2. Electronic Band Structures

CONTENTS

The property which distinguishes semiconductors from other materials concerns the behavior of their electrons, in particular the existence of gaps in their electronic excitation spectra. The microscopic behavior of electrons in a solid is most conveniently specified in terms of the electronic band structure. The purpose of this chapter is to study the band structure of the most common semiconductors, namely, Si, Ge, and related III–V compounds. We will begin with a quick introduction to the quantum mechanics of electrons in a crystalline solid.

The properties of electrons in a solid containing 10^{23} atoms/cm^3 are very complicated. To simplify the formidable task of solving the wave equations for the electrons, it is necessary to utilize the translational and rotational symmetries of the solid. **Group theory** is the tool that facilitates this task. However, not everyone working with semiconductors has a training in group theory, so in this chapter we will discuss some basic concepts and notations of group theory. Our approach is to introduce the ideas and results of group theory when applied to semiconductors without presenting the rigorous proofs. We will put particular emphasis on notations that are often found in books and research articles on semiconductors. In a sense, band structure diagrams are like maps and the group theory notations are like symbols on the map. Once the meaning of these symbols is understood, the band structure diagrams can be used to find the way in exploring the electronic properties of semiconductors.

We will also examine several popular methods of band structure computation for semiconductors. All band structure computation techniques involve approximations which tend to emphasize some aspects of the electronic prop-

P.Y. Yu, M. Cardona, *Fundamentals of Semiconductors*, Graduate Texts in Physics, 4th ed.,
DOI 10.1007/978-3-642-00710-1_2, © Springer-Verlag Berlin Heidelberg 2010

erties in semiconductors while, at the same time, de-emphasizing other aspects. Therefore, our purpose in studying the different computational methods is to understand their advantages and limitations. In so doing we will gain insight into the many different facets of electronic properties in semiconductors.

We note also that within the past two decades, highly sophisticated techniques labeled "*ab initio*" have been developed successfully to calculate many properties of solids, including semiconductors. These techniques involve very few assumptions and often no adjustable parameters. They have been applied to calculate the total energy of crystals including all the interactions between the electrons and with the nuclei. By minimization of this energy as a function of atomic spacing, equilibrium lattice constants have been predicted. Other properties such as the elastic constants and vibrational frequencies can also be calculated. Extensions of these techniques to calculate excited-state properties have led to predictions of optical and photoemission spectra in good agreement with experimental results. It is beyond the scope of the present book to go into these powerful techniques. Interested readers can consult articles in [2.1].

2.1 Quantum Mechanics

The Hamiltonian describing a perfect crystal can be written as

$$\mathcal{H} = \sum_i \frac{p_i^2}{2m_i} + \sum_j \frac{P_j^2}{2M_j} + \frac{1}{2}{\sum_{j',j}}' \frac{Z_j Z_{j'} e^2}{4\pi\varepsilon_0 |\boldsymbol{R}_j - \boldsymbol{R}_{j'}|} - \sum_{j,i} \frac{Z_j e^2}{4\pi\varepsilon_0 |\boldsymbol{r}_i - \boldsymbol{R}_j|} + \frac{1}{2}{\sum_{i,i'}}' \frac{e^2}{4\pi\varepsilon_0 |\boldsymbol{r}_i - \boldsymbol{r}_{i'}|} \tag{2.1}$$

in the cgs system of units. (As mentioned in the preface to this edition, we have printed in red symbols which must be added to the cgs expression to convert them into Si units. ε_0 represents the permittivity of vacuum). In this expression $\boldsymbol{r}_i$ denotes the position of the ith electron, $\boldsymbol{R}_j$ is the position of the jth nucleus, Z_j is the atomic number of the nucleus, $\boldsymbol{p}_i$ and $\boldsymbol{P}_j$ are the momentum operators of the electrons and nuclei, respectively, and $-e$ is the electronic charge. $\sum'$ means that the summation is only over pairs of indices which are not identical.

Obviously, the many-particle Hamiltonian in (2.1) cannot be solved without a large number of simplifications. The first approximation is to separate electrons into two groups: **valence electrons** and **core electrons**. The core electrons are those in the filled orbitals, e. g. the $1s^2$, $2s^2$, and $2p^6$ electrons in the case of Si. These core electrons are mostly localized around the nuclei, so they can be "lumped" together with the nuclei to form the so-called **ion cores**. As a result of this approximation the indices j and j' in (2.1) will, from now on, denote the ion cores while the electron indices i and i' will label only the valence

electrons. These are electrons in incompletely filled shells and in the case of Si include the 3*s* and 3*p* electrons.

The next approximation invoked is the **Born–Oppenheimer** or **adiabatic approximation**. The ions are much heavier than the electrons, so they move much more slowly. The frequencies of ionic vibrations in solids are typically less than 10^{13} s^{-1}. To estimate the electron response time, we note that the energy required to excite electrons in a semiconductor is given by its fundamental bandgap, which, in most semiconductors, is of the order of 1 eV. Therefore, the frequencies of electronic motion in semiconductors are of the order of 10^{15} s^{-1} (a table containing the conversion factor from eV to various other units can be found in the inside cover of this book). As a result, electrons can respond to ionic motion almost instantaneously or, in other words, to the electrons the ions are essentially stationary. On the other hand, ions cannot follow the motion of the electrons and they see only a time-averaged adiabatic electronic potential. With the Born-Oppenheimer approximation the Hamiltonian in (2.1) can be expressed as the sum of three terms:

$$\mathcal{H} = \mathcal{H}_{\text{ions}}(\boldsymbol{R}_j) + \mathcal{H}_{\text{e}}(\boldsymbol{r}_i, \boldsymbol{R}_{j0}) + \mathcal{H}_{\text{e-ion}}(\boldsymbol{r}_i, \delta\boldsymbol{R}_j), \tag{2.2}$$

where $\mathcal{H}_{\text{ion}}(\boldsymbol{R}_j)$ is the Hamiltonian describing the ionic motion under the influence of the ionic potentials plus the time-averaged adiabatic electronic potentials. $\mathcal{H}_{\text{e}}(\boldsymbol{r}_i, \boldsymbol{R}_{j0})$ is the Hamiltonian for the electrons with the ions frozen in their equilibrium positions $\boldsymbol{R}_{j0}$, and $\mathcal{H}_{\text{e-ion}}(\boldsymbol{r}_i, \delta\boldsymbol{R}_j)$ describes the change in the electronic energy as a result of the displacements $\delta\boldsymbol{R}_j$ of the ions from their equilibrium positions. $\mathcal{H}_{\text{e-ion}}$ is known as the **electron–phonon interaction** and is responsible for electrical resistance in reasonably pure semiconductors at room temperature. The vibrational properties of the ion cores and electron-phonon interactions will be discussed in the next chapter. In this chapter we will be mainly interested in the electronic Hamiltonian $\mathcal{H}_{\text{e}}$.

The electronic Hamiltonian $\mathcal{H}_{\text{e}}$ is given by

$$\mathcal{H}_{\text{e}} = \sum_i \frac{p_i^2}{2m_i} + \frac{1}{2}{\sum_{i,i'}}' \frac{e^2}{4\pi\varepsilon_0|\boldsymbol{r}_i - \boldsymbol{r}_{i'}|} - \sum_{i,j} \frac{Z_j e^2}{4\pi\varepsilon_0|\boldsymbol{r}_i - \boldsymbol{R}_{j0}|}. \tag{2.3}$$

Diagonalizing this Hamiltonian when there are $>10^{23}$ electrons/cm^3 in a semiconductor is a formidable job. We will make a very drastic approximation known as the **mean-field approximation**. Without going into the justifications, which are discussed in many standard textbooks on solid-state physics, we will assume that every electron experiences the same average potential $V(\boldsymbol{r})$. Thus the Schrödinger equations describing the motion of each electron will be identical and given by

$$\mathcal{H}_{1\text{e}}\Phi_n(\boldsymbol{r}) = \left(\frac{p^2}{2m} + V(\boldsymbol{r})\right)\Phi_n(\boldsymbol{r}) = E_n\Phi_n(\boldsymbol{r}), \tag{2.4}$$

where $\mathcal{H}_{1\text{e}}$, $\Phi_n(\boldsymbol{r})$ and E_n denote, respectively, the one-electron Hamiltonian, and the wavefunction and energy of an electron in an eigenstate labeled by n.

We should remember that each eigenstate can only accommodate up to two electrons of opposite spin (**Pauli's exclusion principle**).

The calculation of the electronic energies E_n involves two steps. The first step is the determination of the one-electron potential $V(\boldsymbol{r})$. Later in this chapter we will discuss the various ways to calculate or determine $V(\boldsymbol{r})$. In one method $V(\boldsymbol{r})$ can be calculated from *first principles* with the atomic numbers and positions as the only input parameters. In simpler, so-called semi-empirical approaches, the potential is expressed in terms of parameters which are determined by fitting experimental results. After the potential is known, it takes still a complicated calculation to solve (2.4). It is often convenient to utilize the symmetry of the crystal to simplify this calculation. Here by "symmetry" we mean geometrical transformations which leave the crystal unchanged.

2.2 Translational Symmetry and Brillouin Zones

The most important symmetry of a crystal is its invariance under specific translations. In addition to such **translational symmetry** most crystals possess some **rotational** and **reflection** symmetries. It turns out that most semiconductors have high degrees of rotational symmetry which are very useful in reducing the complexity of calculating their energy band structures. In this and the next sections we will study the use of symmetry to simplify the classification of electronic states. Readers familiar with the application of group theory to solids can omit these two sections.

When a particle moves in a periodic potential its wavefunctions can be expressed in a form known as **Bloch functions**. To understand what Bloch functions are, we will assume that (2.4) is one-dimensional and $V(x)$ is a periodic function with the translational period equal to R. We will define a **translation operator** T_R as an operator whose effect on any function $f(x)$ is given by

$$T_R f(x) = f(x + R). \tag{2.5}$$

Next we introduce a function $\Phi_k(x)$ defined by

$$\Phi_k(x) = \exp(\mathrm{i}kx)u_k(x), \tag{2.6}$$

where $u_{\mathrm{k}}(x)$ is a periodic function with the same periodicity as V, that is, $u_k(x + nR) = u_k(x)$ for all integers n. When $\Phi_k(x)$ so defined is multiplied by $\exp[-\mathrm{i}\omega t]$, it represents a plane wave whose amplitude is modulated by the periodic function $u_k(x)$. $\Phi_k(x)$ is known as a Bloch function. By definition, when x changes to $x + R$, $\Phi_k(x)$ must change in the following way

$$T_R\Phi_k(x) = \Phi_k(x + R) = \exp(\mathrm{i}kR)\Phi_k(x). \tag{2.7}$$

It follows from (2.7) that $\Phi_k(x)$ is an eigenfunction of T_R with the eigenvalue $\exp(\mathrm{i}kR)$. Since the Hamiltonian $\mathcal{H}_{1\mathrm{e}}$ is invariant under translation by R, $\mathcal{H}_{1\mathrm{e}}$ commutes with T_R. Thus it follows from quantum mechanics that the

eigenfunctions of $\mathcal{H}_{1e}$ can be expressed also as eigenfunctions of T_R. We therefore conclude that an eigenfunction $\Phi(x)$ of $\mathcal{H}_{1e}$ can be expressed as a sum of Bloch functions:

$$\Phi(x) = \sum_k A_k \Phi_k(x) = \sum_k A_k \exp(\mathrm{i}kx) u_k(x), \tag{2.8}$$

where the A_k are constants. Thus the one-electron wavefunctions can be indexed by constants k, which are the **wave vectors** of the plane waves forming the "backbone" of the Bloch function. A plot of the electron energies in (2.4) versus k is known as the **electronic band structure** of the crystal.

The band structure plot in which k is allowed to vary over all possible values is known as the **extended zone scheme**. From (2.6) we see that the choice of k in indexing a wave function is not unique. Both k and $k + (2n\pi/R)$, where n is any integer, will satisfy (2.6). This is a consequence of the translation symmetry of the crystal. Thus another way of choosing k is to replace k by $k' = k - (2n\pi/R)$, where n is an integer chosen to limit k' to the interval $[-\pi/R, \pi/R]$. The region of k-space defined by $[-\pi/R, \pi/R]$ is known as the **first Brillouin zone**. A more general definition of Brillouin zones in three dimensions will be given later and can also be found in standard textbooks [2.2]. The band structure plot resulting from restricting the wave vector k to the first Brillouin zone is known as the **reduced zone scheme**. In this scheme the wave functions are indexed by an integer n (known as the **band index**) and a wave vector k restricted to the first Brillouin zone.

In Fig. 2.1 the band structure of a "nearly free" electron (i. e., $V \to 0$) moving in a one-dimensional lattice with lattice constant a is shown in both schemes for comparison. Band structures are plotted more compactly in the reduced zone scheme. In addition, when electrons make a transition from one state to another under the influence of a translationally invariant operator, k is conserved in the process within the reduced zone scheme (the proof of this

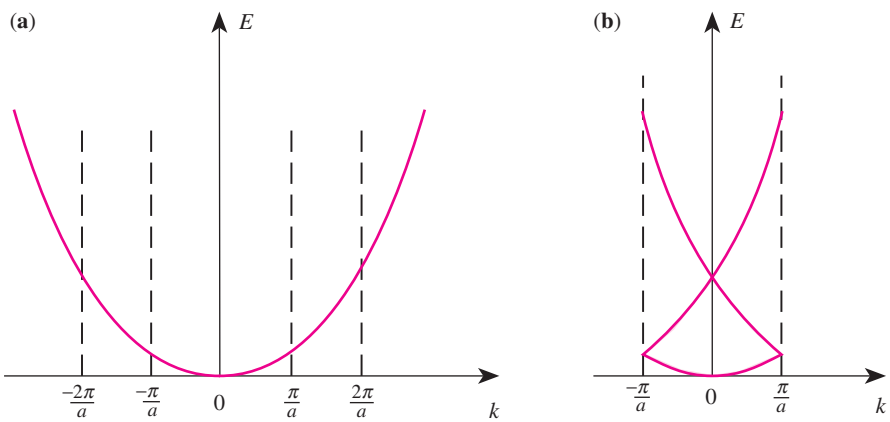


Fig. 2.1. The band structure of a free particle shown in (**a**) the extended zone scheme and (**b**) the reduced zone scheme

statement will be presented when matrix elements of operators in crystals are discussed, Sect. 2.3), whereas in the extended zone scheme k is conserved only to a multiple of (i. e. *modulo*) $2\pi/R$. Hence, the reduced zone scheme is almost invariably used in the literature.

The above results, obtained in one dimension, can be easily generalized to three dimensions. The translational symmetries of the crystal are now expressed in terms of a set of **primitive lattice vectors**: $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$. We can imagine that a crystal is formed by taking a minimal set of atoms (known as a **basis set**) and then translating this set by multiples of the primitive lattice vectors and their linear combinations. In this book we will be mostly concerned with the diamond and zinc-blende crystal structures, which are shown in Fig. 2.2a. In both crystal structures the basis set consists of two atoms. The ba-

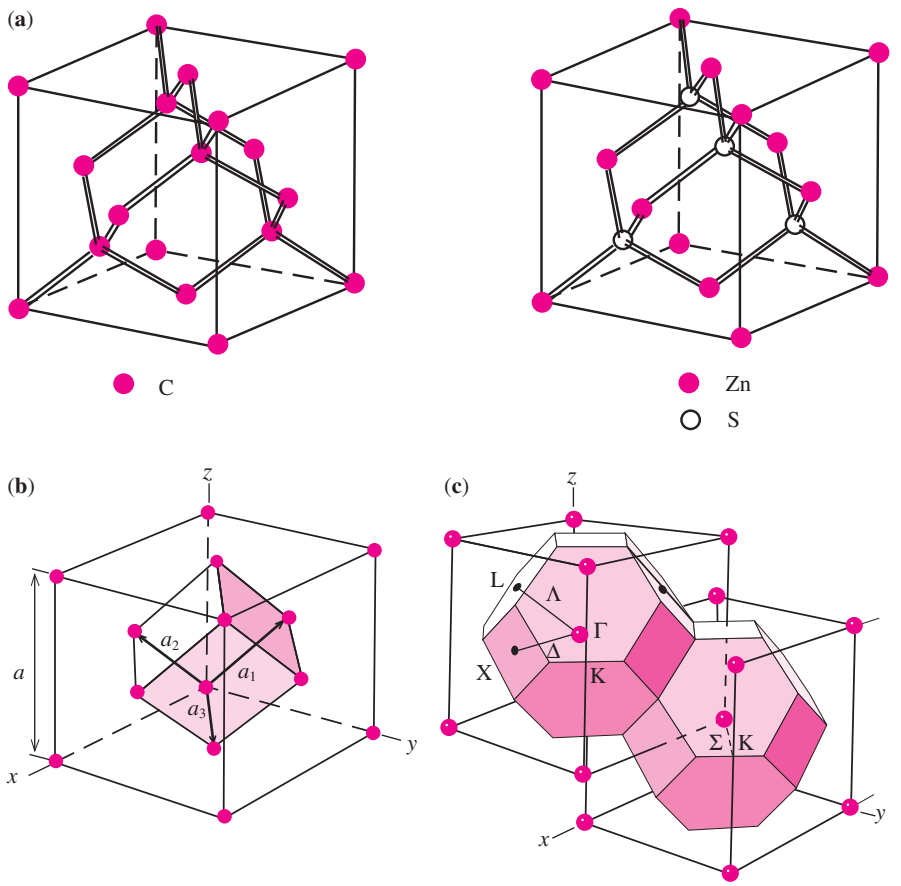


Fig. 2.2. (**a**) The crystal structure of diamond and zinc-blende (ZnS). (**b**) the fcc lattice showing a set of primitive lattice vectors. (**c**) The reciprocal lattice of the fcc lattice shown with the first Brillouin zone. Special high-symmetry points are denoted by Γ, X, and L, while high-symmetry lines joining some of these points are labeled as Λ and Δ

sis set in diamond consists of two carbon atoms while in zinc-blende the two atoms are zinc and sulfur. The lattice of points formed by translating a point by multiples of the primitive lattice vectors and their linear combinations is known as the **direct lattice**. Such lattices for the diamond and zinc-blende structures, which are basically the same, are said to be **face-centered cubic** (fcc) see Fig. 2.2b with a set of primitive lattice vectors. In general, the choice of primitive lattice vectors for a given direct lattice is not unique. The primitive lattice vectors shown in Fig. 2.2b are

$$\boldsymbol{a}_1 = (0, a/2, a/2),$$
$$\boldsymbol{a}_2 = (a/2, 0, a/2),$$

and

$$\boldsymbol{a}_3 = (a/2, a/2, 0),$$

where a is the length of the side of the smallest cube in the fcc lattice. This smallest cube in the direct lattice is also known as the unit cube or the **crystallographic unit cell**.

For a given direct lattice we can define a **reciprocal lattice** in terms of three **primitive reciprocal lattice vectors**: $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$, which are related to the direct lattice vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$ by

$$\boldsymbol{b}_i = 2\pi \frac{(\boldsymbol{a}_j \times \boldsymbol{a}_k)}{(\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{a}_3}, \tag{2.9}$$

where i, j, and k represent a cyclic permutation of the three indices 1, 2, and 3 and $(\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{a}_3$ is the volume of the primitive cell. The set of points generated by translating a point by multiples of the reciprocal lattice vectors is known as the reciprocal lattice. The reason for defining a reciprocal lattice in this way is to represent the wave vector $\boldsymbol{k}$ as a point in **reciprocal lattice space**. The first Brillouin zone in three dimensions can be defined as the smallest polyhedron confined by planes perpendicularly bisecting the reciprocal lattice vectors. It is easy to see that the region $[-\pi/R, \pi/R]$ fits the definition of the first Brillouin zone in one dimension.

Since the reciprocal lattice vectors are obtained from the direct lattice vectors via (2.9), the symmetry of the Brillouin zone is determined by the symmetry of the crystal lattice. The reciprocal lattice corresponding to a fcc lattice is shown in Fig. 2.2c. These reciprocal lattice points are said to form a **body-centered cubic** (bcc) lattice. The primitive reciprocal lattice vectors $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$ as calculated from (2.9) are

$$\boldsymbol{b}_1 = (2\pi/a)\ (-1, 1, 1),$$

$$\boldsymbol{b}_2 = (2\pi/a)\ (1, -1, 1),$$

and

$$\boldsymbol{b}_3 = (2\pi/a)\ (1, 1, -1).$$

[Incidentally, note that all the reciprocal lattice vectors of the fcc lattice have the form $(2\pi/a)(i,j,k)$, where i, j, and k have to be either all odd or all even]. The first Brillouin zone of the fcc structure is also indicated in Fig. 2.2c. The symmetry of this Brillouin zone can be best visualized by constructing a model out of cardboard. A template for this purpose can be found in Fig. 2.27.

In Fig. 2.2c we have labeled some of the high-symmetry points of this Brillouin zone using letters such as X and Γ. We will conform to the convention of denoting high symmetry points and lines *inside* the Brillouin zone by **Greek** letters and points on the *surfaces* of the Brillouin zone by **Roman** letters. The **center** of the Brillouin zone is always denoted by Γ. The three high-symmetry directions [100], [110], and [111] in the Brillouin zone of the fcc lattice are denoted by:

[100] direction : $\dot{\Gamma}\overline{\qquad\Delta\qquad}\dot{X}$

[111] direction : $\dot{\Gamma}\overline{\qquad\Lambda\qquad}\dot{L}$

[110] direction : $\dot{\Gamma}\overline{\qquad\Sigma\qquad}\dot{K}$

The Brillouin zone of the fcc lattice is highly symmetrical. A careful examination of this Brillouin zone shows that it is unchanged by various rotations, such as a 90° rotation about axes parallel to the edges of the body-centered cube in Fig. 2.2c. In addition it is invariant under reflection through certain planes containing the center of the cube. These operations are known as **symmetry operations** of the Brillouin zone. The symmetry of the Brillouin zone results from the symmetry of the direct lattice and hence it is related to the symmetry of the crystal. This symmetry has at least two important consequences for the electron band structure. First, if two wave vectors $\boldsymbol{k}$ and $\boldsymbol{k}'$ in the Brillouin zone can be transformed into each other under a symmetry operation of the Brillouin zone, then the electronic energies at these wave vectors must be identical. Points and axes in reciprocal lattice space which transform into each other under symmetry operations are said to be **equivalent**. For example, in the Brillouin zone shown in Fig. 2.2c there are eight hexagonal faces containing the point labeled L in the center. These eight faces including the L points are equivalent and can be transformed into one another through rotations by 90°. Therefore it is necessary to calculate the energies of the electron at only one of the eight equivalent hexagonal faces containing the L point. The second and perhaps more important consequence of the crystal symmetry is that wave functions can be expressed in a form such that they have definite transformation properties under symmetry operations of the crystal. Such wave functions are said to be symmetrized. A well-known example of **symmetrized wave functions** is provided by the standard wave functions of electrons in atoms, which are usually symmetrized according to their transformation properties under rotations and are classified as s, p, d, f, etc. For example, an s wave function is unchanged by any rotation. The p wave functions are triply degenerate and transform under rotation like the three components of a vector. The d wave functions transform like the five components of a symmetric and traceless second-rank tensor. By classifying the wave functions in this way, some

matrix elements of operators can be shown to vanish, i. e., **selection rules** can be deduced. Similarly, wave functions in crystals can be classified according to their transformation properties under symmetry operations of the crystal and selection rules can be deduced for operators acting on these wave functions. The mathematical tool for doing this is **group theory**. Many excellent textbooks have been written on group theory (see the reference list). It is desirable, but not necessary, to have a good knowledge of group theory in order to study semiconductor physics. Some elementary notions of group theory are sufficient to understand the material covered in this book. The next section contains an introduction to group theoretical concepts and notations. Students familiar with group theory can omit this section.

