potential barrier around the box, (b) the allowable energy levels for $D = 50$ Å, and (c) the degeneracy of the first four energy levels.

2.15 *Wave function in a one-dimensional potential well.* Plot the most probable electron distribution in a one-dimensional infinite potential well for $n = 1, 2, 3$.

2.16 *Degeneracy of electron energy levels in a hydrogen atom.* Prove that the electron degeneracy in a hydrogen atom is $2n^2$.

2.17 *Translational energy level.* A 10 cm^3 box contains a H_2 molecule at 300 K.
(a) Estimate the average translational energy of the H_2 molecule.
(b) How do the first few translational energy levels compare with $\kappa_B T$?
(c) Can you think about a way to count the degeneracy of the translational energy levels corresponding to this average energy?

3

Energy States in Solids

The previous chapter introduced the energy levels in simple potential fields, such as quantum wells, harmonic oscillators, atoms, and molecules. In this chapter, we will discuss energy levels in solids. We focus our discussion on single crystals, which are the simplest form of solids because the atoms are regularly arranged. As we will see, crystal periodicity plays a central role in determining the energy levels. So we will start by discussing crystal structures, including lattices and the potentials binding the atoms into a crystal. Since atoms in solids are packed closely, the electron wavefunctions of individual atoms overlap and form new wavefunctions and, correspondingly, new electron energy levels. The interatomic forces bond nuclei together so that the vibration of the atoms inside the crystal is strongly coupled. The collective atomic vibration can be decomposed into normal modes extending over the crystal and the basic energy quantum of each normal mode is called a phonon, in the same way of that the basic energy quantum of an electromagnetic mode is called a photon. Each electron and phonon wavefunction is characterized by a frequency (or energy) and a wavevector. The relationship between the energy and the wavevector is called the dispersion relation, which plays a central role in determining the properties of the crystal. We limit mathematical derivations to the dispersion relations of electrons and phonons in one-dimensional periodic structures and explain the energy levels in real crystals without a more detailed mathematical derivation, because the dispersion relation in a real crystal can be appreciated on the basis of a sound understanding of the behavior of a one-dimensional periodic system. The energy levels in crystals are highly degenerate. A very useful tool that takes into account the degeneracy of the energy states is the density of states, which will be used repeatedly throughout the book and should be mastered.

3.1 Crystal Structure

A perfect bulk crystal is a three-dimensional periodic arrangement of atoms. To describe a crystal, we use an atomless lattice—a periodic array of mathematical points that replicate the inherent periodicity of the actual crystal. Every point in the lattice is identical to other points. To form an actual crystal, a basis consisting of one or several atoms (or a molecule) is attached to each lattice point, i.e.

$$\text{crystal} = \text{lattice} + \text{basis} \tag{3.1}$$

The exact position of the basis relative to the lattice point is not important, as long as the relative position between the basis and the lattice point is the same for all the lattice points. Many crystals have the same lattice structure; in fact, there is only a limited number of possible lattice types. Thus, we will first discuss the description of lattices in real space, followed by an introduction of the concept of reciprocal lattice, which is the Fourier transform of the real space lattice. The binding between atoms in real crystals will then be discussed.

3.1.1 Description of Lattices in Real Space

Consider a two-dimensional lattice as shown in figure 3.1. From a mathematical point of view, the location of each a point can be described by a vector. Because of the periodic arrangement of lattice points, we can choose a basic set of vectors, called the *primitive lattice vectors*, to construct all other vectors in the lattice. In a three-dimensional lattice, $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are primitive lattice vectors if, from any point, we could

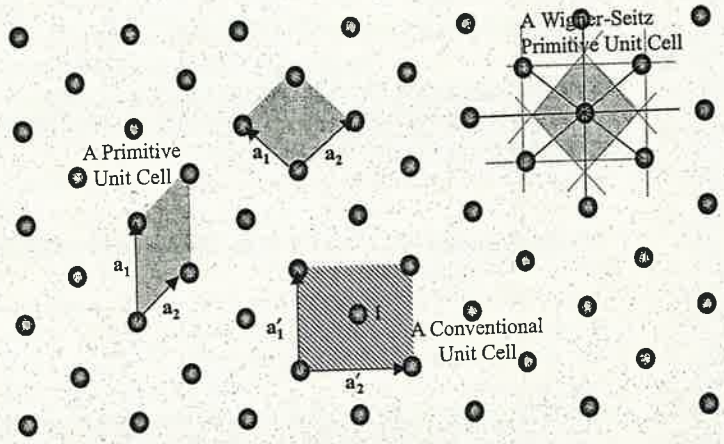


Figure 3.1 A two-dimensional lattice. Different choices of primitive lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ and primitive unit cells (gray areas) are possible. The Wigner-Seitz primitive unit cell is one way to uniquely construct a primitive unit cell. Vectors $\mathbf{a}_1'$ and $\mathbf{a}_2'$ are not a set of primitive lattice vectors and the shaded area is not a primitive unit cell. This area is, however, often used due to its regular shape and is called a conventional unit cell.

reach all other lattice points by a proper choice of integers through the following construction

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \quad (n_1, n_2, n_3 \text{ cover all integers}) \tag{3.2}$$

The magnitudes of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are called the lattice constants. A lattice constructed according to eq. (3.2) is often called a Bravais lattice. Primitive lattice vectors are not unique. We have drawn two sets of primitive lattice vectors in figure 3.1 with primitive unit vectors denoted by $\mathbf{a}_1$ and $\mathbf{a}_2$. The other set of vectors, $\mathbf{a}_1'$ and $\mathbf{a}_2'$, are not primitive lattice vectors because we cannot use them to construct all other lattice points by a two-dimensional version of eq. (3.2). For example, we cannot reach point 1 through any linear integer combination of $\mathbf{a}_1'$ and $\mathbf{a}_2'$.

A *primitive unit cell* is the parallelepiped defined by the primitive lattice vectors. There is only one lattice point (equivalently speaking) per primitive unit cell. For example, each of the four lattice points in the two parallelograms formed by the two sets of primitive lattice vectors in figure 3.1 is shared by four unit cells and thus the number of equivalent lattice points in each parallelogram is one. These are thus primitive unit cells. On the other hand, the shaded rectangle formed by $\mathbf{a}_1'$ and $\mathbf{a}_2'$ is not a primitive unit cell because there are two lattice points in such a rectangle: the center point plus the four corners, each of the latter being shared by four cells. Because the choice of primitive lattice vectors is not unique, there can be different ways to draw a primitive unit cell, as shown by the two examples in figure 3.1. One method to construct a unit cell uniquely is the Wigner–Seitz cell (see figure 3.1), which is constructed by connecting all the neighboring points surrounding an arbitrary lattice point (as shown by the solid lines in figure 3.1) and drawing the bisecting plane (shown by dashed lines in the figure) perpendicular to each connection line. The smallest space formed by all the bisecting planes is a Wigner–Seitz cell, as indicated in the figure.

Sometimes, it is more convenient to describe a lattice by the *conventional unit cell*. For example, in figure 3.1, the rectangle formed by $\mathbf{a}_1'$ and $\mathbf{a}_2'$ is more convenient than the parallelogram formed by the primitive lattice vectors. This unit cell has two lattice points and is called a conventional unit cell. The crystal can also be constructed by repeating such a cell.

A general unit cell in the three-dimensional space is designated by three lattice vectors and the three angles formed between them. In the most general case, these three lattice vectors are of different lengths and the three angles are all oblique, as shown in figure 3.2(a). This lattice is called a triclinic lattice and does not have much symmetry. The symmetry of a lattice is often characterized by the symmetry operations, which include rotation of the unit cell around a fixed lattice point, reflection of the unit cell along a specific plane, and inversion with respect to a lattice point. A fundamental requirement on the lattice is that one can fill the entire space by placing a primitive unit cell at every lattice point. This requirement puts a limitation on the symmetric operations of a lattice. For example, the allowable rotational symmetry operations are 2π, π, $2\pi/3$, $2\pi/4$, and $2\pi/6$. No lattice, however, can have $2\pi/7$ or $2\pi/5$ rotational symmetry.* Given these conditions, it turns out that there are 13 other types of lattice that have special symmetry operations on top of the π and 2π rotational symmetry of

*Some quasicrystals can have five-fold symmetry patterns but they do not satisfy the definition of a crystal discussed in this section (Kittel, 1996).

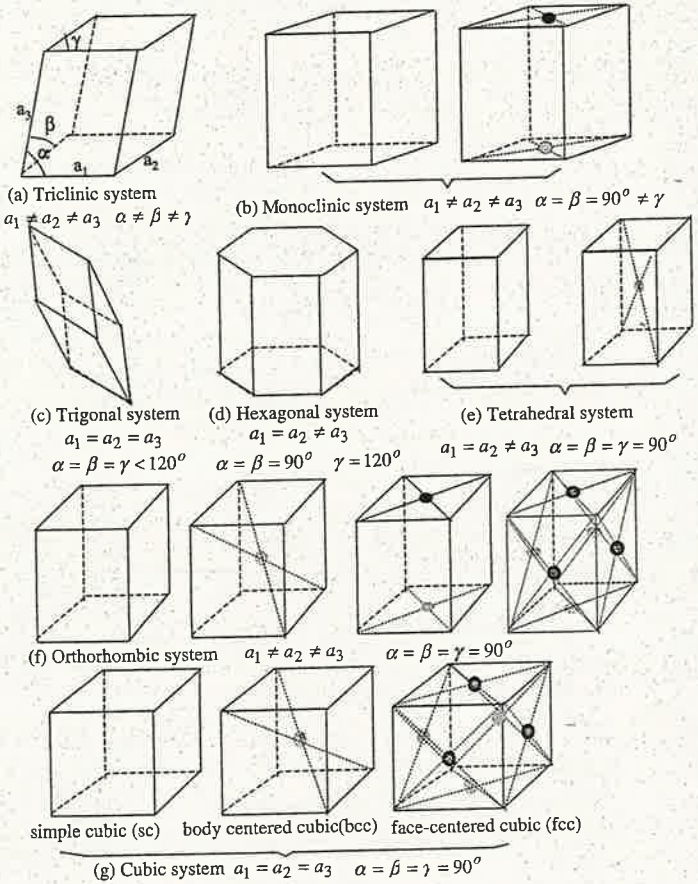


Figure 3.2 Fourteen Bravais lattices and the associated seven systems.

a triclinic lattice, so that there are in total 14 types of Bravais lattice. These 14 lattices can be further grouped into seven types of point symmetry operations (seven crystal systems) or operations around a fixed lattice point (such as rotation and inversion), as shown in figure 3.2. For example, in the cubic system, there are three types of Bravais lattices, the simple cubic (sc), the body-centered cubic (bcc), or the face-centered cubic (fcc). The cube is a primitive unit cell only for the sc lattice and is a conventional unit cell for the bcc and fcc lattices.

Certain crystal planes inside a lattice are identical. These planes are parallel to each other and equally spaced. A common convention for indexing the crystal planes and the high symmetry directions is in terms of the Miller indices. The *Miller indices of crystal planes*, represented by a set of integers in parentheses (*hkl*), are obtained in accordance with the following steps:

1. Find the intercepts of the crystal plane with the axes formed by the lattice vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ in terms of the lattice constants. The origin of the lattice vectors can be at any lattice point. One can choose any crystal plane that is convenient to use. For example, in figure 3.3 we have two crystallographically identical planes; one intercepts the axis at $0.8a_1, 0.4a_2, 0.6a_3$, and the other at $4a_1, 2a_2, 3a_3$.
2. Take the reciprocal of the intercepts and reduce these reciprocals to the three smallest integers that have the same ratio as the original set. The result is enclosed in parentheses (*hkl*) and this set of numbers is called the Miller indices of the plane. The example in figure 3.3 yields

$$(1/0.8, 1/0.4, 5/3) \text{ (for inner plane) or}$$
$$(1/4, 1/2, 1/3) \text{ (for outer plane)} \rightarrow (364)$$

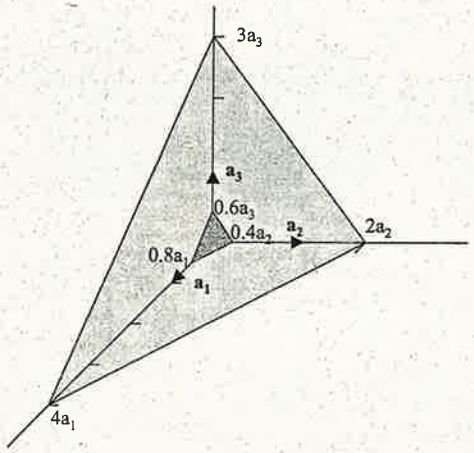


Figure 3.3 Two identical planes with Miller indices (364) in a crystal.

If the plane intersects at the negative side of the chosen primitive lattice vector, a line is placed above the number. For example, the six square faces of a cubic unit cell [figure 3.2(g)] are (100), (010), (001), $(\bar{1}00)$, $(0\bar{1}0)$, $(00\bar{1})$. We can use the sign {100} to denote all the six equivalent planes. The direction index in a crystal is denoted by a set of smallest integers [*uvw*] proportional to the unit vector in the desired direction. All equivalent directions can be denoted by ⟨*uvw*⟩. Based on these definitions, one can prove that the (*hkl*) plane is perpendicular to the [*hkl*] direction.

Crystal planes and directions are often determined by using X-rays or through transmission electron microscopy. Semiconductor wafers are sold with the major crystallographic directions and dopant types marked by the wafer flats. A wafer typically has a primary flat representing a crystal plane and a secondary flat that is positioned to denote the dopant type and the surface crystallographic direction, as shown in figure 3.4 for a 4 in. silicon wafer.

3.1.2 Real Crystals

By attaching a basis, which can be one or several atoms or a molecule, to each lattice point, real crystals are formed. For example, silicon has an fcc structure and the basis is made of two silicon atoms. If we anchor one atom at the fcc lattice point, for example

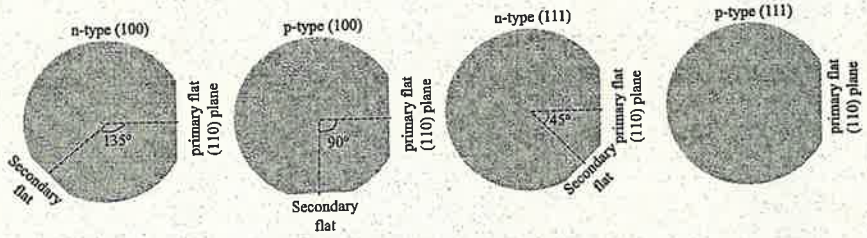


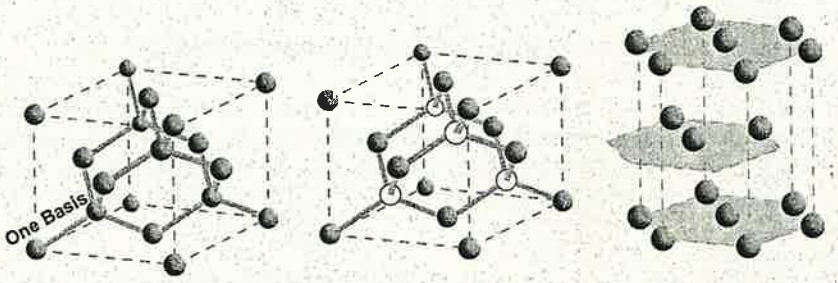
Figure 3.4 Semiconductor wafers are sold with major crystallographic indices and dopant type denoted by primary and secondary flats. These are common notations for 4 in. wafers.

at (0, 0, 0), the other silicon atom is then at $(a/4, a/4, a/4)$ as shown in figure 3.5(a). In a conventional fcc unit cell, there are four lattice points and eight silicon atoms. In a primitive unit cell, there are two silicon atoms. The silicon crystal structure is called diamond structure, which is shared by several other crystals such as germanium and diamond. The zinc blende structure [Figure 3.5(b)], such as that of gallium arsenide (GaAs), is similar to the diamond structure but the basis is made of two different atoms, one Ga and one As atom for GaAs. If we take the Ga atom at (0,0,0), then the As atom is at $(a/4, a/4, a/4)$. Since the Ga and the As atoms are different from one another, the zinc blende crystal structure has fewer symmetry operations than the diamond structure. Graphite has a close-packed hexagonal structure, as shown in figure 3.5(c).

Example 3.1 *Density of Si crystals*

Silicon is an fcc lattice with a lattice constant of 5.43 Å and two atoms per lattice site. Determine the density of Si crystals.

Solution: An fcc lattice has four lattice points. Since there are two Si atoms at each lattice point, there is a total of eight Si atoms per conventional fcc cell. Each atom weighs $28 \times 1.67 \times 10^{-27}$ kg, where 28 is the number of protons and neutrons,



(a) Diamond
C, Si, Ge, ...

(b) Zinc blende
ZnS, GaAs, GaP, SiC, ...

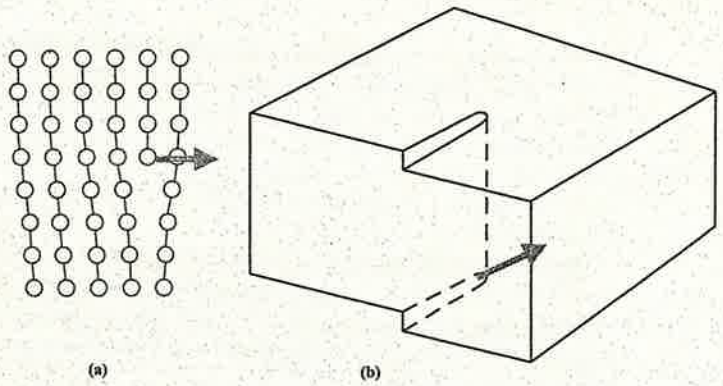
(c) Hexagonal Close-packed (hcp)
Graphite, He, Mg, Zn, Co, ...

