h,y}/\mu_h \tag{6.127}$$

Rewrite Eqs. (6.126) and (6.127) as $n_e\mu_e(E_y - v_{e,x}B) = -n_e v_{e,y}$ and $n_h\mu_h(E_y - v_{h,x}B) = n_h v_{h,y}$, respectively. Compared with Eq. (6.122), we notice that $n_e\mu_e(E_y - v_{e,x}B) + n_h\mu_h(E_y - v_{h,x}B) = 0$, or

$$\frac{E_y}{B} = \frac{n_e\mu_e v_{e,x} + n_h\mu_h v_{h,x}}{n_e\mu_e + n_h\mu_h}$$

Combining it with Eq. (6.121), we obtain the Hall coefficient as follows:

$$\eta_H = \frac{E_y}{J_x B} = \frac{n_e\mu_e v_{e,x} + n_h\mu_h v_{h,x}}{e(n_e\mu_e + n_h\mu_h)(-n_e v_{e,x} + n_h v_{h,x})} \tag{6.128}$$

Substituting $v_{e,x} = -\mu_e E_x$ and $v_{h,x} = \mu_h E_x$ into the previous equation, we obtain

$$\eta_H = \frac{n_h\mu_h^2 - n_e\mu_e^2}{e(n_h\mu_h + n_e\mu_e)^2} \tag{6.129}$$

after canceling E_x. The Hall coefficient for semiconductors may be positive or negative, and becomes zero when $n_h\mu_h^2 = n_e\mu_e^2$. The drift velocities in the y-direction, however, cannot be zero unless $B = 0$ or $J_x = 0$.

6.8.2 *Generation and Recombination*

The generation, recombination, and diffusion processes are directly related to the charge transport in semiconductors and optoelectronic devices. This section takes photoconductivity as an example to illustrate the generation and recombination processes, followed by a brief discussion of luminescence.

Much has been said previously about absorption of light that causes a transition in the electronic states in solids. The bandgap absorption of Si, Ge, and GaAs corresponds to the wavelengths in the visible and near-infrared spectral regions. The excitation of electrons from the valence band to the conduction band by the absorption of radiation increases the conductivity of the semiconductor dramatically. This is known as *photoconductivity* and can be used for sensitive radiation detectors. For

some semiconductors, the bandgap is very narrow so that transitions can happen at longer wavelengths. For example, the bandgap energy of $Hg_{0.8}Cd_{0.2}Te$ is 0.1 eV at 77 K (liquid nitrogen temperature), and the material can be used as infrared detectors, which are commonly referred to as MCT detectors. At very low temperatures, impurities cannot be ionized thermally even though the ionization energy is very small. For boron-doped germanium, the acceptor ionization energy $E_A - E_V \approx 10\,\text{meV}$ corresponds to a wavelength of about 120 μm [125]. Therefore, Ge:B can be used as far-infrared radiation detectors. There are two groups of radiation detectors. The first group is called thermal or bolometric detectors, which rely on the temperature change of the detector as a result of the absorbed radiation. The temperature change can be monitored by a temperature-dependent property, such as the electrical resistance. An example is the superconductive bolometer, which relies on the drastic change in resistance with temperature, near the superconducting-to-normal-state transition or critical temperature T_c. The second group is called nonthermal, nonbolometric, or nonequilibrium detectors. An example is the photoconductive detector in which the conductivity changes as a result of the direct interaction of electrons with photons.

Before the radiation is incident on the photoconductive detector, the conductivity can be expressed as $\sigma_0 = en_{e,0}\mu_e + en_{h,0}\mu_h$ at thermal equilibrium. Under the influence of an incident radiation with photon energies greater than the bandgap, additional electron–hole pairs are created so that the concentration is increased by Δn for both types of carriers. The relative change in the electrical conductance $\Delta\sigma/\sigma_0$ can be expressed as

$$\frac{\Delta\sigma}{\sigma_0} = \frac{\Delta n(\mu_e + \mu_h)}{n_{e,0}\mu_e + n_{h,0}\mu_h} \tag{6.130}$$

Here, Δn is the net increase in carrier concentration as a result of both generation and recombination. The *generation* is associated with the absorbed radiation and depends on the intensity of the incident light and the *quantum efficiency*, which is wavelength dependent. The quantum efficiency is the percentage of the incoming photons that generate an electron–hole pair. The *recombination* is a relaxation process because the excess charges are not at thermal equilibrium. If the incident radiation is blocked off, the semiconductor will quickly reach an equilibrium with the conductivity σ_0. The characteristic time of the recombination process is called the *recombination lifetime* or *recombination time* τ_{rc}. While it is also related to electron scattering, lattice scattering, and/or defect scattering, the recombination time is usually much longer than the relaxation time used in charge transport processes. The net rate of change can be expressed as the rate of generation (creation) minus the rate of recombination (annihilation), viz.,

$$\frac{dn}{dt} = \dot{n}_g - \frac{n - n_0}{\tau_{rc}} \tag{6.131}$$

Under a steady-state incident radiation, we can set $dn/dt = 0$ so that $\Delta n = n - n_0 = \tau_{rc}\dot{n}_g$. Suppose that the incoming photon is of frequency ν in Hz with a

spectral irradiance I_ν in W/m^2 Hz, and the detector has an effective area A, thickness d, and absorptance α_ν. We obtain the rate of generation:

$$\dot{n}_g = \frac{\alpha_\nu I_\nu A}{h\nu A d} = \frac{\alpha_\nu I_\nu}{h\nu d} \tag{6.132}$$

Substituting into Eq. (6.130), we obtain the *sensitivity* of a photoconductive detector as follows:

$$\frac{1}{I_\nu}\frac{\Delta\sigma}{\sigma_0} = \frac{\alpha_\nu \tau_{rc}(\mu_e + \mu_h)}{h\nu d(n_{e,0}\mu_e + n_{h,0}\mu_h)} \tag{6.133}$$

Increasing the recombination time τ_{rc} improves the sensitivity but decreases the speed or response time of the detector. Photoconductivity requires that $h\nu > E_g$ for bandgap absorption to occur. However, the sensitivity decreases toward higher frequencies, or shorter wavelengths, because there are fewer photons per unit radiant power. Consequently, the sensitivity of a photoconductive detector increases with wavelength first and then suddenly drops to zero close to the band edge. For thin films, the absorptance depends on the film thickness when the photon penetration depth is comparable to the thickness.

In photoconductivity, the recombination is not associated with the emission of radiation, and therefore it is said to be nonradiative. The Auger effect and multiphonon emission are two common processes of nonradiative recombination. In the Auger effect, the energy released by a recombining electron–hole pair is absorbed by another electron in the conduction band, which subsequently relaxes to the equilibrium condition by the emission of phonons. In a multiphonon emission process, the recombination of an electron–hole pair is associated with the release of a cascade of phonons, each having a much lower energy. More details on the recombination process and how to calculate the associated lifetime can be found from the texts of Sze [11].

Radiative recombination can also occur and is very important for light-emitting applications, such as *luminescence*, which is essentially the inverse process of absorption. The excitation of electrons may be accomplished by passing through an electrical current. An example is the semiconductor light-emitting diode, in which the electronic transition from the conduction band to the valence band can result in optical radiation. Photoluminescence is often referred to as *fluorescence*, when the emission occurs at the same time as the absorption, or *phosphorescence*, when the emission continues for a while after the excitation.