2.3 A Pedestrian's Guide to Group Theory

Since the purpose of this section is to introduce group theory terminology and notations, no effort will be made to prove many of the statements and theorems mentioned in it. At most we shall illustrate our statements with examples and refer the reader to books on group theory for rigorous proofs.

2.3.1 Definitions and Notations

The first step in studying the symmetry properties of any crystal is to determine its symmetry operations. For example: a square is unchanged under reflection about its two diagonals, or under rotation by 90° about an axis perpendicular to the square and passing through its center. One can generate other symmetry operations for a square which are combinations of these operations. One may say that it is possible to find an infinite number of symmetry operations for this square. However, many of these symmetry operations can be shown to consist of sequences of a few basic symmetry operations. The mathematical tool for systematically analyzing the symmetry operations of any object is group theory.

A **group** $\boldsymbol{G}$ is defined as a set of elements $\{a, b, c, \ldots\}$ for which an operation ab (which we will refer to as **multiplication**) between any two elements a and b of the group is defined. This operation must have these four properties:

- **Closure**: The result of the operation ab on any two elements a and b in $\boldsymbol{G}$ must also belong to $\boldsymbol{G}$.
- **Associativity**: for all elements a, b, and c in $\boldsymbol{G}$ $(ab)c = a(bc)$.
- **Identity**: $\boldsymbol{G}$ must contain an element e known as the **identity** or **unit element** such that $ae = a$ for all elements a in $\boldsymbol{G}$.
- **Inverse element**: for every element a in $\boldsymbol{G}$ there exists a corresponding element a^{-1} such that $a^{-1}a = e$. Element a^{-1} is known as the **inverse** of a.

Notice that the order in which one multiplies two elements a and b is important since ab is not necessarily equal to ba in general. If $ab = ba$ for all elements in $\boldsymbol{G}$, multiplication is **commutative** and $\boldsymbol{G}$ is said to be **Abelian**.

One can easily find many examples of groups. In particular, the set of symmetry operations of a crystal or a molecule can be shown to form a group. As an illustration, we will consider the molecule methane: CH_4. The structure of this molecule is shown in Fig. 2.3. It consists of a carbon atom surrounded by four hydrogen atoms forming the four corners of a regular tetrahedron.

To simplify the description of the symmetry operations of the methane molecule, we will introduce the **Schönflies notation**:

C_2: rotation by 180° (called a **two-fold rotation**);
C_3: rotation by 120° (called a **three-fold rotation**);
C_4: rotation by 90° (called a **four-fold rotation**);
C_6: rotation by 60° (called a **six-fold rotation**);
σ: **reflection** about a plane;
i: **inversion**;
S_n: rotation C_n followed by a reflection through a plane perpendicular to the rotation axis;
E: the **identity operation**.

For brevity, all the above operations are often denoted as rotations. To distinguish between a conventional rotation (such as C_3) from reflections (such as σ) or rotations followed by reflections (such as S_4) the latter two are referred to as **improper rotations**. Notice that the inversion is equal to S_2. This is not the only way to represent symmetry operations. An equally popular system is the **international notation**. The conversion between these two systems can be found in books on group theory [Ref. 2.3, p. 85].

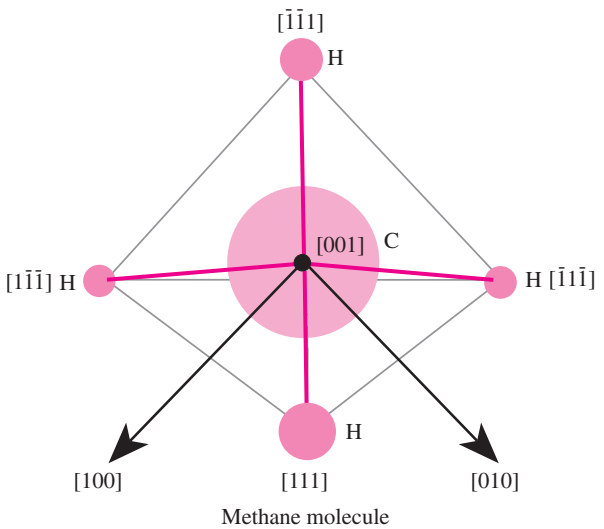


Fig. 2.3. A methane molecule (CH_4) displaying the bonds (*red lines*) and the coordinate axes (*black arrows*). The [001] axis is perpendicular to the paper

To specify a symmetry operation completely, it is also necessary to define the axis of rotation or the plane of reflection. In specifying planes of reflection we will use the notation (kln) to represent a plane that contains the origin and is perpendicular to the vector (k, l, n). (Readers familiar with crystallography will recognize that this notation is an "imitation" of the **Miller indices** for denoting lattice planes in cubic crystals). The corresponding simplified notation for the axis containing this vector is $[kln]$. In Fig. 2.3 we have first chosen the origin at the carbon atom for convenience. Using the coordinate system shown in Fig. 2.3, the four carbon–hydrogen bonds are oriented along the [111], [$1\bar{1}\bar{1}$], [$\bar{1}1\bar{1}$], and [$\bar{1}\bar{1}1$] directions. We will now state without proof (the reader can check these results easily by constructing a balls-and-sticks model of the methane molecule) that the following operations are symmetry operations of the methane molecule:

E: the identity;
C_2: two-fold rotation about one of the three mutually perpendicular [100], [010] and [001] axes (three C_2 operations in total);
C_3: rotation by 120° in clockwise direction about one of the four C–H bonds (four operations in total);
C_3^{-1}: rotation by 120°, counterclockwise, about one of the four C–H bonds (four operations in total);
σ: reflection with respect to one of these six planes: (110), ($\bar{1}$10), (101), ($\bar{1}$01), (011), (0$\bar{1}$1);
S_4: a four-fold clockwise rotation about one of the [100], [010], and [001] axes followed by a reflection on the plane perpendicular to the rotation axis (three operations in total);
S_4^{-1}: a four-fold counterclockwise rotation about one of the [100], [010], and [001] axes followed by a reflection on a plane perpendicular to the rotation axis (three operations in total).

It can be shown easily that the operations C_2 and σ are both the inverse of themselves. The inverse element of C_3 is C_3^{-1}, provided the axis of rotation is the same in the two operations. Similarly, the inverse element of S_4 is S_4^{-1}, provided the rotation axes remain the same. If we now define the multiplication of two symmetry elements a and b as a symmetry operation $c = ab$ consisting of first applying the operation b to the CH_4 molecule followed by the operation a, it can be shown easily that the 24 symmetry operations of CH_4 defined above form a group known as T_d. Such groups of symmetry operations of a molecule are known as **point groups**. As the name implies, point groups consist of symmetry operations in which at least one point remains fixed and unchanged in space. Point groups contain two kinds of symmetry operations: proper and improper rotations.

An infinite crystal is different from a molecule in that it has translational symmetry. Although in real life crystals never extend to infinity, the problems associated with the finite nature of a crystal can be circumvented by applying the so-called **periodic** (**or Born–von Kármán**) **boundary conditions** to the crystal. Equivalently, one can imagine that the entire space is filled with repli-

cas of the finite crystal. It should be no surprise that the set of all symmetry operations of such an infinite crystal also forms a group. Such groups, which contain both translational and rotational symmetry operations, are known as **space groups**. There are 230 non-equivalent space groups in three dimensions.

Besides their translational invariance, crystals also possess rotational symmetries. Space groups can be divided into two types, depending on whether or not the rotational parts of their symmetry operations are also symmetry operations. Let us first consider the purely translational operations of an infinite crystal. It can be shown that these translational symmetry operations form a group (to be denoted by $\boldsymbol{T}$). $\boldsymbol{T}$ is known as **subgroup** of the space group $\boldsymbol{G}$ of the crystal. Let us now denote by R the set of all symmetry operations of $\boldsymbol{G}$ which involve either pure rotations (both proper and improper) only or rotations accompanied by a translation not belonging to $\boldsymbol{T}$. We will denote the elements of R as α, β, τ, etc. Such a subset of $\boldsymbol{G}$ is known as a **complex**. In general R is not a group. For example, if $\boldsymbol{G}$ contains a **screw axis** or **glide plane** (these will be defined later, see Fig. 2.4) then R will not form a group and the space group $\boldsymbol{G}$ is said to be **nonsymmorphic**. If no screw axis or glide planes are present, R is a group (and therefore a subgroup of $\boldsymbol{G}$): the space group $\boldsymbol{G}$ is then said to be **symmorphic**. The symmetry properties of symmorphic groups are simpler to analyze since both translational and rotational operations in such space groups form subgroups. In particular, it can be shown that the rotational symmetry operations of a symmorphic space group form point groups similar to those for molecules. However, there are restrictions on the rotational symmetry of a crystal as a result of its translational symmetry. For example, a crystal cannot be invariant under rotation by 72° (known as a five-fold rotation). However, a molecule can have this rotational symmetry. Point groups which are compatible with a lattice with translational symmetry are called **crystallographic point groups**. It can be shown that there are 32 distinct crystallographic point groups in three-dimensional space (see, e. g. [2.4]).

Of a total of 230 space groups there are only 73 symmorphic space groups. Thus the simpler, symmorphic space groups are more often the exception rather than the norm. We will now consider how to analyze the rotational symmetries of nonsymmorphic space groups. By definition, a nonsymmorphic space group must contain at least one symmetry operation that involves both translation and rotation such that the rotational operation is not a symmetry operation of $\boldsymbol{G}$ by itself. There are two possibilities for such an operation: the rotation can be either proper or improper. The axis for a proper rotation is called a **screw axis** while the plane that corresponds to a twofold improper rotation is known as a **glide plane**. In the case of a screw axis, the crystal is invariant under a rotation about this axis plus a translation along the axis. The crystal is invariant under reflection in a glide plane followed by a translation parallel to the glide plane.

Two simple examples of screw axes for a one-dimensional crystal are shown in Figs. 2.4a and b. From Fig. 2.4b it is clear that a simple three-fold rotation about the vertical axis is not a symmetry operation of this hypothet-

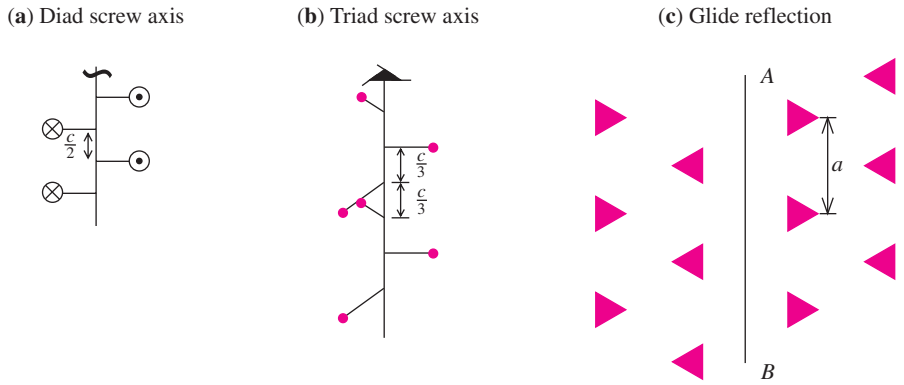


Fig. 2.4. Examples of (**a**) a diad screw axis, (**b**) a triad screw axis and (**c**) a glide plane. The crystals in (**a**) and (**c**) are assumed to be three dimensional, although only one layer of atoms is shown for the purpose of illustration. If they are two dimensional, the glide operation in (**c**) becomes equivalent to that of the diad screw in (**a**). $\odot$ and $\otimes$ represent arrow pointing towards and away from the reader, respectively. Screw axis such as in (**b**) are found in the crystal structure of semiconductors like Se and Te [2.5]

ical crystal. However, if the crystal is translated by an amount ($c/3$) along the vertical axis after the three-fold rotation then the crystal is unchanged. the vertical axis is known in this case as a **triad screw axis**. An example of a diad screw axis is shown in Fig. 2.4a. A glide plane is shown in Fig. 2.4c. The plane labeled A–B in the figure is not a reflection plane. But if after a reflection in the A–B plane we translate the crystal by the amount ($a/2$) parallel to the A–B plane, the crystal will remain unchanged. This symmetry operation is known as a **glide** and the A–B plane is a glide plane. Now suppose R is the set of all pure rotational operations of $\boldsymbol{G}$ plus the glide reflection shown in Fig. 2.4c (which we will denote as m). R defined in this way is not a group since mm is a pure translation and therefore not an element of R.

To study the rotational symmetries of a space group independent of whether it is symmorphic or nonsymmorphic, we will introduce the concept of a **factor group**. Let $\boldsymbol{G}$ be the space group and $\boldsymbol{T}$ its subgroup consisting of all purely translational symmetry operations. Let $C = \{\alpha, \beta, \ldots\}$ be the complex of all the elements of $\boldsymbol{G}$ not in $\boldsymbol{T}$. Unlike the elements of the set R defined earlier, the translation operations in the elements of C can belong to $\boldsymbol{T}$. Next we form the sets $T\alpha$, $T\beta$, etc. The set $T\alpha$ consisting of operations formed by the product of a translation in $\boldsymbol{T}$ and an operation α not in $\boldsymbol{T}$ is known as a **right coset** of $\boldsymbol{T}$. As may be expected, the set αT is called a **left coset** of $\boldsymbol{T}$.

Let us first consider the case when $\boldsymbol{G}$ is a symmorphic group. For a symmorphic group we can decompose any symmetry operation α in C into the product of a translation α_t and a rotation α_r : $\alpha = \alpha_t \alpha_r$. Since multiplication is not necessarily commutative, we may worry about the order in which the two operations α_t and α_r occur. It can be shown that $\boldsymbol{T}$ has the property that the right coset Tx is equal to the left coset xT for every element x in $\boldsymbol{G}$. A subgroup with this property is known as an **invariant subgroup**. When we multi-

ply α by another translation operation to form an element of the coset $T\alpha$ the resultant operation consists of a new translation but multiplied by the same rotation α_r. This suggests that we can establish a correspondence between the set of cosets $\{T\alpha, T\beta, \ldots\}$ and the set of rotational operations $R = \{\alpha_r, \beta_r, \ldots\}$. When $\boldsymbol{G}$ is symmorphic the set R is a subgroup of $\boldsymbol{G}$ so the set $\{T\alpha, T\beta, \ldots\}$ also forms a group. [In order that this set of cosets form a group, we have to define the product of two cosets $(T\alpha)(T\beta)$ as $T\alpha\beta$]. This group is known as the **factor group** of $\boldsymbol{G}$ with respect to $\boldsymbol{T}$ and is usually denoted by $\boldsymbol{G}/\boldsymbol{T}$. In establishing the factor group $\boldsymbol{G}/\boldsymbol{T}$ we have mapped all the elements of a coset $T\alpha$ into a single rotational operation α_r. Such a mapping of many elements in one set into a single element in another set is known as **homomorphism**. On the other hand, the mapping between the factor group $\boldsymbol{G}/\boldsymbol{T}$ and the subgroup $\boldsymbol{R}$ of $\boldsymbol{G}$ is one-to-one, and this kind of correspondence is known as **isomorphism**.

This isomorphism between the factor group $\boldsymbol{G}/\boldsymbol{T}$ and the point group $\boldsymbol{R}$ of a symmorphic space group can be extended to a nonsymmorphic space group. The main difference between the two cases is that while the rotational operations α_r, β_r, etc. are also elements in a symmorphic space group, this is not necessarily true for all rotations in a nonsymmorphic group. If α is a glide or screw then α_r is not an element in $\boldsymbol{G}$. We will still refer to the group $\boldsymbol{R}$ as the point group of a nonsymmorphic space group because $\boldsymbol{R}$ contains all the information about the rotational symmetries of the space group $\boldsymbol{G}$. However, special care must be exercised in studying the point groups of nonsymmorphic space groups since they contain elements which are not in the space group.

We will next study the symmetry operations of the zinc-blende and diamond crystal structures as examples of a symmorphic and a nonsymmorphic space group, respectively.

2.3.2 Symmetry Operations of the Diamond and Zinc-Blende Structures

Figure 2.2a shows the structures of the diamond and zinc-blende crystals. As pointed out in the previous section, both crystal structures consist of a fcc lattice. Associated with every lattice site there are two atoms which are displaced relative to each other by one quarter of the body diagonal along the [111] direction. The volume defined by the primitive lattice vectors and containing these two atoms forms a unit, known as the **primitive cell**, which is repeated at each lattice site. One simple way to construct these crystal structures is to start with two fcc sublattices, each containing only one atom located on every lattice site. Then one sublattice is displaced by one quarter of the body diagonal along the [111] direction with respect to the remaining sublattice. In the resulting crystal structure each atom is surrounded by four nearest neighbors forming a tetrahedron. The space group of the zinc-blende structure is symmorphic and is denoted by T_d^2 (or $F\bar{4}3m$ in international notation). Its translational symmetry operations are defined in terms of the three primitive lattice vectors shown in Fig. 2.2b. Its point group has 24 elements. These 24 elements are identical to the elements of the point group of a tetrahedron

(or the methane molecule discussed in the last section and shown in Fig. 2.3) which is denoted by T_d.

The point group symmetry operations of the zinc-blende crystal are defined with respect to the three mutually perpendicular crystallographic axes with the origin placed at one of the two atoms in the primitive unit cell. With this choice of coordinates, the 24 operations are enumerated below (they are essentially identical to those of the methane molecule):

E:	identity
eight C_3 operations:	clockwise and counterclockwise rotations of 120° about the [111], [$\bar{1}$11], [1$\bar{1}$1], and [11$\bar{1}$] axes, respectively;
three C_2 operations:	rotations of 180° about the [100], [010], and [001] axes, respectively;
six S_4 operations:	clockwise and counterclockwise improper rotations of 90° about the [100], [010], and [001] axes, respectively;
six σ operations:	reflections with respect to the (110), (1$\bar{1}$0), (101), (10$\bar{1}$), (011), and (01$\bar{1}$) planes, respectively.

The diamond structure is the same as the zinc-blende structure except that the two atoms in the primitive unit cell are identical. If we choose the origin at the midpoint of these two identical atoms, we find that the crystal structure is invariant under inversion with respect to this origin. However, for the purpose of studying the point group operations, it is more convenient to choose the *origin at an atom*, as in the case of the zinc-blende structure. The crystal is no longer invariant under inversion with respect to this new choice of origin, but is unchanged under inversion plus a translation by the vector $(a/4)[1,1,1]$, where a is the length of the unit cube. This can be visualized by drawing the carbon atoms in the diamond structure along the [111] direction as shown in Fig. 2.5. **The space group of the diamond structure is nonsymmorphic**: it contains three glide planes. For example, the plane defined by $x = (a/8)$ is a glide plane since diamond is invariant under a translation by $(a/4)[0,1,1]$ followed by a reflection on this plane. In place of the three glide planes defined by $x = (a/8)$, $y = (a/8)$, and $z = (a/8)$, it is possible to use the "glide-like" operations:

$T(1/4,1/4,1/4)\sigma_x$:	reflection on the $x = 0$ plane followed by a translation of the crystal by the vector $a(1/4,1/4,1/4)$;
$T(1/4,1/4,1/4)\sigma_y$:	reflection on the $y = 0$ plane followed by a translation of the crystal by the vector $a(1/4,1/4,1/4)$; and

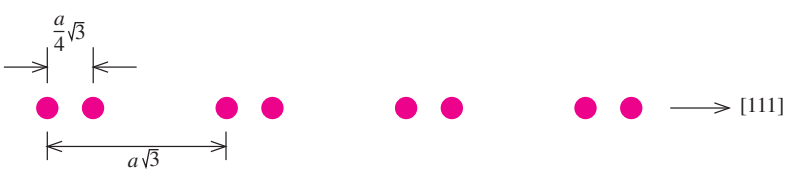


Fig. 2.5. Arrangement of atoms along the [111] direction of the diamond crystal. Notice that the crystal is invariant under inversion either with respect to the midpoint between the atoms or with respect to one of the atoms followed by an appropriate translation along the [111] axis

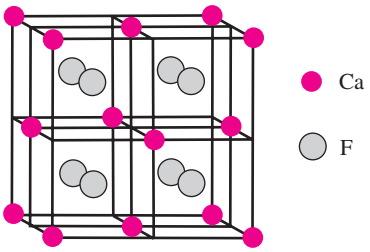


Fig. 2.6. Schematic crystal structure of CaF_2 (fluorite)

$T(1/4, 1/4, 1/4)\sigma_z$: reflection on the $z = 0$ plane followed by a translation of the crystal by the vector $a(1/4, 1/4, 1/4)$.

The factor group of the diamond lattice is isomorphic to the point group generated from the group T_d by adding the inversion operation. This point group has 48 elements and is denoted as O_h. While T_d is the point group of a tetrahedron, O_h is the point group of a cube. The space group of the diamond crystal is denoted by O_h^7 (or $Fd3m$ in international notation).

The CaF_2 (**fluorite**) structure shown in Fig. 2.6 is related to the diamond structure. This is the crystal structure of a family of semiconductors with the formula Mg_2X, where X = Ge, Si, and Sn. The lattice of CaF_2 is fcc as in diamond, but CaF_2 has three sublattices. The two fluorine sublattices are symmetrically displaced by one quarter of the body diagonal from the Ca sublattices, so there is inversion symmetry about each Ca atom. The space group of CaF_2 is symmorphic and its point group is also O_h, like diamond. This space group is denoted by O_h^5 (or $Fm3m$). It is clear that there is a one-to-one correspondence between the elements of the point group of CaF_2 and those of the factor group of diamond.