Figure 3.5 Three-types of common crystal structure: (a) diamond, (b) zinc blende, (c) close-packed hexagonal.

and 1.67×10^{-27} kg is the weight of a proton or neutron. The density of Si crystal is thus

$$\rho = \frac{8 \times 28 \times 1.67 \times 10^{-27}\text{kg}}{(5.43 \times 10^{-10}\,\text{m})^3} = 2.34 \times 10^3\,\text{kg m}^{-3}$$

Real crystals also have *defects*, at which the periodicity of the crystal is disturbed. The defects can be divided into the following three types: points, lines, and planes. Examples of *point defects* are vacancies, where the atoms at the lattice points do not exist, and impurities, where the original atoms are substituted by different atoms. Another form of a point defect is an interstitial defect, where an additional atom is inserted in the space that does not belong to any allowed atomic site in the original lattice. Examples of *line defects* are *dislocations*. The two simplest types of dislocation are the *edge dislocation* and *screw dislocation*, as shown in figures 3.6(a) and (b). The edge dislocation can be constructed by inserting an extra plane of atoms in the upper half of the crystal. The screw dislocation can be thought of as the result of cutting the crystal partway through with a knife and shearing it parallel to the edge of the cut by one lattice vector. Dislocations can only form loops or be terminated at the surfaces (or interfaces), as is vortex in a fluid flow. The number of dislocations is measured by the dislocation density (dislocations/cm^2), which is the number of dislocation lines that intersect a unit area in the crystal. Typical values for the dislocation density are 10^8 cm^{-2}, and they vary significantly depending on how the materials are made. Because the dislocated regions are highly stressed, it takes a relatively small disturbance (shear stress) to move the location of the dislocation to the next lattice plane along the directions drawn in Figure 3.6. Thus crystals with dislocations have a lower plastic deformation limit. But this is not the whole story. In highly dislocated crystals, because there are many defects and dislocations along the path of the dislocation motion, the work needed to move these dislocations over other defects is much higher. Thus, poorly prepared crystals that have many dislocations can be harder than relatively good but yet not perfect (because the existence of dislocations) crystals. Examples of *planar defects* are the grain boundaries between two small crystalline regions inside a crystal, or polycrystals.



(a) (b)

Figure 3.6 Illustrations of (a) edge and (b) screw dislocations. The arrows mark the direction of motion of the dislocation.

3.1.3 Crystal Bonding Potential

What holds the atoms together in a crystal? Fundamentally, it is the electron–electron and electron–nucleus interactions between atoms that hold them together. We touched on the topic of interatomic potential in chapter 1 and will now give a more detailed discussion, focusing on solids. More discussion on the interatomic potential will be given in chapter 10, for molecular dynamic simulations.

The force interaction between atoms *always* consists of a long-range attractive force and a short-range repulsive force. The short-range repulsive force is effective, due to the Pauli exclusion principle, when the inner-shell electrons or the nuclei of the atoms begin to overlap. Two often-used empirical expressions for the repulsive potential between the atoms separated by a distance r are

$$U_R(r) = \frac{B}{r^{12}} \quad \text{(Lennard–Jones)} \tag{3.3}$$

and

$$U_R(r) = U_0 e^{-r/\zeta} \quad \text{(Born–Mayer)} \tag{3.4}$$

where B, ζ, and U_0 are empirical constants determined from experimental data, such as the interatomic spacing and the binding energy. The differences between various types of crystals, however, are mainly due to the long-range attractive interaction, which will be discussed below.

Molecular crystals are characterized by the dipole–dipole interaction between atoms. An isolated atom is not polarized, but, when another atom is close by, the electrical field of electrons from the neighboring atom distorts the positions of the electrons and the nucleus of the current atom, creating an induced dipole. The attractive potential between the induced dipole of two atoms is given by

$$U_A = -\frac{A}{r^6} \tag{3.5}$$

van der Waals potential

Combining this attractive potential (van der Waals potential) with the Lennard–Jones potential for the repulsive force, we obtain the Lennard–Jones interaction potential between a pair of atoms i and j in a crystal as

$$U_{ij} = \frac{B}{r_{ij}^{12}} - \frac{A}{r_{ij}^{6}} \tag{3.6}$$

This potential is most appropriate for crystals of inert atoms (such as argon atoms which form a crystal at low temperatures) that have a neutral, spherically symmetric charge. Such a potential, however, can also be used to describe the interactions between similar atoms or molecules in liquidus or gaseous phases. A crystal structure is stable when the total potential energy of the system

$$U = \frac{1}{2}\sum_{i\neq j}\left(\frac{B}{r_{ij}^{12}} - \frac{A}{r_{ij}^{6}}\right) = \frac{1}{2}\sum_{i\neq j} 4\varepsilon\left(\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right) \tag{3.7}$$

reaches a minimum, as required by the second law of thermodynamics for a stable system. In eq. (3.7), $\sigma = (B/A)^{1/6}$ and $\varepsilon = A^2/(4B)$, and the factor of one-half is there because the potential is shared between two atoms and the summation double-counts the potential. Values of Lennard–Jones parameters for noble-gas crystals are given in table 3.1.

Table 3.1 Lennard–Jones potential parameters for noble-gas crystals

	Neon	Argon	Krypton	Xenon
Crystal structure	fcc	fcc	fcc	fcc
Lattice constant (Å)	4.46	5.31	5.64	6.13
ε (10^{-20} J)	0.050	0.167	0.225	0.320
ε (eV)	0.031	0.0104	0.014	0.0200
σ (Å)	2.74	3.40	3.65	3.98

Source: Ashcroft and Mermin, 1976.

Example 3.2 *Lattice constant*

Determine the lattice constant of an fcc crystal described by the Lennard–Jones potential in terms of σ in eq. (3.7).

Solution: We assume that the nearest neighbor distance is R. We first compute the potential energy for any one atom i interacting with the rest of the atoms in the crystal. The total energy of the crystal with N atoms is thus $N/2$ times this energy. From eq. (3.7), we have

$$U_{\text{tot}} = \frac{(4\varepsilon)N}{2}\sum_j\left(\left(\frac{\sigma}{Rp_{ij}}\right)^{12} - \left(\frac{\sigma}{Rp_{ij}}\right)^{6}\right) \tag{E3.2.1}$$

where p_{ij} is the interatomic distance in terms of the neighbor distance R. For an fcc crystal, we can deduce that

$$\sum_j p_{ij}^{-12} = 12.13188 \quad \sum_j p_{ij}^{-6} = 14.45392 \tag{E3.2.2}$$

The lattice constant is the point when U_{tot} is minimum. Thus,

$$\left[\frac{dU_{\text{tot}}}{dR}\right]_{R=R_0} = -2N\varepsilon\left[12\times 12.13\frac{\sigma^{12}}{R_0^{13}} - 6\times 14.45\frac{\sigma^6}{R_0^7}\right] = 0 \tag{E3.2.3}$$

which leads to

$$R_0/\sigma = 1.09 \tag{E3.2.4}$$

The observed values for Ne, Ar, Kr, Xe are $R_0/\sigma = 1.14, 1.11, 1.10, 1.09$, very close to the calculation.

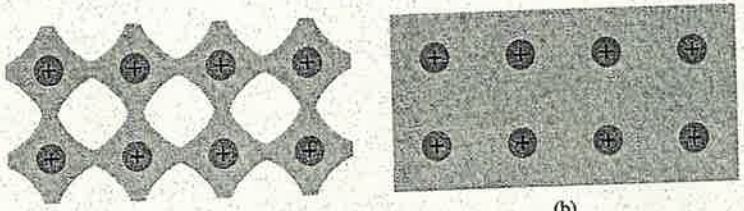
In *ionic crystals*, such as NaCl, the single valence electron in the sodium atom moves to the chlorine atom so that both Na^+ and Cl^- have closed electron shells but, meanwhile, become charged. The Coulomb potential among the ions becomes the major attractive force. The potential energy of any ion i in the presence of other ions j is then

$$U_{i,A} = \sum_{i \neq j} \frac{\pm q^2}{4\pi\varepsilon_0 r_{ij}} = -\frac{\alpha q^2}{4\pi\varepsilon_0 r_0} \tag{3.8}$$

where q is the charge per ion, ε_0 the dielectric permittivity of free space, and r_0 the nearest-neighbor separation. The parameter α is called the Madelung constant and is related to the crystal structure. This attractive potential, combined with an appropriate repulsive potential, constitutes the total potential energy in ionic crystals.

Covalent bonds are formed when electrons from neighboring atoms share common orbitals, rather than being attached to individual ions as in ionic crystals. Biological molecules are often formed through covalent bonding. Many inorganic systems are also covalently bonded, such as the H_2 molecule. The electron in each hydrogen atom of an H_2 molecule shares a common orbital (one spin-up and the other spin down) with the other electron in the other H atom. The covalent bond is strongly directional. In the case of the H_2 molecule, the bond is oriented along the line of the two nuclei. Diamond, silicon, and germanium are all *covalent crystals*. Each atom has four electrons in the outer shell and forms a tetrahedral system of covalent bonds with four neighboring atoms, as indicated in figure 3.5(a). In certain crystals, such as GaAs, both covalent and ionic bonding are important. Fundamentally, the covalent bonding force is also due to charge interaction. However, unlike the van der Waals force in molecular crystals or the electrostatic force in ionic crystals, it is more difficult to construct simple expressions for covalent crystals. Empirical potentials have been developed, such as the Stillinger–Weber potential for silicon (Stillinger and Weber, 1985). Expressions for various empirical potentials will be presented in chapter 10.

In covalent bonds, electrons are preferentially concentrated in regions connecting the nuclei, leaving some regions in the crystal with low charge concentration, as illustrated in figure 3.7(a). Metals and their associated *metallic bonds* can be considered an extreme case of covalent bonds, in which the bonds begin to overlap and all regions of the crystal become filled up with charges [figure 3.7(b)]. In the case of total filling of the empty space, it becomes impossible to tell which electron belongs to which atom. One can imagine the entire crystal as one big molecule with the electrons shared amongst all



(a) (b)

Figure 3.7 Distribution of electrons (gray area) in (a) a covalent bonding crystal and (b) a metallic bonding crystal (after Ashcroft and Mermin, 1970).

the atoms. In this case we say the electrons are delocalized and can wander throughout the crystal; in other words, they become free electrons. Later, we will see how the delocalization of electrons can be explained by solving the Schrödinger equation.

3.1.4 Reciprocal Lattice

We know that any periodic function can be expanded as a Fourier series. For example, if a time-dependent function $f(t)$ is periodic with a period of T, it can be expanded into a Fourier series as

$$\begin{aligned} f(t) &= \sum_{n=-\infty}^{\infty} \left(a_n \sin\left(\frac{2\pi n}{T} t\right) + b_n \cos\left(\frac{2\pi n}{T} t\right) \right) \\ &= \sum_{n=-\infty}^{\infty} (a'_n e^{in\omega t} + b'_n e^{-in\omega t}) \end{aligned} \tag{3.9}$$

where a_n and b_n are the coefficients of the Fourier series and a'_n and b'_n in the second step can be obtained by expressing sine and cosine functions as complex exponential functions. The angular frequency $\omega = 2\pi/T$ is the Fourier conjugate of the temporal periodicity such that $e^{i\omega T} = 1$, which ensures that $f(t)$ is periodic for every $\Delta t = T$; that is, $f(t+T) = f(t)$.

In a crystal, the atoms are periodic in space and, consequently, we expect that some atom-related functions are periodic. The simplest example is the potential energy, which should have a periodicity corresponding to the unit cell. Let's first consider a one-dimensional lattice, with a lattice constant of a. A spatial-dependent function $f(x)$, with a periodicity a, $f(x) = f(x+a)$, can be similarly expanded into a Fourier series,

$$f(x) = \sum_{n=-\infty}^{\infty} (a'_n e^{ink_x x} + b'_n e^{-ink_x x}) \tag{3.10}$$

where the wavevector, $k_x = 2\pi/a$, is the Fourier conjugate to spatial periodicity a. Using eq. (3.10), one can easily show that $f(x)$ repeats for every $\Delta x = a$; that is, $f(x+a) = f(x)$.

The above example is for a one-dimensional lattice. But crystals are three dimensional. How can we expand a function that is periodic over the three-dimensional crystal into a Fourier series? We will use the charge distribution in the crystal, $n(\mathbf{r})$, as an example. It should be invariant with any translational lattice vector $\mathbf{R}$; that is, $n(\mathbf{r}+\mathbf{R}) = n(\mathbf{r})$. We will first give the following answer and then show that the given Fourier expansion indeed satisfies the required periodicity,

$$n(\mathbf{r}) = \sum_{\mathbf{G}} n_{\mathbf{G}} e^{i\mathbf{r}\cdot\mathbf{G}} \tag{3.11}$$

where $\mathbf{G}$ and the inverse transformation are given by

$$\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3 \qquad (m_1, m_2, m_3 \text{ are integers}) \tag{3.12}$$

$$n_{\mathbf{G}} = \frac{1}{V} \int_{\text{unit cell}} n(\mathbf{r}) e^{-i\mathbf{r}\cdot\mathbf{G}} dV \tag{3.13}$$

and $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ are conjugated to the primitive lattice vectors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ through

$$\mathbf{b}_1 = 2\pi(\mathbf{a}_2 \times \mathbf{a}_3)/V, \mathbf{b}_2 = 2\pi(\mathbf{a}_3 \times \mathbf{a}_1)/V, \mathbf{b}_3 = 2\pi(\mathbf{a}_1 \times \mathbf{a}_2)/V \tag{3.14}$$

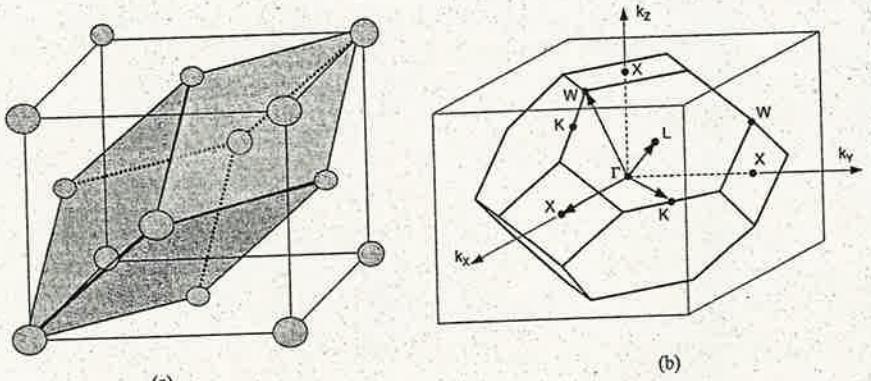
where $V = \mathbf{a}_1 \bullet (\mathbf{a}_2 \times \mathbf{a}_3)$ is the volume of the primitive unit cell in real space. One can easily prove the following relations,

$$\mathbf{a}_i \bullet \mathbf{b}_j = 2\pi\delta_{ij} \text{ where } \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{3.15}$$

and δ_{ij} is the Kronecker delta. With the above definitions, we show now that $n(\mathbf{r})$ is indeed invariant with any translational lattice vector in the real space $\mathbf{R}(= n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3)$, where n_1, n_2, and n_3 are integers:

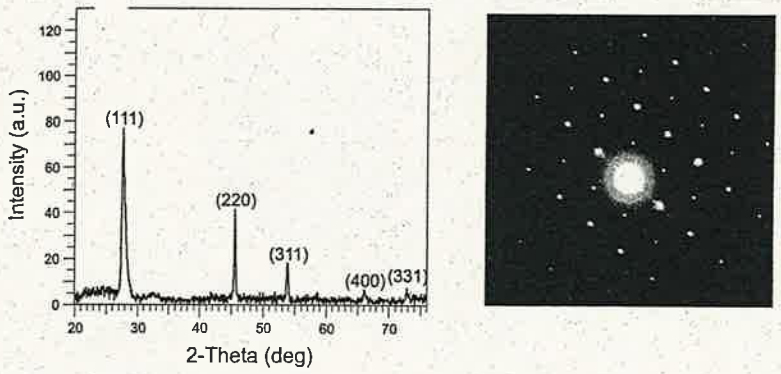
$$\begin{aligned} n(\mathbf{r}+\mathbf{R}) &= \sum_{\mathbf{G}} n_{\mathbf{G}} e^{i(\mathbf{r}+\mathbf{R})\bullet\mathbf{G}} = \sum_{\mathbf{G}} n_{\mathbf{G}} e^{i\mathbf{r}\bullet\mathbf{G}+i\mathbf{R}\bullet\mathbf{G}} \\ &= \sum_{\mathbf{G}} n_{\mathbf{G}} e^{i\mathbf{r}\bullet\mathbf{G}+i2\pi(n_1m_1+n_2m_2+n_3m_3)} = \sum_{\mathbf{G}} n_{\mathbf{G}} e^{i\mathbf{r}\bullet\mathbf{G}} = n(\mathbf{r}) \end{aligned}$$

We used eq. (3.15) in the third step and the fact that m_i and n_i are integers in the fourth step. Thus we see that the new set of vectors introduced, $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$, which has a unit of m^{-1}, is the corresponding Fourier conjugate to the real space lattice vector $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$. We can use $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ and eq. (3.12) to construct a new lattice called the *reciprocal lattice*, with $\mathbf{G}$ as the reciprocal lattice vector. Previous definitions on real space lattices, such as unit cells and the Wigner–Seitz primitive unit cell, are equally applicable to such a reciprocal lattice. This reciprocal space is the Fourier conjugate of the real space. Figures 3.8(a) and (b) show a primitive unit cell of an fcc lattice in real space and its corresponding reciprocal space, based on the Wigner–Seitz construction. The Wigner–Seitz primitive unit cell in reciprocal space is also called the first Brillouin zone, which will be used extensively later to represent the electron and phonon energy levels in solids. Although the representation of crystal properties in the reciprocal lattice may be unfamiliar to some readers, they undoubtedly have seen the spectrum representation of time-varying



(a) (b)

Figure 3.8 Conventional and primitive unit cells in real and reciprocal unit cells of an fcc lattice (a) in real space, (b) in reciprocal space.