6.8.3 *The* p-n *Junction*

The *p-n* junction is familiar to every reader although many of us are unfamiliar with the underlying physics. Let us first take a look at the charge transport by diffusion,

which is a very important process in semiconductor applications. Diffusion takes place when there is a spatial nonuniformity in the carrier concentration. The principle is the same as the diffusion of ideal gas molecules described in Sect. 4.2.3. Using Fick's law, we can write the current densities resulting from the diffusion of electrons and holes as follows:

$$J_e = eD_e \frac{dn_e}{dx} \quad \text{and} \quad J_h = -eD_h \frac{dn_h}{dx} \tag{6.134}$$

where the diffusion coefficient for electrons and holes can be related to the mean free path and the average velocity by $D_e = \frac{1}{3}\Lambda_e \bar{v}_e$ and $D_h = \frac{1}{3}\Lambda_h \bar{v}_h$, according to Eq. (4.42). Assuming $\bar{v}_e \approx \bar{v}_h \approx v_{th}$, we have $D_e = \frac{1}{3}\Lambda_e v_{th} = \frac{1}{3}\tau_e v_{th}^2$. Combined with $\frac{1}{2}m_e^* v_{th}^2 = \frac{3}{2}k_B T$, we obtain

$$D_e = \frac{\tau_e k_B T}{m_e^*} = \frac{\mu_e k_B T}{e} \tag{6.135}$$

which is known as the *Einstein relation*. A similar equation holds also for the holes. In transient heat conduction, the *thermal diffusion length* is usually calculated by $l_{th} = \sqrt{\alpha t}$, where $\alpha = \kappa/\rho c_p$ is the thermal diffusivity and t is a characteristic time. The *diffusion length* for electrons is defined as $l_{dif} = \sqrt{D_e \tau_e}$, which is proportional to τ_e and v_{th}. The diffusion velocity is sometimes defined as $v_{dif} = l_{dif}/\tau_e = v_{th}/\sqrt{3}$. The factor of $\sqrt{3}$ reduction arises because the diffusion velocity is the average thermal velocity along one direction only. In semiconductors, charge transfer is a combined effect of the carrier drift and diffusion. Electron diffusion is not important for metals because of the large drift velocity given by the Fermi velocity, which changes little with temperature at moderate temperatures.

Through oxidation, lithography, diffusion and ion implantation, and metallization, semiconductor *p-n* junctions can be fabricated with microelectronics manufacturing technology [11]. A *p-n* junction consists of a *p*-type semiconductor, with a high hole concentration, joined with an *n*-type semiconductor, with a high electron concentration, as shown in Fig. 6.26.

If one compares Fig. 6.26a with Fig. 4.7b, the process looks similar to a binary diffusion. Because of the concentration gradient, holes will diffuse right and electrons will diffuse left. Diffusion causes the region near the interface to be depleted, so that there are fewer free holes on the left side and fewer free electrons on the right side of the depletion region. Keep in mind that electrons and holes are charged particles. As they leave the host material, ions of opposite charges are left behind. This results in a charge accumulation, as shown in Fig. 6.26a, that leads to a built-in potential in the depletion region that will inhibit further diffusion. As a consequence of this built-in potential, the energy in the *p*-doped region is raised relative to that in the *n*-doped region, as shown in Fig. 6.26b. The Fermi level is the same everywhere, and it is closer to the conduction band for the *n*-type and the valence band for the *p*-type.

Example 6.7 Prove that the Fermi energy in a *p-n* junction is independent of x at thermal equilibrium, as shown in Fig. 6.26b.

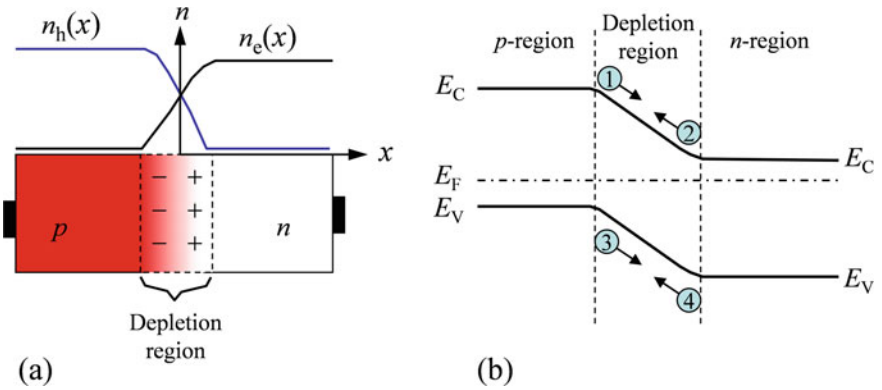

Fig. 6.26 Schematic of a p-n junction at thermal equilibrium. **a** The device and carrier concentrations, including the charge distribution in the depletion region. The width of the depletion region is exaggerated for clarity. **b** The energy band diagram for the $p - n$ junction near the depletion region. The dash-dotted line is the Fermi level. The four processes are (1) electron drift, (2) electron diffusion, (3) hole diffusion, and (4) hole drift

Solution Without any externally applied voltage, the current densities become

$$J_e = J_{e,drif} + J_{e,diff} = -en_e\mu_e\frac{dV_0}{dx} + eD_e\frac{dn_e}{dx} \tag{6.136}$$

$$J_h = J_{h,drif} + J_{h,diff} = -en_h\mu_h\frac{dV_0}{dx} - eD_h\frac{dn_h}{dx} \tag{6.137}$$

where V_0 is the built-in potential and the electric field is $-dV_0/dx$. Because a high potential V_0 means a smaller electron kinetic energy, we have

$$\frac{dV_0}{dx} = -\frac{1}{e}\frac{dE_C}{dx} = -\frac{1}{e}\frac{dE_V}{dx} \tag{6.138}$$

From Fig. 6.26, we see that the built-in electric field in the depletion region points toward the negative x-direction and thus $dV_0/dx > 0$; consequently, $dE_C/dx < 0$ and $dE_V/dx < 0$. Employing Eq. (6.110), we notice that

$$\frac{1}{n_e}\frac{dn_e}{dx} = \frac{1}{k_BT}\left(\frac{dE_F}{dx} - \frac{dE_C}{dx}\right) \quad \text{and} \quad \frac{1}{n_h}\frac{dn_h}{dx} = \frac{1}{k_BT}\left(\frac{dE_V}{dx} - \frac{dE_F}{dx}\right) \tag{6.139}$$

Substituting Eqs. (6.135), (6.138), and (6.139) into Eq. (6.136) and setting $J_e = 0$, we end up with

$$\frac{dE_F}{dx} = 0 \tag{6.140}$$

This equation can also be derived using Eq. (6.137) in the p-region; hence, the Fermi energy E_F is independent of x at thermal equilibrium.

A popular application of p-n junction is as a diode rectifier, which allows current to flow easily with a forward bias but becomes highly resistive when the bias is reversed. For the configuration shown in Fig. 6.26, a forward bias means that the electrical field is in the positive x-direction, opposite to the built-in field. Qualitatively, this can be understood as a forward bias removes the barrier for holes to diffuse right and for electrons to diffuse left. On the other hand, a reverse bias creates an even stronger barrier for these diffusion processes. Quantitatively, it can be shown that for an externally applied voltage V (positive for forward bias and negative for reverse bias), the current density can be expressed as

$$J = J_0\left[\exp\left(\frac{eV}{k_B T}\right) - 1\right] \tag{6.141}$$

where J_0 is the saturation current density, which depends on the diffusion coefficient, scattering time, number density, and other factors. Noting that

$$\frac{dJ}{dV} = \frac{eJ_0}{k_B T}\exp\left(\frac{eV}{k_B T}\right) \tag{6.142}$$

The electrical conductance increases with V for forward bias, and decreases to zero as $V \to -\infty$. It should be noted that in practice, the width of the depletion region is often less than 0.5 μm and the built-in potential may be around 1 V through the depletion region. There is actually a very large built-in field.