2.3.3 Representations and Character Tables

The effect of a symmetry operation, such as a rotation, on a coordinate system (x, y, z) can be represented by a transformation matrix. For example, under a four-fold rotation about the x axis the axes x, y, and z are transformed into x', y', and z' with $x' = x$, $y' = z$ and $z' = -y$. This transformation can be represented by the matrix $\boldsymbol{M}$:

$$\boldsymbol{M} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}.$$

Similarly a three-fold rotation about the [111] axis will transform the axes x, y, and z into $x' = z$, $y' = x$, and $z' = y$. This transformation can be represented by

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

In the rest of this chapter we will use the abbreviated notation $(xyz) \rightarrow (xz\overline{y})$ to denote the transformation matrix for the four-fold rotation and $(xyz) \rightarrow (zxy)$ for the three-fold rotation. All the symmetry operations in a point group can be represented by transformation matrices similar to $\boldsymbol{M}$. It is easy to prove that the set of such transformation matrices corresponding to a group of symmetry operations is also a group. This group of matrices is said to form a **representation** of the group. There are actually an infinite number of such groups of matrices for a given group. The correspondence between a group and its representation is not, in general, an isomorphism but rather a homomorphism. A representation of a group $\boldsymbol{G}$ is defined as any group of matrices onto which $\boldsymbol{G}$ is homomorphic. Since representations of a group are not unique, we will be interested only in those of their properties that are common to all the representations of this group.

One way to generate a representation for a group is to choose some function $f(x, y, z)$ and then generate a set of functions $\{f_i\}$ by applying the symmetry operations O_i of the group to $f(x, y, z)$ so that $f_i = O_i[f]$.[1] By definition a group has to satisfy the closure requirement. This means that when the operation O is applied to f_i the resultant function $O[f_i]$ can be expressed as a linear combination of the functions f_i:

$$O[f_i] = \sum_j f_j a_{ji}. \tag{2.10a}$$

The coefficients a_{ji} form a square matrix, which will be referred to as a **transformation matrix**. The set of transformation matrices of the form $\{a_{ji}\}$ corresponding to all the operations in the group now forms a representation of the group. The functions $\{f_i\}$ used to generate this representation are said to form a set of **basis functions** for this representation. Clearly the choice of basis functions for generating a given representation is not unique.

If one uses the above method to generate a representation, then the dimension of the resulting transformation matrices will always be equal to the number of elements in the group (known as the **order** of a group). Some of the matrices will, however, be equal. Also many of the elements in these matrices will be zero. If the matrices in a given representation *for all the operations in a group* can be expressed in the following block form:

$$\begin{pmatrix} & & & 0 & & 0 & \dots & 0 \\ & \alpha & & 0 & & 0 & \dots & 0 \\ & & & 0 & & 0 & \dots & 0 \\ 0 & 0 & 0 & \beta & & 0 & \dots & 0 \\ 0 & 0 & 0 & & \gamma & & \dots & 0 \\ 0 & 0 & 0 & & & & \dots & 0 \\ 0 & 0 & 0 & & & & \dots & 0 \\ 0 & 0 & 0 & 0 & & 0 & \dots & \tau \end{pmatrix}, \tag{2.10b}$$

[1] In our applications $\{f_i\}$ will usually be a set of *degenerate* eigenfunctions corresponding to a given eigenvalue.

where α, β, ..., τ are square matrices, obviously the symmetry operations in this group can also be represented by the smaller matrices α and β, etc. While the matrices α_i and β_i for an operation i may not necessarily have the same dimension, the matrices α_i for all the operations i in the group must have the same dimension. A representation of the form of (2.10b) is said to be **reducible** otherwise the representation may be irreducible. As pointed out earlier, the choice of matrices to form a representation for a given group is not unique. Given one set of transformation matrices $\{A_i\}$, we can generate another set $\{A_i'\}$ by a **similarity transformation**: $A_i' = TA_iT^{-1}$, where T is an arbitrary nonsingular matrix with the same dimensionality as A_i. The transformed set of matrices $\{A_i'\}$ will also form a representation. The two sets of matrices $\{A_i\}$ and $\{A_i'\}$ are then said to be **equivalent**. Often the matrices of a representation may not appear to have the form given by (2.10b) and hence be regarded as irreducible. However, if by applying similarity transformation it is possible to express these matrices in the form of (2.10b) then this representation is also called reducible. Otherwise it is called **irreducible**.

Two axes of rotation or two reflection planes which transform into each other under a symmetry operation of a point group are said to be equivalent. It can be shown that the matrices of a representation which correspond to such equivalent rotations have identical **traces** (the trace of a matrix is the sum of the diagonal elements). Although the choice of irreducible representations for a group is not unique, the set of traces of these irreducible representations is unique since unitary transformations preserve the trace. This suggests that the set of all equivalent irreducible representations of a given group can be specified uniquely by their traces. For this reason the traces of the matrices in a representation are called its **characters**. The representations obviously contain more information than their characters; however, to utilize the symmetry of a given group it often suffices to determine the number of inequivalent irreducible representations and their characters.

The determination of the characters of an irreducible representation is simplified by these properties of a group:

- Elements in a group can be grouped into **classes**. A set of elements T in a group is said to form a class if for any element a in the group, $aT = Ta$. In a given representation all the elements in a class have the same character.
- The number of inequivalent irreducible representations of a group is equal to the number of classes.

These two properties suggest that if the elements of a group can be divided into j classes the characters of its j irreducible representations can be tabulated to form a table with j columns and j rows, which is known as a **character table**. Assume that a group has N elements and these elements are divided into j classes denoted by $C_1, C_2, \ldots, C_j$. The number of elements in each class will be denoted by $N_1, N_2, \ldots, N_j$. The identity operation E forms a class with only one element and, by convention, it is labeled C_1. This group also has j inequivalent irreducible representations (from now on the set of irreducible

Table 2.1. The character table of a group

	Classes			
Representations	$\{E\}$	$\{N_2C_2\}$	· · ·	$\{N_jC_j\}$
R_1	$\chi_1(E)$	$\chi_1(2)$	· · ·	$\chi_1(j)$
R_2	$\chi_2(E)$	$\chi_2(2)$	· · ·	$\chi_2(j)$
·	·	·	· · ·	·
·	·	·	· · ·	·
R_j	$\chi_j(E)$	$\chi_j(2)$	· · ·	$\chi_j(j)$

representations of a group will be understood to contain only inequivalent ones), which will be denoted by $R_1, R_2, \ldots, R_j$. The character of C_k in R_i will be denoted by $\chi_i(k)$. Since the identity operation E leaves any basis function invariant, its representations always consist of **unit matrices** (that is, diagonal matrices with unity as the diagonal elements). As a result, the character $\chi_i(E)$ is equal to the **dimension of the representation** R_i. Thus the character table of this group will have the form of Table 2.1.

In principle, the character table for the point group of a crystal can be calculated from the transformation matrices using a suitable set of basis functions. In practice, the character table can be obtained, in most cases, by inspection using the following two **orthogonality relations**:

$$\sum_k \chi_i(C_k)^* \chi_j(C_k) N_k = h\delta_{ij} \tag{2.11}$$

$$\sum_i \chi_i(C_k)^* \chi_i(C_l) = (h/N_l)\delta_{kl}, \tag{2.12}$$

where * denotes the complex conjugate of a character, h is the order of the group, N_k is the number of elements of class C_k, and δ_{ij} is the Kronecker delta.

As an illustration of the procedure used to obtain character tables we will consider two examples.

EXAMPLE 1 Character Table of the Point Group T_d

As we showed in Sect. 2.3.1, the point group T_d consists of 24 elements representing the proper and improper rotational symmetry operations of a tetrahedral methane molecule. In Sect. 2.3.2 we showed that this group is also the point group of the zinc-blende crystal. The 24 elements of this group can be divided into five classes

$$\{E\}, \quad \{8C_3\}, \quad \{3C_2\}, \quad \{6S_4\} \quad \text{and} \quad \{6\sigma\}$$

by noting that:

- rotations by the same angle with respect to **equivalent** axes belong to the same class and
- reflections on **equivalent** planes also belong to the same class.

Since the number of irreducible representations is equal to the number of classes, T_d has five irreducible representations, which are usually denoted by A_1, A_2, E, T_1 and T_2. Notice that the capital letter E has been used in the literature to denote a large number of entities varying from energy, electric field, the identity operation in group theory to an irreducible representation in the T_d group! To avoid confusion we will always specify what E stands for.

The next step is to construct the 5×5 character table using (2.11) and (2.12). First, we note again that the character of the class containing the identity operation $\{E\}$ is equal to the dimension of the representation. Substituting this result into (2.12) we find

$$\sum_i |\chi_i(E)|^2 = h. \tag{2.13}$$

Since the number of classes is usually small, this equation can often be solved by inspection. For T_d it is easily shown that the only possible combination of five squares which add up to 24 is: $2 \times 1^2 + 2^2 + 2 \times 3^2$. This result means that the group T_d has two irreducible representations of dimension one (denoted by A_1 and A_2), one irreducible representation of dimension two (denoted by E), and two irreducible representations of dimension three (denoted by T_1 and T_2). Next we note that a scalar will be invariant under all operations, so there is always a trivial **identity representation** whose characters are all unity. By convention this representation is labeled by the subscript 1, A_1 in the present case. So without much effort we have already determined one row and one column of the character table for T_d (Table 2.2).

The remaining characters can also be determined by inspection with the application of (2.12). For the classes other than $\{E\}$ the characters can be either positive or negative. The sign can be determined by inspection with some practice. For example, for the class $\{6\sigma\}$ the only combination of sums of squares satisfying (2.12) is $4 \times 1^2 + 0^2 = 24/6 = 4$. Applying (2.12) to the characters of $\{E\}$ and $\{6\sigma\}$ it can be easily seen that $A_2(6\sigma) = -1$, $E(6\sigma) = 0$

Table 2.2. Determining the character table for the T_d group by inspection

	$\{E\}$	$\{3C_2\}$	$\{6S_4\}$	$\{6\sigma\}$	$\{8C_3\}$
A_1	1	1	1	1	1
A_2	1	·	·	·	·
E	2	·	·	·	·
T_1	3	·	·	·	·
T_2	3	·	·	·	·

Table 2.3. Character table and basis functions of the T_d group

	$\{E\}$	$\{3C_2\}$	$\{6S_4\}$	$\{6\sigma\}$	$\{8C_3\}$	Basis functions
A_1	1	1	1	1	1	xyz
A_2	1	1	−1	−1	1	$x^4(y^2-z^2)+y^4(z^2-x^2)+z^4(x^2-y^2)$
E	2	2	0	0	−1	$\{(x^2-y^2), z^2-\frac{1}{2}(x^2+y^2)\}$
T_1	3	−1	1	−1	0	$\{x(y^2-z^2), y(z^2-x^2), z(x^2-y^2)\}$
T_2	3	−1	−1	1	0	$\{x, y, z\}$

while the two remaining characters for T_1 and T_2 contain 1 and -1. The final result for the character table of T_d is given in Table 2.3.

It is instructive to examine some possible basis functions for the irreducible representations of T_d. One choice of basis functions for the A_1 representation is a constant, as we have mentioned earlier. Another possibility would be the function xyz, which is also invariant under all symmetry operations of T_d. A_2 is very similar to A_1 except that under the operations S_4 and σ the character of A_2 is -1 rather than 1. This implies that the basis function for A_2 must change sign under interchange of any two coordinate axes, such as interchanging x and y. One choice of basis function for A_2 is $x^4(y^2-z^2)+y^4(z^2-x^2)+z^4(x^2-y^2)$. Similarly, the three-dimensional representations T_1 and T_2 differ only in the sign of their characters under interchange of any two coordinates. It can be shown that the three components x, y, and z of a vector transform as T_2. A corresponding set of basis functions for the T_1 representation would be $x(y^2-z^2)$, $y(z^2-x^2)$, and $z(x^2-y^2)$. The reader should verify these results by calculating the characters directly from the basis functions (Problem 2.2).

At the beginning of this chapter we pointed out the importance of notation in group theory. The notation we have used so far to label the irreducible representations of the T_d group: A_1, E, T_1, etc. is more commonly found in literature on molecular physics. We now introduce another notation used frequently in articles on semiconductor physics. The wave functions of a crystal with wave vector $\boldsymbol{k}$ at the center of the Brillouin zone (Γ point) always transform in the way specified by the irreducible representations of the point group of the crystal. Hence the Bloch functions at Γ of a zinc-blende crystal can be classified according to these irreducible representations. In semiconductor physics literature it is customary to use Γ plus a subscript i to label these irreducible representations of T_d. Unfortunately there are two different conventions in the choice of the subscript i for labeling the same irreducible representation. One of these conventions is due to Koster (more commonly used in recent research articles) while the other was proposed by Bouckaert, Smoluchowski and Wigner (BSW) and tends to be found in older articles. The correspondence between the different notations for the T_d point group is shown in Table 2.4.

Table 2.4. Commonly used notations for the irreducible representations of the T_d point group

Koster notation[a]	BSW notation	Molecular notation
Γ_1	Γ_1	A_1
Γ_2	Γ_2	A_2
Γ_3	Γ_{12}	E
Γ_4	Γ_{15}	T_2
Γ_5	Γ_{25}	T_1

[a] Note that Γ_4 and Γ_5 are sometimes reversed in the literature. We recommend the student to check it whenever he encounters this notation [2.4].

EXAMPLE 2 Character Table of O_h

We mentioned earlier in this section that the factor group of the diamond structure is O_h and that it is isomorphic to the point group derived from the T_d group by including the inversion operation i. It has therefore 48 elements: the 24 symmetry operations of T_d plus those of T_d followed by i. These include all 48 symmetry operations of a cube. From the properties of the group T_d, one can deduce that O_h has ten classes:

$\{E\}$: identity;
$\{3C_2\}$: C_2 rotation about each of the three equivalent [100] axes;
$\{6S_4\}$: two four-fold improper rotations about each of the three equivalent [100] axes;
$\{6\sigma_d\}$: reflection on each of the six equivalent (110) planes;
$\{8C_3\}$: two C_3 rotations about each of the four equivalent [111] axes;
$\{i\}$: inversion;
$\{3\sigma_h\}$: reflection on each of the three equivalent (100) planes;
$\{6C_4\}$: two C_4 rotations about each of the three equivalent [100] axes;
$\{6C_2'\}$: C_2 rotation about each of the six equivalent [110] axes;
$\{8S_6\}$: two three-fold improper rotations about each of the four equivalent [111] axes.

The first five classes are the same as those of T_d while the remaining five are obtained from the first five by multiplication with the inversion.

Correspondingly, there are ten irreducible representations. Five of them correspond to even transformations under those operations obtained from the T_d group operation followed by inversion, while the other five correspond to odd ones. Similarly, the basis functions of the irreducible representations of O_h are either even or odd under those operations. In the terminology of quantum mechanics, these basis functions are said to have even or odd **parity**. The characters for the O_h group are listed in Table 2.5, while a set of basis functions for its irreducible representations is given in Table 2.6. Table 2.5 has been purposedly presented in a way to show the similarity between the "unprimed" representations in the O_h group and those of the T_d group. For example, a scalar

Table 2.5. Character table of the O_h group presented in a way to highlight the similarity with Table 2.3 for the T_d group. BSW notation

	$\{E\}$	$\{C_2\}$	$\{S_4\}$	$\{\sigma_d\}$	$\{C_3\}$	$\{i\}$	$\{\sigma_h\}$	$\{C_4\}$	$\{C_2'\}$	$\{S_6\}$
Γ_1	1	1	1	1	1	1	1	1	1	1
Γ_2	1	1	−1	−1	1	1	1	−1	−1	1
Γ_{12}	2	2	0	0	−1	2	2	0	0	−1
Γ_{25}	3	−1	1	−1	0	−3	1	−1	1	0
Γ_{15}	3	−1	−1	1	0	−3	1	1	−1	0
$\Gamma_{1'}$	1	1	−1	−1	1	−1	−1	1	1	1
$\Gamma_{2'}$	1	1	1	1	1	−1	−1	−1	−1	−1
$\Gamma_{12'}$	2	2	0	0	−1	−2	−2	0	0	1
$\Gamma_{25'}$	3	−1	−1	1	0	3	−1	−1	1	0
$\Gamma_{15'}$	3	−1	1	−1	0	3	−1	1	−1	0

Table 2.6. Basis functions for the irreducible representations of the O_h group

Representation	Basis functions
Γ_1 :	1
Γ_2 :	$x^4(y^2 - z^2) + y^4(z^2 - x^2) + z^4(x^2 - y^2)$
Γ_{12} :	$\{[z^2 - (x^2 + y^2)/2],\ x^2 - y^2\}$
Γ_{25} :	$\{x(y^2 - z^2),\ y(z^2 - x^2),\ z(x^2 - y^2)\}$
Γ_{15} :	$\{x,\ y,\ z\}$
$\Gamma_{1'}$:	$xzy[x^4(y^2 - z^2) + y^4(z^2 - x^2) + z^4(x^2 - y^2)]$
$\Gamma_{2'}$:	xyz
$\Gamma_{12'}$:	$\{xyz[z^2 - (x^2 + y^2)/2],\ xyz(x^2 - y^2)\}$
$\Gamma_{25'}$:	$\{xy,\ yz,\ zx\}$
$\Gamma_{15'}$:	$\{yz(y^2 - z^2),\ zx(z^2 - x^2),\ xy(x^2 - y^2)\}$

still belongs to the Γ_1 representation while a vector belongs to the Γ_{15} representation in the O_h group. However, the relation between the "primed" and "unprimed" representations is not so clear. For example, a pseudo-scalar belongs to the $\Gamma_{2'}$ representation, while a pseudo-vector belongs to the $\Gamma_{15'}$ representation. Furthermore, some of the primed representations, e. g. $\Gamma_{15'}$ and $\Gamma_{25'}$, are even while others are odd under inversion.

When Table 2.5 is rearranged into Table 2.7, the correlations between the first five representations and the remaining five become clear. Note that sometimes a hybrid of the K and BSW notations is used: the primes are omitted and replaced by $+$, $-$ superscripts to denote the parity. The student will find this notation in Chaps. 6 and 7.

Table 2.7. Character table of the O_h group rearranged to show the relationship between the even and odd parity representations. Both the Koster (K) and BSW notations are given [2.6]

K	BSW	$\{E\}$	$\{C_2\}$	$\{C_4\}$	$\{C_2'\}$	$\{C_3\}$	$\{i\}$	$\{\sigma_h\}$	$\{S_4\}$	$\{\sigma_d\}$	$\{S_6\}$
Γ_1^+	Γ_1	1	1	1	1	1	1	1	1	1	1
Γ_2^+	Γ_2	1	1	−1	−1	1	1	1	−1	−1	1
Γ_3^+	Γ_{12}	2	2	0	0	−1	2	2	0	0	−1
Γ_4^+	$\Gamma_{15'}$	3	−1	1	−1	0	3	−1	1	−1	0
Γ_5^+	$\Gamma_{25'}$	3	−1	−1	1	0	3	−1	−1	1	0
Γ_1^-	$\Gamma_{1'}$	1	1	1	1	1	−1	−1	−1	−1	−1
Γ_2^-	$\Gamma_{2'}$	1	1	−1	−1	1	−1	−1	1	1	−1
Γ_3^-	$\Gamma_{12'}$	2	2	0	0	−1	−2	−2	0	0	1
Γ_4^-	Γ_{15}	3	−1	1	−1	0	−3	1	−1	1	0
Γ_5^-	Γ_{25}	3	−1	−1	1	0	−3	1	1	−1	0

2.3.4 Some Applications of Character Tables

We will now describe some of the applications of character tables. Further applications will be found throughout this book.

a) Decomposition of Representation into Irreducible Components

A problem one often faces is this: when given a group $\boldsymbol{G}$ and a representation τ, how does one determine whether τ is reducible? If τ is reducible then how can it be decomposed into its irreducible components? These questions can be answered with the help of the character table of $\boldsymbol{G}$. Suppose $\chi_\tau(i)$ is the character of the given representation τ corresponding to the class $\{i\}$. If τ is an irreducible representation, the set of characters $\chi_\tau(i)$ must be equal to the characters of one of the irreducible representations of $\boldsymbol{G}$. If this is not the case then τ is reducible. Suppose τ is reducible into two irreducible representations α and β and $\chi_\alpha(i)$ and $\chi_\beta(i)$ are the characters of α and β, respectively. By definition $\chi_\alpha(i)$ and $\chi_\beta(i)$ must satisfy

$$\chi_\tau(i) = \chi_\alpha(i) + \chi_\beta(i) \tag{2.14}$$

for all classes $\{i\}$ in the group $\boldsymbol{G}$. The representation τ is said to be the **direct sum** of the two irreducible representations α and β. The direct sum will be represented by the symbol $\oplus$ as in $\tau = \alpha \oplus \beta$.

When the dimension of a reducible representation is not very large, it can often be reduced into a direct sum of irreducible representations by inspection. As an example, let us consider the group T_d with its character table given in Table 2.3 and a second-rank tensor $\{T_{ij}\}$ with components T_{xx}, T_{xy}, T_{xz}, T_{yx}, T_{yy}, T_{yz}, T_{zx}, T_{zy}, and T_{zz}. Using these components as basis functions we can generate a nine-dimensional representation of T_d, which we will denote as Γ. Obviously Γ must be reducible since no irreducible representation in T_d has

dimensions larger than three. The way to decompose Γ into irreducible representations of T_d is to first determine the characters of Γ for all the classes in T_d. In principle this can be accomplished by applying the symmetry operations of T_d to the nine basis functions to produce the 9×9 matrices forming the representation Γ. A simpler and more direct approach is possible for this second-rank tensor. We note that a vector with three components x, y, and z forms a set of basis functions for the three-dimensional irreducible representation T_2 of T_d. Therefore, the 3×3 transformation matrices of a vector form a T_2 representation. By taking the **matrix product** of two such 3×3 transformation matrices we obtain a set of 9×9 matrices forming a representation for Γ. This suggests that the characters of Γ are equal to the squares of the characters for T_2:

	$\{E\}$	$\{3C_2\}$	$\{6S_4\}$	$\{6\sigma\}$	$\{8C_3\}$
χ_Γ :	9	1	1	1	0

When the matrices of a representation τ are equal to the matrix product of the matrices of two representations α and β, τ is said to be the **direct product** of α and β. Direct products are represented by the symbol $\otimes$ as in

$$\Gamma = T_2 \otimes T_2. \tag{2.15}$$

After determining the characters of Γ, the next step is to find the irreducible representations of T_d whose characters will add up to those of Γ. The systematic way of doing this is to apply the orthogonality relation (2.11). It is left as an exercise (Problem 2.3) to show that

$$T_2 \otimes T_2 = T_1 \oplus T_2 \oplus E \oplus A_1 \tag{2.16}$$

With practice this result can also be derived quickly by inspection. In the present example, one starts by writing down various combinations of representations with total dimensions equal to nine. Next one eliminates those combinations whose characters for the other classes do not add up to χ_Γ. Very soon it is found that the only direct sum with characters equal to that of Γ for all five classes of T_d is the one in (2.16). Once we realize that Γ can be decomposed into the direct sum of these four irreducible representations, we can use the basis functions for these representations given in Table 2.3 as a guide to deduce the correct linear combinations of the nine components of the second-rank tensor which transform according to these four irreducible representations:

$$\begin{aligned}
A_1 &: \; T_{xx} + T_{yy} + T_{zz} \\
E &: \; \{T_{xx} - T_{yy}, T_{zz} - (T_{xx} + T_{yy})/2\} \\
T_1 &: \; \{(T_{xy} - T_{yx})/2, \quad (T_{zx} - T_{xz})/2, \quad (T_{yz} - T_{zy})/2\} \\
T_2 &: \; \{(T_{xy} + T_{yx})/2, \quad (T_{xz} + T_{zx})/2, \quad (T_{yz} + T_{zy})/2\}.
\end{aligned}$$

b) Symmetrization of Long Wavelength Vibrations in Zinc-Blende and Diamond Crystals

The process we have described above is known as the **symmetrization** of the nine components of the second-rank tensor. This method can also be applied

to symmetrize wave functions. When a Hamiltonian is invariant under the symmetry operations of a group, its wave functions can be symmetrized so as to belong to irreducible representations of this group. Just as atomic wave functions are labeled s, p, and d according to their symmetry under rotation, it is convenient to label the electronic and vibrational wave functions of a crystal at the point $\boldsymbol{k}$ in reciprocal space by the irreducible representations of the group of symmetry operations appropriate for $\boldsymbol{k}$. We will now explain this statement with an example drawn from the vibrational modes of zinc-blende and diamond crystals.