(a) (b)

Figure 3.9 (a) X-ray 2-theta scan of a germanium crystal and (b) an electron diffraction pattern for a bismuth nanowire (Courtesy of Dr. Z.F. Ren and Dr. M.S. Dresselhaus).

signals in electrical engineering, or the spectral-dependent radiative properties of matter, which conveniently express time-dependent electrical signals or electromagnetic fields into stationary spectral properties through Fourier transformations. The representation of properties of crystals in the reciprocal space assumes a similar role. Specific symbols have been designated for different directions of the reciprocal lattice. For example, in figure 3.8(b), the Γ point is the center of the Brillouin zone. The Γ–X direction represents the [100] direction of the real lattice, and Γ–L the [111] direction of the real lattice.

Although a very abstract concept, the reciprocal lattice can actually be easily mapped out with diffraction experiments. When electron waves or X-rays (electromagnetic waves) with the proper energy are directed onto a crystal, the reflection or transmission occurs only along specific directions, as shown in figure 3.9(a) (X-ray reflection) and figure 3.9(b) (electron transmission). Such phenomena, known as diffraction, can be attributed to the superposition of the incident waves and scattered waves. Consider an incident wave (an electron beam or an X-ray beam) from the source along direction $\mathbf{k}$. At any point in the crystal, the magnitude of the wave is proportional to $e^{i\mathbf{k}\bullet(\mathbf{r}-\mathbf{r}_s)}$, where $\mathbf{r}_s$ is the location of the source relative to the origin of coordinates, as shown in figure 3.10(a). The wave scattered into the detector is then proportional to $n(\mathbf{r})e^{i\mathbf{k}'\bullet(\mathbf{r}_d-\mathbf{r})}$, where $\mathbf{k}'$ is the propagation direction of the scattered wave and $\mathbf{r}_d$ is the position of the detector. Because each atom scatters the incident wave, the total magnitude of the wave, S, at the detector is

$$\begin{aligned} S &\propto \int_{\substack{\text{whole}\\\text{crystal}}} e^{i\mathbf{k}\bullet(\mathbf{r}-\mathbf{r}_s)} n(\mathbf{r}) e^{i\mathbf{k}'\bullet(\mathbf{r}_d-\mathbf{r})} dV = e^{i(\mathbf{k}'\bullet\mathbf{r}_d-\mathbf{k}\bullet\mathbf{r}_s)} \int_{\substack{\text{whole}\\\text{crystal}}} n(\mathbf{r}) e^{i(\mathbf{k}-\mathbf{k}')\bullet\mathbf{r}} dV \\ &= \sum_{\mathbf{G}} e^{i(\mathbf{k}'\bullet\mathbf{r}_d-\mathbf{k}\bullet\mathbf{r}_s)} \int_{\substack{\text{whole}\\\text{crystal}}} n_{\mathbf{G}} e^{i(\mathbf{G}+\mathbf{k}-\mathbf{k}')\bullet\mathbf{r}} dV \end{aligned} \tag{3.16}$$

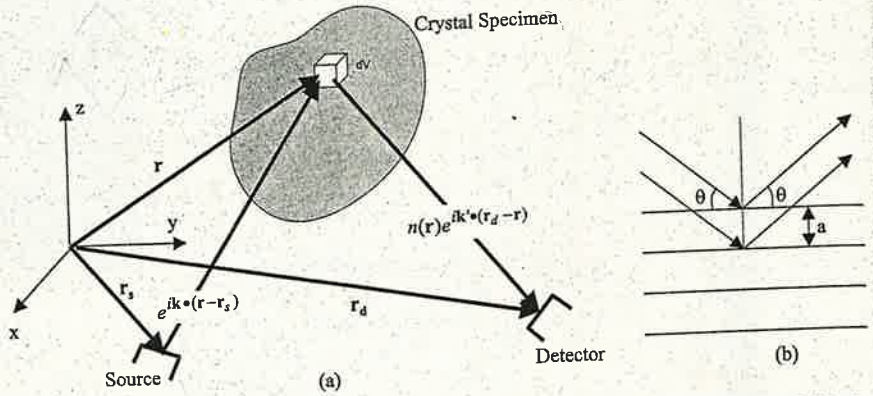


Figure 3.10 Derivation of the Bragg diffraction condition for (a) the general case and (b) the one-dimensional case.

where we have used eq. (3.11) to expand $n(\mathbf{r})$ into a Fourier series. Because the exponential function $e^{i(\mathbf{G}+\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}$ is a rapidly varying function in the crystal with both negative and positive values, the above integral will be close to zero except when the exponent vanishes, that is, when

$$\mathbf{G}+\mathbf{k}-\mathbf{k}'=\mathbf{0} \tag{3.17}$$

Equation (3.17) is called the Bragg condition for diffraction and determines the relationship between incident ($\mathbf{k}$) and diffracted ($\mathbf{k}'$) waves. The wavevectors $\mathbf{k}$ and $\mathbf{k}'$ are determined by the relative positions of the source, the sample, and the detector, and therefore the reciprocal lattice vectors $\mathbf{G}$, and thus the crystal structure, can be determined from diffraction experiments.

The Bragg condition, eq. (3.17), was derived by considering the scattering of individual atoms. We can treat each crystal plane as a continuous sheet and establish an equivalent condition based on the interference of the waves reflected from all the parallel crystal planes in a specific direction. Consider the special set of crystal planes separated by a distance a, as shown in figure 3.10(b), and an incident wave (photon or electron) of wavelength λ at an angle θ. Constructive interference between waves reflected from crystal planes occurs when the phase difference of the waves scattered between two consecutive planes is πn. From figure 3.10(b), we see that the path difference is $2a\sin\theta$. Thus diffraction peaks will be observed when the path difference is multiples of the wavelength,

$$2a\sin\theta = n\lambda \tag{3.18}$$

It can be shown that eq. (3.18) is a special case of eq. (3.17). The proof is left as an exercise.

Example 3.3 *X-ray diffraction*

One way of using X-ray diffraction to determine the crystal structure is the rotating crystal method. In this method, an X-ray of fixed wavelength λ is directed onto the sample at a fixed direction, and the diffracted X-ray is measured by a fixed detector; refer to figure 3.10(a). The crystal is rotated to change the angle of incidence θ [figure 3.10(b)] with respect to a special crystal plane. When the Bragg condition is not satisfied, the detector will register very little signal. But when crystals are rotated to the positions where the Bragg condition, eq. (3.18), is satisfied, the detector will register a peak. A typical scan curve is shown in figure 3.9(a). Different peaks correspond to different crystal planes. For an X-ray of wavelength 1 Å and a first-order diffraction peak at $\theta = 30°$, the corresponding spacing between the two crystal planes is

$$a = \frac{1\text{Å}}{2\sin 30°} = 1\text{Å}$$

3.2 Electron Energy States in Crystals

In the previous chapter, we discussed the energy levels of single atoms and harmonic oscillators. These energy levels are typically discrete. In solids, the wavefunctions of closely spaced atoms begin to overlap and form new wavefunctions and, correspondingly, new energy levels. We will see that the energy levels become more continuous than those of individual atoms. This trend can be thought of as the result of the broadening of the energy levels of individual atoms to avoid the overlapping of wavefunctions because, according to the Pauli exclusion principle, each quantum state can have only a maximum of one electron. In crystals, the most fundamental characteristic is the periodicity of the lattice. Such periodicity brings in many new features to the allowable energy levels of electrons as well as phonons. In this section, we will start from a simple one-dimensional model to examine the effect of periodicity on the electronic energy levels and then extend the discussion to three-dimensional crystals.

3.2.1 One-Dimensional Periodic Potential (Kronig–Penney Model)

Let us consider a simple one-dimensional lattice. The potential field is a periodic function, as sketched in figure 3.11(a). At the location of each ion, the electrons are attracted by the ion and have the lowest potential. As an approximation to the actual atomic potential distribution in a crystal as in figure 3.11(a), we consider a square periodic potential as shown in figure 3.11(b) and want to find out the energy levels, assuming there is only one electron inside such a periodic potential. As in the case that the hydrogen energy level can be used to explain the periodic table, the existence of many electrons in a

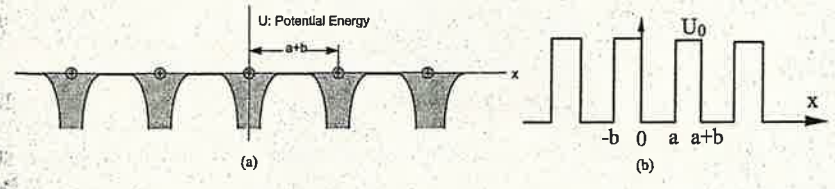


Figure 3.11 One-dimensional periodic potential model: (a) sketch of atomic potential; (b) Kronig-Penney model.

crystal does not change the main picture obtained from the one-electron assumption. This one-electron, rectangular periodic potential model is called the Kronig–Penney model. The Schrödinger equation is then

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} + (U - E)\Psi = 0 \tag{3.19}$$

The potential distribution $U(x)$ is given by

$$U(x) = \begin{cases} 0 & 0 < x \le a \\ U_0 & -b < x \le 0 \end{cases} \tag{3.20}$$

subject to the following periodicity requirement

$$U(x + a + b) = U(x) \tag{3.21}$$

Solutions for eq. (3.19) are

$$\Psi = Ae^{iKx} + Be^{-iKx} (0 < x \le a) \tag{3.22}$$

$$\Psi = Ce^{Qx} + De^{-Qx} (-b \le x \le 0) \tag{3.23}$$

where

$$E = \frac{\hbar^2 K^2}{2m} \text{ and } U_0 - E = \frac{\hbar^2 Q^2}{2m} \tag{3.24}$$

and K and Q are to be determined, from which the eigen energy E of the electron inside such a periodic potential is to be extracted.

Four boundary conditions are needed to determine the unknown coefficients A, B, C, and D. We can use the continuity of the wavefunction and its derivative at $x = 0$, which gives

$$A + B = C + D \tag{3.25}$$

$$iK(A - B) = Q(C - D) \tag{3.26}$$

Two more boundary conditions are necessary to determine the four unknown coefficients. We can consider the continuity of the wavefunction and its derivative at $x = a$, but this requires that we know the wavefunction in $a < x \le a + b$. The wavefunction in this region can be related to that in the region $-b < x < a$ because the potential is periodic. Due to the periodicity in the potential, the wavefunction at any two points separated by a lattice vector is related through the *Bloch theorem*,

$$\Psi(\mathbf{r} + \mathbf{R}) = \Psi(\mathbf{r}) \exp(i\mathbf{k} \bullet \mathbf{R}) \tag{3.27}$$

where $\mathbf{R}$ is a lattice vector and $\mathbf{k}$ is the wavevector of the crystal. The Bloch theorem implies that the wavefunction values at two equivalent points ($\mathbf{r}$ and $\mathbf{r} + \mathbf{R}$) inside a crystal differ by only a phase factor $\exp(i\mathbf{k} \bullet \mathbf{R})$ and that we need to know only the

wavefunction inside one unit cell. For the one-dimensional problem being considered, denoting k as the magnitude of $\mathbf{k}$ in the x-direction, eq. (3.27) is

$$\Psi[x + (a + b)] = \Psi(x) \exp[ik(a + b)] \tag{3.28}$$

We should distinguish the wavevector k from the propagation vector of the solution, K in eq. (3.22). The latter contains the energy of the electrons that we want to find. We will explain later, in more detail, the meaning of wavevector k. We want to find a relation between k and E, which is equivalent to a relation between k and K.

From Bloch's theorem, we know that if the wavefunction for $-b < x < 0$ is given by eq. (3.23), the wavefunction for $a < x < a + b$ is then given by eq. (3.23) multiplied by $\exp[ik(a + b)]$. The continuity requirements for the wavefunction and its derivative at $x = a$ are then

$$Ae^{iKa} + Be^{-iKa} = (Ce^{-Qb} + De^{Qb}) \exp[ik(a + b)] \tag{3.29}$$

$$iK(Ae^{iKa} - Be^{-iKa}) = Q(Ce^{-Qb} - De^{Qb}) \exp[ik(a + b)] \tag{3.30}$$

Now we have four equations, eqs. (3.25), (3.26), (3.29), (3.30), and four unknowns, A, B, C, D. Examining these equations indicates that this is a set of linear homogeneous equations and is thus again an eigenvalue problem, and a solution exists only when the determinant of the coefficients A, B, C, and D equals zero. From this condition, we arrive at the following equation

$$\frac{Q^2 - K^2}{2KQ} \sinh(Qb) \sin(Ka) + \cosh(Qb) \cos(Ka) = \cos[k(a + b)] \tag{3.31}$$

where "sinh(x)" and "cosh(x)" are hyperbolic sine and cosine functions. For a given wavevector k, the only unknown in the above equation is the electron energy E, which is embedded in both K and Q. Thus the above equation can be used to determine a relationship between E and k. To get a better idea of what the solution looks like, let's assume $b \to 0$ and $U_0 \to \infty$, but keep $Q^2ba/2 (= P)$ equal to a constant. Under this approximation, $\sinh(Qb) \approx Qb$, and $\cosh(Qb) \approx 1$. Equation (3.31) reduces to

$$\frac{P}{Ka} \sin Ka + \cos Ka = \cos ka \tag{3.32}$$

We can solve the above equation for (Ka) as a function of (ka) and use eq. (3.24) to find out allowable energy E from K. One important observation is that the magnitude of the left-hand side of eq. (3.32) can be larger than 1 whereas the right-hand side cannot. Therefore, the equation has no solution for those values of K (and thus energy) where the absolute values of the left-hand side are larger than one. A graphical representation of the left-hand side is shown in figure 3.12, where the right-hand side is bounded within $[-1, 1]$. In the shaded region, there is no solution for K, and thus no electrons with energies corresponding to such K values exist. We can convert the solution for K into energy, and redraw the graph as a function of ka as shown in figure 3.13(a). The figure shows that, for each wavevector k, there are multiple values for the electron energy E.

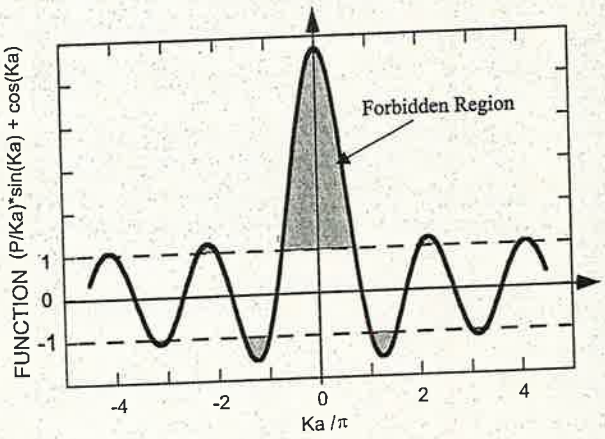


Figure 3.12 Left-hand side of eq. (3.32) as a function of Ka/π. Because the right-hand side is always less than or equal to one, there are regions (the shaded area) where no solution for Ka/π exists, and thus no electrons exist with energy corresponding to the values of K in these regions.

The electron energy forms quasi-continuous bands (because k itself is quasi-continuous) separated from each other by a minimum gap that occurs at $ka = s\pi$ ($s = 0, \pm1, \pm2, \ldots$), or $k = s\pi/a$, at which the right-hand side of eq. (3.32) is ±1. Figure 3.13(a) implies that there are multiple values of k for each energy E. However, the Bloch theorem, eq. (3.28), says that wavefunctions correspond to the wavevectors k separated by $m(2\pi/a)$ (since $b = 0$) are identical, they are the same quantum state and should be counted only once. Thus, rather than plotting the energy eigenvalues for all the wavevectors, we can plot them in one period, as shown in figure 3.13(b). This

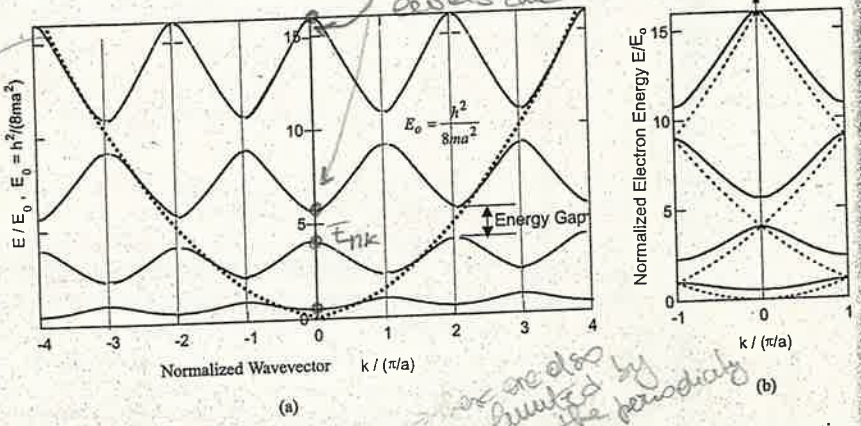


Figure 3.13 Electron energy as a function of its wavevector: (a) extended zone representation; (b) reduced zone representation. Dashed lines are free electron energy levels. Solid lines from Kronig–Penney model.