Heterojunction is a junction of dissimilar semiconductors with different bandgap energies. The energy band diagram can be very different from that shown in Fig. 6.26b. The Fermi energies can be different on each side. Bipolar transistors were invented in 1947 at Bell Labs. It is based on two p-n junctions arranged in a p-n-p or n-p-n configuration. Field-effect transistors (FETs) work on a different principle. Referring to Fig. 1.3 in Chap. 1, free electrons cannot move from the source to the drain because of the lack of free carriers in the p-type wafer. If a negative voltage is applied to the gate, electrons below the gate will be pushed even further, and there is still little chance for the electron to flow from the source to the drain. However, as soon as a positive voltage is applied to the gate, electrons will be attracted to the region below it and form a path for electricity to flow from the source to the drain. Furthermore, a transistor can amplify the signal since only a weak signal is necessary to the gate. Metal-oxide-semiconductor field-effect transistors (MOSFETs) have become the most important device in contemporary integrated circuits. Thermal management is important for such devices because of the local heating or hot spots where Fourier's law often fails to predict the temperature history. More discussion on nonequilibrium heat conduction will be given in Chap. 7. A brief discussion on photovoltaic devices will be given next.

6.8.4 Optoelectronic Applications

The photovoltaic effect is a direct energy conversion process in which electromagnetic radiation, incident upon a *p*-*n* junction, generates electron–hole pairs. The built-in electric field in the *p*-*n* junction tends to push the generated holes to the *p*-region and the generated electrons to the *n*-region, resulting in a reverse photocurrent. Solar cells and photovoltaic detectors have been developed and applied for over half a century. Thermophotovoltaic (TPV) devices have also been considered as energy conversion systems that allow recycling of the waste heat [126]. Figure 6.27 shows a typical TPV cell and the associated electrical circuit. When the incident radiation with a photon energy greater than the bandgap energy E_g of the cell material strikes the *p*-*n* junction, an electron–hole pair is generated at the location as each photon is absorbed. Carriers generated in the depletion region are swept by the built-in electric field and then collected by the electrodes at the ends of the cell, resulting in a drift current. For radiation absorbed near the depletion region, the minority carriers (electrons in the *p*-region and holes in the *n*-region) tend to diffuse toward the depletion region, yielding a diffusion current. If the load resistance R_L is zero, i.e., in the case of a short circuit, there is a photocurrent I_{sc} flowing in the circuit due to the combination of the diffusion and drift of charge carriers. The direction of this current is indicated

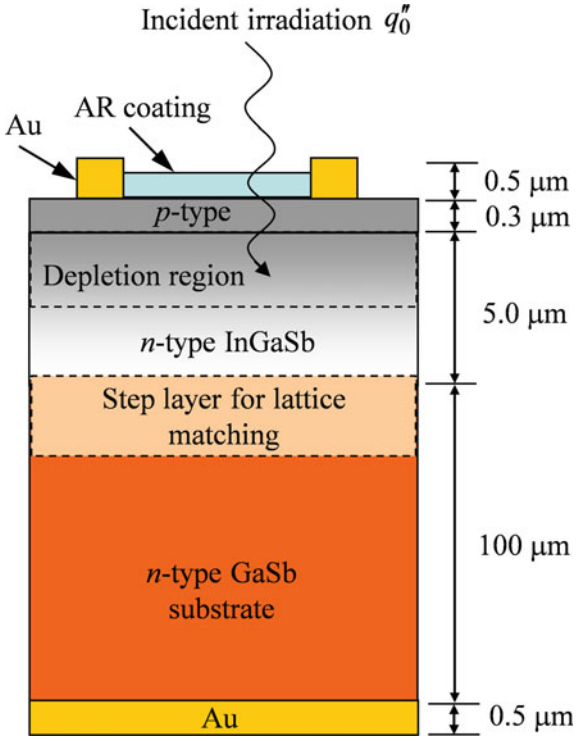

Fig. 6.27 Schematic of a typical TPV cell [126]

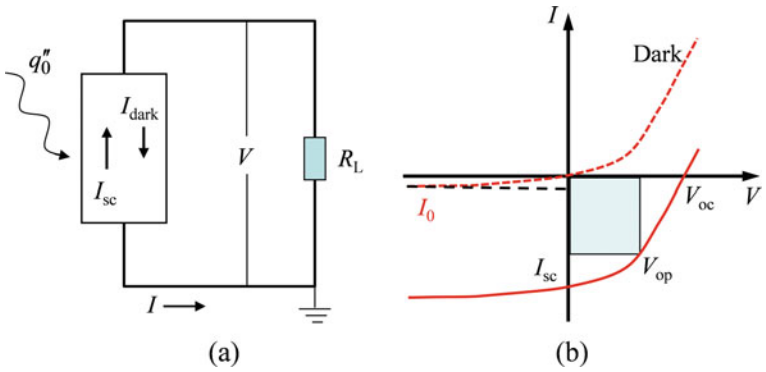

Fig. 6.28 Schematic of **a** the circuit diagram and **b** the I–V characteristics of a PV cell

on Fig. 6.28a. If the circuit is open, or the load resistance R_L approaches infinite, a positive open-circuit voltage is built up due to irradiation. This gives the maximum voltage $V = V_{max}$, when no current flows through the load, i.e., $I = 0$. When the load has a finite resistance R_L, a voltage V is developed, not only across the load but also across the photovoltaic cell. This voltage reduces the built-in potential of the cell as if a forward bias is applied to the *p-n* junction. Subsequently, the diffusion of minority carriers produces a forward current, which is called the *dark current* in photovoltaic devices. The current I flowing through the load resistor becomes

$$I = -I_{sc} + I_0\left[\exp\left(\frac{eV}{k_B T}\right) - 1\right] \tag{6.143}$$

The first term on the right is the photocurrent, or the short-circuit current, which depends on the incident photon flux, quantum efficiency, as well as the transport properties. The second term on the right is the dark current I_{dark} with I_0 being the saturation current, as previously given in Eq. (6.141) based on the current density. The dark current is zero, when $V = 0$, as shown in Fig. 6.28b. Consider the photovoltaic cell, shown in Figs. 6.27 and 6.28a. If the total (net) irradiance $q_0'' = 0$, then without bias voltage, $I = I_{sc} = 0$, which is the condition of thermal equilibrium. When $q_0'' \neq 0$, which can be the situation where the photovoltaic cell is exposed to a high-temperature emitter, the I–V curve is shifted down by I_{sc} as shown in Fig. 6.28b. The output power is determined by the product of $|IV|$, which can be optimized to yield the maximum output at the optimal point. Basu et al. [126] provided an extensive review of the operation principle and the state of the art in TPV technology, as well as the potential application of microscale radiative heat transfer for performance improvement. Near-field TPV will be further discussed in Chap. 10.

Light-emitting diodes (LEDs) are based on *p-n* junctions as well but with direct gap semiconductors. At low forward bias voltages, the recombination processes are

essentially nonradiative. At high forward bias voltages, however, radiative recombination results in the emission of photons; this phenomenon is called electroluminescence. The emission is a spontaneous process and is therefore incoherent. Depending on the materials used and their bandgaps, LEDs can emit in the ultraviolet, visible, and infrared regions. It should be noted that electroluminescence can also be employed to create a refrigeration effect [127].

Semiconductor lasers are based on the stimulated emission process, as discussed in Chap. 3, and have numerous important applications due to their small size, portability, and ease of operation. Semiconductor lasers have been used in laser printers, optical fiber communication, CD reading/writing, and so forth. The key is to create population inversion so that lasing can occur. Quantum well lasers, based on quantum confinement, offer significant advantages over conventional semiconductor lasers, such as low threshold current, high output power, high speed, and so forth. Further explanation of the optical and electronic characteristics of semiconductor lasers can be found from the books of Sze [11] and Zory [128], for example.