Although we will not discuss lattice vibrations in semiconductors until the next chapter, it is easier to demonstrate their symmetry properties than those of electrons. First, we can argue that vibrations of atoms in a crystal can be described by waves based on its translational symmetry, just as its electrons can be described by Bloch functions. For example, sound is a form of such vibration. Thus atomic motions in a crystal can be characterized by their **displacement vectors** (in real space) plus their wave vectors $\boldsymbol{k}$ (in reciprocal lattice space). The symmetry of a vibration is therefore determined by the effects of symmetry operations of the crystal on both vectors. Due to the discrete location of atoms in a crystal, a wave with wave vector equal to $\boldsymbol{k}$ or $\boldsymbol{k}$ plus a reciprocal lattice vector are indistinguishable (this point will be discussed further in Chap. 3). Thus an operation which transforms $\boldsymbol{k}$ into another wave vector $\boldsymbol{k}'$ differing from $\boldsymbol{k}$ by a reciprocal lattice vector also belongs to the group of symmetry operations of $\boldsymbol{k}$. This group is known as the **group of the wave vector** $\boldsymbol{k}$. In particular, the group of the Γ point or zone center is always the same as the point group of the crystal.

A long wavelength (that is, $\boldsymbol{k}$ near the Brillouin zone center) vibration in a crystal involves nearly uniform displacements of identical atoms in different unit cells. For a zinc-blende crystal with two atoms per primitive unit cell, a zone-center vibrational mode can be specified by two vectors representing the displacements of these two atoms. We have already pointed out that the three components of a vector transform under the symmetry operations of T_d according to the T_2, also called Γ_4 representation (see Tables 2.3 and 2.4). To discuss properties in the zinc-blende crystal we will switch to the Koster notation. For brevity we will refer to the vector as "**belonging**" to the Γ_4 representation. Two vectors, one associated with each atom in the primitive cell, give rise to a six-dimensional representation. Since irreducible representations in T_d have at most three dimensions, this representation is reducible. To reduce it one can calculate its characters by applying the symmetry operations of T_d to the two vectors. An alternative method is to consider the two atoms as the basis of a two-dimensional representation R. The characters of R are obtained by counting the number of atoms which are unchanged by the symmetry operations of T_d (since each atom satisfying this condition contributes one unity diagonal element to the representation matrix). The two atoms in the zinc-blende lattice are not interchanged by the operations of T_d, therefore all the characters of R are simply two. Thus R is reducible to two Γ_1 representations. The representations of the two displacement vectors in the unit cell

of the zinc-blende crystal are equal to the direct product of R and Γ_4, which is equal to $2\Gamma_4$. These two Γ_4 representations correspond to the **acoustic** and **optical phonon** modes (see Chap. 3 for further details). In the acoustic mode the two atoms in the primitive cell move in phase while in the optical mode they move 180° out of phase.

As pointed out in Sect. 2.3.2, the factor group of the diamond crystal is isomorphic to the point group O_h. We should remember that the origin has been chosen to be one of the carbon atoms. The space group operation which corresponds to inversion in O_h is inversion about the origin plus a translation by $(a/4)(1,1,1)$ (for brevity this operation will be denoted here by i'). From Table 2.6 one finds that a vector belongs to the Γ_{15} representation of the O_h point group. As in the case of the zinc-blende crystal, we can obtain the characters of the six-dimensional representation by determining the characters of R and then calculating the direct product of R and Γ_{15}. The characters of R now depend on whether the symmetry operations include i'. For all symmetry operations which already exist for the zinc-blende structure and therefore do not involve i', the characters are equal to two, as in the zinc-blende crystal. For all other operations, the two atoms inside the primitive unit cell are interchanged by i', so their characters are zero. By inspection of Table 2.5 one concludes that R reduces to $\Gamma_1 \oplus \Gamma_{2'}$. Thus the displacement vectors of the two atoms in the primitive unit cell of diamond transform as Γ_{15} and $\Gamma_{25'}$. The displacement vectors of the acoustic phonon change sign (the parity is said to be odd) under i' and therefore have symmetry Γ_{15}. On the other hand the optical phonon parity is even and has symmetry $\Gamma_{25'}$. The effects of i' on the long-wavelength acoustic and optical phonons propagating along the body diagonal of the diamond crystal are shown in Fig. 2.7.

c) Symmetrization of Nearly Free Electron Wave Functions in Zinc-Blende Crystals

As an example of application of character tables in symmetrizing electronic wave functions, we will consider a **nearly free electron** in a zinc-blende crystal. By nearly free we mean that the electron is moving inside a crystal with a vanishingly small periodic potential of T_d symmetry, so that its energy E and wave function Φ are essentially those of a free particle:

$$\Phi(x, y, z) = \exp\left[\mathrm{i}(k_x x + k_y y + k_z z)\right] \tag{2.17}$$

and

$$E = \hbar^2 k^2 / 2m. \tag{2.18}$$

However, because of the periodic lattice, its wave vector $\boldsymbol{k}$ can be restricted to the first Brillouin zone in the reduced zone scheme.

We will assume that the crystal has the zinc-blende structure and that $\boldsymbol{k} = (2\pi/a)(1,1,1)$, where a is the length of an edge of the unit cube in the zinc-blende lattice. By applying the C_3 symmetry operations of zinc-blende we can show that all the eight points $(2\pi/a)(\pm 1, \pm 1, \pm 1)$ in the Brillouin zone are equivalent. Furthermore, from the definition of the primitive reciprocal lattice vectors given in Sect. 2.2 all eight points differ from the zone center by

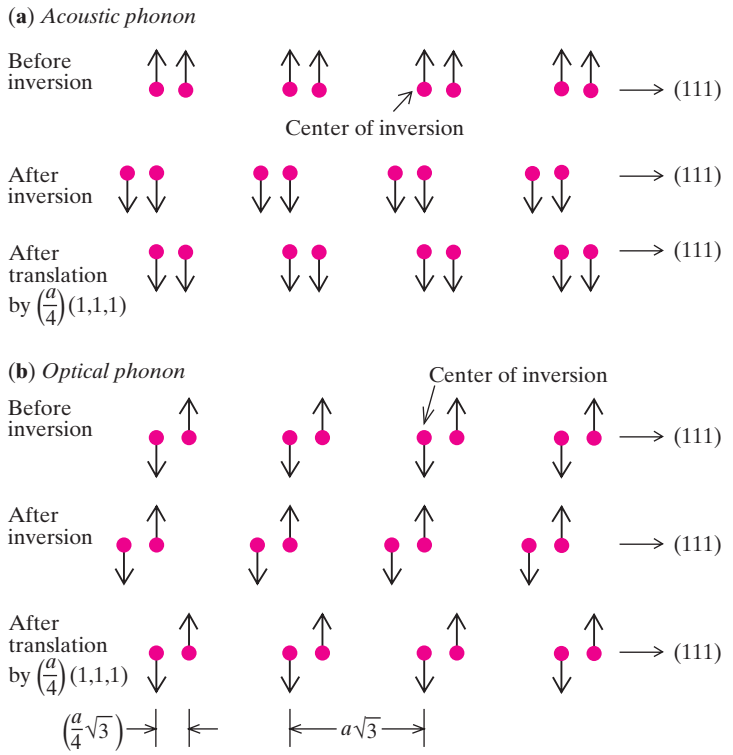


Fig. 2.7. Schematic diagrams of the transformation of (**a**) the acoustic phonon and (**b**) the zone-center optical phonon in diamond under inversion plus translation by $(a/4)(1,1,1)$ where a is the size of the unit cube of diamond

a primitive reciprocal lattice vector. Hence all eight points will map onto the zone center in the reduced zone scheme. The group of the wave vector $\boldsymbol{k} = (2\pi/a)(1,1,1)$ is therefore T_d. To simplify the notation we will represent the electronic wave functions $\exp[\mathrm{i}(k_x x + k_y y + k_z z)]$ as $\{k_x k_y k_z\}$. The eight wave functions $\{111\}, \{\bar{1}11\}, \{1\bar{1}1\}, \{11\bar{1}\}, \{\bar{1}\bar{1}1\}, \{\bar{1}1\bar{1}\}, \{1\bar{1}\bar{1}\}$ and $\{\bar{1}\bar{1}\bar{1}\}$ are degenerate but the degeneracy will be lifted by perturbations such as a nonzero crystal potential. Our goal now is to form symmetrized linear combinations of these eight wave functions with the aid of the character table of T_d in Table 2.3.

We note first that these eight wave functions form the basis functions of an eight-dimensional representation. Obviously this representation is reducible. Unlike the cases given in (a) and (b), there are no shortcuts in determining the characters of this eight-dimensional representation. Since characters are the sums of the diagonal elements, they can be deduced by determining the number of wave functions unchanged by the symmetry operations. The characters calculated in this way are given in Table 2.8.

Table 2.8. Characters of the representations formed by the nearly free electron wave functions with wave vectors equal to $(2\pi/a)(1,1,1)$ and $(2\pi/a)(2,0,0)$ in a zinc-blende crystal

Class	Transformation	Characters	
		[111]	[200]
E	xyz	8	6
$3C_2$	$x\bar{y}\bar{z}$	0	2
$6S_4$	$\bar{x}z\bar{y}$	0	0
6σ	yxz	4	2
$8C_3$	yzx	2	0

Table 2.9. Symmetrized nearly free electron wave functions in a zinc-blende crystal with wave vectors equal to $(2\pi/a)(\pm1,\pm1,\pm1)$

Representation	Wave function
Γ_1	$(1/\sqrt{8})(\{111\}+\{1\bar{1}\bar{1}\}+\{\bar{1}1\bar{1}\}+\{\bar{1}\bar{1}1\}+\{\bar{1}\bar{1}\bar{1}\}+\{\bar{1}11\}+\{1\bar{1}1\}+\{11\bar{1}\}) = (\sqrt{8})\cos(2\pi x/a)\cos(2\pi y/a)\cos(2\pi z/a)$
Γ_1	$(\sqrt{8})\sin(2\pi x/a)\sin(2\pi y/a)\sin(2\pi z/a)$
Γ_4	$(\sqrt{8})\{\sin(2\pi x/a)\sin(2\pi y/a)\cos(2\pi z/a);$ $\sin(2\pi x/a)\cos(2\pi y/a)\sin(2\pi z/a);$ $\cos(2\pi x/a)\sin(2\pi y/a)\sin(2\pi z/a)\}$
Γ_4	$(\sqrt{8})\{\sin(2\pi x/a)\cos(2\pi y/a)\cos(2\pi z/a);$ $\cos(2\pi x/a)\sin(2\pi y/a)\cos(2\pi z/a);$ $\cos(2\pi x/a)\cos(2\pi y/a)\sin(2\pi z/a)\}.$

By using the orthogonality relations or the method of "inspection", we found from Table 2.3 (using Table 2.4 to convert to the Koster notation) that the only combination of irreducible representations giving rise to the set of characters in Table 2.8 is the direct sum $2\Gamma_1 \oplus 2\Gamma_4$. Thus the eight $\{(\pm)1(\pm)1(\pm)1\}$ free electron wave functions can be expressed as two wave functions belonging to the one-dimensional Γ_1 representation and two wave functions belonging to the three-dimensional Γ_4 representation. The proper linear combinations of wave functions which transform according to these irreducible representations can be obtained systematically by using **projection operators** (see any one of the references on group theory for further details). In many simple cases this can be done by inspection also. The proper linear combinations of the [111] wave functions can be shown to be (see Problem 2.4) those in Table 2.9.

Similarly one can show that the six degenerate $\{(\pm)200\}$, $\{0(\pm)20\}$ and $\{00(\pm)2\}$ wave functions form a six-dimensional representation whose characters are given in Table 2.8. Using these characters one can decompose this six-

Table 2.10. Symmetrized nearly free electron wave functions in a zinc-blende crystal with wave vectors equal to $(2\pi/a)(\pm 2, 0, 0)$, $(2\pi/a)(0, \pm 2, 0)$, and $(2\pi/a)(0, 0, \pm 2)$

Representation	Wave function
Γ_1	$\cos(4\pi x/a) + \cos(4\pi y/a) + \cos(4\pi z/a)$
Γ_3	$\cos(4\pi y/a) - \cos(4\pi z/a)$; $\cos(4\pi x/a) - (1/2)[\cos(4\pi y/a) + \cos(4\pi z/a)]$
Γ_4	$\sin(4\pi x/a)$; $\sin(4\pi y/a)$; $\sin(4\pi z/a)$

dimensional representation into the direct sum $\Gamma_1 \oplus \Gamma_3 \oplus \Gamma_4$. The symmetrized wave functions are given in Table 2.10, while the proof is left as an exercise (Problem 2.4).

d) Selection Rules

In atomic physics one learns that optical transitions obey **selection rules** such as: in an electric-dipole transition the orbital angular momentum can change only by ± 1. These selection rules result from restrictions imposed on matrix elements of the **electric-dipole operator** [see (6.29,30)] by the rotational symmetry of the atomic potential. One may expect similar selection rules to result from the symmetry of potentials in crystals. To see how such selection rules can be derived, we will consider the following example.

Let $\boldsymbol{p}$ be the electron momentum operator and Ψ_1 be a wave function in a zinc-blende-type crystal with the point group T_d. Since $\boldsymbol{p}$ is a vector its three components p_x, p_y, and p_z belong to the irreducible representation T_2 (Γ_{15} or Γ_4 according to Table 2.4). Let us assume that Ψ_1 is a triply degenerate wave function belonging to T_2 also. Operating with $\boldsymbol{p}$ on Ψ_1 results in a set of nine wave functions, which we will label Ψ_3. These nine wave functions generate a nine-dimensional reducible representation which can be reduced to the direct sum $T_1 \oplus T_2 \oplus E \oplus A_1$ as shown in (2.16). Next we form the matrix element $M = \langle \Psi_2 | \boldsymbol{p} | \Psi_1 \rangle = \langle \Psi_2 | \Psi_3 \rangle$ between Ψ_3 and another wave function Ψ_2. Suppose Ψ_2 belongs to an irreducible representation B which is *not* one of the irreducible representations in the direct sum $T_1 \oplus T_2 \oplus E \oplus A_1$ of the wave function Ψ_3. From the orthogonality of the basis functions for different irreducible representations one concludes that the matrix element M is zero. In general, it can be proved that the matrix element between an operator p and two wave functions Ψ_1 and Ψ_2 can differ from zero only when the direct product of the representations of p and Ψ_1 contains an irreducible representation of Ψ_2. This important group theoretical result is known as the **matrix-element theorem**.

When applied to atoms the matrix-element theorem leads to the familiar selection rules for electric-dipole transitions. For instance, if Ψ_1 and Ψ_2 are atomic wave functions they will have definite parities under inversion, and the parity of their direct product is simply the product of their parities. The electric-dipole operator has odd parity so its matrix element is zero between two states of the same parity according to the matrix-element theorem. If Ψ_1 and Ψ_2 both have

s symmetry then their direct product also has *s* symmetry. Since the dipole operator has *p*-symmetry its matrix element between two *s* states is zero. On the other hand if one of these two wave functions has *p* symmetry its direct product will contain a component with *p* symmetry, and the electric-dipole transition will be nonzero. Thus, application of the matrix-element theorem leads to selection rules for optical transitions in systems with spherical symmetry.

Using the matrix-element theorem we can also obtain very general selection rules for optical transitions in zinc-blende-type and diamond-type crystals. In Chap. 6 we will show that electric-dipole transitions in a crystal are determined by the matrix element of the electron momentum operator $\boldsymbol{p}$. In a zinc-blende-type crystal $\boldsymbol{p}$ belongs to the Γ_4 irreducible representations. To derive the selection rules for optical transitions involving zone-center wave functions we need to know the direct product between Γ_4 and all the irreducible representations of T_d. The results are summarized in Table 2.11.

From this table we can easily determine whether or not electric-dipole transitions between any two bands at the zone center of the zinc-blende crystal are allowed. For example, dipole transitions from a Γ_4 valence band to conduction bands with Γ_1, Γ_3, Γ_4, and Γ_5 symmetries are all allowed. Using Table 2.11 one can derive selection rules for optical excitation of phonons by photons in the infrared (to be discussed further in Chap. 6). The ground state of the crystal with no phonons should have Γ_1 symmetry. In zinc-blende crystals only Γ_4 optical phonons can be directly excited by an infrared photon via an electric-dipole transition. Such phonons are said to be **infrared-active**. On the other hand the $\Gamma_{25'}$ optical phonon of the diamond structure is not infrared-active because of the parity selection rule (Ge, Si, and diamond are highly transparent in the infrared!). The ionic momentum operator has symmetry Γ_{15} for the O_h group and odd parity under the operation i' of the diamond crystal. Hence electric-dipole transitions can only connect states with opposite parity.

Selection rules for higher order optical processes, such as Raman scattering can also be obtained from Table 2.11. As will be shown in Chap. 7, Raman scattering involves the excitation of a phonon via two optical transitions. If both optical transitions are of the electric-dipole type in a zinc-blende crystal, the excited phonon must belong to one of the irreducible representations of the direct product $\Gamma_4 \otimes \Gamma_4 = \Gamma_4 \oplus \Gamma_5 \oplus \Gamma_3 \oplus \Gamma_1$. Phonons which can be excited optically in Raman scattering are said to be **Raman-active**. Thus the Γ_4 optical phonon in the zinc-blende crystal is Raman-active in addition to being

Table 2.11. Direct products of the Γ_4 representation with all the representations of T_d

Direct product	Direct sum
$\Gamma_4 \otimes \Gamma_1$	Γ_4
$\Gamma_4 \otimes \Gamma_2$	Γ_5
$\Gamma_4 \otimes \Gamma_3$	$\Gamma_4 \oplus \Gamma_5$
$\Gamma_4 \otimes \Gamma_4$	$\Gamma_4 \oplus \Gamma_5 \oplus \Gamma_3 \oplus \Gamma_1$
$\Gamma_4 \otimes \Gamma_5$	$\Gamma_4 \oplus \Gamma_5 \oplus \Gamma_3 \oplus \Gamma_2$

infrared-active. Similarly, the symmetries of Raman-active phonons in crystals with the O_h point group can be shown to be $\Gamma_{25'}$, Γ_{12}, and Γ_1 (see Chap. 7 for further details). Hence the $\Gamma_{25'}$ optical phonon of the diamond structure, while not infrared-active, is Raman-active. In crystals with inversion symmetry (said to be **centrosymmetric**), an infrared-active phonon must be odd while a Raman-active phonon must be even under inversion, therefore a phonon cannot be both infrared-active and Raman-active in such crystals.

2.4 Empty Lattice or Nearly Free Electron Energy Bands

We now apply the group theoretical notations to the electron energy band structure of the diamond- and zinc-blende-type semiconductors. Since the electrons move in the presence of a crystal potential, their wave functions can be symmetrized to reflect the crystal symmetry, i. e., written in a form such that they belong to irreducible representations of the space group of the crystal. However, in order to highlight the symmetry properties of the electron wave function, we will assume that the crystal potential is vanishingly small. In this **empty lattice** or **nearly free electron model**, the energy and wave functions of the electron are those of a free particle as given by (2.18) and (2.17), respectively. The electron energy band is simply a parabola when plotted in the extended zone scheme. This parabola looks much more complicated when replotted in the reduced zone scheme. It looks especially intimidating when the wave functions are labeled according to the irreducible representations of the point group of the crystal. Such complications have resulted from using the crystal symmetry which was supposed to simplify the problem! The simplification, however, occurs when we consider the band structure of electrons in a non-empty lattice in the remaining sections of this chapter. In this section we will use group theory to analyze the symmetry properties of nearly free electron band structures in both zinc-blende- and diamond-type crystals.

2.4.1 Nearly Free Electron Band Structure in a Zinc-Blende Crystal

Figure 2.8 shows the energy band of a nearly free electron plotted in the reduced zone scheme for wave vectors along the [111] and [100] directions only. To analyze this band diagram we will consider the symmetry and wave functions at a few special high-symmetry points in reciprocal space.

$\boldsymbol{k} = (0, 0, 0)$

As pointed out in Sect. 2.3.4, the group of the $\boldsymbol{k}$ vector at the Γ point is always isomorphic to the point group of the lattice. Since the wave function is a constant for $k = 0$, it has the symmetry Γ_1.