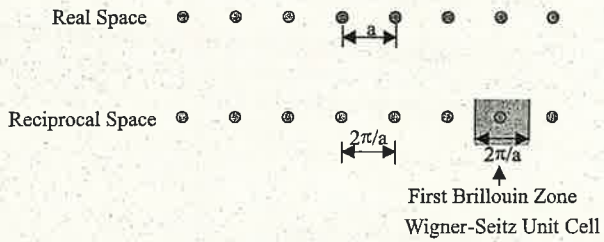


Figure 3.14 The Brillouin zone and Wigner–Seitz unit cell of a one-dimensional lattice.

way of representation is called the *reduced-zone* representation. Often, only half of the band, $[0, \pi/a]$, needs to be drawn because the band is symmetric for both positive and negative wavevector values. The relationship between the energy and the wavevector, as examplified in figure 3.13(b), is the dispersion relation.

We see that the minimum separation between two energy bands occurs at $k_m (= s\pi/a,\ s = 0, \pm1, \pm2, \ldots)$. What do these k_m stand for and why do the minimum separations occur at k_m? For the one-dimensional lattice being considered with a lattice constant equal to a, its reciprocal lattice is also one-dimensional with a lattice constant equal to $2\pi/a$. The Wigner–Seitz cell in the reciprocal lattice, which is the first Brillouin zone as we explained before, is shown in figure 3.14. The boundaries of this primitive unit cell in the reciprocal space are at $\pm\pi/a$. Thus k_m represents the lattice vectors constructed using the primitive lattice vector of the Wigner–Seitz cell in the reciprocal space for the one-dimensional lattice. When we generalize to three-dimensional crystals, k_m will be replaced by the reciprocal lattice vector $\mathbf{G}$. In most cases, the energy gap occurs at the Brillouin zone boundaries, that is, when $\mathbf{k} = \mathbf{G}$. This is not a coincidence since it results from the interference effects of electrons in periodic structures. This mechanism is not very different from the observation of diffraction peaks by X-ray and electron beams that we discussed in section 3.1.4. We also plotted the energy dispersion of a free electron in the reduced-zone representation in figure 3.13, which does not show an energy jump at k_m but is otherwise similar to that of an electron inside the periodic potential. The main effect of the periodic potential is to modify the band structure near k_m, as a result of the diffraction of the electron waves. More discussion on wave interference will be given in chapter 5.

We now determine the value of the wavevector k in the Bloch theorem, using the *Born–von Karman* periodic boundary condition. This boundary condition deals with the end points of a crystal. Ordinarily, we would think that the two end points are different from the internal points. For many applications, however, it is not necessary to distinguish the boundary points from the internal points, because a crystal usually has a tremendous number of lattice points (this is not true for quantum wells, quantum wires, and quantum dots). The Born–von Karman boundary condition requires that the wave functions at the two end points be equal to each other; that is, the two end points [points 1 and $N + 1$ in figure 3.15(a)] are overlapped to form a lattice loop as shown in figure 3.15(b),

$$\Psi[x + N(a + b)] = \Psi(x) \tag{3.33}$$

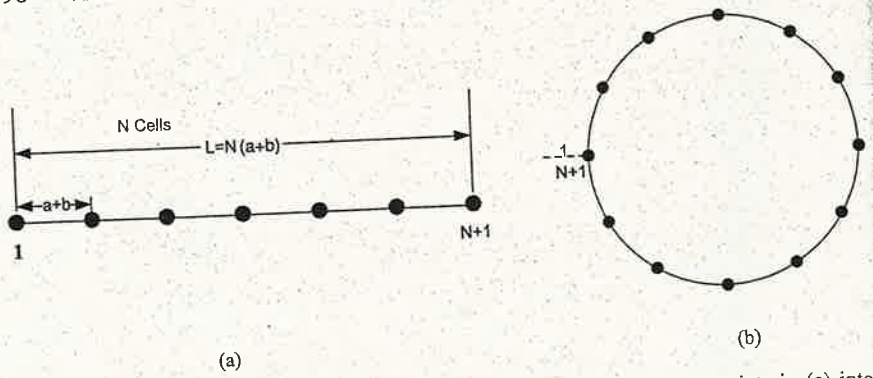


Figure 3.15 The von Karman boundary condition joins the two boundary points in (a) into a periodic loop in (b).

Using Bloch's theorem, eq. (3.23) can be written as

$$\Psi(x) = \Psi(x)\exp[ikN(a+b)] \tag{3.34}$$

The above equation imposes conditions on the allowable wavevectors of the Bloch wave inside the periodic potential

$$k = \frac{2\pi n}{N(a+b)} = \frac{2\pi n}{L} \quad (n = 0, \pm 1, \pm 2, \ldots) \tag{3.35}$$

where L is the length of the crystal.

What is the meaning of the wavevector k in the Bloch theorem? The exponential factor e^{ikx} in eq. (3.28) resembles that for a free electron, as we discussed in section 2.3.1 and eq. (2.34). In the latter case, $\hbar k$ represents the momentum of the free electron. The momentum of an electron in the crystal, however, should be calculated from the wavefunction using the momentum operator $-i\hbar\nabla\Psi$. Such a calculation would show that $\hbar k$ is not the momentum of the electron in a crystal. Nevertheless, in many ways $\hbar k$ for a periodic potential behaves as the momentum of a free electron and thus it is called the *crystal momentum*. The momentum conservation rule during the collision of several particles (now electrons, and later to be generalized to photons and phonons) is

$$\sum_i \hbar\mathbf{k}_i = \sum_f \hbar(\mathbf{k}_f + \mathbf{G}) \tag{3.36}$$

where the indices i and f mean the states before and after the collision, respectively. The addition of the reciprocal lattice vector maintains both k_i and k_f into the first Brillouin zone, and is a consequence of the identical wavefunctions and energy eigenvalues for waves with wavevector $\mathbf{k}$ and $\mathbf{k} + \mathbf{G}$.

The simple model of one electron inside a periodic potential carries many important messages which we will discuss below. This model shows that the electron does not belong to an individual atom. Its wavefunction extends over the whole crystal; therefore, it can be considered a free electron. In reality, the scattering due to the distortion of potential from the perfect periodicity reduces the spatial extension of the wavefunction.

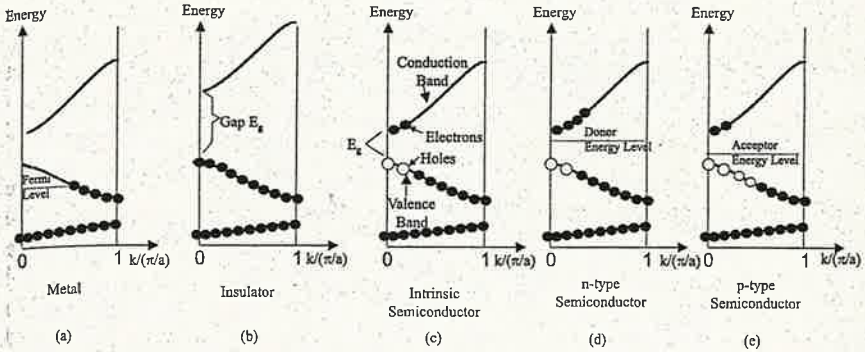


Figure 3.16 Explanation of metals, insulators, and semiconductors based on the one-dimensional band structure. (a) Electrons in metal partially fill a band. The top-most level (E_f) is called the Fermi level. (b) Electrons fill to the top of the band. When the energy gap E_g is large, no electrons can be excited to the next higher energy band and the material is an electrical insulator. (c) When the energy gap E_g is relatively small, some electrons can be thermally excited to the next higher energy band (called the conduction band), leaving the same number of empty states (holes) in the valence band. The material is an intrinsic semiconductor. (d) Impurities (more commonly called dopants) may have an energy level close to that of the conduction band. Electrons can be excited from the impurities and fall into the conduction band, resulting in more electrons than holes. Such a semiconductor is called an n-type semiconductor and the dopants are called donors. (e) When the impurity energy levels are close to the valence band, electrons are excited from the valence band into the impurity level, leaving more holes behind. Such semiconductors are called p-type and the impurities are called acceptors.

Yet still the mean free path of an electron can be as long as thousands of angstroms, and the number of atoms in a cube on the order of one mean free path is enormous, $\sim 10^6$ to $\sim 10^8$ atoms. It is amazing that an electron can zigzag through these atoms without getting scattered. Because of this behavior, we often treat electrons as a gas and neglect the ions completely, except when considering their occasional scattering effect.

Although the above solution is valid only for one electron, the existence of multiple electrons does not affect the qualitative picture of the energy bands, as long as the Coulomb potential between electrons is small compared to the potentials between electrons and ions. With such a simple picture of the energy bands, we can begin to understand the difference between insulators, metals, and semiconductors. In the first Brillouin zone, there are N allowable wavevectors for a lattice chain with N lattice points. Because each wavevector represents a wavefunction and each wavefunction can have a maximum of two electrons with different spins, each band can have a maximum of $2N$ electrons for a one-dimensional lattice. At zero temperature, the filling rule for the electrons is that they always fill the lowest energy level first, as required by thermodynamics. For alkali metals and noble metals that have one valence (free) electron per primitive cell, the band is only half filled since there are only N valence electrons in this case, as shown in figure 3.16(a). The topmost energy level that is filled with electrons at zero kelvin is called the *Fermi level*. The electron energy and momentum can be changed (almost) continuously within the same band because the separation between successive energy levels is small. Thus, the electrons in these metals can flow freely, making the

materials good electrical conductors. If the valence electrons exactly fill one or more bands, leaving others empty [figure 3.16(b)], the crystal will be an insulator at zero kelvin and can be an insulator or a semiconductor at other temperatures, depending on the value of the minimum energy gap between the filled and the empty band. If a filled band is separated by a large energy gap (>3 eV) from the next higher unfilled band, one cannot change the energy and the momentum of an electron in the filled band easily; that is, these electrons cannot move freely and the material is an insulator [figure 3.16(b)]. A semiconductor is essentially similar to an insulator in terms of bandgap. The difference between them is that for a semiconductor, the gap between the filled and the empty bands is not so large (<3 eV), such that some electrons have enough thermal energy to jump across the gap to the empty band above (called the conduction band) [figure 3.16(c)], and these electrons can conduct electricity (these materials are called intrinsic semiconductors). The unoccupied states left behind also leave room for the electrons in the original band (called the valence band) to move. It turns out that the motion of these electrons in the valence band is equivalent to vacant states moving as positive electrons, or holes. The energy of these holes is opposite to that of electrons; it is a minimum at the peak in the valence band and increases as the electron energy becomes smaller. Impurities are added to most semiconductors and these impurities have energy levels somewhere within the bandgap; some are close to the bottom edge of the conduction band or the top edge of the valence band (or band edge). The electrons in the impurity levels can be thermally excited to the conduction band if their level is close to the bottom of the conduction band, thus creating more electrons than holes in the semiconductor. Such semiconductors are called n-type and the impurities are called donors [figure 3.16(d)]. Similarly, if the impurity energy level is close to the valence band edge, electrons in the valence band can be excited to the impurity levels, leaving more empty states or holes behind. Such semiconductors are called p-type and the impurities are called acceptors. Apparently, the number of electrons that are excited to the conduction band depends on the thermal energy of these electrons. This is why semiconductors are especially sensitive to temperature variation. Refer to Appendix B for more discussion on semiconductors.

3.2.2 Electron Energy Bands in Real Crystals

The energy band structures for real materials are more complex but have similarities to the one-dimensional potential model. In three-dimensional crystals, there are different crystallographic directions. Each of the directions has a different periodicity and a correspondingly different potential profile. Taking a cubic conventional cell as an example, the period is equal to the lattice parameter a in the ⟨100⟩ direction, $\sqrt{2}a$ in the ⟨110⟩ direction, and $\sqrt{3}a$ in the ⟨111⟩ direction. In each direction, due to the difference in the potential and periodicity, the energy bands will be different. The energy bands can be plotted along each of these directions and, particularly, along the major crystallographic directions, as in figure 3.17 for Cu, Si, and GaAs, the latter two being common semiconductors. The special points on the first Brillouin zone surface, such as X and L, which are indicated in figure 3.8(b), represent the major crystallographic directions. In copper, the Fermi level at zero kelvin falls inside a band. Therefore, electrons close to the Fermi level are free to move with minimum thermal energy disturbance. For Si and GaAs, the Fermi level is

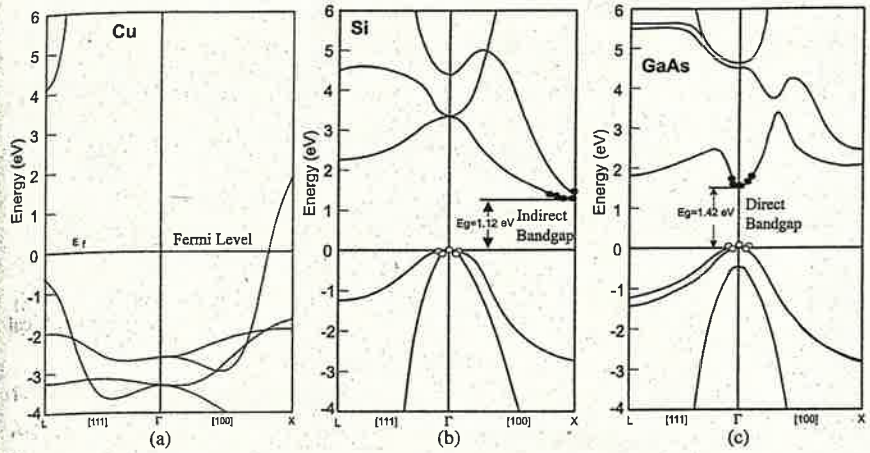


Figure 3.17 Electron band structures of (a) copper (after Mattheiss, 1964), (b) silicon, and (c) GaAs (after Chelikowsky and Cohen, 1976). Copper is a metal because the Fermi level falls inside the bands. The Fermi level for Si and GaAs at zero temperature is at the top of the valence band ($E = 0$). Silicon is an indirect gap semiconductor since the minimum of the conduction band and that of the valence band are not at the same wavevector location. GaAs is a direct gap semiconductor because the minima occur at the same wavevector ($k = 0$ for this case). All bandgap values are those at 300 K.

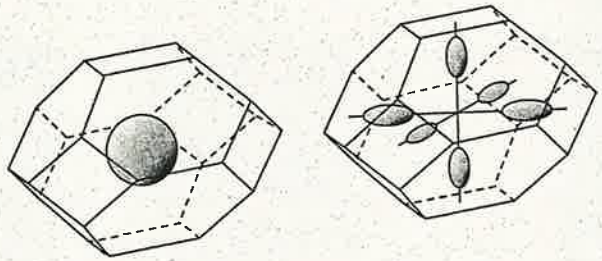
at $E = 0$ so that all bands below this level are filled. Above the filled bands, an energy gap exists in which no electrons are allowed at $T = 0$ K. The values and locations of the energy gap are different for dissimilar crystallographic directions, and the absolute minimum gap is called the *bandgap*. GaAs is a direct gap semiconductor because the minima of the conduction and the valence bands occur at the same wavevector. Si is an indirect gap semiconductor because the two minima do not occur at the same wavevector. Direct and indirect gap semiconductors have major differences in their optical properties. Direct bandgap semiconductors are more efficient photon emitters, semiconductor lasers are made of direct gap semiconductors such as GaAs, whereas most electronic devices are built on silicon technology.

For semiconductors, since most electrons are close to the minimum of the conduction band and holes are close to the minimum of the valence band, it is convenient to express the band structure near the minima in analytical form. Since the minima typically mean that the first-order derivative, $\partial E/\partial k$, is zero (as long as the first-order derivative exists), the second-order terms often are used. For the conduction band, the expansion of the electron allowable energy level near the minimum is often written in the form

$$E - E_c = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_{11}} + \frac{k_y^2}{m_{22}} + \frac{k_z^2}{m_{33}}\right) \tag{3.37}$$

where

$$m_{ij} = \frac{\hbar^2}{(\partial^2 E/\partial k_i \partial k_j)} \tag{3.38}$$



(a) (b)

Figure 3.18 Constant electron energy surfaces in reciprocal space (or k-space): (a) a spherical band such as in GaAs; (b) an elliptical band such as in Si (after Shur, 1990).

is representative of the local curvature of the E–$\mathbf{k}$ relation and is called the effective mass, and E_c is the bottom of the conduction band, or band edge. The effective mass derives its name from the fact that the free electron energy can be expressed as $E = \hbar^2 k^2/2m$. The effective mass is a second-order tensor, according to eq. (3.38). Equation (3.37) is valid only when x, y, and z are the principal directions of the effective mass tensor and E_c is at $\mathbf{k} = 0$. For a spherical band, $m_{11} = m_{22} = m_{33}$ and the energy–wavevector relation, that is, the dispersion relation, reduces to the free electron form, except that the free electron mass is replaced by the effective mass. Using this effective mass, we can treat the motion of electrons in the conduction band (or the motion of holes in the valence band, which depends on the hole effective mass) as free electrons (or holes).