6.9 Summary

This chapter began with an introduction to the atomic structures, chemical bonds, and crystal lattices. Emphasis was placed on electronic band structures and phonon dispersion relations, allowing one to gain a deeper knowledge of solid-state physics, beyond the previous chapter. Electronic band structures of metal, semiconductor, and 2D materials (especially graphene) are introduced. Phonon dispersion and scattering mechanisms are presented, along with some coverage of coherent phonons. The basic concepts of atomistic simulations based on first principles and molecular dynamics methods are delineated with extensive references of recent literature. Photoelectric effect, thermionic emission, and field emission were described in subsequent sections to stress the interrelation between these phenomena. The basic electrical transport processes in semiconductors, such as number density, mobility, electrical conductivity, charge diffusion, and photoconductivity, were explained. The *p-n* junction was discussed along with applications, such as photovoltaic cells, thermophotovoltaics, LEDs, luminescent refrigeration, and semiconductor lasers.

Problems

6.1. Consider a phosphorus-doped 250-μm-thick silicon wafer, with a doping concentration of 10^{17} cm^{-3}. The applied current is 10 mA and the magnetic induction is 0.5 T.

(a) Determine the Hall coefficient and the Hall voltage, assuming there is only one type of carriers.

(b) For a chip of area 1×1 cm^2 and resistivity $0.075\,\Omega$ cm, what is the voltage drop along the direction of current flow?

6.2. Consider the Hall experiment arranged in Fig. 6.1, under steady-state operation and with uniform magnetic field. Assume a current is flowing in the y-direction.

(a) Show that $v_x = -(e\tau/m_e)E_x - \omega_c \tau v_y$ and $v_y = -(e\tau/m_e)E_y + \omega_c \tau v_x$, when the current is carried by electrons. Here, τ is the relaxation time, v_x and v_y are the electron drift velocities in the x- and y-directions, respectively, and $\omega_c = eB/m_e$ is called the *cyclotron frequency*.

(b) Prove Eq. (6.1) by setting $v_y = 0$.

6.3. Express the electron configurations for Ag and Au. Based on the orbital occupation of outer electrons, discuss the similarities in their chemical and electrical properties.

6.4. Express the electron configurations for Ca and Zn. Based on the orbital occupation of outer electrons, discuss the similarities in their chemical and electrical properties.

6.5. Give a general discussion of insulators, semiconductors, and metals. Explain why glass (SiO_2) is transparent, silicon wafers appear dark, and aluminum foils look bright. What are the types of chemical bonds in SiO_2, Si, and Al?

6.6. How many billiard balls can you pack in a basket with a volume of 0.25 m^3? Assume that the balls are rigid spheres with a diameter $d = 43$ mm and mass $m = 46$ g. Arrange the spheres in a crystal lattice according to the diamond, simple cubic, bcc, fcc, and hcp structures. What is the total weight for each arrangement? [Hint: Show that for close-packed spheres, the fraction of volume occupied by the spheres is $\sqrt{3}\pi/16 \approx 0.340$ (diamond), $\pi/6 \approx 0.524$ (simple cubic), $\sqrt{3}\pi/8 \approx 0.680$ (bcc), and $\sqrt{2}\pi/6 \approx 0.740$ (fcc or hexagonal close - packed).]

6.7. (a) Count the number of atoms inside a unit cell of $YBa_2Cu_3O_7$ as shown in Fig. 6.6d, and confirm that it is the same as that in the basis.

(b) Find the density of $YBa_2Cu_3O_7$ crystal based on the dimensions of the unit cell, noting that the molecular weight $M = 88.9$ (Y), 137.3 (Ba), 63.6 (Cu), and 16.0 (O) kg/kmol.

6.8 (a) Calculate the diameter and cross-sectional area for CNTs with chiral indices $(m, n) = (5, 5)$, (8, 8), (10, 10), (10, 20), and (20, 40).

(b) Take (40, 40) SWNTs of 10-μm length, with a thermal conductivity $\kappa = 3200$ W/m K at room temperature. Align sufficient nanotubes to make a bundle with a diameter of 1 μm; how many CNTs are needed?

(c) Neglect the effect of interface and defects on the thermal conductivity. What is the heat transfer rate if the temperatures at both ends are 320 and 300 K?

(d) Compare the heat transfer rate if the CNT is replaced by a Si nanowire of 1-μm diameter and 10-μm length.

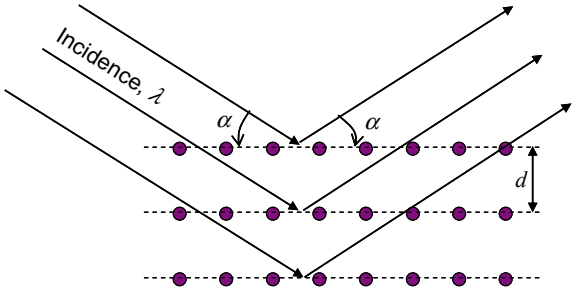

Fig. 6.29 Schematic of Bragg's x-ray diffraction experiment

6.9. The interatomic potential for a KBr crystal can be expressed as $\phi(r) = -\frac{\alpha_M e^2}{4\pi\varepsilon_0 r} + C\left(\frac{a}{r}\right)^m$, where α_M is the Madelung constant, which is 1.748 for crystals with NaCl structure, $\varepsilon_0 = 8.854 \times 10^{-12}$ C^2/J m is the electric permittivity of vacuum, $a = 6.60 \times 10^{-10}$ m is the lattice constant, $m = 8.85$, and $C = 2.65 \times 10^{-21}$ J for KBr. Note that r is in meter.

(a) Plot the attractive potential, the repulsive potential, and the combined potential in eV as a function of r in Å.
(b) Find the equilibrium distance, which should be the nearest distance between K^+ and Br^- ions.
(c) At the equilibrium distance, what are the attractive and repulsive forces between each ion pair?

6.10. Bragg's x-ray diffraction formula relates the angle α of diffraction maximum and the x-ray wavelength λ as follows: $2d \sin\alpha = n\lambda$, where n is the refractive index that can be taken as unity in the x-ray region, d is the spacing between adjacent layers of atoms, and α is measured between the incidence and the crystal plane, as shown in Fig. 6.29. This formula can be understood by the constructive interference between the two layers.

(a) To measure a spacing $d = 3.12$ Å, what is the maximum wavelength λ that can still be used to perform the experiment successfully?
(b) In an x-ray experiment, $\lambda = 1.5$ Å. Assume that the errors in λ and n are negligible. How accurately must one determine α in order to measure the spacing with an uncertainty of 0.01 Å?

6.11. Using Eq. (6.10) to show that the reciprocal lattice of a hexagon is also a hexagon, as shown in Fig. 6.3d. Calculate the volumes of the direct and reciprocal lattices in terms of a and c.

6.12. Use the Kronig–Penney model to solve the Schrödinger equation for an electron in a square-well array. Referring to Fig. 6.10, assume that the potential function is $U(x) = 0$ at $0 \le x \le (a-b)/2$ and $(a+b)/2 \le x \le a$, and $U(x) = U_0 > 0$ at $(a-b)/2 \le x \le (a+b)/2$. Note that $x = 0$ at the core of atom location and the potential is periodic. Find the conditions for the solutions to exist. For simplicity, you may now assume $b \to 0$ and $U_0 \to \infty$

to obtain the relation for $E(k)$. Plot this function to illustrate the electronic band structure.

6.13. Discuss the difference between interband transitions and transitions that occur within a band for copper. Explain why copper appears reddish brown.

6.14. What is the difference between a direct-bandgap semiconductor and an indirect bandgap semiconductor? Why are Si and GaAs wafers opaque to the visible light?