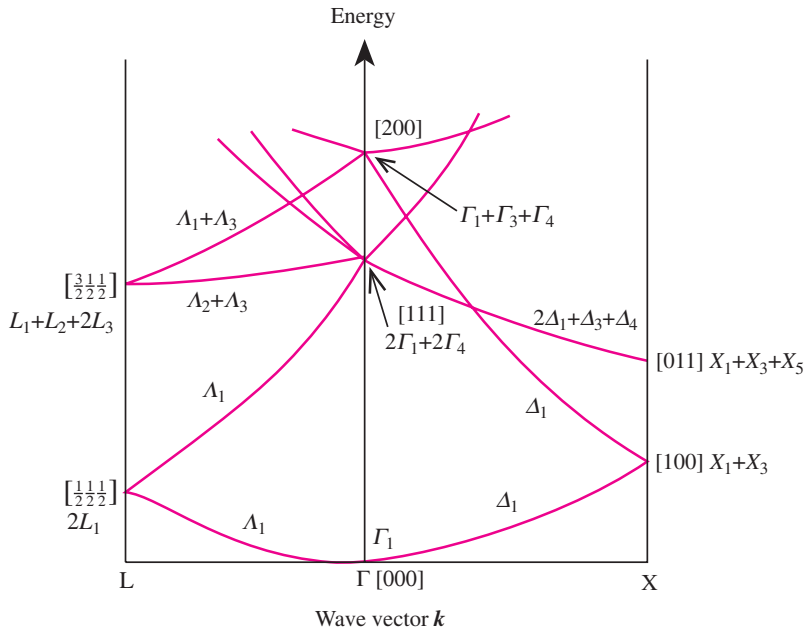


Fig. 2.8. Band structure of nearly free electrons in a zinc-blende-type crystal in the reduced zone scheme. The numbers in *square brackets* denote corresponding reciprocal lattice vectors in the extended zone scheme in units of $(2\pi/a)$, a being the size of the unit cube. *Note*: To conform to notations used in the literature, we will use + instead of $\oplus$ to represent the direct sum of two representations in all figures

$\boldsymbol{k} = (b, b, b)$, $b \neq (\pi/a)$

In the eight equivalent [111] directions the bands are labeled Λ, according to the Brillouin zone notations in Fig. 2.2c. The wave functions for $\boldsymbol{k} \neq 0$ are classified according to the group of the wave vector $\boldsymbol{k}$. The group of a wave vector along the [111] direction inside the Brillouin zone is C_{3v} and contains six elements divided into three classes:

$\{E\}$: identity;
$\{C_3, C_3^{-1}\}$: two three-fold rotations about the [111] direction;
$\{m_1, m_2, m_3\}$: three reflections in the three equivalent (110) planes containing the [111] axis.

The characters and basis functions for the irreducible representations of Λ are summarized in Table 2.12.

The free-electron wave function given by $\exp[(i\pi\zeta/a)(x + y + z)]$, where $0 < \zeta < 1$, is invariant under all the symmetry operations of the group of Λ so it belongs to the Λ_1 representation. We can also obtain this symmetry of the electron wave function by using the so-called **compatibility relations**. The symmetry of the endpoints of an axis in the Brillouin zone is higher than or equal to that of a point on the axis. Therefore the group of a point on an axis

Table 2.12. Characters and basis functions of the irreducible representations of the group of Λ (C_{3v}) in a zinc-blende-type crystal

	$\{E\}$	$\{2C_3\}$	$\{3m\}$	Basis functions
Λ_1	1	1	1	1 or $x+y+z$
Λ_2	1	1	−1	$xy(x-y)+yz(y-z)+zx(z-x)$
Λ_3	2	−1	0	$\{(x-y); \sqrt{\frac{2}{3}}(z-\frac{1}{2}[x+y])\}$

is either equal to or constitutes a subgroup of the group of the endpoints. In the latter case, a representation belonging to the group of the endpoints of an axis can be reduced to irreducible representations of the group of the axis. The procedure for this reduction is the same as that described in Sect. 2.3.4. The difference is that only symmetry operations common to both groups need be considered now. When a representation of the group of an axis is contained in a representation of one of the group's endpoints, the two representations are said to be **compatible**. For points lying on the [111] axis of a zinc-blende-type crystal, the group of Λ is a subgroup of Γ but is identical to the group of L. From the character tables for Γ and Λ it is clear that Γ_1 is compatible with Λ_1 only. Thus when the band starts out at the zone center with symmetry Γ_1, the symmetry of the band along the [111] direction must be Λ_1. This case illustrates a rather trivial application of the compatibility relations. Compatibility relations provide very useful consistency checks on band-structure calculations. Further applications of the compatibility relations can be found in Problem 2.6 at the end of this chapter.

$\boldsymbol{k} = (\pi/a)(1,1,1)$

In the zinc-blende structure the symmetry operations in the group of the L point are identical to those of the Λ axis. So the Λ_1 representation is compatible with the L_1 representation only. For free electrons the wave function is doubly degenerate at $(\pi/a)(1,1,1)$ since $(\pi/a)(1,1,1)$ and $(-\pi/a)(1,1,1)$ differ by $(2\pi/a)(1,1,1)$, a reciprocal lattice vector of the zinc-blende structure. Using the compatibility relations one can show that the next higher energy band along the Λ axis also has Λ_1 symmetry.

$\boldsymbol{k} = (2\pi/a)(1,1,1)$

The point $\boldsymbol{k} = (2\pi/a)(1,1,1)$ is equivalent to Γ since it differs from Γ by a reciprocal lattice vector. As shown in Sect. 2.3.4, the eight degenerate wave functions of the form $\exp[(\mathrm{i}2\pi/a)(\pm x \pm y \pm z)]$ can be symmetrized into two wave functions with Γ_1 symmetry and two sets of three wave functions with Γ_4 symmetry. The symmetries of the higher energy bands in the [111] direction are given in Fig. 2.8. They can be deduced using Table 2.12 and checked by the compatibility relations. The reader is urged to verify this as an exercise.

Table 2.13. Symmetry operations and classes of the group of Δ (C_{2v}) in the zinc-blende structure

Class	Symmetry operations
$\{E\}$	xyz
$\{C_4^2\}$	$x\bar{y}\bar{z}$
$\{m_d\}$	xyz
$\{m_d'\}$	$x\bar{z}\bar{y}$

Table 2.14. Characters of the irreducible representations of the group of Δ (C_{2v})

	$\{E\}$	$\{C_4^2\}$	$\{m_d\}$	$\{m_d'\}$
Δ_1	1	1	1	1
Δ_2	1	1	−1	−1
Δ_3	1	−1	1	−1
Δ_4	1	−1	−1	1

$\boldsymbol{k} = (c, 0, 0)$, $c \neq (2\pi/a)$

Wave vectors in the [100] and equivalent directions are denoted by Δ. The group of Δ (C_{2v}) contains four elements divided into the four classes listed in Table 2.13. The irreducible representations and characters of the group of Δ are summarized in Table 2.14. The symmetry of the wave function in the [100] direction is Δ_1 since this is the only representation compatible with Γ_1. The Δ axis ends at the X point on the surface of the Brillouin zone.

$\boldsymbol{k} = (2\pi/a)(1, 0, 0)$

The group of X contains twice as many symmetry operations as the group of Δ since the wave vectors $(2\pi/a)(1, 0, 0)$ and $(2\pi/a)(-1, 0, 0)$ differ by the reciprocal lattice vector $(2\pi/a)(2, 0, 0)$. The eight elements of the group of X (D_{2d}) are divided into five classes:

$\{E\}$: identity;
$\{C_4^2(x)\}$: two-fold rotation about the x axis;
$\{2C_4^2(y, z)\}$: two-fold rotations about the y and z axes;
$\{2S_4\}$: two four-fold improper rotations about the x axis;
$\{2m_d\}$: two mirror reflections on the [011] and [0$\bar{1}$1] planes.

The irreducible representations of the group of X and their characters are given in Table 2.15. The wave functions at the X point with $\boldsymbol{k} = (2\pi/a)(\pm 1, 0, 0)$ are doubly degenerate in the nearly-free electron model. From the compatibility relations it can be found that these wave functions belong to either the X_1 or X_3 representations.

Table 2.15. Characters of the irreducible representations of the group of X (D_{2d}) in the zinc-blende structure

	$\{E\}$	$\{C_4^2(x)\}$	$\{2C_4^2(y,z)\}$	$\{2S_4\}$	$\{2m_d\}$
X_1	1	1	1	1	1
X_2	1	1	1	−1	−1
X_3	1	1	−1	−1	1
X_4	1	1	−1	1	−1
X_5	2	−2	0	0	0

$\boldsymbol{k} = (2\pi/a)(0,0,2)$

The points $\boldsymbol{k} = (2\pi/a)(\pm 2,0,0)$, $(2\pi/a)(0,\pm 2,0)$, and $(2\pi/a)(0,0,\pm 2)$ differ from the zone center by reciprocal lattice vectors. As already shown in example (c) in Sect. 2.3.4, the six degenerate wave functions

$$\exp[\pm \mathrm{i}4\pi x/a];\ \exp[\pm \mathrm{i}4\pi y/a];\quad \text{and}\quad \exp[\pm \mathrm{i}4\pi z/a]$$

can be symmetrized to transform like the Γ_1, Γ_3, and Γ_4 irreducible representations.

2.4.2 Nearly Free Electron Energy Bands in Diamond Crystals

Obviously, the band structure of a free electron is the same whether it is in a zinc-blende or a diamond crystal. Therefore, in order to obtain the symmetrized wave functions specific to the diamond structure, we have to assume first that the diamond crystal potential is nonzero and symmetrize the electron wave functions accordingly. Afterwards the crystal potential is made to approach zero. The band structure of nearly free electrons in a diamond-type crystal obtained in this way is shown in Fig. 2.9. It serves as an important guide to the band structure of Si (shown in Fig. 2.10 for comparison) calculated by more sophisticated techniques to be discussed later in this chapter.

The symmetries of the bands in diamond are very similar to those of zinc-blende because both crystals have a fcc lattice and tetrahedral symmetry. However, there are also important differences resulting from the existence of glide planes in the diamond structure as discussed in Sect. 2.3.2. We pointed out in that section that, if we choose the origin at one of the carbon atoms in diamond, the crystal is invariant under all the symmetry operations of the point group T_d plus three "glide-like" operations: $T(1/4,1/4,1/4)\sigma_x$, $T(1/4,1/4,1/4)\sigma_y$, and $T(1/4,1/4,1/4)\sigma_z$ (for brevity, we will now denote these three operations as $T\sigma_x$, $T\sigma_y$, and $T\sigma_z$, respectively). However, the factor group of the space group of diamond is isomorphic to the point group O_h. In symmetrizing the electronic wave functions in the diamond structure, one has to consider the effect of $T\sigma_x$ on the Bloch functions. In this subsection we shall pay special attention to the electron wave functions at the points Γ, L, and X of the Brillouin zone of the diamond crystal.

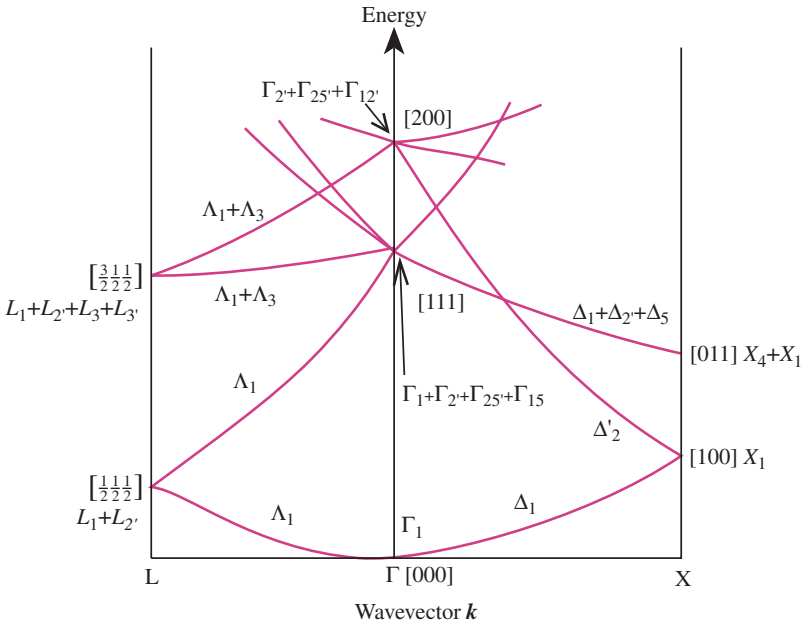


Fig. 2.9. Band structure of nearly free electrons for a diamond-type crystal in the reduced zone scheme

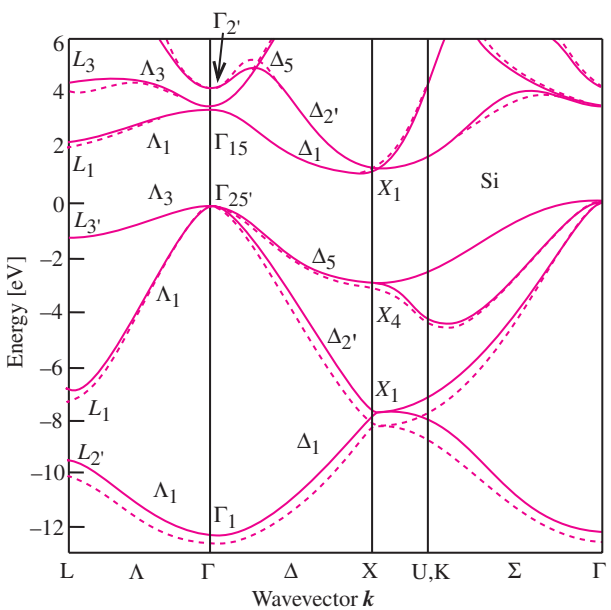


Fig. 2.10. Electronic band structure of Si calculated by the pseudopotential technique. The *solid* and the *dotted lines* represent calculations with a **nonlocal** and a **local pseudopotential**, respectively. [Ref. 2.8, p. 81]

$\boldsymbol{k} = (0, 0, 0)$

From (2.8) the Bloch functions at the zone center can be written as $u(\boldsymbol{r})$, where u has the periodicity of the lattice. We define C as a set formed from the group T_d plus all the operations obtained by multiplying each element of T_d by $T\sigma_x$. C defined this way is not a group because operations involving the glide, such as $(T\sigma_x)^2 = T(0, 1/2, 1/2)$ are not a member of C [for brevity, we will denote the operation $T(0, 1/2, 1/2)$ by Q]. Let us now generate a set $\{Cu\}$ consisting of 48 functions by applying the operations of C to $u(\boldsymbol{r})$. For any two symmetry operations, a and b of C we define the operation multiplication between the corresponding two elements au and bu in $\{Cu\}$ as $(au)(bu) = (ab)u$. The set of operations in $\{Cu\}$, defined by their effect on the function u, can be easily shown to form a group. In particular, $Qu(\boldsymbol{r}) = u(\boldsymbol{r})$ because $u(\boldsymbol{r})$ has the translational symmetry of the crystal and hence Qu is now an element of $\{Cu\}$. In this group, it is convenient to introduce the element $i'u$ where $i' = T(1/4, 1/4, 1/4)i$ was introduced in Sect. 2.3.4 (i is the inversion operation with respect to the origin). As pointed out in Sect. 2.3.4, the diamond crystal is not invariant under inversion with respect to one of the carbon atoms; it is however invariant under the combined operation of inversion followed by the translation $T(1/4, 1/4, 1/4)$. One can show that the 48 operations in $\{Cu\}$ are isomorphic to the O_h group. The character table of the group of wave functions of Γ is given in Table 2.16. It can be compared with the character table for the O_h group (Table 2.5). Note that the classes are listed in different orders in Tables 2.5 and 2.16. In Table 2.16 the five classes of symmetry operations in the point group T_d are listed first. The remaining five classes are obtained by multiplying the T_d operations by i'.

The effects of i' on the symmetry of wave functions at different high-symmetry points of the diamond crystal are not the same. For example, points along Λ are not invariant under i', so their symmetries are the same as in the zinc-blende crystal. On the other hand, the L point is invariant under i', therefore the wave functions at L have definite parity under i'.

Table 2.16. Characters of the irreducible representations of the group of Γ in the diamond structure. The notation is that of Koster (BSW notation in parentheses)

	$\{E\}$	$\{C_2\}$	$\{S_4\}$	$\{\sigma_d\}$	$\{C_3\}$	$\{i'\}$	$\{i'C_2\}$	$\{i'S_4\}$	$\{i'\sigma_d\}$	$\{i'C_3\}$
$\Gamma_1^+(\Gamma_1)$	1	1	1	1	1	1	1	1	1	1
$\Gamma_2^+(\Gamma_2)$	1	1	−1	−1	1	1	1	−1	−1	1
$\Gamma_3^+(\Gamma_{12})$	2	2	0	0	−1	2	2	0	0	−1
$\Gamma_4^+(\Gamma_{15'})$	3	−1	1	−1	0	3	−1	1	−1	0
$\Gamma_5^+(\Gamma_{25'})$	3	−1	−1	1	0	3	−1	−1	1	0
$\Gamma_1^-(\Gamma_{1'})$	1	1	−1	−1	1	−1	−1	1	1	−1
$\Gamma_2^-(\Gamma_{2'})$	1	1	1	1	1	−1	−1	−1	−1	−1
$\Gamma_3^-(\Gamma_{12'})$	2	2	0	0	−1	−2	−2	0	0	1
$\Gamma_4^-(\Gamma_{15})$	3	−1	−1	1	0	−3	1	1	−1	0
$\Gamma_5^-(\Gamma_{25})$	3	−1	1	−1	0	−3	1	−1	1	0

$\boldsymbol{k} = (\pi/a)(1,1,1)$

The group of the L point in the diamond structure is isomorphic to the group of L in the fcc Bravais lattice (that is, a crystal formed by putting only one atom at each lattice point of a fcc lattice). The characters and basis functions for the irreducible representations in the group of L (D_{3d}) are shown in Table 2.17.

$\boldsymbol{k} = (2\pi/a)(1,1,1)$

It has been pointed out already in the case of the zinc-blende crystal that $\boldsymbol{k} = (2\pi/a)(1,1,1)$ is equivalent to Γ. This is, of course, also true for the diamond crystal. From the eight symmetrized wave functions for the zinc-blende crystal given in Table 2.9 it can be shown readily that, for the diamond crystal, the eight equivalent (111) wave functions are symmetrized to transform according to the irreducible representations in Table 2.18.

The symmetry of the wave functions in the diamond structure along the [001] directions are quite different from those of zinc-blende. We will first consider the X point since it presents an especially interesting case.

Table 2.17. Characters and basis functions for the irreducible representations of the group of L in the diamond structure

	$\{E\}$	$\{2C_3\}$	$\{3C_2\}$	$\{i'\}$	$\{2i'C_3\}$	$\{3i'C_2\}$	Basis functions
L_1	1	1	1	1	1	1	1
L_2	1	1	−1	1	1	−1	$xy(x^2-y^2)+yz(y^2-z^2)+zx(z^2-x^2)$
L_3	2	−1	0	2	−1	0	$\{z^2-1/2(x^2+y^2);\,(x^2-y^2)\}$
$L_{1'}$	1	1	1	−1	−1	−1	$(x-y)(y-z)(z-x)$
$L_{2'}$	1	1	−1	−1	−1	1	$x+y+z$
$L_{3'}$	2	−1	0	−2	1	0	$\{(x-z);(y-1/2[x+z])\}$

Table 2.18. Symmetrized nearly free electron wave functions in the diamond crystal with wave vectors equal to $(2\pi/a)(\pm1,\ \pm1,\ \pm1)$. The origin of coordinates has been taken to coincide with an atomic site.

Representation	Wave function
$\Gamma_1^+\,(\Gamma_1)$	$\cos(2\pi x/a)\cos(2\pi y/a)\cos(2\pi z/a)$ $+\sin(2\pi x/a)\sin(2\pi y/a)\sin(2\pi z/a)$
$\Gamma_2^-\,(\Gamma_{2'})$	$\cos(2\pi x/a)\cos(2\pi y/a)\cos(2\pi z/a)$ $-\sin(2\pi x/a)\sin(2\pi y/a)\sin(2\pi z/a)$
$\Gamma_5^+\,(\Gamma_{25'})$	$\sin(2\pi x/a)\cos(2\pi y/a)\cos(2\pi z/a)$ $+\cos(2x/a)\sin(2\pi y/a)\sin(2\pi z/a)$; plus two cyclic permutations
$\Gamma_4^-\,(\Gamma_{15})$	$\sin(2\pi x/a)\cos(2\pi y/a)\cos(2\pi z/a)$ $-\cos(2x/a)\sin(2\pi y/a)\sin(2\pi z/a)$; plus two cyclic permutations

$\boldsymbol{k} = (2\pi/a)(0,0,1)$

A very special property of the wave functions at the X point of the diamond structure is that all *relevant* irreducible representations of the group of the X point are doubly degenerate, but they do not have definite parity under i'. To understand this peculiar property, let us first enumerate all the symmetry operations of the group of the X point. We will start with the eight symmetry operations of the group of the point $(2\pi/a)(0,0,1)$ in the Brillouin zone of the zinc-blende structure:

$$\{E,\ C_4^2(z),\ 2C_4^2(x,y),\ 2S_4,\ 2m_d\}.$$

Next we will consider the combined effect of these operations and the operation $T\sigma_z$ on a wave function at the X point:

$$\phi = \exp(\mathrm{i}2\pi z/a)u(\boldsymbol{r}). \tag{2.19}$$

At first we may expect that we can construct a group for the X point by taking the above eight elements and adding to them their products with the operation $T\sigma_z$. This should result in a set of sixteen elements. It turns out that these sixteen elements do not form a group because translation and rotation do not necessarily commute. For example, consider the combined effect of $C_4^2(x)T\sigma_z$ on a vector (x,y,z):

$$(x,y,z) \overset{\sigma_z}{\rightarrow} (x,y,\overline{z}) \xrightarrow{T(1/4,1/4,1/4)} (x+1/4a, y+1/4a, -z+1/4a)$$
$$\xrightarrow{C_4^2(x)} (x+1/4a, -y-1/4a, z-1/4a).$$

If we interchange the order of $C_4^2(x)$ and $T\sigma_z$ we find that

$$[T\sigma_z C_4^2(x)](x,y,z) = (x+1/4a, -y+1/4a, z+1/4a), \tag{2.20a}$$

so the operation C_4^2 does not commute with $T\sigma_z$. In particular,

$$[T\sigma_z C_4^2(x)]\phi = T(0,1/2,1/2)[C_4^2]T\sigma_z\phi = Q[C_4^2]T\sigma_z\phi. \tag{2.20b}$$

In order that the set $\{E\phi, C_4^2(z)\phi, \ldots, 2m_d\phi, T\sigma_z\phi, \ldots, T\sigma_z 2m_d\phi\}$ forms a group, the operation Q has to be included also. Taking the 16 operations mentioned above and their products with Q, a group with 32 elements is obtained. This group can be divided into 14 classes:

$$\begin{aligned}
C_1 &= \{E\}\\
C_2 &= \{C_4^2(x),\ C_4^2(y),\ QC_4^2(x),\ QC_4^2(y)\}\\
C_3 &= \{C_4^2(z)\}\\
C_4 &= \{QT\sigma_z\sigma_x,\ T\sigma_z\sigma_y\}\\
C_5 &= \{T\sigma_z S_4,\ T\sigma_z S_4^{-1},\ QT\sigma_z S_4,\ QT\sigma_z S_4^{-1}\}\\
C_6 &= \{T\sigma_z,\ QT\sigma_z\}\\
C_7 &= \{T\sigma_z C_4^2(x),\ T\sigma_z C_4^2(y),\ QT\sigma_z C_4^2(x),\ QT\sigma_z C_4^2(y)\}\\
C_8 &= \{T\sigma_z C_4^2(z),\ QT\sigma_z C_4^2(z)\}\\
C_9 &= \{\sigma_x,\ \sigma_y\}\\
C_{10} &= \{S_4,\ S_4^{-1},\ QS_4,\ QS_4^{-1}\}\\
C_{11} &= \{Q\sigma_x,\ Q\sigma_y\}
\end{aligned}$$

$C_{12} = \{QT\sigma_z\sigma_y,\ T\sigma_z\sigma_x\}$
$C_{13} = \{QC_4^2(z)\}$
$C_{14} = \{Q\}$

The characters of the corresponding 14 irreducible representations are given in Table 2.19. However, not all of these representations are acceptable for wave functions at the X point of the Brillouin zone in the diamond crystal. Since $(a/2)(0,1,1)$ is a lattice vector of the fcc lattice the operation Q will leave the periodic part of the X-point wave function invariant. The sinusoidal envelope $\exp(i2\pi z/a)$ of the Bloch function changes sign under the translation Q, so overall the X-point wave functions must be odd under Q. Of the 14 irreducible representations only four are odd under the translation Q (or C_{14}). These are labeled X_1, X_2, X_3, and X_4 in Table 2.19. The interesting point is that these four representations are all doubly degenerate. This degeneracy results from the glide reflection and the fact that the two atoms in the unit cell of the diamond structure are identical. The degeneracy in the X_1 and X_2 states is lifted in the zinc-blende structure, where the two atoms in the primitive cell are different (see Problem 2.8). Some examples of symmetrized wave functions at the X point are

$\boldsymbol{k} = (2\pi/a)(0,0,1)$:
$X_1 : \{\cos(2\pi z/a);\ \sin(2\pi z/a)\}$

$\boldsymbol{k} = (2\pi/a)(\pm1,\pm1,0)$:
$X_1 : \{\cos(2\pi x/a)\ \cos(2\pi y/a);\ \sin(2\pi x/a)\sin(2\pi y/a)\}$
$X_4 : \{\sin(2\pi x/a)\ \cos(2\pi y/a);\ \cos(2\pi x/a)\sin(2\pi y/a)\}$

$\boldsymbol{k} = (\xi\pi/a)(0,0,1)$ where $0 < \xi < 2$.