By setting the energy in eq. (3.37) to a constant, we can plot the constant energy surface in k-space, that is, inside the first Brillouin zone. Examples of constant energy surfaces are given in figure 3.18.

3.3 Lattice Vibration and Phonons

The previous section deals with the electron energy levels in solids. We now turn our attention to the vibrational energy levels of atoms, or the lattice vibration. Here, the term "lattice" really means the lattice together with its basis (we recall that the lattice in crystallography is a mathematical abstraction of periodic points in space). Similar to what we have done for the electronic energy levels, a simple one-dimensional model will render the fundamental characteristics of lattice vibrations. Therefore, we will start our discussion with a one-dimensional model and then move on to more complicated cases.

3.3.1 One-Dimensional Monatomic Lattice Chains

Let us consider first a one-dimensional monatomic chain, as shown in figure 3.19. By limiting our consideration to monatomic chains, we have made the following simplifications: (1) the mass at each lattice point is the same; (2) the separation between adjacent atoms is the same; and (3) the force interactions between adjacent atoms are the same. Such simplifications are no longer valid when the basis comprises more than one atom.

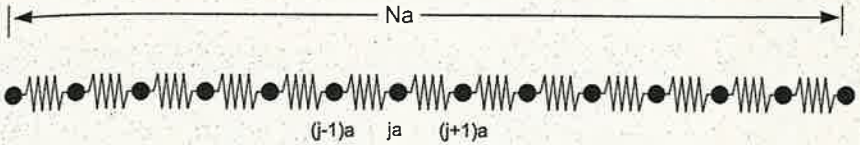


Figure 3.19 One-dimensional monatomic lattice chain model.

We will make the following assumptions in the analysis. First, we consider the force interaction between the nearest neighbors only. Second, the interaction force between atoms is assumed to be a harmonic force (which obeys Hooke's law) with spring constant K. This assumption can be justified in a similar fashion to that done for harmonic oscillators. Now consider a typical atom j. The displacement of atom j from its equilibrium position x_j^0 is

$$u_j = x_j - x_j^0 \tag{3.39}$$

The force acting on atom j comes from two components. One is due to the relative displacement between atom $(j-1)$ and atom j, and the other is due to the relative displacement between atom j and $(j+1)$. The net force is then

$$F_j = K(u_{j+1} - u_j) - K(u_j - u_{j-1}) \tag{3.40}$$

At this point, we have two choices for continuing this discussion. The first choice includes writing out the potential and the Schrödinger equation, and then solving for the allowable energy. But for our understanding here, a classical approach is easier to grasp and the quantum effect can be added into the classical solution later. Let's apply Newton's second law to atom j to obtain

$$m\frac{d^2u_j}{dt^2} = K(u_{j+1} - u_j) - K(u_j - u_{j-1}) \tag{3.41}$$

The above equation is a special form of the differential wave equation

$$m\frac{\partial^2 u}{\partial t^2} = Ka^2\frac{\partial^2 u}{\partial x^2} \tag{3.42}$$

which has a solution of the form $u \propto e^{-i(\omega t - kx)}$. Such a similarity suggests a wave type of solution for Eq. (3.41),

$$u_j = A\exp[-i(\omega t - kja)] \tag{3.43}$$

where ja is the discrete equilibrium location of atom j, ω is the frequency, and k is the wavevector. The major difference between this "guessed" solution and the conventional continuous wave, $\exp[-i(\omega t - kx)]$, lies in the use of "ja" as a discrete lattice coordinate rather than an infinitely divisible continuum coordinate "x" used in a continuous medium, because talking about the vibration at locations other than the atomic sites is meaningless.

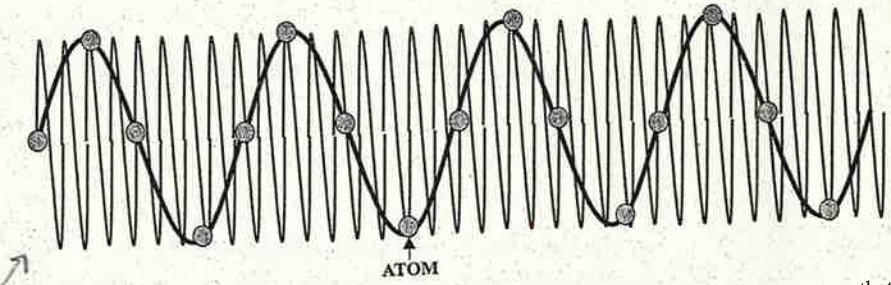


Figure 3.20 Snapshots of the atom displacements represented by two different wavevectors that differ by $2\pi/a$. The displacements of the atoms are the same for the two given wavevectors drawn in the figure. The short wavelength mode means that atom displacements would exist in the empty space between two closest atoms, which is not possible. Thus, the allowable wavevectors for phonons are limited to the first Brillouin zone.

Substituting the guessed solution (3.43) into eq. (3.41), we get

$$-m\omega^2 = K[e^{ika} + e^{-ika} - 2] \tag{3.44}$$

or

$$\omega = 2\sqrt{\frac{K}{m}}\left|\sin\frac{ka}{2}\right| \tag{3.45}$$

The allowable wavevector can similarly be determined from the von Karman periodic boundary condition, as we did for electrons,

$$k = \frac{2\pi n}{Na} \quad (n = 0, \pm1, \pm2, \ldots) \tag{3.46}$$

When treating the electronic energy levels, we assert that the states for k and $[k + m(2\pi/a)]$ are identical, as shown in figure 3.13(a). For the lattice vibration, we can argue physically that the allowable values of n for lattice vibrations lie between $-N/2$ and $N/2$ and the extension of k beyond the first Brillouin zone is meaningless. From $k = 2\pi/\lambda$, where λ is the wavelength, we observe that if $n > N/2$, λ will be less than twice the atom spacing; this means that there are no atoms between one period, as shown in figure 3.20. Talking about atomic displacement in empty spaces with no atoms is meaningless. Thus, the allowable wavevector for a lattice vibration is naturally confined to the first Brillouin zone.

Equation (3.46), together with the limitation on n, states that for a monatomic, one-dimensional lattice chain with N atoms there are N allowable wavevectors. Each of these wavevectors corresponds to one mode of the lattice vibration. Each mode is mathematically described by Eq. (3.43) and is called a *normal mode*. Normal modes are a familiar concept in the vibration analysis of structures, drums, and musical instruments.

Figure 3.21(a) shows the dispersion relationship between the vibrational frequency and the wavevector for a monatomic lattice chain. Although this relationship is nonlinear, very often the Debye approximation is used, which assumes a linear dispersion relation

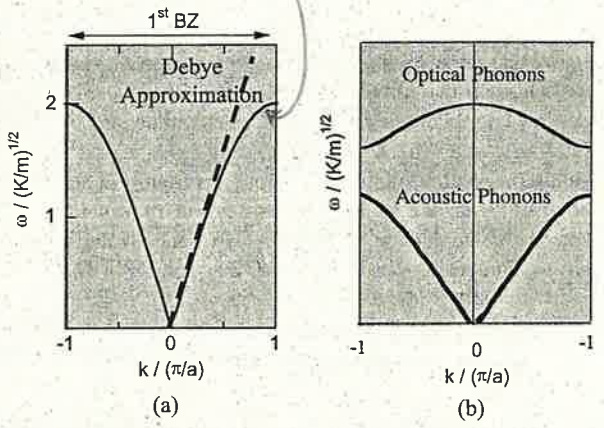


Figure 3.21 Phonon dispersion of a one-dimensional (a) monatomic lattice chain and (b) diatomic lattice chain. The Debye approximation use a linear relationship between the frequency and the wavevector. BZ stands for Brillouin zone.

between the frequency and the wavevector. This approximation is valid at low frequencies but is not a good approximation at high frequencies. In the very low frequency region the lattice vibration carries the sound wave. We will return to a discussion of the Debye approximation in the next section.

3.3.2 Energy Quantization and Phonons

The above treatment is based on classical mechanics. In the classical solution, the vibrational energy at each frequency is determined by the amplitude A in eq. (3.43). If we use the Schrödinger equation to solve the same problem, we will witness that the dispersion relation is the same. The difference is that the energy of the lattice vibration is quantized. This result is similar to the case of a simple harmonic oscillator that we dealt with in chapter 2, for which the vibrational frequency is the same as that of the classical mass–spring system [eq. (2.56)], but the energy of the oscillator is quantized. At each frequency determined by eq. (3.45), the allowable energy levels are

$$E_n = h\nu\left(n + \frac{1}{2}\right) \quad (n = 0, 1, 2, 3, \ldots) \tag{3.47}$$

This quantization is identical to that of a single harmonic oscillator, eq. (2.54), and to Planck's relation between photon energy and its frequency. The similarity leads to the concept of phonons, parallel to the concept of photons, as the minimum quantum of lattice vibration. Each phonon has an energy of $h\nu$. Its momentum is, according to the de Broglie relation, $h/\lambda = \hbar k$. Under the phonon picture, we can forget about atoms and consider the lattice waves or the phonon particles in a crystal just as we treat the electromagnetic waves and photons in a box. For many materials (nonmetals), heat is conducted by lattice vibrations. From the analogy between photons and phonons, it is logical to expect that lattice heat conduction should obey a law similar to the

Stefan–Boltzmann law, which is proportional to the fourth power of the temperature. This temperature dependence is not evident in most of (if not all) the heat conduction calculations that we learned in an undergraduate heat transfer course. The reason is that in most cases the phonon scattering is so strong that these waves cannot travel a long distance. At very low temperatures, in fact, we can calculate heat conduction just as we do with radiation (Casimir, 1938). A similar situation arises when dealing with very thin films, for which internal scattering does not occur as often as in bulk materials.

Although eq. (3.47) is identical to the expression of a single spring-mass harmonic oscillator, there are indeed differences. For an isolated harmonic oscillator only one vibrational frequency exists, while for a lattice chain N wavevectors (where N is the total number of atoms in the chain) present, and thus N frequencies are possible. Each of these allowable frequencies supports an energy that must be a multiple of $h\nu$. How many phonons actually occupy each state at a specific frequency is a point that we will address in the next chapter.

We should emphasize that although photons and phonons can be thought of as particles with an energy $h\nu$, these particles are fundamentally different from electrons. Some of the differences are discussed here. Unlike electrons, phonons and photons at rest do not have mass; and are fictitious particles since they are the quantization of the normal mode of a field. Electrons obey the Pauli exclusion principle, which says that each quantum state can have only one electron at most. Photons and phonons are not limited by the Pauli exclusion principle. Each quantum state, which corresponds to one set of wavevectors, can have many phonons and photons, as is indicated by eq. (3.47). The differences in behavior between the various quantum species will later be reflected in their statistical behavior, which we will discuss in more detail in chapter 4. Due to the differences in their statistical behavior, electrons are called *fermions*, while phonons and photons are called *bosons*.

3.3.3 One-Dimensional Diatomic and Polyatomic Lattice Chains

If there is more than one atom at each lattice point, several changes should be made to the above monatomic model. First, the masses of two adjacent atoms may be different (such as in a GaAs crystal) or the same (such as the two silicon atoms in the basis of a silicon crystal). Second, the distances between two adjacent atoms may be different. And third, the spring constants between adjacent atoms can also be different. Taking these differences into consideration and following a similar analysis as for the one-dimensional lattice chain, one can derive the dispersion relation between the vibrational frequency and the wavevector. For a diatomic lattice chain, a typical solution looks like figure 3.21(b). The lower branch is similar to that of a monatomic lattice chain, and the upper branch is similar to the folded representation of the electronic energy levels. Note that a gap exists between the two branches. The upper branch is due to the additional degrees of vibrational freedom caused by the additional atom in a unit cell. In the long wavelength limit, the vibration of the two atoms at each lattice point can be either in-phase or out-of phase, as illustrated in figure 3.22. Clearly, the out-of-phase modes require more energy. The lower frequency branch is called the acoustic branch, and the higher frequency one is called the optical branch, because the high-frequency phonons in the optical branch can interact with electromagnetic waves more easily. In general, for a one-dimensional lattice chain with m atoms in a basis and N lattice points

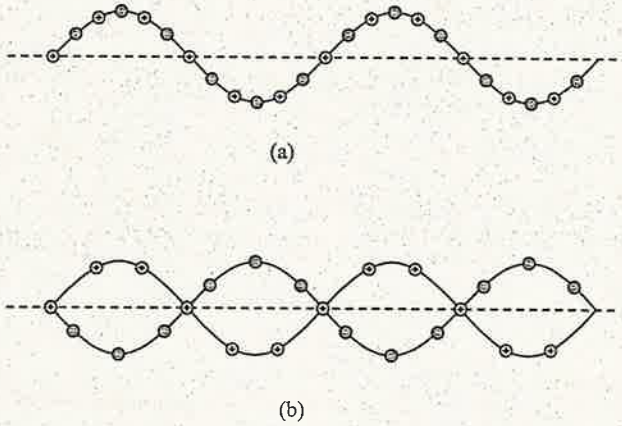


Figure 3.22 In the long wavelength limit, the two atoms in the acoustic phonon branch vibrate in phase (a) but are out of phase in the optical branch (b).

in the chain, there exists one acoustic branch with N acoustic modes and $(m-1)$ optical branches with $(m-1)N$ optic modes.

3.3.4 Phonons in Three-Dimensional Crystals

In a one-dimensional lattice, the phonon waves are longitudinal waves. In a three-dimensional crystal, the atoms can vibrate in three dimensions. Thus, we will have three vibrational branches for the acoustic modes—one longitudinal and two transverse branches. Furthermore, if there are m atoms per lattice point, then $3(m-1)$ optical branches exist, of which $2(m-1)$ are transverse optical phonons and the rest are longitudinal optical phonons. In a transverse wave, the atomic displacement direction is perpendicular to the wave propagation direction. The two transverse branches will coincide with each other if the two vibrational directions are symmetric. As with electrons, phonon dispersions along different crystallographic directions are different. Figure 3.23(a) shows the phonon dispersion relations for lead, which has an fcc structure and one atom as a basis, and thus three acoustic branches along each crystallographic direction. Figure 3.23(b) shows the phonon dispersion for Si, which has an fcc structure but two atoms as the basis. Thus, Si has three acoustic branches and three optical branches in each direction. In some high-symmetry directions, such as the [100] directions, the two transverse phonon branches collapse onto one curve.

3.4 Density of States

In the previous chapter, we introduced the concept of degeneracy for quantum states that have the same energy levels. From the one-dimensional electron model, we see that each wavevector k corresponds to one wavefunction and thus two quantum states after we include the electron spin. Because the dispersion relation is symmetric for

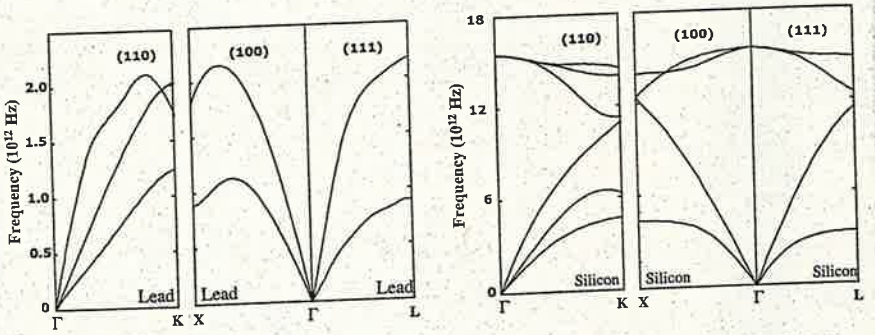


Figure 3.23 Phonon dispersion relations for lead and silicon; both have an fcc lattice. Lead has one atom as the basis, so only acoustic phonons are present. Silicon has two atoms as a basis, and thus there are three acoustic phonons and three optical phonons (lead after Brockhouse et al., 1962, and silicon after Giannozzi et al., 1991).

both positive and negative k values, the degeneracy for electrons at each energy level can be regarded as 4. In three-dimensional crystals, however, the dispersion picture is totally different because each crystallographic direction has its own dispersion relation between the energy and the wavevector. There exist potentially many combinations of wavevectors that have the same energy. As shown in figure 3.24, for the constant energy surface in the wavevector space for a spherical band, that is, equal effective mass [eq. (3.37)] but in two dimensions. Clearly, there can be many wavevectors on each constant energy surface. Because the energy levels in solids are quasi-continuous, we use the concept of density of states to describe the energy degeneracy. We will discuss next how the density of states is defined, and derive expressions for the density of states for electrons, phonons, and photons, using simplified dispersion relations derived earlier.