6.15. Plot the band structure of graphene according to Eq. (6.32) and discuss the main features.

6.16. Why we say electrons in pure graphene or graphene whose Fermi level is zero are massless? How does the free-electron mass in graphene changes with bias voltage?

6.17. Derive Eqs. (6.41) and (6.42) first. Then, plot the phonon dispersion curves for a diatomic chain with mass ratio m_1/m_2 equal to 1, 2, 3, and 4. What happens when $m_1/m_2 = 1$?

6.18. Approximate $k_{\max} = \pi/a$, and find the group velocities of LA and TA phonons for Si and SiC at $k = 0.3k_{\max}$, using Fig. 6.19. What is the phase speed at $k = 0.3k_{\max}$ for LO phonon in SiC? [Hint: Convert the unit of ω from cm^{-1} to rad/s first.]

6.19. Perform a literature search to discuss phonon–phonon scattering mechanism. When will four-phonon scattering be important? When can four-phonon scattering be neglected?

6.20. Prove Eqs. (6.96) and (6.97). Assume that $\psi = 0.4\,\mathrm{eV}$ and $E_F = 3\,\mathrm{eV}$, estimate the error in Eq. (6.96) caused by approximating the Fermi–Dirac distribution with the Maxwell–Boltzmann distribution in the numerical evaluation.

6.21. Clearly explain the differences between thermionic emission and field emission.

6.22. For a gallium-doped silicon with $N_A = 5 \times 10^{16}\ \mathrm{cm}^{-3}$, use the information from Example 6.6 to calculate the number density of electrons and holes from 300 to 1000 K. Assume the effect of impurity on the mobility can be neglected, so that $\mu_e = 1450\,\mathrm{cm}^2/\mathrm{V\ s}$ for electrons, and $\mu_h = 500\,\mathrm{cm}^2/\mathrm{V\ s}$ for holes at 300 K. Determine the electrical resistivity of the doped silicon from 300 to 1000 K.

6.23. For a single-type doped silicon with $\mu_e = 1350\,\mathrm{cm}^2/\mathrm{V\ s}$ and $\mu_h = 450\,\mathrm{cm}^2/\mathrm{V\ s}$ at 400 K, the Hall coefficient is zero. Is this semiconductor n-type or p-type? What is the impurity concentration? [Hint: Use the parameters given in Example 6.6.]

6.24. For a single-type doped silicon with $\mu_e = 1350\,\mathrm{cm}^2/\mathrm{V\ s}$, $\mu_h = 450\,\mathrm{cm}^2/\mathrm{V\ s}$, and $N_{th} = 2\times10^{10}\ \mathrm{cm}^{-3}$, calculate and plot the Hall coefficient for p-type doping, with N_A ranging from 0 to $2 \times 10^{12}\ \mathrm{cm}^{-3}$. Discuss, without calculation, the trend with n-type doping.

6.25. For a phosphorus-doped silicon, $N_D = 2 \times 10^{15}\ \mathrm{cm}^{-3}$, $\mu_e = 1350\,\mathrm{cm}^2/\mathrm{V\ s}$, and $\mu_h = 450\,\mathrm{cm}^2/\mathrm{V\ s}$ at 300 K. Use the parameters from Example 6.6 as needed.

(a) Calculate the thermal velocity and the diffusion length for the electrons and holes at room temperature.
(b) Find the electrical conductivity at room temperature.
(c) Plot the thermal velocity and wavelength as a function of temperature.

6.26. Plot the J–V curve of a p-n junction based on Eq. (6.141), using dimensionless groups J/J_0 and eV/k_BT as the axes. Discuss the meaning of saturation current (density).

6.27. Based on the I–V curve for a photovoltaic cell shown in Fig. 6.28b, explain how to determine the open voltage. How to determine the optimal operating point? The I–V curve with irradiation is shifted downward, what if the cell is facing a heat sink at lower temperatures?

References

1. N.W. Ashcroft, N.D. Mermin, *Solid State Physics* (Harcourt College Publishers, Fort Worth, TX, 1976)
2. C. Kittel, *Introduction to Solid State Physics*, 7th edn. (Wiley, New York, 1996)
3. J. E. Avron, D. Osadchy, R. Seiler, A topological look at the quantum Hall effect. Phys. Today, 38–42 (2003)
4. G.A. Prinz, Magnetoelectronics. Science **282**, 1660–1663 (1998)
5. K. von Klitzing, G. Dorda, M. Pepper, New method for high-accuracy determination of the fine-structure constant based on quantized hall resistance. Phys. Rev. Lett. **45**, 494–497 (1980); K. von Klitzing, The quantized Hall effect. Rev. Mod. Phys. **58**, 519–530 (1986)
6. A. Hartland, The quantum Hall effect and resistance standards. Metrologia **29**, 175–190 (1992)
7. J.P. Eisenstein, H.L. Strörmer, The fractional quantum Hall effect. Science **248**, 1510–1516 (1990)
8. C. Strohm, G.L.J.A. Rikken, P. Wyder, Phenomenological evidence for the phonon Hall effect. Phys. Rev. Lett. **95**, 155901 (2005)
9. J.E. Lay, *Statistical Mechanics and Thermodynamics of Matter* (Harper Collins Publishers, New York, 1990)
10. F. Yang, J.H. Hamilton, *Modern Atomic and Nuclear Physics* (McGraw-Hill, New York, 1996)
11. S.M. Sze, *Physics of Semiconductor Devices*, 2nd edn. (Wiley, New York, 1981); S.M. Sze, *Semiconductor Devices: Physics and Technology*, 2nd edn. (Wiley, New York, 2002)
12. R.J. Cava, Structure chemistry and the local charge picture of copper oxide superconductors. Science **247**, 656–662 (1990)
13. B.I. Choi, Z.M. Zhang, M.I. Flik, T. Siegrist, Radiative properties of Y-Ba-Cu–O films with variable oxygen content. J. Heat Transfer **114**, 958–964 (1992)
14. M.S. Dresselhaus, P.C. Eklund, Phonons in carbon nanotubes. Adv. Phys. **49**, 705–814 (2000)
15. M.A. Omar, *Elementary Solid State Physics: Principles and Applications* (Addison-Wesley, New York, 1975)
16. B. Segall, Fermi surface and energy band of copper. Phys. Rev. **125**, 109–122 (1962)
17. G.A. Burdick, Energy band structure of copper. Phys. Rev. **129**, 138–150 (1963)
18. R.E. Hummel, *Electronic Properties of Materials* (Springer, Berlin, 1993)
19. M.L. Cohen, T.K. Bergstresser, Band structure and pseudopotential from factors for fourteen semiconductors of the diamond and zinc-blende structures. Phys. Rev. **141**, 789–796 (1966)
20. F. Herman, W.E. Spicer, Spectral analysis of photoemission yields in GaAs and related crystals. Phys. Rev. **174**, 906–908 (1968)