Table 2.19. Irreducible representations and characters of the group of symmetry operations on the wave functions at the X point $(2\pi/a)(0,0,1)$ of the Brillouin zone of the diamond structure [Ref. 2.7, p. 162]

	C_1	$4C_2$	C_3	$2C_4$	$4C_5$	$2C_6$	$4C_7$	$2C_8$	$2C_9$	$4C_{10}$	$2C_{11}$	$2C_{12}$	C_{13}	C_{14}
M_1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M_2	1	1	1	−1	−1	1	1	1	−1	−1	−1	−1	1	1
M_3	1	−1	1	−1	1	1	−1	1	−1	1	−1	−1	1	1
M_4	1	−1	1	1	−1	1	−1	1	1	−1	1	1	1	1
M_5	2	0	−2	0	0	2	0	−2	0	0	0	0	−2	2
M_1'	1	1	1	1	1	−1	−1	−1	−1	−1	−1	1	1	1
M_2'	1	1	1	−1	−1	−1	−1	−1	1	1	1	−1	1	1
M_3'	1	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1	1
M_4'	1	−1	1	1	−1	−1	1	−1	−1	1	−1	1	1	1
M_5'	2	0	−2	0	0	−2	0	2	0	0	0	0	−2	2
X_1	2	0	2	0	0	0	0	0	2	0	−2	0	−2	−2
X_2	2	0	2	0	0	0	0	0	−2	0	2	0	−2	−2
X_3	2	0	−2	2	0	0	0	0	0	0	0	−2	2	−2
X_4	2	0	−2	−2	0	0	0	0	0	0	0	2[a]	2	−2

[a] An error in [2.7] has been corrected.

We will denote the Bloch function along the Δ direction as $\psi = \exp(i\xi\pi z/a)u(\boldsymbol{r})$ as in (2.19). It is invariant under the following space group operations of the diamond structure:

$$\{E,\ \sigma_x,\ C_x^2(z),\ \sigma_y,\ T\sigma_z C_4^2(x),\ T\sigma_z S_4,\ T\sigma_z C_4^2(y),\ T\sigma_z S_4^{-1}\},$$

which can be divided into 5 classes:

$$\begin{aligned} C_1 &= \{E\} \\ C_2 &= \{C_4^2(z)\} \\ C_3 &= \{T\sigma_z S_4,\ T\sigma_z S_4^{-1}\} \\ C_4 &= \{\sigma_x,\ \sigma_y\} \\ C_5 &= \{T\sigma_z C_4^2(x),\ T\sigma_z C_4^2(y)\} \end{aligned}$$

The representations generated by these operations acting on ψ are isomorphic with the group of Δ for a cubic lattice. The corresponding characters are shown in Table 2.20.

Table 2.20. Irreducible representations and characters of the group of symmetry operations on the wave functions at the Δ point of the Brillouin zone of the diamond structure [Ref. 2.7, p. 158]

	C_1	C_2	$2C_3$	$2C_4$	$2C_5$
Δ_1	1	1	1	1	1
Δ_2	1	1	−1	1	−1
$\Delta_{2'}$	1	1	−1	−1	1
$\Delta_{1'}$	1	1	1	−1	−1
Δ_5	2	−2	0	0	0

2.5 Band Structure Calculations by Pseudopotential Methods

In Fig. 2.10 we have shown the electronic band structure of Si. It has been calculated with a sophisticated method known as the **pseudopotential technique**, which will be discussed in this section. Comparing these results with the nearly free electron band structure in Fig. 2.9 we notice that there are many similarities between the two. The nearly free electron band structure is basically a parabola redrawn in the reduced zone scheme. In the other case the band structure is computed by large-scale numerical calculations using supercomputers. The question is now: why do the two band structures, obtained by completely different methods, look so similar qualitatively? The answer to this question lies in the concept of pseudopotentials.

The electronic configuration of a Si atom is $1s^2 2s^2 2p^6 3s^2 3p^2$. When Si atoms form a crystal we can divide their electrons into core electrons and valence electrons as pointed out in Sect. 2.1. In crystalline Si the $1s$, $2s$, and $2p$

orbitals are completely occupied and form the core shells. The outer 3*s* and 3*p* shells are only partially filled. Electrons in these shells are called valence electrons because they are involved in bonding with neighboring Si atoms. The crystal structure of Si at ambient pressure is similar to that of diamond. The tetrahedral arrangement of bonds between a Si atom and its four nearest neighbors can be understood if one of the electrons in the 3*s* shell is "promoted" to the 3*p* shell so that the four valence electrons form **hybridized sp^3 orbitals**. This sp^3 hybridization is well known from the bonding of carbon atoms and is responsible for the tetrahedral structure found in many organic molecules. But carbon atoms are more versatile than silicon atoms in that they can form double and triple bonds also. As a result, carbon atoms are crucial to all known forms of life while silicon atoms are important only to the highest form of life, namely human beings. It is these valence electrons in the outermost shells of a Si atom that are nearly free. These electrons are not affected by the full nuclear charge as a result of screening of the nucleus by the filled core shells. In the core region the valence electron wave functions must be orthogonal to those of the core. Thus the true wave functions may have strong spatial oscillations near the core, which make it difficult to solve the wave equation. One way to overcome this difficulty is to divide the wave functions into a smooth part (the **pseudo-wave function**) and an oscillatory part. The kinetic energy from the latter provides an "effective repulsion" for the valence electrons near the core (alternatively one can regard the valence electrons as being expelled from the core due to Pauli's exclusion principle). Thus we can approximate the strong true potential by a weaker "effective potential" or **pseudopotential** for the valence electrons. Since the "smooth" parts of the valence electron wave functions have little weight in the core region, they are not very sensitive to the shape of the pseudopotential there. Figure 2.11 shows qualitatively how the pseudopotential in Si varies with distance *r* from the nucleus. At large values of *r* the pseudopotential approaches the unscreened Coulomb potential of the Si^{4+} ion. This concept of replacing the true potential with a pseudopotential can be justified mathematically. It can

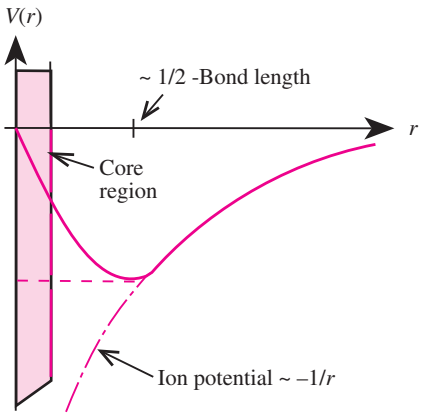


Fig. 2.11. Schematic plot of the atomic pseudopotential of Si in real space [Ref. 2.8, p. 17]. The *solid curve* in which $V(r) \rightarrow 0$ in the core region is said to be a "soft core" pseudopotential. The *broken curve* in which $V(r) \rightarrow$ constant is a "hard core" pseudopotential

be shown to reproduce correctly both the conduction and valence band states while eliminating the cumbersome, and in many cases irrelevant, core states [Ref. 2.8, p. 16].

Using the pseudopotential concept, the one-electron Schrödinger equation (2.4) can be replaced by the **pseudo-wave-equation**

$$\left[\frac{p^2}{2m} + V(\boldsymbol{r}_i)\right]\psi_{\boldsymbol{k}}(\boldsymbol{r}_i) = E_{\boldsymbol{k}}\psi_{\boldsymbol{k}}(\boldsymbol{r}_i), \tag{2.21}$$

where ψ is the pseudo-wave-function. This function is a good approximation to the true wave function outside the core region and therefore can be used to calculate the physical properties of the semiconductors which are dependent on the valence and conduction electrons only. Since pseudopotentials are weak perturbations on the free-electron band structure, a good starting point for diagonalizing (2.21) is to expand $\psi_{\boldsymbol{k}}$ as a sum of plane waves:

$$\psi_{\boldsymbol{k}} = \sum_{\boldsymbol{g}} a_{\boldsymbol{g}} |\boldsymbol{k} + \boldsymbol{g}\rangle, \tag{2.22}$$

where the vectors $\boldsymbol{g}$ are the reciprocal lattice vectors and $|\,\boldsymbol{k}\,\rangle$ represents a plane wave with wave vector $\boldsymbol{k}$. The coefficients $a_{\boldsymbol{g}}$ and the eigenvalues $E_{\boldsymbol{k}}$ can be determined by solving the secular equation

$$\det\,|\,[(\hbar^2k^2/2m) - E_{\boldsymbol{k}}]\delta_{\boldsymbol{k},\,\boldsymbol{k}+\boldsymbol{g}} + \langle \boldsymbol{k}\,|\,V(\boldsymbol{r})\,|\,\boldsymbol{k} + \boldsymbol{g}\rangle\,|\,| = 0, \tag{2.23}$$

The matrix elements of the pseudopotential $V(\boldsymbol{r})$ are given by

$$\langle \boldsymbol{k}\,|\,V(\boldsymbol{r})\,|\,\boldsymbol{k} + \boldsymbol{g}\rangle = \left[\frac{1}{N}\sum_{\boldsymbol{R}} \exp(-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{R})\right]\frac{1}{\Omega}\int_{\Omega} V(\boldsymbol{r})\exp[-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}]d\boldsymbol{r}, \tag{2.24}$$

where $\boldsymbol{R}$ is a direct lattice vector and Ω the volume of a primitive cell. As a result of summation over all the lattice vectors inside the bracket, the pseudopotential matrix element is zero unless $\boldsymbol{g}$ is a reciprocal lattice vector. In other words, the matrix elements of the pseudopotential are determined by Fourier components of the pseudopotential ($V_{\boldsymbol{g}}$) defined by

$$V_{\boldsymbol{g}} = \frac{1}{\Omega}\int_{\Omega} V(\boldsymbol{r})\exp[-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}]d\boldsymbol{r}, \tag{2.25}$$

where $\boldsymbol{g}$ is a reciprocal lattice vector.

If there is only one atom per primitive cell these Fourier components of the pseudopotential are known as the **pseudopotential form factors**. When there are several different atoms in the primitive cell, it is convenient to define for each kind of atom a pseudopotential form factor and a structure factor which depends only on the positions of one particular kind of atom in the primitive cell. For example, let there be two kinds of atoms α and β in the crystal and let their positions inside the primitive cell be denoted by $\boldsymbol{r}_{\alpha i}$ and

$\boldsymbol{r}_{\beta i}$. The **structure factor** $S_{\boldsymbol{g}\alpha}$ of atom α is defined as

$$S_{\boldsymbol{g}\alpha} = \frac{1}{N_\alpha} \sum_i \exp(-\mathrm{i}\boldsymbol{g} \cdot \boldsymbol{r}_{\alpha i}), \tag{2.26}$$

where N_α is the number of α atoms in the primitive cell. The structure factor of atom β is defined similarly. The pseudopotential form factor $V_{\boldsymbol{g}\alpha}$ for atom α can be defined as in (2.25) except that V is now the potential of one α atom and the integration is performed over Ω_α, which is the volume corresponding to one α atom. The pseudopotential $V(\boldsymbol{r})$ can be expressed in terms of the structure and form factors by

$$V(\boldsymbol{r}) = \sum_{\boldsymbol{g}} (V_{\boldsymbol{g}\alpha} S_{\boldsymbol{g}\alpha} + V_{\boldsymbol{g}\beta} S_{\boldsymbol{g}\beta}) \exp(\mathrm{i}\boldsymbol{g} \cdot \boldsymbol{r}). \tag{2.27}$$

From (2.24) we conclude that the pseudopotential mixes the free-electron states whose $\boldsymbol{k}$'s differ by a reciprocal lattice vector. If these states are degenerate, the degeneracy may be split by the pseudopotential provided the corresponding form factor is nonzero. For example, consider the free-electron states with $\boldsymbol{k} = (2\pi/a)(\pm 1, \pm 1, \pm 1)$ at the Γ point in the diamond structure. The $\boldsymbol{k}$'s of these eight-fold degenerate states differ by reciprocal lattice vectors $(2\pi/a)(2,0,0)$, $(2\pi/a)(2,2,0)$, and $(2\pi/a)(2,2,2)$. These eight states are degenerate when the electron is free. With the introduction of the pseudopotential, they become coupled and their degeneracy is partly lifted, producing energy gaps (compare Figs. 2.9 and 2.10). When an energy gap opens up at the Fermi level (highest occupied energy level) a semiconductor is obtained. This opening of energy gaps in the nearly-free-electron band structure by the pseudopotential form factors can be explained by Bragg reflection of the free-electron plane waves by the crystal potential with the formation of standing waves. When the pseudopotential form factors are small, their effect on the band structure is weak so the actual band structure is not too different from the free-electron band structure. This is the reason why the nearly free electron bands drawn in the reduced zone scheme are a good starting point for understanding the band structure of most semiconductors.

2.5.1 Pseudopotential Form Factors in Zinc-Blende- and Diamond-Type Semiconductors

The main reason why pseudopotentials are so useful is because only a small number of these form factors are sufficient for calculating a band structure. In semiconductors with the diamond structure, such as Si and Ge, just three pseudopotential form factors are sufficient. In semiconductors with the zinc-blende structure the number of required pseudopotential form factors doubles to six. To show this we first note that there are two atoms a and b in the unit cell. We will denote the **atomic pseudopotentials** of these two atoms by $V_\mathrm{a}(r - r_\mathrm{a})$ and $V_\mathrm{b}(r - r_\mathrm{b})$, where r_a and r_b are the positions of the two atoms in the unit cell. Substituting these potentials into (2.25) we obtain the Fourier

components of the crystal pseudopotential

$$V_{\boldsymbol{g}} = \frac{1}{\Omega}\int [V_{\mathrm{a}}(\boldsymbol{r}-\boldsymbol{r}_{\mathrm{a}}) + V_{\mathrm{b}}(\boldsymbol{r}-\boldsymbol{r}_{\mathrm{b}})]\ \exp[-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}]d\boldsymbol{r} \tag{2.28}$$

$$= \frac{1}{\Omega}\int [V_{\mathrm{a}}(\boldsymbol{r})\exp(-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}_{\mathrm{a}}) + V_{\mathrm{b}}(\boldsymbol{r})\exp(-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}_{\mathrm{b}})] \times \exp[-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}]d\boldsymbol{r}. \tag{2.29}$$

Without loss of generality we can take the midpoint between the two atoms in the unit cell as the origin, so that $\boldsymbol{r}_{\mathrm{a}} = (a/8)(1,1,1) = \boldsymbol{s}$ and $\boldsymbol{r}_{\mathrm{b}} = (-a/8)(1,1,1) = -\boldsymbol{s}$. We can now write

$$V_{\mathrm{a}}(\boldsymbol{r})\exp(-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}_{\mathrm{a}}) + V_{\mathrm{b}}(\boldsymbol{r})\exp(-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}_{\mathrm{b}}) = (V_{\mathrm{a}}+V_{\mathrm{b}})\cos(\boldsymbol{g}\cdot\boldsymbol{s}) - \mathrm{i}(V_{\mathrm{a}}-V_{\mathrm{b}})\sin(\boldsymbol{g}\cdot\boldsymbol{s}). \tag{2.30}$$

Next we define the symmetric and antisymmetric components of the pseudopotential form factor by

$$V_{\boldsymbol{g}}^{\mathrm{s}} = \frac{1}{\Omega}\int (V_{\mathrm{a}}+V_{\mathrm{b}})\exp(-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r})d\boldsymbol{r} \tag{2.31}$$

and

$$V_{\boldsymbol{g}}^{\mathrm{a}} = \frac{1}{\Omega}\int (V_{\mathrm{a}}-V_{\mathrm{b}})\exp(-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r})d\boldsymbol{r}. \tag{2.32}$$

Substituting the results in (2.30–32) back into (2.29) we arrive at

$$V_{\boldsymbol{g}} = V_{\boldsymbol{g}}^{\mathrm{s}}\cos(\boldsymbol{g}\cdot\boldsymbol{s}) - \mathrm{i}V_{\boldsymbol{g}}^{\mathrm{a}}\sin(\boldsymbol{g}\cdot\boldsymbol{s}). \tag{2.33}$$

By symmetrizing the pseudopotential form factors in this way, it is clear that the antisymmetric form factors $V_{\boldsymbol{g}}^{\mathrm{a}}$ vanish in the diamond structure. The factor $\cos(\boldsymbol{g}\cdot\boldsymbol{s})$ is just the structure factor of diamond defined in (2.26). In the III–V semiconductor, where the difference in the potentials of the anion and cation is small, $V_{\boldsymbol{g}}^{\mathrm{a}}$ is expected to be smaller than $V_{\boldsymbol{g}}^{\mathrm{s}}$ and furthermore $V_{\boldsymbol{g}}^{\mathrm{s}}$ should be almost the same as in their neighboring group–IV semiconductors. For example, consider the pseudopotential form factors in Ge and the III–V semiconductor GaAs formed from its neighbors in the periodic table. In the diamond and zinc-blende structures, the reciprocal lattice vectors in order of increasing magnitude are (in units of $2\pi/a$):

$$\begin{aligned}
\boldsymbol{g}_0 &= (0,0,0);\\
\boldsymbol{g}_3 &= (1,1,1),\ (1,-1,1),\ \ldots,\ (-1,-1,-1);\\
\boldsymbol{g}_4 &= (2,0,0),\ (-2,0,0),\ \ldots,\ (\ 0,\ \ 0,-2);\\
\boldsymbol{g}_8 &= (2,2,0),\ (2,-2,0),\ \ldots,\ (\ 0,-2,-2);\\
\boldsymbol{g}_{11} &= (3,1,1),\ (-3,1,1),\ \ldots,\ (-3,-1,-1).
\end{aligned}$$

We can neglect pseudopotential form factors with $g^2 > 11(2\pi/a)^2$ because typically $V_{\boldsymbol{g}}$ decreases as g^{-2} for large $\boldsymbol{g}$. Figure 2.12 shows a schematic plot of a pseudopotential as a function of the magnitude of $\boldsymbol{g}$ ($\boldsymbol{g}$ is assumed to be spherically symmetrical as in the case of a free atom).

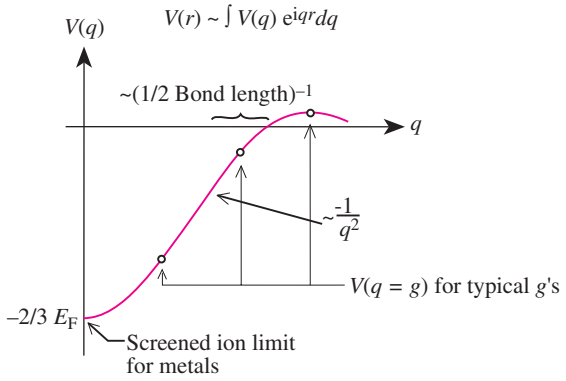


Fig. 2.12. Schematic plot of a pseudopotential form factor in reciprocal space [Ref. 2.8, p. 21]

The pseudopotential form factor V_0 corresponding to $\boldsymbol{g}_0$ is a constant potential, which merely shifts the entire energy scale: it can therefore be set equal to zero or any other convenient value (see Fig. 2.12). The pseudopotential form factors for all the equivalent reciprocal lattice vectors with the form $(\pm1, \pm1, \pm1)$ and magnitude $3(2\pi/a)$ are equal by symmetry and will be denoted by V_3. The structure factor corresponding to $\boldsymbol{g}_4$ is zero because $\cos(\boldsymbol{g}\cdot\boldsymbol{s}) = 0$ for $\boldsymbol{g} = (2\pi/a)(2, 0, 0)$. Thus we conclude that there are only three important pseudopotential form factors for Ge: V_3^{s}, V_8^{s} and V_{11}^{s}. In GaAs, V_8^{a} vanishes because $\sin(\boldsymbol{g}\cdot\boldsymbol{s}) = 0$, so only six pseudopotential form factors are required: V_3^{s}, V_8^{s}, V_{11}^{s}, V_3^{a}, V_4^{a}, and V_{11}^{a}. The pseudopotential form factors of Ge, GaAs, and a few other semiconductors are listed in Table 2.21. One should keep in mind that the sign of the antisymmetric form factors depends on whether the anion or cation is designated as atom a. The sign of the antisymmetric form factors in Table 2.21 are all positive because the cation has been chosen to be atom a and the anion (which has a more negative atomic pseudopotential) to be atom b. Note also that the magnitude of the form factor V_3^{s} is the largest and furthermore it is negative in sign, as shown schematically in Fig. 2.12. In the III–V and II–VI compounds, V_3^{s} is comparable to the corresponding V_3^{s} in the group-IV semiconductors

Table 2.21. Pseudopotential form factors of several group-IV, III–V and II–VI semiconductors (in units of Rydbergs = 13.6 eV) [2.9,10]. Note that the sign of $V_{\mathrm{i}}^{\mathrm{a}}$ depends on the positions chosen for the anion and the cation (see text).