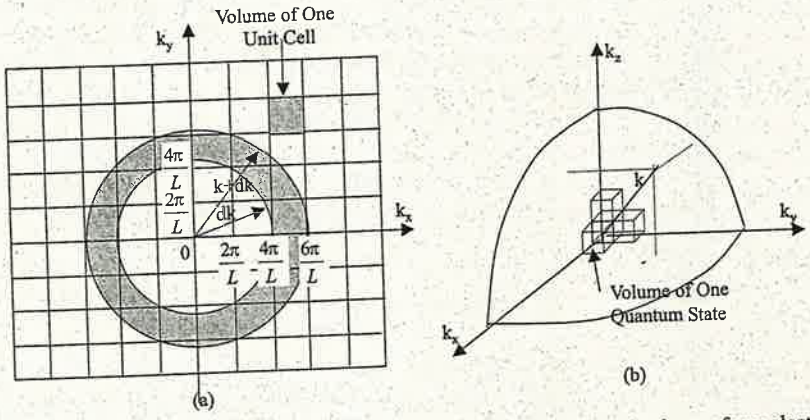


Figure 3.24 Constant energy surface in k-space for a spherical band, and volume of one electron state in the k-space: (a) two-dimensional projection; (b) three-dimensional view.

see p. 99

3.4.1 Electron Density of States

Consider a spherical parabolic band with the following relationship between the energy and the wavevector,

$$E - E_c = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2 + k_z^2) = \frac{\hbar^2 k^2}{2m^*} \left(k_x, k_y, k_z = 0, \pm\frac{2\pi}{Na}, \pm\frac{4\pi}{Na}, \ldots, \pm\frac{\pi}{a}\right) \quad (3.48)$$

where k is the magnitude of the wavevector,

$$k^2 = k_x^2 + k_y^2 + k_z^2$$

For each constant E value, there can be potentially many combinations of k_x, k_y, k_z satisfying eq. (3.48). To find the density of states, we refer to figure 3.24. The allowable wavevectors for k_x, k_y, and k_z are integral multiples of $2\pi/L$, where $L(= Na)$ is the length of the crystal along the x, y, and z directions. The volume of each quantum mechanical state in three-dimensional k-space is $(2\pi/L)^3$. The number of states between k and $k + dk$ in three-dimensional space is then

$$\#\text{ states} = 2 \times \frac{4\pi k^2 dk}{(2\pi/L)^3} = \frac{Vk^2 dk}{\pi^2} \quad (3.50)$$

where the factor of two accounts for the electron spin and $V(= L^3)$ is the crystal volume. In the above treatment, we have implicitly assumed that k and E are continuous functions. This should be valid as long as the number of atoms in the system is large enough (N is large).

On the basis of eq. (3.50), we can define the density of states as the number of quantum states per unit interval of the wavevector and per unit volume,

$$D(k) = \frac{\#\text{ states between } k \text{ and } k + dk}{V dk} = \frac{k^2}{\pi^2} \quad (3.51)$$

We can also define the *density of states* as the number of states per unit volume and per unit energy interval

$$D(E) = \frac{\#\text{ states between } E \text{ and } E + dE}{V dE} = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}(E - E_c)^{1/2} \quad (3.52)$$

where we have used eq. (3.48) to replace k and dk in terms of E and dE. Sometimes, we also define the density of states on the basis of a frequency interval (as we will do for phonons). For electrons, eq. (3.52) is used most often. A schematic of eq. (3.52) is shown in figure 3.25.

The density of states is a purely mathematical convenience, but nevertheless, it is central for correctly counting the number of electrons and the energy (or charge and momentum) that they carry. As a simple example of how the density of states is needed, let's evaluate the electron energy of the topmost level at $T = 0$ K, that is, the Fermi level E_f. At 0 K, the filling of electron quantum states starts from the lowest energy

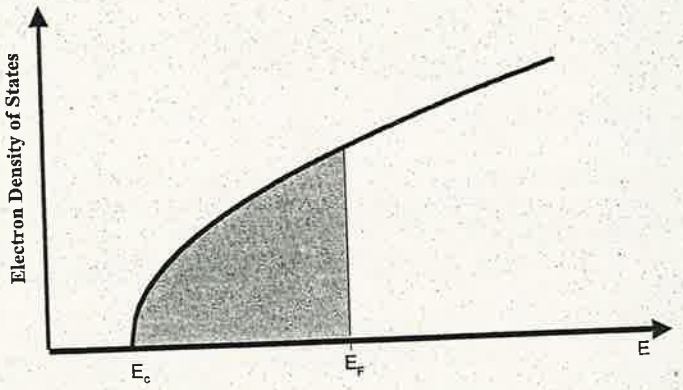


Figure 3.25 Density of states of electrons in a bulk crystal.

level and moves up from one energy level to the next until all electrons are placed into distinct quantum states. The number of electrons per unit volume at $T = 0$ K is

$$n = \int_{E_c}^{E_f} D(E)dE = \frac{1}{3\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}(E_f - E_c)^{3/2} \tag{3.53}$$

Often, E_c is taken as a reference point and set to zero. From the above relation, we can calculate easily the Fermi level from the given electron number density and the effective mass.

Example 3.4 *Fermi level*

A gold crystal has an fcc lattice with one gold atom per lattice point and a lattice constant of 4.08 Å. Every gold atom has one valence electron. Estimate the electron Fermi level in a gold crystal.

Solution: We assume that the electron effective mass in gold is identical to that of free electrons. For an fcc lattice, there are four lattice points and thus four valence electrons. Consequently, the electron number density is

$$n = 4/(4.08 \times 10^{-10})^3 = 5.89 \times 10^{28}\ \text{m}^{-3} = 5.89 \times 10^{22}\ \text{cm}^{-3}$$

From eq. (3.53), the Fermi level at 0 K is

$$E_f = \frac{\hbar^2}{2m}(3\pi^2 n)^{2/3} = 8.66 \times 10^{-19}\ \text{J} = 5.4\ \text{eV}$$

Comment: We give the number density in terms of cm^{-3} because such a unit is commonly used in semiconductors. Typical dopant levels for silicon-based semiconductor devices are $\sim 10^{17}$–$10^{19}\ \text{cm}^{-3}$, which is much smaller than those in metals.

3.4.2 Phonon Density of States

The phonon dispersion relation differs considerably from the electron dispersion relation, as is seen by comparing eq. (3.45) with eq. (3.37). Rather than having spins, phonons have different polarizations (branches) in crystals. One common simplification to the phonon dispersion is the Debye approximation, which assumes a linear relation between the frequency and the wavevector,

$$\omega = v_D|\mathbf{k}| = v_D k \tag{3.54}$$

Following a similar procedure to our previous treatment of electrons, we calculate the density of states of phonons. The volume of one phonon state in **k**-space is $(2\pi/L)^3 = (2\pi)^3/V$. Under the Debye approximation, the density of states for phonons per unit volume and per unit frequency interval is then

$$D(\omega) = \frac{dN}{Vd\omega} = 3\frac{4\pi k^2 dk/(2\pi/L)^3}{Vd\omega} = \frac{3\omega^2}{2\pi^2 v_D^3} \tag{3.55}$$

where a factor of 3 has been added to account for the three polarizations of phonons.

The Debye approximation, as represented by eq. (3.54), implies that all phonon branches have the same speed in all directions; in other words, the medium is isotropic. In reality, even a cubic crystal does not have the same velocity in different crystallographic directions. This isotropic medium will therefore have a different lattice constant a_D compared to real crystals. To calculate the equivalent lattice constant of a Debye crystal, we require that the total number of states in this isotropic crystal equal that of a real crystal. We assume that k_D is the wavevector at the boundary of the Brillouin zone, that is, $k_D = \pi/a_D$ in the Debye model. The number of states existing in a real crystal having N ions is equal to $3N$.* A Debye crystal should contain the same number of states,

$$3N = 3 \times \frac{4\pi k_D^3/3}{(2\pi)^3/V} \tag{3.56}$$

which gives

$$k_D = \left(\frac{6\pi^2 N}{V}\right)^{1/3} \quad \text{or} \quad a_D = \left(\frac{\pi V}{6N}\right)^{1/3} \tag{3.57}$$

where k_D is called the Debye cutoff wavevector. Correspondingly, the Debye frequency is $\omega_D = v_D k_D$.

It should be noted that the Debye approximation does not represent the reality at the Brillouin zone boundary, where the dispersion is flat, as shown in figures 3.21 and 3.23, as well as eq. (3.45). It is also not suitable for optical phonons. For the latter, a much better approximation is to set all the frequencies to the same value, that is, $\omega = \omega_E$ (for each branch). If there are N' lattice points in the crystal, there should be a total of N' modes for each branch, with the degeneracy equal to N'. This approximation was first used by Einstein to explain the specific heat, and the resultant theory is called the Einstein model. A more thorough discussion regarding the Einstein model will be given in the next chapter.

*This includes both acoustic and optical modes. In a Debye model, however, all the modes are approximated by the three identical acoustic branches.

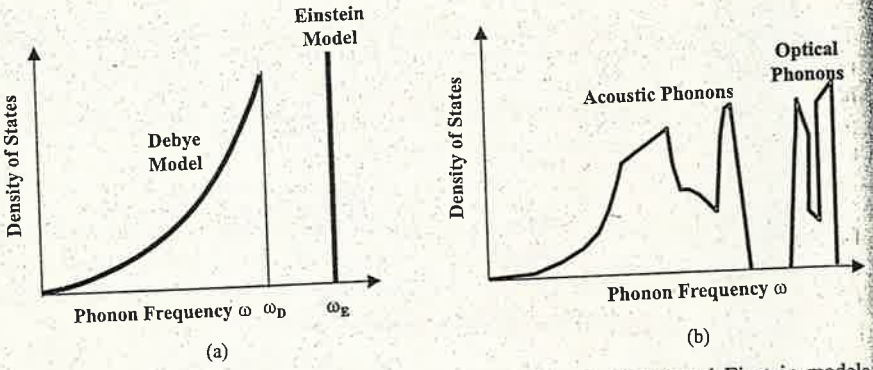


Figure 3.26 Phonon density of states, (a) approximated by the Debye and Einstein models and (b) in real crystals.

The densities of states for the Debye and Einstein models are illustrated in figure 3.26. The Debye model gives $D \propto \omega^2$, while the Einstein model gives a spike at ω_E. The densities of states in real crystals can be quite different from the predictions of these simple models, as illustrated in figure 3.26(b). At each frequency that the phonons intersect the zone boundary, a singularity, called the van Hove singularity, appears in the density of states because the dispersion curve is perpendicular to the zone boundary.

3.4.3 Photon Density of States

Photons also have a linear dispersion between frequency and its wavevector, $\omega = ck$, which is identical to that of phonons under the Debye approximation. Consider an electromagnetic wave in a cubic box of length L. The electromagnetic fields can be decomposed into normal modes using Fourier series, as we did for phonons. The allowable wavevectors are then

$$k_x, k_y, k_z = 0, \pm 2\pi/L, \pm 4\pi/L, \ldots \tag{3.58}$$

Hence, as before, photons share much commonality with phonons. However, significant differences exist: unlike phonon waves in a crystal, which have a minimum wavelength as imposed by the interatomic distance, no such a limit presents on the wavevector for photons. Following a derivation similar to phonons, we can obtain the density of states for an electromagnetic wave as

$$D(\omega) = \frac{dN}{V d\omega} = \frac{\omega^2}{\pi^2 c^3} \tag{3.59}$$

One difference in the above equation from eq. (3.55) is that a factor of two rather than three is used to reflect the fact that electromagnetic waves (photons) have two transverse polarizations, whereas phonons can be longitudinally polarized as well. The other difference is that while the phonon density of states has a cut-off frequency given by the Debye frequency, photons do not have such a cutoff frequency.

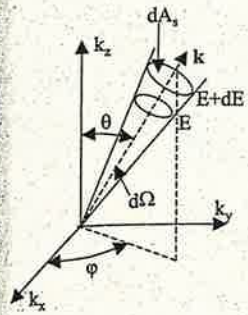


Figure 3.27 Differential density of states and solid angle.

3.4.4 Differential Density of States and Solid Angle

Although the density of states is usually defined on the basis of the magnitude of wavevector k or energy, we found that it is useful in the study of transport process to define a differential density of states. We first define the solid angle Ω in k-space as (figure 3.27)

$$d\Omega = \frac{dA_s}{k^2} = \sin\theta \, d\theta \, d\varphi \tag{3.60}$$

where dA_s is a differential area perpendicular to the $\mathbf{k}$ direction and θ and ϕ are polar and azimuthal angles, defined in figure 3.27. With this definition, it is easy to show that the solid angle over the entire space is 4π. The differential density of states, along a specific wave vector direction $\mathbf{k}$ is defined as

$$dD(E, \mathbf{k}) = \frac{\text{No. of states within}(E, E + dE) \text{ and } d\Omega}{V dE d\Omega} = \frac{D(E)}{4\pi} \tag{3.61}$$

where the second equality applies to isotropic dispersions only.

3.5 Energy Levels in Artificial Structures

We touched upon quantum wells and quantum dots in the previous chapters. These structures can be made by various synthesis routes such as molecular beam epitaxy and self-assembly. The energy states of electrons, phonons, and photons in such structures are often different from those in their bulk counterparts. Many of the novel properties of these artificial structures originate from the different energy states and, consequently, different densities of states. These artificial structures can be categorized into two groups. One imposes new boundary conditions, as in quantum wells and quantum dots, and the other creates new periodicity, as in superlattices. We will briefly illustrate some examples in this section.

3.5.1 Quantum Wells, Wires, Dots, and Carbon Nanotubes

A quantum well can be formed by sandwiching a thin film between two other materials. For example, a thin layer of GaAs (typically < 200 Å can be sandwiched between

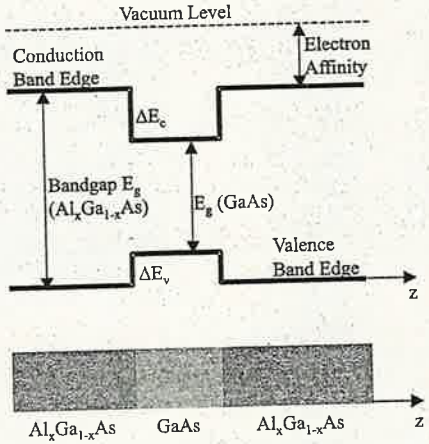


Figure 3.28 A quantum well can be formed by sandwiching one material (GaAs) between two materials ($Al_xGa_{1-x}As$). The top figure shows the band-edge alignment. The band-edge offset provides the potential barrier to confine electrons in the GaAs layer.

two $Al_xGa_{1-x}As$ layers, where the AlAs volume fraction x can be controlled to a high precision between 0 and 1. Both GaAs and AlAs are semiconductors. AlAs (bandgap 2.17 eV) and its alloy with GaAs, $Al_xGa_{1-x}As$, have larger bandgaps than GaAs (bandgap 1.42 eV). When two materials form an interface, a general rule for the alignment of the bands is that the vacuum level must be the same. The energy required to bring an electron from the conduction band edge to the vacuum level, that is, the energy needed to take an electron out of the conduction band to vacuum, is called the electron affinity [figure 3.28], which is different for different materials. For the $Al_xGa_{1-x}As/GaAs/Al_xGa_{1-x}$ As sandwich structure, the final band-edge alignment is shown in figure 3.28. A potential difference exists at the interface in both the conduction band and the valence band, called the band-edge offset. For electrons in the conduction band, the conduction band-edge offset ΔE_c provides the potential barrier to form a quantum well. Similarly, the valence band-edge offset provides a potential well for holes. In chapter 2, we solved the energy levels for a one-dimensional quantum well. In a realistic quantum well constructed of a thin film, the electrons are not constrained in the x–y plane (the film plane) and the potential barrier is not infinitely high. Hence, the solution is more complicated and will not be pursued here (see exercise 2.12). Instead, we assume for simplicity an infinite potential barrier height in the z-direction. We then can obtain the following energy–wavevector relation from the Schrödinger equation,

$$E(k_x, k_y, n) = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2) + n^2\frac{\hbar^2\pi^2}{2m^*d^2} \quad (n = 1, 2, \ldots, N) \tag{3.62}$$

where d is the width of the quantum well and m^* the electron effective mass. In the above relation, the dispersion relations in the k_x and k_y directions are the same as in the bulk material, but in the z-direction, the energy becomes discrete as given by the one-dimensional particle-in-a-box model presented in chapter 2.

Example 3.5 *Quantum well density of states*

For the energy dispersion relation given by eq. (3.62), determine the corresponding electron density of states.

Solution: The allowable wavevectors for k_x and k_y are

$$k_x, k_y = \pm\frac{2\pi n}{Na} \ (n = 0, 1, 2, \ldots) \tag{E3.5.1}$$

where a is the lattice constant. Thus, the area per state in the k_x, k_y plane is $(2\pi/L)^2$. We can rewrite eq. (3.62) as

$$E(k_x, k_y, n) = \frac{\hbar^2 k_{xy}^2}{2m^*} + n^2\frac{\hbar^2\pi^2}{2m^*d^2} \tag{E3.5.2}$$

where $k_{xy}^2 = k_x^2 + k_y^2$. Examining eq. (E3.5.2), we see that for energy E larger than $E_n = n^2\hbar^2\pi^2/(2m^*d^2)$ but smaller than $E_{(n+1)}$, there exist n series of k_{xy} values that can satisfy eq. (E3.5.2) because the value n in the last term can be any integer between 1 and n. For each of these n series, the number of states between k_{xy} and $k_{xy} + dk_{xy}$ is

$$N = \text{No. of states} = \frac{4\pi k_{xy}dk_{xy}}{(2\pi/L)^2} = \frac{Ak_{xy}dk_{xy}}{\pi} \text{ (for each allowable series of } k_{xy})$$

where $A = L^2$ is the area along the x–y plane. From eq. (E3.5.2) we get

$$dE = \frac{\hbar^2}{m^*}k_{xy}dk_{xy} \text{ (for each allowable series of } k_{xy}) \tag{E3.5.3}$$

So the electron density of states per energy interval and per unit area of film for each allowable k_{xy} series is

$$D_1(E) = \frac{N}{AdE} = \frac{m^*}{\pi\hbar^2} \tag{E3.5.4}$$

For an energy state $E_n < E < E_{(n+1)}$, the total number of states is $D(E) = nD_1(E)$. The electron density of states for such a dispersion relation is a staircase, as illustrated in figure E3.5.