21. M.L. Cohen, J.R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors* (Springer, Berlin, 1988)
22. A.H. Castro Neto, F. Guinea, N.M.R. Peres, K.S.Novoselov, A.K. Geim, The electronic properties of graphene. Rev. Mod. Phys. **81**, 109–162 (2009)
23. D. Akinwande, N. Petrone, J. Hone, Two-dimensional flexible nanoelectronics. Nat. Commun. **5**, 5678 (2014)
24. P. Avouris, C. Dimitrakopoulos, Graphene: synthesis and applications. Mater. Today **15**, 86–97 (2012)
25. A.A. Balandin, Thermal properties of graphene and nanostructured carbon materials. Nat. Mater. **10**, 569–581 (2011)
26. F. Bonaccorso, L. Colombo, G. Yu, M. Stoller, V. Tozzini, A.C. Ferrari, R.S. Ruoff, V. Pellegrini, Graphene, related two-dimensional crystals, and hybrid systems for energy conversion and storage. Science **347**, 1246501 (2015)
27. X. Luo, T. Qiu, W. Lu, Z. Ni, Plasmons in graphene: recent progress and applications. Mater. Sci. Eng. R **74**, 351–376 (2013)
28. S.B. Trickey, F. Müller-Plathe, G.H. Diercksen, J.C. Boettger, Interplanar binding and lattice relaxation in a graphite delayer. Phys. Rev. B **45**, 4460–4468 (1992)
29. D.W. Boukhvalov, M.I. Katsnelson, A.I. Lichtenstein, Hydrogen on graphene: electronic structure, total energy, structural distortions and magnetism from first-principles calculations. Phys. Rev. B **77**, 035427 (2008)
30. K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, M.I. Katsnelson, I.V. Grigorieva, S.V. Dubonos, A.A. Firsov, Two-dimensional gas of massless Dirac fermions in graphene. Nature **438**, 197–200 (2005)
31. T. Stauber, N.M.R. Peres, F. Guinea, Electronic transport in graphene: a semiclassical approach including midgap states. Phys. Rev. B **76**, 205423 (2007)
32. Y.-W. Tan, Y. Zhang, K. Bolotin, Y. Zhao, S. Adam, E.H. Hwang, S. Das Sarma, H.L. Stormer, P. Kim, Measurement of scattering rate and minimum conductivity in graphene. Phys. Rev. Lett. **99**, 246803 (2007)
33. S.V. Morozov, K.S. Novoselov, M.I. Katsnelson, F. Schedin, D.C. Elias, J.A. Jaszczak, A.K. Geim, Giant intrinsic carrier mobilities in graphene and its bilayer. Phys. Rev. Lett. **100**, 016602 (2008)
34. M. Xu, T. Liang, M. Shi, H. Chen, Graphene-like two-dimensional materials. Chem. Rev. **113**, 3766–3798 (2013)
35. M. Ashton, J. Paul, S.B. Sinnott, R.G. Hennig, Topology-scaling identification of layered solids and stable exfoliated 2D materials. Phys. Rev. Lett. **118**, 106101 (2017)
36. X. Gu, Y. Wei, X. Yin, B. Li, R. Yang, Colloquium: phononic thermal properties of two-dimensional materials. Rev. Mod. Phys. **90**, 041002 (2018)
37. Q. Cai, D. Scullion, W. Gan et al., High thermal conductivity of high-quality monolayer boron nitride and its thermal expansion. Sci. Adv. **5**, eaav0129 (2019)
38. F. Xia, H. Wang, D. Xiao, M. Dubey, A. Ramasubramaniam, Two-dimensional material nanophotonics. Nat. Photon. **8**, 899–907 (2014)
39. D. Chimene, D.L. Alge, A.K. Gaharwar, Two-dimensional nanomaterials for biomedical applications: emerging trends and future prospects. Adv. Mater. **27**, 7261–7284 (2015)
40. M. Born, K. Huang, *Dynamic Theory of Crystal Lattices* (Oxford University Press, London, 1954)
41. G.P. Srivastava, *The Physics of Phonons* (Adam Hilger, Bristol, 1990)
42. J.M. Ziman, *Electrons and Phonons* (Oxford University Press, Oxford, 1960)
43. B.N. Brockhouse, Lattice vibration in silicon and germanium. Phys. Rev. Lett. **2**, 256–258 (1959)
44. G. Dolling, in *Second Symposium on Inelastic Scattering of Neutrons in Solids and Liquids*, vol. II, Chalk River, Canada (IAEA, Vienna, 1963), p. 37
45. R. Tubino, L. Piseri, G. Zerbi, Lattice dynamics and spectroscopic properties by a valence force potential of diamondlike crystals: C, Si, Ge, and Sn. J. Chem. Phys. **56**, 1022–1039 (1972)

46. D.W. Feldman, J.H. Parker Jr., W.J. Choyke, L. Patrick, Phonon dispersion curves by Raman scattering in SiC, polytypes 3C, 4H, 6H, 15R, and 21R. Phys. Rev. **173**, 787–793 (1968)
47. A.M. Greenstein, S. Graham, Y.C. Hudiono, S. Nair, Thermal properties and lattice dynamics of polycrystalline MFI zeolite films. Nanoscale Microscale Thermophys. Eng. **10**, 321–331 (2006)
48. S.L. Shindé, G.P. Srivastava (eds.), *Length-Scale Dependent Phonon Interactions* (Springer, New York, 2014)
49. C. Dames, Ultrahigh thermal conductivity confirmed in boron arsenide. Science **361**, 549–550 (2018)
50. T.L. Feng, L. Lindsay, X.L. Ruan, Four-phonon scattering significantly reduces intrinsic thermal conductivity of solids. Phys. Rev. B **96**, 161201(R) (2017)
51. D.J. Ecsedy, P.G. Klemens, Thermal conductivity of dielectric crystals due to four-phonon processes and optical modes. Phys. Rev. B **15**, 5957–5962 (1977)
52. Z.M. Zhang, Surface temperature measurement using optical techniques. Annu. Rev. Heat Transfer **11**, 351–411 (2000)
53. J. Lee, T. Beechem, T.L. Wright, B.A. Nelson, S. Graham, W.P. King, Electrical, thermal, and mechanical characterization of silicon microcantilever heaters. J. Microelectromech. Sys. **15**, 1644–1655 (2006)
54. M. Maldovan, Sound and heat revolutions in phononics. Nature **503**, 209–217 (2013); M. Maldovan, Narrow low-frequency spectrum and heat management by thermocrystals. Phys. Rev. Lett. **110**, 025902 (2014); M. Maldovan, Phonon wave interference and thermal bandgap materials. Nat. Mater. **14**, 667–674 (2015)
55. R. Anufriev, A. Ramiere, J. Maire, M. Nomura, Heat guiding and focusing using ballistic phonon transport in phononic nanostructures. Nat. Commun. **8**, 15505 (2017)
56. G. Xie, D. Ding, G. Zhang, Phonon coherence and its effect on thermal conductivity of nanostructures. Adv. Phys: X **3**, 1480417 (2018)
57. Y. Xiao, Q. Chen, D. Ma, N. Yang, Q. Hao, Phonon transport within periodic porous structures—from classical phonon size effects to wave effects. ES Mater. Manuf. **5**, 2–18 (2019)
58. R. Merlin, Generating coherent THz phonons with light pulses. Solid State Commun. **102**, 207–220 (1997)
59. C. Colvard, R. Merlin, M.V. Klein, A.C. Gossard, Observation of folded acoustic phonons in a semiconductor superlattice. Phys. Rev. Lett. **45**, 298–301 (1980)
60. Y. Ezzahri, S. Grauby, J.M. Rampnoux, H. Michel, G. Pernot, W. Claeys, S. Dilhaire, C. Rossignol, G. Zeng, A. Shakouri, Coherent phonons in Si/SiGe superlattices. Phys. Rev. B **75**, 195309 (2007)
61. A.A. Maznev, F. Hofmann, A. Jandl, K. Esfarjani, M.T. Bulsara, E.A. Fitzgerald, G. Chen, K.A. Nelson, Lifetime of sub-THz coherent acoustic phonons in a GaAs-AlAs superlattice. Appl. Phys. Lett. **102**, 041901 (2013)
62. F. He, W. Wu, Y. Guo, Direct measurement of coherent thermal phonons in Bi_2Te_3/Sb_2Te_3 superlattice. Appl. Phys. A **122**, 777 (2016)
63. R. Anufriev, M. Nomura, Reduction of thermal conductance by coherent phonon scattering in two-dimensional phononic crystals of different lattice types. Phys. Rev. B **93**, 045410 (2016); *ibid.* Heat conduction engineering in pillar-based phononic crystals. Phys. Rev. B **95**, 155432 (2017)
64. M.N. Luckyanova, J. Garg, K. Esfarjani, A. Jandl, M.T. Bulsara, A.J. Schmidt, A.J. Minnich, S. Chen, M.S. Dresselhaus, Z. Ren, E.A. Fitzgerald, G. Chen, Coherent phonon heat conduction in superlattices. Science **338**, 936–939 (2012)
65. J. Ravichandran, A.K. Yadav, R. Cheaito, P.B. Rossen, A. Soukiassian, S.J. Suresha, J.C. Duda, B.M. Foley, C.-H. Lee, Y. Zhu, A.W. Lichtenberger, J.E. Moore, D.A. Muller, D.G. Schlom, P.E. Hopkins, A. Majumdar, R. Ramesh, M.A. Zurbuchen, Crossover from incoherent to coherent phonon scattering in epitaxial oxide superlattices. Nat. Mater. **13**, 168–172 (2014)
66. J.K. Yu, S. Mitrovic, D. Tham, J. Varghese, J.R. Heath, Reduction of thermal conductivity in phononic nanomesh structures. Nat. Nanotech. **5**, 718–721 (2010)