	V_3^{s}	V_8^{s}	V_{11}^{s}	V_3^{a}	V_4^{a}	V_{11}^{a}
Si	−0.211	0.04	0.08	0	0	0
Ge	−0.269	0.038	0.035	0	0	0
GaAs	−0.252	0	0.08	0.068	0.066	0.012
GaP	−0.249	0.017	0.083	0.081	0.055	0.003
InAs	−0.27	0.02	0.041	0.078	0.038	0.036
InSb	−0.25	0.01	0.044	0.049	0.038	0.01
ZnSe	−0.23	0.01	0.06	0.18	0.12	0.03
CdTe	−0.245	−0.015	0.073	0.089	0.084	0.006

and always larger than V_3^a in magnitude. Nevertheless, as the ionicity increases in going from the III–V semiconductors to the II–VI semiconductors, the antisymmetric pseudopotential form factors become larger. Some band structures of diamond- and zinc-blende-type semiconductors calculated by the pseudopotential method are shown in Figs. 2.13–15. These band structure calculations include the effect of **spin–orbit coupling**, which will be discussed in Sect. 2.6. As a result of this coupling, the irreducible representations of the electron wave functions must include the effects of symmetry operations on the spin wave function. (For example, a rotation by 2π will change the sign of the wave function of a spin-1/2 particle). The notations used in Figs. 2.13–15, including this feature, are known as the **double group notations** and will be discussed in Sect. 2.6.

The effect of ionicity on the band structures of the compound semiconductors can be seen by comparing the band structure of Ge with those of GaAs and ZnSe as shown in Figs. 2.13–15. Some of the differences in the three band structures result from spin–orbit coupling. Otherwise most of these differences can be explained by the increase in the antisymmetric components of the pseudopotential form factors as the ionicity increases along the sequence Ge, GaAs, ZnSe. One consequence of this increase in ionicity is that the energy gap between the top of the valence band and the bottom of the conduc-

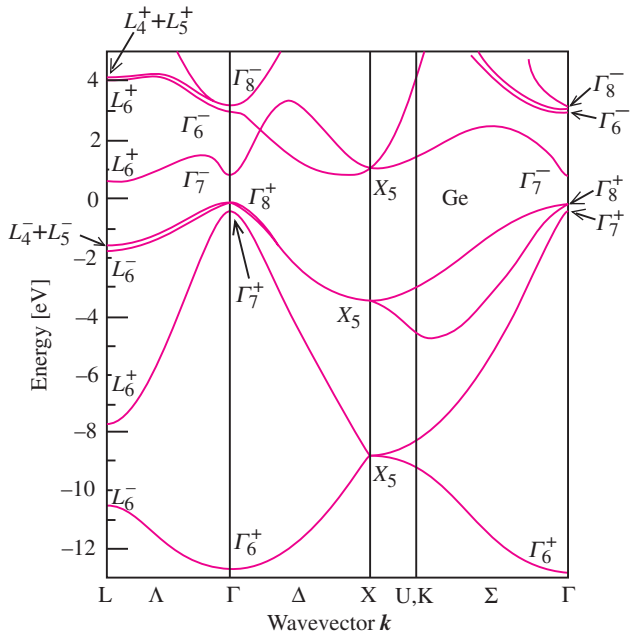


Fig. 2.13. Electronic band structure of Ge calculated by the pseudopotential technique. The energy at the top of the filled valence bands has been taken to be zero. Note that, unlike in Fig. 2.10, the double group symmetry notation is used [Ref. 2.8, p. 92]

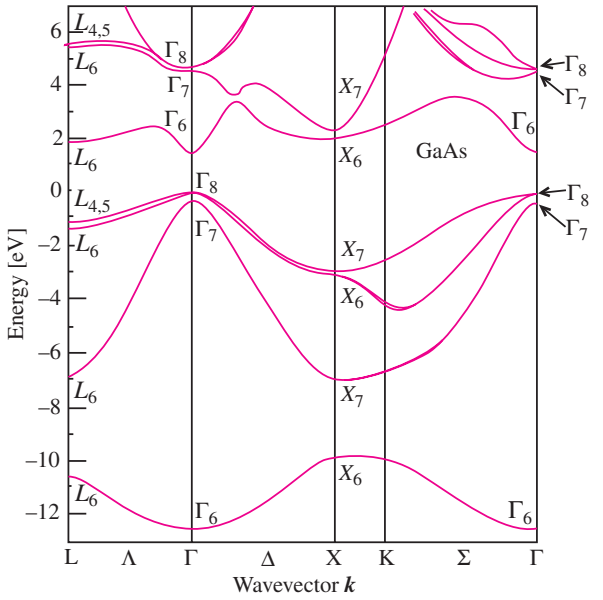


Fig. 2.14. Electronic band structure of GaAs calculated by the pseudopotential technique. The energy scale and notation (double group) are similar to those for Fig. 2.13 [Ref. 2.8, p. 103]

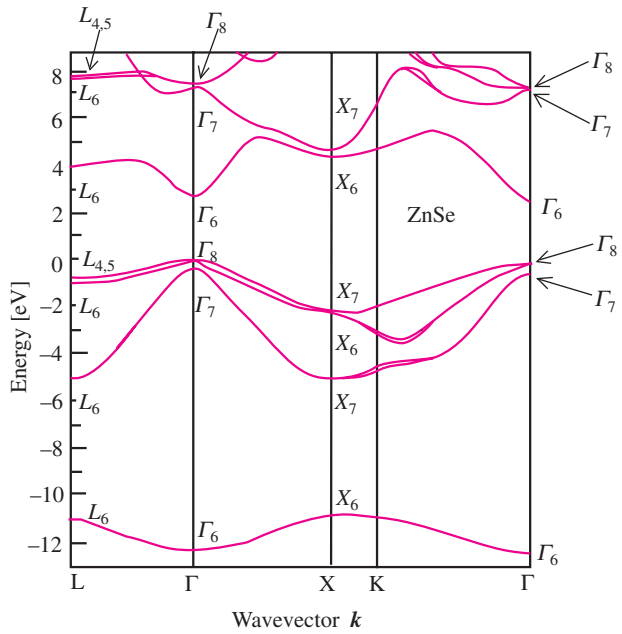


Fig. 2.15. Electronic band structure of ZnSe calculated by the pseudopotential technique. The energy scale and notation (double group) are similar to those for Fig. 2.14 [Ref. 2.8, p. 113]

tion band at Γ increases monotonically in going from Ge to ZnSe. Another consequence is that some of the doubly degenerate states in Ge at the X point of the Brillouin zone are split in the III–V and II–VI compounds, as pointed out in Sect. 2.4.2. For example, the lowest energy X_1 conduction band state in Ge is split into two spin doublets of X_6 and X_7 symmetry (X_1 and X_3 without spin–orbit coupling) in GaAs and ZnSe. The explicit dependence of this splitting on the antisymmetric pseudopotential form factors is calculated in Problem 2.8.

2.5.2 Empirical and Self-Consistent Pseudopotential Methods

There are two approaches to calculating pseudopotential form factors. Since the number of relevant pseudopotential form factors is small, they can be determined by fitting a small number of experimental data, such as the position of peaks in optical reflectivity spectra (Chap. 6) or features in the photoelectron spectra (Chap. 8). This approach is known as the **Empirical Pseudopotential Method** (EPM). The flow diagram for calculating the band structure with the EPM is as follows:

$$V_{\boldsymbol{g}}$$
$$\downarrow$$
$$V(\boldsymbol{r}) = \sum_{\boldsymbol{g}} V_{\boldsymbol{g}} \exp(-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r})$$
$$\downarrow$$
$$H = (p^2/2m) + V(\boldsymbol{r})$$

Solve $H\psi_{\boldsymbol{k}}(\boldsymbol{r}) = E_{\boldsymbol{k}}\psi_{\boldsymbol{k}}(\boldsymbol{r})$ to obtain $\psi_{\boldsymbol{k}}(\boldsymbol{r})$ and $E_{\boldsymbol{k}}$

$$\downarrow$$

Calculate reflectivity, density of states, etc., and compare with experiments

$$\downarrow$$

Alter $V_{\boldsymbol{g}}$ if agreement between theory and experiment is not satisfactory.

The disadvantage of the EPM is that it requires experimental inputs. However, this is not a major disadvantage since atomic pseudopotential form factors are often "transferable" in the sense that once they are determined in one compound they can be used (sometimes after suitable interpolation) in other compounds containing the same atom. For example, the atomic pseudopotential form factors for Ga determined empirically from GaAs can be used to calculate the band structure of other Ga compounds such as GaSb and GaP. With the availability of high-speed computers, however, it is possible to determine the pseudopotential form factors from first principles without any experimental input. These first-principles pseudopotential methods are known as **self-consistent** or ***ab initio* pseudopotential methods**. These meth-

ods use atomic pseudopotentials and a model for the crystal structure (from which an ionic potential V_{ion} can be constructed) as the starting point of the calculation. After the wave functions have been obtained the contribution of the valence electrons to the potential is calculated. It is then used to evaluate the total one-electron potential, which is compared with the starting potential. Self-consistency is achieved when the calculated one-electron potential agrees with the starting potential. The flow diagram for such a calculation is shown below. The **exchange and correlation** term V_{xc}, which takes into account the many-body effects, is usually calculated with approximations such as the **Local Density Approximation** (LDA)[2]. In this approximation, V_{xc} is assumed to be a function of the local charge density only. The LDA gives good results for the ground state properties such as the cohesive energies and charge density of the valence electrons. However, it gives poor results for the excitation energies. For instance, it typically underestimates the fundamental energy gap by about 1 eV. Thus it predicts semiconductors like Ge to be semimetals. The band structures shown in Figs. 2.10 and 2.13–15 have been calculated with the EPM since this method gives better overall agreement with experiments. This shortcoming of the LDA can be overcome by many-body techniques such as the **quasiparticle approach** [2.13].

Choose $V(r)$

$\downarrow$

Solve $(H + V)\psi = E\psi$

$\downarrow$

Calculate charge density $\varrho = \psi^*\psi$

$\downarrow$

Solve $\nabla^2 V_{\text{Hartree}} = 4\pi\varrho \left(\frac{1}{4\pi\varepsilon_0}\right)$

$\downarrow$

Calculate $V_{\text{xc}} = f[\varrho(r)]$

$\downarrow$

$V_{\text{sc}} = V_{\text{Hartree}} + V_{\text{xc}}$

$\downarrow$

Model structure $V_{\text{ion}} \rightarrow V = V_{\text{sc}} + V_{\text{ion}}$

In recent years the *ab initio* pseudopotential method has been refined so as to be able to handle *semicore* electrons such as the 3d electrons of the copper halides. The pseudopotentials used are very smooth near the core (*ultrasoft pseudopotentials*) and reduce the number of plane waves required for the expansion of the wavefunctions to converge. The method is particularly useful for CuCl and diamond [2.14].

[2] For his development of the local density functional method of calculating electronic structures W. Kohn was awarded the Nobel prize for Chemistry in 1998 [2.11, 12].

2.6 The $k \cdot p$ Method of Band-Structure Calculations

The pseudopotential method is not the only method of band structure calculation which requires a small number of input parameters obtainable from experimental results. In the empirical pseudopotential method the inputs are usually energy gaps. In optical experiments one typically determines both energy gaps and oscillator strengths of the transitions. Thus it can be an advantage if the optical matrix elements can also be used as inputs in the band structure calculation. In the $\boldsymbol{k} \cdot \boldsymbol{p}$ method the band structure over the entire Brillouin zone can be extrapolated from the zone center energy gaps and optical matrix elements. The $\boldsymbol{k} \cdot \boldsymbol{p}$ method is, therefore, particularly convenient for interpreting optical spectra. In addition, using this method one can obtain analytic expressions for band dispersion and effective masses around high-symmetry points.

The $\boldsymbol{k} \cdot \boldsymbol{p}$ method can be derived from the one-electron Schrödinger equation given in (2.4). Using the Bloch theorem the solutions of (2.4) are expressed, in the reduced zone scheme, as

$$\Phi_{n\boldsymbol{k}} = \exp(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r}) u_{n\boldsymbol{k}}(\boldsymbol{r}), \tag{2.34}$$

where n is the band index, $\boldsymbol{k}$ lies within the first Brillouin zone, and $u_{n\boldsymbol{k}}$ has the periodicity of the lattice. When $\Phi_{n\boldsymbol{k}}$ is substituted into (2.4) we obtain an equation in $u_{n\boldsymbol{k}}$ of the form[3]

$$\left(\frac{p^2}{2m} + \frac{\hbar \boldsymbol{k} \cdot \boldsymbol{p}}{m} + \frac{\hbar^2 k^2}{2m} + V \right) u_{n\boldsymbol{k}} = E_{n\boldsymbol{k}} u_{n\boldsymbol{k}}. \tag{2.35}$$

At $\boldsymbol{k}_0 = (0,0,0)$, (2.35) reduces to

$$\left(\frac{p^2}{2m} + V \right) u_{n\boldsymbol{0}} = E_{n\boldsymbol{0}} u_{n\boldsymbol{0}} \quad (n = 1,2,3,\ldots). \tag{2.36}$$

Similar equations can also be obtained for $\boldsymbol{k}$ equal to any point $\boldsymbol{k}_0$. Equation (2.36) is much easier to solve than (2.4) since the functions $u_{n\boldsymbol{0}}$ are periodic. The solutions of (2.36) form a complete and orthonormal set of basis functions. Once $E_{n\boldsymbol{0}}$ and $u_{n\boldsymbol{0}}$ are known, we can treat the terms $\hbar \boldsymbol{k} \cdot \boldsymbol{p}/m$ and $\hbar^2 k^2/(2m)$ as perturbations in (2.35) using either **degenerate** or **nondegenerate perturbation theory**. This method for calculating the band dispersion is known as the $\boldsymbol{k} \cdot \boldsymbol{p}$ **method**. Since the perturbation terms are proportional to k, the method works best for small values of k [2.15]. In general, the method can be applied to calculate the band dispersion near any point $\boldsymbol{k}_0$ by expanding (2.35) around $\boldsymbol{k}_0$ provided the wave functions (or the matrix elements of $\boldsymbol{p}$ between these wave functions) and the energies at $\boldsymbol{k}_0$ are known. Furthermore, by using a sufficiently large number of $u_{n\boldsymbol{0}}$ to approximate a complete set of basis

[3] Equation (2.35) is rigorously valid only if V is a *local* potential, i.e., it depends only on one spatial coordinate r. This is not strictly true in the case of pseudopotential [2.8]

functions, (2.35) can be diagonalized with the help of computers to calculate the band structure over the entire Brillouin zone [2.16]. Only a limited number of energy gaps and matrix elements of $\boldsymbol{p}$ determined experimentally are used as input in the calculation.

As examples of application of the $\boldsymbol{k}\cdot\boldsymbol{p}$ method we will derive the band dispersion and effective mass for a nondegenerate band and for a three-fold degenerate (or nearly degenerate) p-like band. The nondegenerate band case is applicable to the conduction band minimum in direct-bandgap semiconductors with the zinc-blende and wurtzite structures (examples of the latter semiconductor are CdS and CdSe). The nearly degenerate band is a model for the top valence bands in many semiconductors with the diamond, zinc-blende, or wurtzite structures.

2.6.1 Effective Mass of a Nondegenerate Band Using the $k\cdot p$ Method

Let us assume that the band structure has an extremum at the energy $E_{n\boldsymbol{0}}$ and the band is nondegenerate at this energy. Using standard nondegenerate perturbation theory, the eigenfunctions $u_{n\boldsymbol{k}}$ and eigenvalues $E_{n\boldsymbol{k}}$ at a neighboring point $\boldsymbol{k}$ can be expanded to second order in k in terms of the unperturbed wave functions $u_{n\boldsymbol{0}}$ and energies $E_{n\boldsymbol{0}}$ by treating the terms involving k in (2.35) as perturbations.

$$u_{n\boldsymbol{k}} = u_{n\boldsymbol{0}} + \frac{\hbar}{m}\sum_{n'\neq n}\frac{\langle u_{n\boldsymbol{0}}\,|\,\boldsymbol{k}\cdot\boldsymbol{p}\,|\,u_{n'\boldsymbol{0}}\rangle}{E_{n\boldsymbol{0}} - E_{n'\boldsymbol{0}}}u_{n'\boldsymbol{0}} \tag{2.37}$$

and

$$E_{n\boldsymbol{k}} = E_{n\boldsymbol{0}} + \frac{\hbar^2k^2}{2m} + \frac{\hbar^2}{m^2}\sum_{n'\neq n}\frac{|\langle u_{n\boldsymbol{0}}\,|\,\boldsymbol{k}\cdot\boldsymbol{p}\,|\,u_{n'\boldsymbol{0}}\rangle|^2}{E_{n\boldsymbol{0}} - E_{n'\boldsymbol{0}}}. \tag{2.38}$$

The linear terms in k vanish because $E_{n\boldsymbol{0}}$ has been assumed to be an extremum. It is conventional to express the energy $E_{n\boldsymbol{k}}$, for small values of k, as

$$E_{n\boldsymbol{k}} = E_{n\boldsymbol{0}} + \frac{\hbar^2k^2}{2m^*}, \tag{2.39}$$

where m^* is defined as the **effective mass** of the band. Comparing (2.38) and (2.39) we obtain an expression for this effective mass:

$$\frac{1}{m^*} = \frac{1}{m} + \frac{2}{m^2k^2}\sum_{n'\neq n}\frac{|\langle u_{n\boldsymbol{0}}\,|\,\boldsymbol{k}\cdot\boldsymbol{p}\,|\,u_{n'\boldsymbol{0}}\rangle|^2}{E_{n\boldsymbol{0}} - E_{n'\boldsymbol{0}}}. \tag{2.40}$$

Formula (2.40) can be used to calculate the effective mass of a nondengenerate band. Also it shows that an electron in a solid has a mass different from that of a free electron because of coupling between electronic states in different bands via the $\boldsymbol{k}\cdot\boldsymbol{p}$ term. The effect of neighboring bands on the effective mass of a band depends on two factors.

- A wave function $u_{n'\mathbf{0}}$ can couple to u_{n0} only if the matrix element $\langle u_{n'\mathbf{0}} \,|\, \boldsymbol{p} \,|\, u_{n\mathbf{0}} \rangle$ is nonzero. In Sect 2.3.4 we pointed out that, using the matrix element theorem and group theory, it is possible to enumerate all the symmetries $u_{n'\mathbf{0}}$ can have. For example, $\boldsymbol{p}$ has Γ_4 symmetry in the zinc-blende structure. If the conduction band has Γ_1 symmetry, as in GaAs, its effective mass will be determined only by coupling with bands having Γ_4 symmetry. On the other hand a valence band with Γ_4 symmetry can be coupled via $\boldsymbol{p}$ to bands with Γ_1, Γ_3, Γ_4, and Γ_5 symmetries.
- The energy separation $E_{n'\mathbf{0}} - E_{n\mathbf{0}}$ between the two bands n and n' determines the relative importance of the contribution of n' to the effective mass of n. Furthermore, bands with energies less than $E_{n\mathbf{0}}$ will contribute a positive term to $1/m^*$, making m^* smaller than the free electron mass. Conversely, bands with energies higher than $E_{n\mathbf{0}}$ tend to increase m^* or even cause m^* to become *negative* as in the case of the top valence bands in the diamond- and zinc-blende-type semiconductors.

These two simple results can be used to understand the trend in the conduction band effective mass m_c^* in many of the group-III–V and II–VI semiconductors with direct bandgaps. In these semiconductors the lowest conduction band at the zone center has Γ_1 symmetry. From the above considerations, its effective mass will be determined mainly by its coupling, via the $\boldsymbol{k} \cdot \boldsymbol{p}$ term, to the nearest bands with Γ_4 symmetry. They include both valence and conduction bands. As we will show in the next section, the conduction bands in the group-IV, III–V, and II–VI semiconductors have antibonding character, while the valence bands have bonding character. What this means is that in the diamond-type structure the $\Gamma_{2'}$ (or Γ_2^-) conduction band and its nearest Γ_{15} (or Γ_4^-) conduction band both have odd parity and the momentum matrix element between them vanishes because of the parity selection rule. In III–V semiconductors, the antisymmetric pseudopotential breaks the inversion symmetry. As a result, the momentum matrix element between the Γ_1 conduction band and its nearest Γ_4 conduction band in III–V semiconductors is nonzero, but still much smaller than its momentum matrix element with the top Γ_4 valence bands [2.17]. The separation between the Γ_1 conduction band and the Γ_4 valence band is just the direct band gap E_0, so m_c^* can be approximated by

$$\frac{1}{m_c^*} = \frac{1}{m} + \frac{2\,|\langle \Gamma_{1c} \,|\, \boldsymbol{k} \cdot \boldsymbol{p} \,|\, \Gamma_{4v} \rangle|^2}{m^2 E_0 k^2}. \tag{2.41}$$

It is customary to represent the three Γ_4 wave functions as $|\,X\rangle$, $|\,Y\rangle$, and $|\,Z\rangle$. From the T_d symmetry it can be shown that the only nonzero elements of $\langle \Gamma_{1c} \,|\, \boldsymbol{k} \cdot \boldsymbol{p} \,|\, \Gamma_{4v} \rangle$ are

$$\langle X \,|\, p_x \,|\, \Gamma_1 \rangle = \langle Y \,|\, p_y \,|\, \Gamma_1 \rangle = \langle Z \,|\, p_z \,|\, \Gamma_1 \rangle = \mathrm{i}P. \tag{2.42}$$

Without loss of generality we can assume that the wave functions $|\,X\rangle$, $|\,Y\rangle$, $|\,Z\rangle$, and $|\,\Gamma_1\rangle$ are all real. Since the operator $\boldsymbol{p}$ is equal to $-\mathrm{i}\hbar\nabla$ the matrix element in (2.42) is purely imaginary and P is real. With these results (2.41) simplifies to

$$\frac{m}{m_c^*} \approx 1 + \frac{2P^2}{mE_0}. \tag{2.43}$$

It turns out that the matrix element P^2 is more or less constant for most group-IV, III–V and II–VI semiconductors, with $2P^2/m \approx 20$ eV. The reason is that the values of P^2 for these semiconductors are very close to those calculated for nearly free electron wave functions: $P = 2\pi\hbar/a_0$ (see Problem 2.9). Since E_0 is typically less than 2 eV, $2P^2/(mE_0) \gg 1$ and (2.43) further simplifies to

$$\frac{m}{m_c^*} \approx \frac{2P^2}{mE_0}. \tag{2.44}$$

In Table 2.22 we compare the values of m_c^* calculated from (2.44) with those determined experimentally for several group-IV, III–V, and II–VI semiconductors. The values of E_0 are from experiment.