As with electrons, phonon energy states in quantum structures are also altered because of new boundary conditions. Figure 3.29 compares the phonon dispersion and the phonon density of states in a freestanding thin film (Yang and Chen, 2000). The phonon spectrum

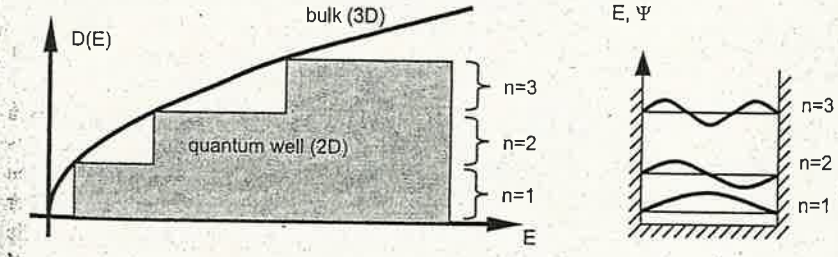


Figure E3.5 Density of states of electrons in a quantum well.

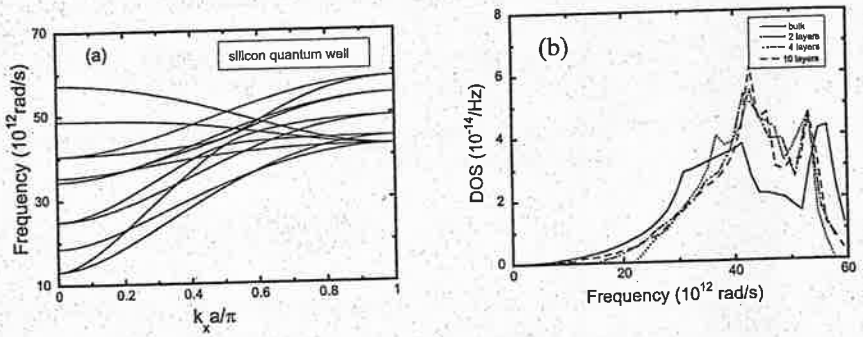


Figure 3.29 (a) Phonon dispersion and (b) density of states in quantum wells (Yang and Chen, 2000).

change can be seen experimentally through, for example, Raman spectroscopy, which probes the phonons through the frequency shift of a photon that interacts with a phonon (Weisbuch and Vinter, 1991). Numerous studies have been devoted to the effects of phonon confinement in quantum structures (Bannov et al., 1995). Recent applications include the use of phonon confinement to reduce thermal conductivity and thus, increase the thermoelectric energy conversion efficiency (Chen, 2001).

The quantum effects for nanometer-scale wires (quantum wires) and nanometer-scale dots (quantum dots) are expected to be even stronger than in quantum wells because of the additional boundary conditions on the electron or phonon motion in one or two more directions. A recent discovery is that of nanoscale tubular structures, particularly carbon nanotubes (Iijima, 1991). A carbon nanotube can be considered as the rolling of an atomic sheet (or several atomic sheets) of graphite carbon (Dresselhaus et al., 2001). Graphite has a close-packed hexagonal structure, as shown in figure 3.5(c). The bonding between different layers is through the van der Waals bond, which is weaker than the covalent bonds within each layer. If only one atomic layer rolls up, the nanotube thus formed is called a single-walled carbon nanotube. If several layers roll up, the nanotube formed is called multiwalled. Depending on the nanotube diameter and the orientation of the major crystallographic directions with the nanotube axis, the nanotube can be a semiconductor or a metal, due to quantum size effects. The electron and phonon energy states in carbon nanotubes are very different from those in their bulk materials, leading to some special properties. The mechanical strength and thermal conductivity of these tubes are expected to be very high (Kim et al., 2001). Research is actively exploring various properties and applications of carbon nanotubes (Dresselhaus et al., 2001).

3.5.2 Artificial Periodic Structures

We have observed that the periodicity that naturally exists in bulk crystals plays a crucial rule in determining the electron and phonon energy levels. Natural systems are three dimensional, with a periodicity determined by the lattice constants. One can also create artificial periodic structures, for example, by repeatedly growing a thin layer of GaAs and a thin layer of AlAs on the same substate. In fact, artificial periodic structures have been used widely in optical coatings, such as in the making of optical

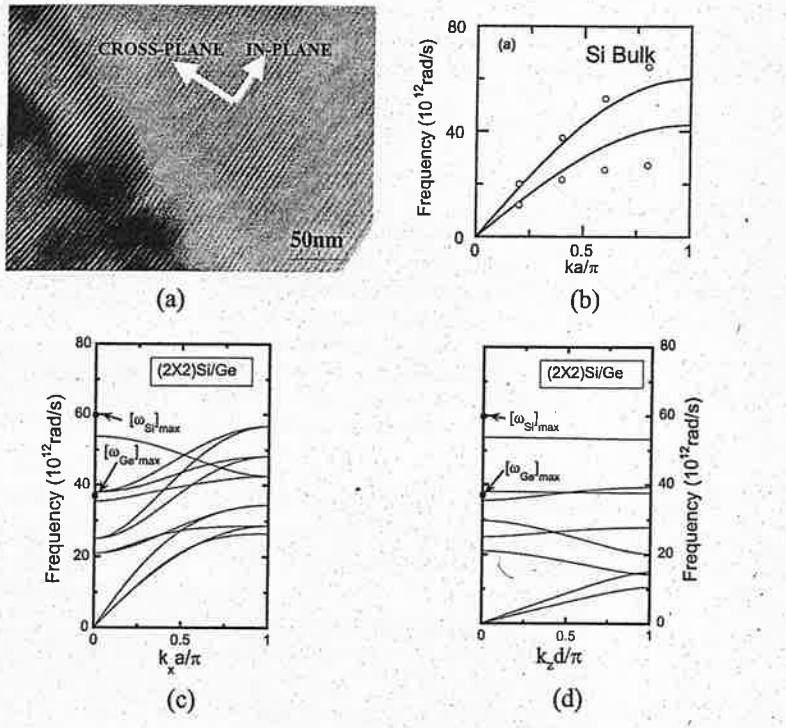


Figure 3.30 (a) A Si/Ge superlattice (Borca-Tasciuc et al., 2000); (b) acoustic phonons in bulk silicon; and phonon dispersion (c) along and (d) perpendicular to the superlattice plane (Yang and Chen, 2004).

interference filters that consist of alternating layers of quarter-wavelength thin films (Heavens, 1965). An optical wave can be totally reflected when the wavelength matches the period thickness, as the Bragg condition in figure 3.10(b) and eq. (3.18) dictates. We will discuss optical interference filters further in chapter 5. In an analogy to the optical interference filters, Esaki and Tsu (1970) proposed superlattices, which are periodic structures with the thickness of each layer less than the electron or phonon mean free path. Since the electron and phonon mean free path and wavelength are very short, such an analogy requires much thinner films than those used in optical interference filters. Consequently, the concept was demonstrated only after the invention of advanced thin film deposition techniques such as molecular beam epitaxy (Chang et al., 1973). Electron wave propagation and energy states in superlattices can be modeled using the Kronig–Penney model, leading to drastically different electrical and optical properties from those of their constituent materials (Weisbuch and Vinter, 1991). Phonons can also exhibit similar behavior, forming bandgaps and new energy states (Narayanamurti et al., 1979). Figure 3.30 shows an example of a model Si/Ge superlattice and the calculated phonon spectra for acoustic phonons in a model Si/Ge-like superlattice, together with the acoustic phonon dispersion relations in bulk Si (Yang and Chen, 2001). We can

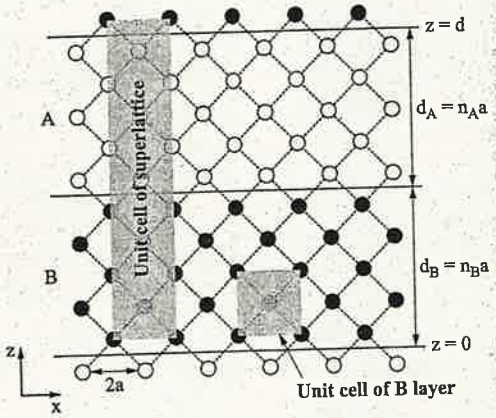


Figure 3.31 Unit cell of a superlattice made of two cubic crystals, being a tetrahedron with a much larger number of atoms as the basis.

see that, along the superlattice film plane, the period is still a and thus the maximum reciprocal lattice vector in the first Brillouin zone is π/a. In the cross-plane direction, the lattice constant of the original bulk lattice no longer represents the true periodicity in this direction. It is the total thickness of one period, d, that represents the lattice periodicity in this direction. Thus the maximum width of the first Brillouin zone is π/d rather than π/a. The superlattice clearly has very different dispersion relations to those of the bulk materials. In the cross-plane direction, for example, a small gap, called a minigap, forms at every $k_z = \pi/d$ in the phonon spectrum, similar to the electronic gap formation in a one-dimensional Kronig–Penney model. The dispersion of high-frequency acoustic phonons in bulk silicon becomes very flat inside a superlattice because there are no corresponding phonons in adjacent germanium layers; these phonons are confined inside the silicon layer. Another way to think of these confined "acoustic" phonons is that the unit cell of a Si/Ge superlattice is no longer a cube as in bulk silicon or germanium, but a tetrahedron as shown in figure 3.31. It is a new material with a new unit cell that has more than one atom as the basis (at least one silicon and one germanium atom). However, we should remind ourselves that to form the new phonon or electron spectra, as theoretically predicted using idealized models (such as Kronig–Penney or harmonic lattice dynamics), the mean free path must be much longer than one period thickness. We will discuss this point in more detail in chapter 5.

Superlattice structures made of alternating layers of thin films have artificial periodicity only in one direction. By periodically arranging quantum wires and quantum dots, one can also make quantum wire and quantum dot superlattices that have artificial periodicity in two or three directions.

The periodicity in naturally existing crystals creates electronic bandgaps and phonon branches, partially due to the fact that electrons and phonons can sample and feel individual atoms and potential barriers and thus experience diffraction and interference effects. This is not the case for optical waves (except X-rays) in naturally existing crystals, since visible light usually has a quite long wavelength and thus averages over a large volume of the crystal. The artificial optical interference filters that are made of alternating quarter-wavelength layers create frequency ranges (called stop bands) along certain directions that can completely reflect all incident photons. This phenomenon corresponds to the

photon bandgap, similar to that in the Kronig–Penney model for electrons. By extending such a concept to three dimensions to make three-dimensional periodic structures with periods comparable to optical wavelength, Yablonovitch (1986) proposed the concept of three-dimensional photonic bandgap structures. These photonic crystals have become a very active research field and have potential applications in lasers, telecommunication, and optical coatings (Joannopoulos et al., 1995, 1997).

3.6 Summary of Chapter 3

The contents of this chapter are often covered, in a solid-state physics course, in at least three individual chapters: crystal structure, electronic energy states, and phonon energy states (Kittel, 1996; Ashcroft and Mermin, 1976). This condensed chapter introduces the terminology and often-used methodologies for the analysis of energy states in crystalline structures.

The most important characteristic of crystals is their periodicity, which is described by a lattice. Real crystals are obtained by attaching a basis to each lattice point. The basis can consist of one atom or a cluster of atoms. Lattices are described by the primitive lattice vectors. A primitive unit cell contains one lattice point, but a conventional unit cell can have more than one lattice point. One way to construct a primitive unit cell unambiguously is to form the Wigner–Seitz cell. In three-dimensional space, a total of 14 lattice types exists. The Miller index method is commonly used to denote crystal planes and directions.

A lattice is periodic in real space, and we often express a periodic function in terms of its Fourier transformation. The Fourier conjugate of real space is called the reciprocal space. The primitive lattice vectors in reciprocal space can be calculated from the primitive lattice vectors in real space. Diffraction experiments provide an image of reciprocal space. A Wigner–Seitz cell in reciprocal space is called the first Brillouin zone. Later, we express the energy dispersion of electrons and phonons in the frist Brillouin zone.

In a periodic structure, the electronic energy levels form energy bands. The band formation is demonstrated by the solution of the Schrödinger equation based on the Kronig–Penney model. In real crystals, each crystallographic direction has its own dispersion relation. The electronic band structure determines whether a material is metal, semiconductor, or insulator. In metals, an electronic band is only partially filled and electrons can move to the empty quantum states within the same band. The topmost electron energy level at 0 K is called the Fermi level. If a band is totally filled and the next band has an energy gap from this band, electrons cannot move within the band. Whether the electron can go to the next energy band depends on the magnitude of the bandgap compared to the thermal energy which is 26 meV at 300 K. A material can be a semiconductor if the bandgap is relatively small such that there exist some electrons with high enough energy to jump to the conduction band, leaving some vacant quantum states behind. If the bandgap is very large, no electrons can jump to the conduction band and the material is an insulator.

In a semiconductor, the motion of electrons in the valence band can be described by the motion of equivalent positive charges, called holes, that occupy the empty states in the band. Semiconductors can be intrinsic or extrinsic. Extrinsic semiconductors are obtained by adding impurities that have an energy level close to the conduction or

valence band such that they can contribute electrons to the conduction band or holes to the valence band. Semiconductors can also be divided into direct gap or indirect gap, depending on whether the minima of the conduction band and the valence band have identical or different wavevectors, respectively. The motion of electrons or holes in a semiconductor is similar to the motion of free electrons, except that their masses should be replaced by their effective mases.

The vibration of atoms in a crystal was investigated using Newtonian mechanics, assuming harmonic force interaction between nearest neighbors. The vibration can be decomposed into normal modes, each having a specific wavevector and frequency. The dispersion relation obtained between the wavevector and the frequency is identical to the result from quantum mechanics. Quantum mechanics requires, however, that each normal mode be quantized and that the minimum quantum has an energy of $h\nu$. This minimum energy quantum is called a phonon, similar to a photon which is the quantized normal mode of an electromagnetic wave. Phonons can exist only in the first Brillouin zone since larger wavevectors are meaningless in terms of atomic displacement. In three-dimensional crystals, one longitudinal acoustic mode and two transverse acoustic modes exist along each crystallographic direction. If there are m atoms in a basis and N lattice points, $3N$ acoustic modes and $3(m-1)N$ optical modes exist.

An important counting method for the degeneracy of energy levels is the density of states. The density of states can be defined on the basis of per unit magnitude of the wavevector, per unit frequency interval, or per unit energy interval. The most often used definition is based on the per unit energy interval. We have shown how to compute the density of states for electrons, phonons, and photons, for three-dimensional structures as well as structures of lower dimensions.

Artificial nanostructures, such as quantum wells, superlattices, carbon nanotubes, and photonic crystals, can have artificial energy levels and densities of states. These artificial structures break down the analysis for bulk materials by imposing new boundary conditions and creating new periodicities that do not exist naturally in bulk materials. The new energy states and densities of states lead to novel properties that are often hard to find in bulk materials. To form the new energy spectra, the mean free path of the carriers (electrons, phonons, or photons) usually should be much larger than the characteristic length; thus the most interesting regime is often at nanoscale, when these conditions are relatively easily satisfied, which speaks strongly for the current interest in nanotechnology.

3.7 Nomenclature for Chapter 3

a	lattice constant or length in figure 3.11, m
$\mathbf{a}$	primitive or conventional lattice vector, m
A	constant; area, m^2
b	length in figure 3.11, m
$\mathbf{b}$	reciprocal space primitive lattice vector, m^{-1}
B	constant
d	thickness of quantum well or period of superlattice, m
D	density of states per unit volume, m^{-3}
E	energy, J
E_c	conduction band edge, J
f	arbitrary function, periodic in time or space

F	interatomic force, N
$\mathbf{G}$	reciprocal space lattice vector, m^{-1}
h	Planck constant, J s
$\hbar$	Planck constant divided by 2π, J s
k	magnitude of wavevector, m^{-1}
$\mathbf{k}$	wavevector, m^{-1}
K	spring constant between atoms, $N\,m^{-1}$; quantity defined by eq. (3.24)
L	L-point of the Brillouin zone of an fcc lattice, [111] direction; length of crystal, m
m	mass, kg
m_{ij}	effective mass tensor, kg
m^*	effective mass, kg
n	integer; local electron density, m^{-3}
N	total number of atoms in the crystal; number of states
P	constant in eq. (3.32)
q	charge per ion, C
Q	quantity defined by eq. (3.24)
r	separation between atoms
r_0	nearest neighbor separation, m
$\mathbf{r}$	atom position
$\mathbf{R}$	translational vector
s	integer
S	electron or photon amplitude at detector
t	time, s
T	period in time, s
u	atom displacement from its equilibrium position
U	interatomic potential, J
v	velocity, $m\,s^{-1}$
V	volume of primitive unit cell, m^3; volume of crystal
x	atom coordinate
X	X-point of fcc reciprocal lattice, [100] direction
α	Madelung constant for ionic crystals
Γ	center of the first Brillouin zone for an fcc lattice
δ	delta function, eq. (3.15)
ΔE	band edge offset, J
ε	parameter in Lennard–Jones potential, J
ε_0	electrical permittivity in vacuum, $C^2\,m^{-2}\,N^{-1}$
ζ	parameter in Born–Meyer repulsive potential, m, eq. (3.4)
θ	polar angle
κ_B	Boltzmann constant, $J\,K^{-1}$
λ	wavelength, m
ν	frequency of phonons and photons, Hz
ρ	density, $kg\,m^{-3}$
σ	parameter in Lennard–Jones potential, m
ϕ	azimuthal angle
Ψ	wavefunction
ω	angular frequency, rad.Hz
Ω	solid angle, srad
( )	Miller index of a crystal plane
{ }	Miller index for identical crystal planes
[ ]	vector along a crystallographic direction

Subscripts

A	attractive
c	conduction band
D	Debye
f	Fermi level
G	bandgap
ij	between atom i and atom j
R	repulsive
x, y, z	Cartesian coordinate direction
v	valence band

Superscript

$*$	complex conjugate

3.8 References

Ashcroft, N.W., and Mermin, N.D., 1976, *Solid State Physics*, Saunders College, Philadelphia.