67. P.E. Hopkins, C.M. Reinke, M.F. Su, R.H. Olsson III, E.A. Shaner, Z.C. Leseman, J.R. Serrano, L.M. Phinney, I. El-Kady, Reduction in the thermal conductivity of single crystalline silicon by phononic crystal patterning. Nano Lett. **11**, 107–112 (2011)
68. E. Dechaumphai, R. Chen, Thermal transport in phononic crystals: the role of zone folding effect. J. Appl. Phys. **111**, 073508 (2012)
69. A. Jain, Y.J. Yu, A.J.H. McGaughey, Phonon transport in periodic silicon nanoporous films with feature sizes greater than 100 nm. Phys. Rev. B **87**, 195301 (2013)
70. N.K. Ravichandran, A.J. Minnich, Coherent and incoherent thermal transport in nanomeshes. Phys. Rev. B. **89**, 205432 (2014)
71. J. Lee, W. Lee, G. Wehmeyer, S. Dhuey, D.L. Olynick, S. Cabrini, C. Dames, J.J. Urban, P.D. Yang, Investigation of phonon coherence and backscattering using silicon nanomeshes. Nat. Commun. **8**, 14054 (2017)
72. N. Zen, T.A. Puurtinen, T.J. Isotalo, S. Chaudhuri, I.J. Maasilta, Engineering thermal conductance using a two-dimensional phononic crystal. Nat. Commun. **5**, 3435 (2014)
73. J. Maire, R. Anufriev, R. Yanagisawa, A. Ramiere, S. Volz, M. Nomura, Heat conduction tuning by wave nature of phonons. Sci. Adv. **3**, e1700027 (2017)
74. M.R. Wagner, B. Graczykowski, J.S. Reparaz, A. El Sachat, M. Sledzinska, F. Alzina, C.M.S. Torres, Two-dimensional phononic crystals: disorder matters. Nano Lett. **16**, 5661–5668 (2016)
75. D. Broido, M. Malorny, G. Birner, N. Mingo, D. Stewart, Intrinsic lattice thermal conductivity of semiconductors from first principles. Appl. Phys. Lett. **91**, 231922 (2007)
76. T. Luo, G. Chen, Nanoscale heat transfer—from computation to experiment. Phys. Chem. Chem. Phys. **15**, 3389–3412 (2013)
77. G. Chen, Multiscale simulation of phonon and electron thermal transport. Annu. Rev. Heat Transfer **17**, 1–8 (2014)
78. K. Esfarjani, J. Garg, G. Chen, Modeling heat conduction from first principles. Annu. Rev. Heat Transfer **17**, 9–47 (2014)
79. L. Lindsay, First principles Peierls-Boltzmann phonon thermal transport: a topical review. Nanosc. Microsc. Thermophys. Eng. **20**, 67–84 (2016)
80. H. Bao, J. Chen, X. Gu, B. Cao, A review of simulation methods in micro/nanoscale heat conduction. ES Energy Environ. **1**, 16–55 (2018)
81. A.J.H. McGaughey, A. Jain, H.Y. Kim, B. Fu, Phonon properties and thermal conductivity from first principles, lattice dynamics, and the Boltzmann transport equation. J. Appl. Phys. **125**, 011101 (2019)
82. P.K. Schelling, S.R. Phillpot, P. Keblinski, Comparison of atomic-level simulation methods for computing thermal conductivity. Phys. Rev. B **65**, 144306 (2002); ibid, Phonon wave-packet dynamics at semiconductor interfaces by molecular dynamics simulation. Appl. Phys. Lett. **80**, 2484–2487 (2002)
83. A.J.H. McGaughey, M. Kariany, Phonon transport in molecular dynamics simulations: formulation and thermal conductivity prediction. Adv. Heat Transfer **39**, 169–255 (2006)
84. A.J.H. McGaughey, J.M. Larkin, Prediction of phonon properties from equilibrium molecular dynamics simulations. Annu. Rev. Heat Transfer **17**, 49–87 (2014)
85. J. Shiomi, Nonequilibrium molecular dynamics method for lattice heat conduction calculations. Annu. Rev. Heat Transfer **17**, 177–203 (2014)
86. P. Hohenberg, W. Kohn, Inhomogeneous electron gas. Phys. Rev. **136**, B864–B871 (1964); W. Kohn, L.J. Sham, Self-consistent equation including exchange and correlation effects. Phys. Rev. **140**, A1133–A1138 (1965)
87. R.O. Jones, Density functional theory: its origins, rise to prominence, and future. Rev. Mod. Phys. **87**, 897–923 (2015)
88. P.J. Hasnip, K. Refson, M.I.J. Probert, J.R. Yates, S.J. Clark, C.J. Pickard, Density functional theory in the solid state. Phil. Trans. R. Soc. A **372**, 20130270 (2014)
89. B.G. Johnson, P.M.W. Gill, J.A. Pople, The performance of a family of density functional methods. J. Chem. Phys. **98**, 5612–5626 (1993)
90. M. Kaviany, *Heat Transfer Physics*, 2nd edn. (Cambridge University Press, Cambridge, 2014)

91. S. Baroni, S. de Gironcoli, A. Dal Corso, Phonons and related crystal properties from density-functional perturbation theory. Rev. Mod. Phys. **73**, 515–562 (2001)
92. X. Gonze, J.-P. Vigneron, Density-functional approach to nonlinear-response coefficients of solids. Phys. Rev. B **39**, 13120–13128 (1989)
93. D. Vanderbilt, S.H. Taole, S. Narasimhan, Anharmonic elastic and phonon properties of Si. Phys. Rev. B **40**, 5657–5668 (1989)
94. K. Parlinski, Z.Q. Li, Y. Kawazoe, First-principles determination of the soft mode in cubic ZrO_2. Phys. Rev. Lett. **78**, 4063–4066 (1997)
95. K. Esfarjani, H.T. Stokes, Method to extract anharmonic force constants from first principles calculations. Phys. Rev. B **77**, 144112 (2008)
96. A. Togo, I. Tanaka, First principles phonon calculations in materials science. Scr. Mater. **108**, 1–5 (2015); A. Togo, L. Chaput, I. Tanaka, Distributions of phonon lifetimes in Brillouin zones. Phys. Rev. B **91**, 094306 (2015)
97. T. Tadano, S. Tsuneyuki, First-principles lattice dynamics method for strongly anharmonic crystals. J. Phys. Soc. Jpn. **87**, 041015 (2018); *ibid*, Self-consistent phonon calculations of lattice dynamical properties in cubic $SrTiO_3$ with first-principles anharmonic force constants. Phys. Rev. B **92**, 054301 (2015)
98. B. Liao, B. Qiu, J. Zhou, S. Huberman, K. Esfarjani, G. Chen, Significant reduction of lattice thermal conductivity by the electron-phonon interaction in silicon with high carrier concentrations: a first-principles study. Phys. Rev. Lett. **114**, 115901 (2015)
99. Y.Z. Liu, Y.H. Feng, Z. Huang, X.X. Zhang, Thermal conductivity of 3D boron-based covalent organic frameworks from molecular dynamics simulations. J. Phys. Chem. C **120**, 17060–17068 (2016)
100. LAMMPS Molecular Dynamics Simulator, Sandia National Laboratory. https://lammps.sandia.gov/, Accessed 13 Dec 2019
101. Z. Wang, S. Safarkhani, G. Lin, X. Ruan, Uncertainty quantification of thermal conductivities from equilibrium molecular dynamics simulations. Int. J. Heat Mass Transfer **112**, 267–278 (2017)
102. B.-L. Huang, M. Kaviany, Ab initio and molecular dynamics predictions for electron and phonon transport in bismuth telluride. Phys. Rev. B **77**, 125209 (2008)
103. K. Esfarjani, G. Chen, H.T. Stokes, Heat transport in silicon from first-principles calculations. Phys. Rev. B **84**, 085204 (2011)
104. A. Marcolongo, P. Umari, S. Baroni, Microscopic theory and quantum simulation of atomic heat transport. Nat. Phys. **12**, 80–84 (2015)
105. C. Carbogno, R. Ramprasad, M. Scheffler, Ab initio Green-Kubo approach for the thermal conductivity of solids. Phys. Rev. Lett. **118**, 175901 (2017)
106. J.S. Tse, N.J. English, K. Yin, T. Iitaka, Thermal conductivity of solids from first-principles molecular dynamics calculations. J. Phys. Chem. **122**, 10682–10690 (2018)
107. S. Hüfner, *Photoelectron Spectroscopy: principles and Applications*, 3rd edn. (Springer, Berlin, 2003)
108. G. N. Hatsopoulos, E.P. Gyftopoulos, *Thermionic Energy Conversion*, vol. 1 (1973); Vol. 2. (MIT Press, Cambridge, MA, 1979)
109. G.D. Mahan, Thermionic refrigeration. J. Appl. Phys. **76**, 4362–4366 (1994); G.D. Mahan, L.M. Woods, Multilayer thermionic refrigeration. Phys. Rev. Lett. **80**, 4016–4019 (1998); G.D. Mahan, J.O. Sofo, M. Barkowiak, Multilayer thermionic refrigerator and generator. J. Appl. Phys. **83**, 4683–4689 (1998)
110. A. Shakouri, J.E. Bowers, Heterostructure integrated thermionic coolers. Appl. Phys. Lett. **71**, 1234–1236 (1997); A. Shakouri, C. LaBounty, J. Piprek, P. Abraham, J.E. Bowers, Thermionic emission cooling in single barrier heterostructures. Appl. Phys. Lett. **74**, 88–89 (1999)
111. D. Vashaee, A. Shakouri, Nonequilibrium electrons and phonons in thin film thermionic coolers. Microscale Thermophys. Eng. 8, 91–100 (2004); D. Vashaee, A. Shakouri, Electronic and thermoelectric transport in semiconductor and metallic superlattices. J. Appl. Phys. 95, 1233–1245 (2004)

112. T. Zeng, G. Chen, Interplay between thermoelectric and thermionic effects in heterostructures. J. Appl. Phys.**92**, 3152–3161 (2002); T. Zeng, G. Chen, Nonequilibrium electron and phonon transport and energy conversion in heterostructures. Microelectron. J. **34**, 201–206 (2003)
113. Y. Hishinuma, T.H. Geballe, B.Y. Moyzhes, T.W. Kenny, Refrigeration by combined tunneling and thermionic emission in vacuum: use of nanometer scale design. Appl. Phys. Lett. **78**, 2572–2574 (2001); Y. Hishinuma, T.H. Geballe, B.Y. Moyzhes, T.W. Kenny, Measurements of cooling by room-temperature thermionic emission across a nanometer gap. J. Appl. Phys. **94**, 4690–4696 (2003)
114. T. Zeng, Thermionic-tunneling multilayer nanostructures for power generation. Appl. Phys. Lett. **88**, 153104 (2006)
115. R.H. Fowler, L. Nordheim, Electron emission in intense electric field. Proc. Royal Soc. Lond. A **119**, 173–181 (1928)
116. D.J. Griffiths, *Introduction to Quantum Mechanics*, 2nd edn, Chap. 8 (Pearson Prentice Hall, Upper Saddle River, NJ, 2005)
117. R. Tsu, L. Esaki, Tunneling in a finite superlattice. Appl. Phys. Lett. **22**, 562–564 (1973); L.L. Chang, L. Esaki, R. Tsu, Resonant tunneling in semiconductor double barriers. Appl. Phys. Lett. **24**, 593–595 (1974)
118. J.W. Gadzuk, E.W. Plummer, Field emission energy distribution (FEED). Rev. Mod. Phys. **45**, 487–548 (1973)
119. L. Nilsson, O. Groening, P. Groening, O. Kuettel, L. Schlapbach, Characterization of thin film electron emitters by scanning anode field emission spectroscopy. J. Appl. Phys. **90**, 768–780 (2001)
120. J.B. Xu, K. Läuger, R. Möller, K. Dransfeld, I.H. Wilson, Energy-exchange processes by tunneling electrons. Appl. Phys. A **59**, 155–161 (1994)
121. T.S. Fisher, Influence of nanoscale geometry on the thermodynamics of electron field emission. Appl. Phys. Lett. **79**, 3699–3701; T.S. Fisher, D.G. Walker, Thermal and electrical energy transport and conversion in nanoscale electron field emission process. J. Heat Transfer **124**, 954–962 (2002)
122. C. Trinkle, P. Kichambare, R.R. Vallance, B. Wong, M.P. Mengüç, B. Sadanadan, et al., Thermal transport during nanoscale machining by field emission of electrons from carbon nanotubes. J. Heat Transfer **125**, 546 (2003); R.R. Vallance, A.M. Rao, M.P. Mengüç, Processes for nanomachining using carbon nanotubes. US Patent No. 6,660,959, 9 Dec 2003
123. B.T. Wong, M.P. Mengüç, R.R. Vallance, Nano-scale machining via electron beam and laser processing. J. Heat Transfer **126**, 566–576 (2004)
124. W.I. Milne, K.B.K. Teo, G.A.J. Amaratunga, P. Legagneux, L. Gangloff, J.-P. Schnell, V. Semet, V. Thien Binh, O. Groening, Carbon nanotubes as field emission sources. J. Mater. Chem. **14**, 933–943 (2004)
125. R.J. Keyes (ed.), *Optical and Infrared Detectors* (Springer, Berlin, 1980)
126. S. Basu, Y.-B. Chen, Z.M. Zhang, Microscale radiation in thermophotovoltaic devices—A review. Intl. J. Ener. Res. **31** (2007)
127. T.P. Xiao, K. Chen, P. Santhanam, S. Fan, E. Yablonovitch, Electroluminescent refrigeration by ultra-efficient GaAs light-emitting diodes. J. Appl. Phys. **123**, 173104 (2018)
128. P.S. Zory Jr. (ed.), *Quantum Well Lasers* (Academic Press, San Diego, 1993)