Bannov, N., Aristov, V., Mitin, V., and Stroscio, M.A., 1995, "Electron Relaxation Time due to the Deformation-Potential. Interaction of Electron with Confined Acoustic Phonons in a Free-Standing Quantum Well," *Physical Review B*, vol. 51, pp. 9930–9942.

Borca-Tasciuc, T., Liu, W.L., Zeng, T., Song, D.W., Moore, C.D., Chen, G., Wang, K.L., Goorsky, M.S., Radetic, T., Gronsky, R., Koga, T., and Dresselhaus, M.S., 2000, "Thermal Conductivity of Symmetrically Strained Si/Ge Superlattices," *Superlattices and Microstructures*, vol. 28, pp. 119–206.

Brockhouse, B.N., Arase, T., Caglioti, G., Rao, K.R., and Woods, D.B., 1962, "Crystal Dynamics of Lead. I. Dispersion Curves at 100°K," *Physical Review*, vol. 128, pp. 1099–1111.

Casimir, H.B.G., 1938, "Note on the Conduction of Heat in Crystals" *Physica*, vol. 5, pp. 495–500.

Chang, L.L., Esaki, L., Howard, W.E., and Ludeke, R., 1973, "The Growth of GaAs–GaAlAs Superlattice," *Journal of Vacuum Science and Technology*, vol. 10, pp. 11–16.

Chelikowsky, J.R., and Cohen, M.L., 1976, "Nonlocal Pseudopotential Calculations for the Electronic Structure of Eleven Zinc-Blende Structures," *Physical Review B*, vol. 14, pp. 556–582.

Chen, G., 2001, "Phonon Heat Conduction in Low-Dimensional Structures," in *Semiconductors and Semimetals, Recent Trends in Thermoelectric Materials Research III*, vol. 71, pp. 203–259, Academic Press, San Diego.

Dresselhaus, M.S., Dresselhaus, G., and Phaedon, A., 2001, *Carbon Nanotubes: Synthesis, Structure, Properties, and Applications*, Springer, New York.

Esaki, L., and Tsu, R., 1970, "Superlattice and Negative Differential Conductivity in Semiconductors," *IBM Journal of Research and Development*, vol. 14, pp. 61–65.

Giannozzi, P., de Cironocoli, S., Pavone, P., and Baroni, S., 1991, "Ab Initio Calculation of Phonon Dispersions in Semiconductors," *Physical Review B*, vol. 43, pp. 7231–7242.

Heavens, O.S., 1965, *Optical Properties of Thin Solid Films*, Dover Publications, New York.

Iijima, S., 1991, "Helical Microtubules of Graphitic Carbon," *Nature*, vol. 354, pp. 56–58.

Joannopoulos, J.D., Meade, R.D., and Winn, J., 1995, *Photonic Crystals*, Princeton University Press, Princeton, NJ.

Joannopoulos, J.D., Villeneuve, P.R., and Fan, S., 1997, "Photonic Crystals: Putting a New Twist on Light," *Nature*, vol. 386, pp. 143–149.

Kim, P., Shi, L., Majumdar, A., and McEuen, P.L., 2001, "Thermal Transport Measurements of Individual Multiwalled Nanotubes," *Physical Review Letters*, vol. 87, pp. 215502/1–4.

Kittel, C., 1996, *Introduction to Solid State Physics*, 7th ed., Wiley, New York.

Mattheiss, L.F., 1964, "Energy Bands for Iron Transition Series," *Physical Review*, vol. 134, pp. A970–A973.

Narayanamurti, V., Stormer, H.L., Chin, M.A., Gossard, A.C., and Wiegmann, W., 1979, "Selective Transmission of High-Frequency Phonons by a Superlattice: The 'Dielectric' Phonon Filter," *Physical Review Letters*, vol. 43, pp. 2012–2016.

Shur, M., 1990, *Physics of Semiconductor Devices*, Prentice Hall, Englewood Cliffs, NJ.

Stillinger, F.H., and Weber, T., 1985, "Computer Simulation of Local Order in Condensed Phases of Silicon," *Physical Review B*, vol. 31, pp. 5262–5271.

Weisbuch, C., and Vinter, B., 1991, *Quantum Semiconductor Structures*, Academic Press, Boston.

Yablonovitch, E., 1986, "Inhibited Spontaneous Emission in Solid-State Physics and Electronics," *Physics Review Letters*, vol. 58, pp. 2059–2062.

Yang, B., and Chen, G., 2000, "Lattice Dynamics Study of Phonon Heat Conduction in Quantum Wells," *Physics of Low-Dimensional Structures Journal*, vol. 5/6, pp. 37–48.

Yang, B., and Chen, G., 2001, "Anisotropy of Heat Conduction in Superlattices," *Microscale Thermophysical Engineering*, vol. 5, pp. 107–116.

3.9 Exercises

3.1 *Number of atoms.* How many silicon atoms are there in a cube of 100 Å, 1000 Å, and 1 μm?

3.2 *Density of crystals.* The lattice constants of germanium and GaAs are 5.66 Å and 5.65 Å, respectively. Ge has a diamond structure and GaAs has a zinc-blende structure. Calculate their density.

3.3 *Unit cell in real and reciprocal space.* A body-centered cubic lattice has the following primitive translation vector:

$$\mathbf{a}_1 = \frac{1}{2}a(-\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}}); \quad \mathbf{a}_2 = \frac{1}{2}a(\hat{\mathbf{x}} - \hat{\mathbf{y}} + \hat{\mathbf{z}}); \quad \mathbf{a}_3 = \frac{1}{2}a(\hat{\mathbf{x}} + \hat{\mathbf{y}} - \hat{\mathbf{z}})$$

(a) Construct the Wigner–Seitz cell in real space.

(b) Find out the corresponding primitive translation vector in the reciprocal space and prove that the reciprocal lattice is an fcc structure.

(c) Sketch the Wigner–Seitz cell in the reciprocal space, that is, the first Brillouin zone.

3.4 *Lennard–Jones potential.* The values of the Lennard–Jones potential for noble gas crystals are given in table 3.1. For argon crystal,

(a) Calculate the interatomic distance.

(b) Calculate the energy at the minimum (called cohesive energy).

(c) Calculate the effective spring constant between two argon atoms.

3.5 *Miller index.* Index the following planes in a silicon crystal: (100), (110), (111), (121).

3.6 *X-ray diffraction.* In an X-ray diffraction experiment a, the angle formed between the incident ray and the detector (2θ) is 90° and first-order diffraction peaks are observed with the X-ray wavelength at 1 Å. Determine the distance of the crystal planes in this specific direction.

3.7 *Kronig–Penney model.* For $P = 5$, find out possible solutions for Ka in eq. (3.32). Convert the solution into a relationship between wavevector and energy and plot the solution in

(a) extended zone representation,

(b) reduced zone representation.

3.8 *Phonon spectra of a diatomic lattice chain.* Consider a diatomic chain of atoms as shown in figure P3.8. The masses of the two atoms are different but the spacing and the spring constant between them are the same. Derive the following given expression for the phonon dispersion in this diatomic lattice chain. Schematically draw the phonon dispersion you obtained

$$\omega^2 = K\left(\frac{M_1 + M_2}{M_1 M_2}\right) \pm \left[K^2\left(\frac{M_1 + M_2}{M_1 M_2}\right)^2 - \frac{4K^2}{M_1 M_2}\sin^2\left(\frac{1}{2}ka\right)\right]^{1/2}$$

where K is the spring constant and k the wavevector with the following values

$$k = \pm\frac{2\pi}{Na}, \pm\frac{4\pi}{Na}, \ldots, \frac{\pi}{a}$$

and N is the total number of lattice points in the chain.

3.9 *Electron density of states.* For an elliptical electronic band given by eq. (3.37), derive an expression for the electron density of states.

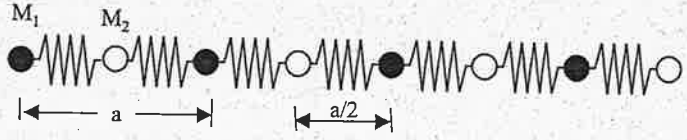


Figure P3.8 Figure for problem 3.8.

3.10 *Electron density of states inside quantum wires.* The electron energy dispersion in an infinite potential barrier quantum wire can be expressed as

$$E(k_x, \ell, n) = \frac{\hbar^2 k_x^2}{2m^*} + \frac{\hbar^2 \pi^2}{2m^*}\left[\left(\frac{\ell}{L_y}\right)^2 + \left(\frac{n}{L_z}\right)^2\right]$$

where ℓ, n can take integer values $1, 2, \ldots$. Derive an expression for the electron density of states and plot this expression for $L_y = L_z = 50$ Å.

3.11 *Electron density of states inside quantum dots.* Determine the electron density of states of a cubic quantum dot with side length $d = 20$ Å, assuming that the electron effective mass equals the free electron mass.

3.12 *Phonon density of states.* Assuming that phonons of a three-dimensional crystal obey the following isotropic dispersion relation,

$$\omega = 2\sqrt{\frac{K}{m}}\left|\sin\frac{ka}{2}\right|$$

where a is the lattice constant, derive an expression for the phonon density of states.

3.13 *Debye approximation.* Derive an expression for sound velocity from eq. (3.45). Calculate the sound velocity for a monatomic fcc crystal along (100) and (111) directions, using this simplified expression. Assume that the mass of the atom is 9.32×10^{-23} kg, the lattice constant of the conventional fcc unit cell is 5.54×10^{-10} m, and the spring constant is 7600 N m^{-1}.

3.14 *Transverse and longitudinal phonons.* Consider three separate acoustic phonons in a three-dimensional isotropic medium with an effective lattice constant of 2.5 Å. The dispersion for each branch is $\omega_L = v_L k$, $\omega_t = v_t k$ (degenerate). For $v_L = 8000$ m s^{-1} and $v_t = 5000$ m s^{-1}, plot the density of states as a function of frequency.

3.15 *Size effects on density of states.* The density of states expressions we derived are valid when the separations between states are small and the number of states is large, such that we can calculate the number of states by eq. (3.50). In small geometries, the energy separations between different states can be large and the number of states at each energy level can be small, so that eq. (3.50) is no longer valid. As an example, consider a cubic cavity of $(2\ \mu\text{m})^3$ size. Find out how many states are allowed to exist inside the cavity for electromagnetic waves with a wavelength in the range of 0.5–1 μm, using the following two methods:

(a) by finding out how many sets of (k_x, k_y, k_z) are allowed in this cavity that fall into the given wavelength range;

(b) by integrating eq. (3.59) over the given wavelength range.

4

Statistical Thermodynamics and Thermal Energy Storage

The quantum mechanics principles covered in the previous two chapters give the allowable energy states of matter. The number of allowable states in typical macroscopic matter is usually very large, and at any instant, the matter can be at any one of these states. Although our mathematical treatments in the previous two chapters were based on solving the steady-state Schrödinger equation for the energy states and the wavefunctions, matter will not stay at one particular quantum state (a microscopic state) for long because of the interactions among particles (atoms, molecules, electrons, and phonons) in the matter. For example, we assumed a harmonic potential between atoms to obtain the phonon dispersion relation. In reality, the interatomic potential is not harmonic, as one can easily infer from examining the Lennard–Jones potential. When the anharmonicity (the deviation from the harmonic potential) is small, the solutions of the Schrödinger equation for the quantum states based on the harmonic potential are approximately correct. Yet a small degree of anharmonicity can cause a rapid ($\sim 10^{-9}$–10^{-13} s) change of the matter from one quantum state to another. Due to the large number of quantum states available in matter, it is impractical to follow the real time evolution of matter among its allowable quantum states. The bridge connecting the allowable quantum states to the macroscopic behavior is provided by statistical thermodynamics, which determines the probability that matter will be at a particular quantum state when it is at equilibrium. Through statistical thermodynamics, temperature enters into the picture of energy storage and transport.

In this chapter, we focus on the equilibrium state of a system and discuss different probability functions for systems under different constraints, such as an isolated system or a system at constant temperature. From the probability distribution functions, we will show how to calculate the internal energy and specific heat of a system, including

nanostructures, using what we learned in previous chapters about the energy levels, degeneracy, and the density of states.

4.1 Ensembles and Statistical Distribution Functions

In an actual experiment, we often follow the time history of a system and observe its time-averaged behavior. In analysis, however, following the time history requires the solution of master equations that govern the motion of a large number of particles, such as the Newton equations of motion and the time-dependent Schrödinger equation. Although, with increasing computational power, such computation is becoming feasible for limited situations, as in the molecular dynamics simulations to be introduced in chapter 10, for most applications direct computation of the time history is impractical. Statistical thermodynamics avoids the time averaging by introducing *ensembles*, which are large collections of systems, each representing a microscopic state that satisfies the macroscopic constraints. Examples of the macroscopic constraints of a system are its total energy, temperature, and volume. The quantities to be measured are averaged over the ensemble at a fixed time, rather than, as in an experimental situation, over a time period of a single system. A fundamental assumption made in statistical mechanics is that the ensemble average of an observed quantity is equal to the time average of the same quantity. This assumption is called the *ergodic hypothesis*. The study of conditions necessary for this hypothesis to be valid is an ongoing research area (Kubo et al., 1998). Our analysis will assume that all systems are ergodic. Depending on the macroscopic constraints, various ensembles are developed, each has a probability distribution for the microscopic states in the ensemble that differs from other ensembles. In the following, we will discuss three ensembles: microcanonical, canonical, and grand canonical ensembles.

4.1.1 Microcanonical Ensemble and Entropy

Unlike classical thermodynamics, which completely neglects the microscopic processes in a system, statistical thermodynamics builds the system properties from its microscopic states (Kittel and Kroemer, 1980; Callen, 1985). We consider an *isolated* macroscopic system with a volume V, a total number of particles N, and a total energy of U (macroscopic constraints). The quantum states of the system that satisfy these macroscopic constraints are called the accessible quantum states (or simply accessible states). Given these macroscopic constraints, together with detailed information about the interatomic potentials between the particles in the system and the initial conditions, one could in principle solve the Schrödinger equation to follow the temporal evolution of the system among the accessible quantum states. The macroscopic properties of the system, such as temperature and pressure, are a measure of the average corresponding microscopic properties over a certain amount of time.

How can we calculate the average values of this system? If we performed an experiment, we would measure these values as a function of time and carry out a time average. In statistical mechanics, instead of tracing the time evolution of the system, we focus on the probability of a system being at a specific accessible quantum state. A *fundamental*

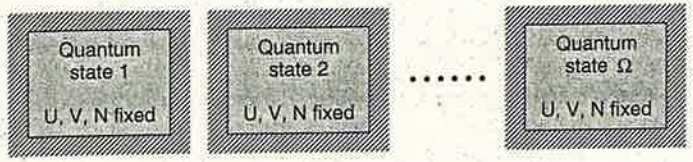


Figure 4.1 A microcanonical ensemble is made of isolated systems with fixed U, V, and N. Each system corresponds to one accessible quantum state of the original system.

postulate in statistical mechanics is that an isolated macroscopic system samples every accessible quantum state with equal probability. This postulate is also called the *principle of equal probability*. If Ω is the total number of accessible quantum states, the probability of each accessible quantum state, denoted by s, being sampled is

$$P(s) = 1/\Omega \tag{4.1}$$

Once the probability of each accessible quantum state is known, we can construct a way to calculate a desired macroscopic quantity $\langle X \rangle$ of a macroscopic system. We first evaluate the corresponding property X (such as temperature, pressure) for each accessible quantum state, and then calculate the average according to

$$\langle X \rangle = \sum_{s=1}^{\Omega} P(s)X(s) \tag{4.2}$$

Note that the summation is over all accessible quantum states.

Because each accessible quantum state is a state of the system that satisfies the macroscopic constraints and each one has equal probability to be sampled, we are effectively dealing with a collection of Ω stationary systems, as shown in figure 4.1. These systems are identical from the macroscopic points of view; that is, they have the same U, V, and N and are all isolated from their surroundings. This collection of Ω systems is called an *ensemble*. A fixed U, V, N ensemble is called a *microcanonical* ensemble. The principle of equal probability is valid only for each system in a microcanonical ensemble. Later, we will introduce canonical and grand canonical ensembles and derive the probability of each system in such ensembles on the basis of results we obtained from the microcanonical ensemble. Equation (4.2) means that each of the stationary systems in the ensemble is sampled once in the computing of the average. Such an average is called the ensemble average. For a microcanonical ensemble, $P(s) = 1/\Omega$, thus

$$\langle X \rangle = \sum_{s=1}^{\Omega} X(s)/\Omega \tag{4.3}$$

The idea of equal probability for each accessible quantum state in an isolated system may seem unreasonable for some readers. For example, for a system of 10^{23} particles, one accessible quantum state might be that one particle has energy U and the rest have zero energy. This distribution of energy among N particles seems to be a quite unlikely