Equation (2.44) can be extended to estimate the increase in m_c^* away from the band minimum (non-parabolicity) which can be qualitatively described by an increase in E_0. See problem 6.15.

Table 2.22. Experimental values of the Γ_1 conduction band effective masses in diamond- and zinc-blende-type semicondutors compared with the values calculated from (2.44) using the values of E_0 obtained from experiment [2.18]

	Ge	GaN	GaAs	GaSb	InP	InAs	ZnS	ZnSe	ZnTe	CdTe
E_0 [eV]	0.89	3.44	1.55	0.81	1.34	0.45	3.80	2.82	2.39	1.59
m_c^*/m (exp)	0.041	0.17	0.067	0.047	0.073	0.026	0.20	0.134	0.124	0.093
m_c^*/m ((2.44))	0.04	0.17	0.078	0.04	0.067	0.023	0.16	0.14	0.12	0.08

2.6.2 Band Dispersion near a Degenerate Extremum: Top Valence Bands in Diamond- and Zinc-Blende-Type Semiconductors

To apply the $\boldsymbol{k}\cdot\boldsymbol{p}$ method to calculate the band dispersion near a degenerate band extremum we consider the highest energy $\Gamma_{25'}$ (Γ_4) valence bands at the zone center of semiconductors with the diamond (zinc-blende) structure. As pointed out in the previous section, these valence band wave functions are p-like, and they will be represented by the eigenstates $|X\rangle$, $|Y\rangle$, and $|Z\rangle$. The electron **spin** is 1/2, so the spin states will be denoted by α and β to correspond to spin-up and spin-down states, respectively. In atomic physics it is well-known that the electron spin can be coupled to the **orbital angular momentum** via the **spin–orbit interaction**. The spin–orbit coupling is a relativistic effect (inversely proportional to c^2) which scales with the atomic number of the atom. Thus for semiconductors containing heavier elements, such as Ge, Ga, As, and Sb, one expects the spin–orbit coupling to be significant and

has to include it in the unperturbed Hamiltonian, in particular for states near $k = 0$. The Hamiltonian for the spin–orbit interaction is given by

$$H_{\mathrm{so}} = \frac{\hbar}{4c^2m^2}(\nabla V \times \boldsymbol{p}) \cdot \boldsymbol{\sigma}, \tag{2.45a}$$

where the components of $\boldsymbol{\sigma}$ are the **Pauli spin matrices**:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}; \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.45b}$$

(In crystals with the diamond structure the "vector" $\nabla V \times \boldsymbol{p}$ is an example of a pseudovector with symmetry $\Gamma_{15'}$). The Hamiltonian H_{so} operates on the spin wave functions so the symmetry of H_{so} should depend also on the symmetry properties of the spin matrices. As is known from quantum mechanics, spin behaves differently than classical properties of particles such as the orbital angular momentum. For example, a spatial wave function is invariant under a rotation of 2π about any axis. However, under the same rotation the spin wave functions of a spin-1/2 particle will change sign. Let us denote a rotation of 2π about a unit vector $\hat{\boldsymbol{n}}$ as $\hat{E}$ (Problem 2.10). For a spinless particle $\hat{E}$ is equal to the identity operation. For a spin-1/2 particle $\hat{E}$ is an additional symmetry operation in the point group of its spin-dependent wave function. Thus, if $\boldsymbol{G}$ is the point group of a crystal neglecting spin, then the corresponding point group including spin effects will contain $\boldsymbol{G}$ plus $\hat{E}\boldsymbol{G}$ and is therefore twice as large as $\boldsymbol{G}$. Groups containing symmetry operations of spin wave functions are known as **double groups**. It is beyond the scope of this book to treat double groups in detail. Interested readers should refer to references listed for this chapter at the end of the book [Refs. 2.4, p. 103; 2.5; 2.7, p. 258].

Although many band diagrams in this book use the double group notation (for example, Figs. 2.13–15), in most cases it is sufficient to know only the irreducible representations for the double group at the zone center (Γ point) of zinc-blende-type crystals. Since the single group of Γ in zinc-blende-type crystals contains 24 elements, one expects the double group to contain 48 elements. However, the number of classes in a double group is not necessarily twice that of the corresponding "single group". The reason is that a class C in the single group may or may not belong to the same class as $\hat{E}C$ in the double group. For example, two sets of operations C_i and $\hat{E}C_i$ belong to the same class if the point group contains a two-fold rotation about an axis perpendicular to $\hat{\boldsymbol{n}}_i$ (the rotation axis of C_i). In the case of the group of Γ of a zinc-blende-type crystal, elements in $\{3C_2\}$ and $\{3\hat{E}C_2\}$ belong to the same class in the double group. This is also true for the elements in $\{6\sigma\}$ and $\{6\hat{E}\sigma\}$. As a result, the 48 elements in the double group of Γ in zinc-blende-type crystals are divided into eight classes. These eight classes and the eight irreducible representations of the double group of Γ are listed in Table 2.23.

Instead of using Table 2.23 to symmetrize the p-like valence band wave functions in zinc-blende-type crystals including spin–orbit coupling, we will utilize their similarity to the atomic p wave functions. We recall that, in atomic physics, the orbital electronic wave functions are classified as s, p, d, etc., ac-

Table 2.23. Character table of the double group of the point Γ in zinc-blende-type semiconductors

	$\{E\}$	$\{3C_2/3\hat{E}C_2\}$	$\{6S_4\}$	$\{6\sigma/6\hat{E}\sigma\}$	$\{8C_3\}$	$\{\hat{E}\}$	$\{6\hat{E}S_4\}$	$\{8\hat{E}C_3\}$
Γ_1	1	1	1	1	1	1	1	1
Γ_2	1	1	−1	−1	1	1	−1	1
Γ_3	2	2	0	0	−1	2	0	−1
Γ_4	3	−1	−1	1	0	3	−1	0
Γ_5	3	−1	1	−1	0	3	1	0
Γ_6	2	0	$\sqrt{2}$	0	1	−2	$-\sqrt{2}$	−1
Γ_7	2	0	$-\sqrt{2}$	0	1	−2	$\sqrt{2}$	−1
Γ_8	4	0	0	0	−1	−4	0	1

cording to the orbital angular momentum l. The p states correspond to $l = 1$ and are triply degenerate. The three degenerate states can be chosen to be eigenstates of l_z, the z component of $\boldsymbol{l}$. The eigenvalues of l_z are known as the **magnetic quantum numbers** (usually denoted as m_l). For the p states $m_l = 1, 0, -1$. The wave functions of the orbital angular momentum operator are known as **spherical harmonics**. The spherical harmonics corresponding to the $l = 1$ states can be represented as (except for a trivial factor of $(x^2 + y^2 + z^2)^{-1/2}$):

$$|lm_l\rangle = \begin{cases} |1\ 1\rangle & = -(x + \mathrm{i}y)/\sqrt{2}, \\ |1\ 0\rangle & = z, \\ |1{-}1\rangle & = (x - \mathrm{i}y)/\sqrt{2}. \end{cases} \tag{2.46}$$

The spin-orbit interaction in atomic physics is usually expressed in terms of $\boldsymbol{l}$ and the spin $\boldsymbol{s}$ as

$$H_{\mathrm{so}} = \lambda \boldsymbol{l} \cdot \boldsymbol{s} \tag{2.47}$$

The constant λ is referred to as the spin–orbit coupling. The eigenfunctions of (2.47) are eigenstates of the **total angular momentum** $\boldsymbol{j} = \boldsymbol{l} + \boldsymbol{s}$ and its z component j_z. For $l = 1$ and $s = 1/2$ the eigenvalues of $\boldsymbol{j}$ can take on two possible values: $j = l + s = 3/2$ and $j = l - s = 1/2$. The eigenvalues of j_z (denoted by m_z) can take on the $2j + 1$ values $j, j-1, \ldots, -j+1, -j$. The eigenfunctions of j and j_z can be expressed as linear combinations of the eigenfunctions of the orbital angular momentum and spin (α = spin-up, β = spin-down):

$$|jm_j\rangle = \begin{cases} |3/2,\ 3/2\rangle = |1,\ 1\rangle\alpha \\ |3/2,\ 1/2\rangle = (1/\sqrt{3})(|1,\ 1\rangle\beta + \sqrt{2}\,|1,\ 0\rangle\alpha) \\ |3/2,\ -1/2\rangle = (1/\sqrt{3})(|1,\ -1\rangle\alpha + \sqrt{2}\,|1,\ 0\rangle\beta) \\ |3/2,\ -3/2\rangle = |1,\ -1\rangle\beta \end{cases} \tag{2.48}$$

$$\begin{cases} |1/2,\ 1/2\rangle = (1/\sqrt{3})(|1,\ 0\rangle\alpha - \sqrt{2}\,|1,\ 1\rangle\beta) \\ |1/2,\ -1/2\rangle = (1/\sqrt{3})(|1,\ 0\rangle\beta - \sqrt{2}\,|1,\ -1\rangle\alpha) \end{cases} \tag{2.49}$$

The spin–orbit interaction in (2.47) splits the $j = 3/2$ states in (2.48) from the $j = 1/2$ states in (2.49). This splitting Δ_0 is known as the **spin–orbit splitting** of the valence band at Γ, and in the case of the $j = 3/2$ and $j = 1/2$ states $\Delta_0 = 3\lambda/2$.

Using the atomic physics results as a guideline we can similarly symmetrize the six electronic states $|X\rangle\alpha$, $|X\rangle\beta$, $|Y\rangle\alpha$, $|Y\rangle\beta$, $|Z\rangle\alpha$, and $|Z\rangle\beta$ in the diamond- and zinc-blende-type semiconductors. First, we make use of the similarity between the p-like Γ_4 states and the atomic p states to define three "$(l = 1)$-like" states in the zinc-blende-type crystals:

$$\begin{aligned} |1,\quad 1\rangle &= -(|X\rangle + \mathrm{i}\,|Y\rangle)/\sqrt{2}, \\ |1,\quad 0\rangle &= |Z\rangle, \\ |1,-1\rangle &= (|X\rangle - \mathrm{i}\,|Y\rangle)/\sqrt{2}. \end{aligned} \tag{2.50}$$

Next we define ($j = 3/2$)-like and ($j = 1/2$)-like states in the diamond- and zinc-blende-type crystals by substituting the expressions in (2.50) into (2.48) and (2.49). From now on we will refer to these ($j = 3/2$)-like and ($j = 1/2$)-like states as the $j = 3/2$ states and $j = 1/2$ states in the case of semiconductors.

From the characters of the double group of Γ in Table 2.23, one easily concludes that the four-fold degenerate $j = 3/2$ states belong to the Γ_8 representation, since this is the only four-dimensional representation. The two-fold degenerate $j = 1/2$ states must belong to either the Γ_6 or Γ_7 representations. A way to decide between these two representations is to calculate the character of the representation matrix generated by $j = 1/2$ states under an S_4 operation. Using the result of Problem 2.10 it can be shown that the $j = 1/2$ states belong to the Γ_7 representation. As in the atomic case, the Γ_8 and Γ_7 states are split by the spin–orbit Hamiltonian in (2.45a). Typically, the magnitude of the spin–orbit splitting Δ_0 in a semiconductor is comparable to the Δ_0 of its constituent atoms. For example, semiconductors containing heavier atoms, such as InSb and GaSb, have $\Delta_0 \approx 1$ eV, which is as large as or larger than the bandgap. When the anion and cation in the compound semiconductor have different Δ_0 the anion contribution tends to be weighted more, reflecting its larger influence on the p-like valence bands. In semiconductors containing lighter atoms, such as Si and AlP, $\Delta_0 (\approx 0.05\ \mathrm{eV})$ is negligible for many purposes. The values of Δ_0 in some diamond- and zinc-blende-type semiconductors are given in Table 2.24. The values of Δ_0 in Table 2.24 are all positive, and as a result the $j = 3/2$ (Γ_8) valence band has higher energy than the $j = 1/2$ (Γ_7) valence band states (Figs. 2.13–15). In some zinc-blende-type crystals, such as CuCl, where there is a large contribution to the valence bands from the core d-electrons, Δ_0 can be negative, leading to a reversal in the ordering of the Γ_8 and Γ_7 valence bands.

In Sect. 2.3.4d it was shown that the operator $\boldsymbol{p}$ couples a state with Γ_4 symmetry to states with Γ_1, Γ_3, Γ_4, and Γ_5 symmetries. By examining the band structure of several semiconductors calculated by the pseudopotential method (Figs. 2.10, 2.13–15) we find that the bands which have the above symmetries and are close to the Γ_4 valence bands are typically the lowest

Table 2.24. Valence band parameters A and B in units of $(\hbar^2/2m)$ and $|C|^2$ in units of $(\hbar^2/2m)^2$. The spin–orbit splitting of the valence bands Δ_0 is given in units of eV. The averaged experimental [exp] and theoretical [th, obtained from A, B, C^2 with (2.69a,b)] values of the effective masses of the heavy hole (hh), light hole (lh) and spin–orbit split-off hole (so) valence bands [(2.59), p. 268] are in units of the free electron mass. [2.16, 18]

	A	B	$\lvert C\rvert^2$	Δ_0	$m_{\rm hh}/m_0$		$m_{\rm lh}/m_0$		$m_{\rm so}/m_0$	
				[eV]	exp	th	exp	th	exp	th
C[b]	−2.5	0.2	4.6	0.013[a]		0.66[e]		0.29[e]		0.39[b]
Si[e]	−4.28	−0.68	24	0.044	0.54	0.50	0.15	0.15	0.23	0.24
Ge	−13.38	−8.5	173	0.295	0.34	0.43	0.043	0.041	0.095	0.1
SiC[c]	−2.8	−1.016	5.8	0.014		0.6		0.25		0.36
GaN[d]	−5.05	−1.2	34	0.017		0.5[e]		0.13[e]		0.2
GaP[e]	−4.05	−0.98	16	0.08	0.57	0.51	0.18	0.16		0.25
GaAs	−6.9	−4.4	43	0.341	0.53	0.73	0.08	0.08	0.15	0.17
GaSb	−13.3	−8.8	230	0.75	0.8	0.98	0.05	0.04		0.15
InP[e]	−5.15	−1.9	21	0.11	0.58	0.44	0.12	0.11	0.12	0.2
InAs	−20.4	−16.6	167	0.38	0.4	0.4	0.026	0.026	0.14	0.10
InSb	−36.41	−32.5	43	0.81	0.42	0.48	0.016	0.013		0.12
ZnS	−2.54	−1.5		0.07						
ZnSe	−2.75	−1.0	7.5	0.43		1.09		0.145		
ZnTe	−3.8	−1.44	14.0	0.93						
CdTe	−4.14	−2.18	30.3	0.92						

[a] See: J. Serrano, M. Cardona, and T. Ruf, Solid State Commun. **113**, 411 (2000)
[b] See: M. Willatzen, M. Cardona, N. E. Christensen, *Linear Muffin-tin-orbital and* $\boldsymbol{k}\cdot\boldsymbol{p}$ *calculation of band structure of semiconducting diamond.* Phys. Rev. B**50**, 18054 (1994)
[c] See: M. Willatzen, M. Cardona, N. E. Christensen: *Relativistic electronic structure of 3C-SiC*. Phys. Rev. B**51**, 13150 (1995).
[d] See [1.1].
[e] The theoretical values of $m_{\rm hh}/m_{\rm o}$ and $m_{\rm lh}/m_{\rm o}$ in these materials are calculated with the equations in Problem 4.4 on p. 201 since the assumption of small warping of the valence bands is not valid for them.

conduction bands with symmetries Γ_1 and Γ_4. For the conduction band Γ_{1c} we have already shown that the only significant momentum matrix elements are $\langle X\,|\,p_x\,|\,\Gamma_1\rangle = \langle Y\,|\,p_y\,|\,\Gamma_1\rangle = \langle Z\,|\,p_z\,|\,\Gamma_1\rangle = {\rm i}P$, see (2.42). One can also use symmetry arguments to show that the nonzero matrix elements of $\boldsymbol{p}$ between the Γ_4 valence bands and the Γ_4 conduction band states are

$$\begin{aligned}&\langle X\,|\,p_y\,|\,\Gamma_{4c}(z)\rangle = \langle Y\,|\,p_z\,|\,\Gamma_{4c}(x)\rangle = \langle Z\,|\,p_x\,|\,\Gamma_{4c}(y)\rangle = {\rm i}Q,\\ &\langle X\,|\,p_z\,|\,\Gamma_{4c}(y)\rangle = \langle Y\,|\,p_x\,|\,\Gamma_{4c}(z)\rangle = \langle Z\,|\,p_y\,|\,\Gamma_{4c}(x)\rangle = {\rm i}Q\end{aligned} \tag{2.51}$$

(details of the proof are left for Problem 2.11).

The Γ_{4v} valence bands together with the Γ_{1c} and Γ_{4c} conduction bands now form a set of 14 unperturbed wave functions which are coupled together by the $\boldsymbol{k}\cdot\boldsymbol{p}$ term of (2.35). The resultant 14×14 determinant can be diagonalized either with the help of computers or by using approximations. Löwdin's perturbation method is most commonly used to obtain analytic expressions

for the dispersion of the valence bands. In this method the 14×14 matrix is divided into two parts: the wave functions of interest and their mutual interactions are treated exactly while the interaction between this group of wave functions and the remaining wave functions is treated by perturbation theory. For example, in the present case the six Γ_{4v} valence bands (including spin degeneracy) are of interest and their mutual coupling via the $\boldsymbol{k}\cdot\boldsymbol{p}$ and spin–orbit interactions will be treated exactly. The coupling between these valence band states and the conduction bands will be treated as a perturbation by defining an effective matrix element between any two valence band wave function as

$$H'_{ij} = H_{ij} + \sum_{\substack{k\neq \text{ the } \Gamma_4 \\ \text{valence bands}}} \frac{H_{ik}H_{kj}}{E_i - E_k}. \tag{2.52}$$

Within this approximation the 14×14 matrix reduces to a 6×6 matrix of the form $\{H'_{ij}\}$, where i and j run from 1 to 6. To simplify the notation we will number the six Γ_{4v} valence band wave functions as

$$\begin{aligned}
\Phi_1 &= |3/2,\ 3/2\rangle \\
\Phi_2 &= |3/2,\ 1/2\rangle \\
\Phi_3 &= |3/2,\ -1/2\rangle \\
\Phi_4 &= |3/2,\ -3/2\rangle \\
\Phi_5 &= |1/2,\ 1/2\rangle \\
\Phi_6 &= |1/2,\ -1/2\rangle
\end{aligned}$$

and the doubly degenerate Γ_{1c} and six-fold degenerate Γ_{4c} conduction band wave functions as Φ_7 to Φ_{14}.

The calculation of all the matrix elements H'_{ij} is left for Problem 2.14a. Here we will calculate only the matrix element H'_{11} as an example. According to (2.52) the effective matrix element H'_{11} is given by

$$\begin{aligned}
H'_{11} = &\left\langle \Phi_1 \left| \frac{\hbar^2 k^2}{2m} + \frac{\hbar \boldsymbol{k}\cdot\boldsymbol{p}}{m} \right| \Phi_1 \right\rangle \\
&+ \sum_j \left| \left\langle \Phi_1 \left| \frac{\hbar^2 k^2}{2m} + \frac{\hbar \boldsymbol{k}\cdot\boldsymbol{p}}{m} \right| \Phi_j \right\rangle \right|^2 \frac{1}{(E_1 - E_j)}.
\end{aligned} \tag{2.53}$$

To simplify the notation again we introduce the following symbols: E_0, energy separation between Γ_{1c} and the $j = 3/2$ valence bands; and E'_0, energy separation between Γ_{4c} and the $j = 3/2$ valence bands. Using these symbols we can express H'_{11} as

$$\begin{aligned}
H'_{11} = &\frac{\hbar^2 k^2}{2m} + \left\langle \Phi_1 \left| \frac{\hbar \boldsymbol{k}\cdot\boldsymbol{p}}{m} \right| \Phi_1 \right\rangle - \left(\left| \left\langle \Phi_1 \left| \frac{\hbar \boldsymbol{k}\cdot\boldsymbol{p}}{m} \right| \Gamma_{1c} \right\rangle \right|^2 \frac{1}{E_0} \right) \\
&- \left(\left| \left\langle \Phi_1 \left| \frac{\hbar \boldsymbol{k}\cdot\boldsymbol{p}}{m} \right| \Gamma_{4c} \right\rangle \right|^2 \frac{1}{E'_0} \right).
\end{aligned} \tag{2.54}$$

In principle, the term $\hbar\boldsymbol{k}\cdot\boldsymbol{p}/m$ can give rise to a term linear in k in the band dispersion. In the diamond-type semiconductors this term vanishes ex-

actly because of the parity selection rule. In zinc-blende-type crystals the linear $\boldsymbol{k}\cdot\boldsymbol{p}$ term can be shown to be zero within the basis used. While the k linear term is strictly zero in diamond-type crystals because of the parity selection rule, this is *not* true in crystals without a center of inversion symmetry. In zinc-blende- and wurtzite-type crystals, it has been demonstrated [2.19, 20] that both the conduction and valence bands can possess small k-linear terms. However, these k-linear terms do not come from the $\boldsymbol{k}\cdot\boldsymbol{p}$ term alone, instead they involve also spin-dependent terms which have been neglected here. Since the k linear terms are relatively unimportant for the valence bands of most semiconductors they will not be considered further here.

To simplify the notation we define

$$\begin{aligned}
L &= \frac{-\hbar^2 P^2}{m^2 E_0};\\
M &= \frac{-\hbar^2 Q^2}{m^2 E_0'};\\
N &= L + M;\\
L' &= \frac{-\hbar^2 P^2}{m^2 (E_0 + \Delta_0)};\\
M' &= \frac{-\hbar^2 Q^2}{m^2 (E_0' + \Delta_0)}.
\end{aligned}$$

With these definitions, the term

$$-\left(\left|\left\langle \Phi_1 \left| \frac{\hbar\boldsymbol{k}\cdot\boldsymbol{p}}{m} \right| \Gamma_{1c} \right\rangle\right|^2 \frac{1}{E_0}\right)$$

in (2.54) can easily be shown to be equal to

$$-\left(\left|\left\langle \Phi_1 \left| \frac{\hbar\boldsymbol{k}\cdot\boldsymbol{p}}{m} \right| \Gamma_{1c} \right\rangle\right|^2 \frac{1}{E_0}\right) = \frac{1}{2} L (k_x^2 + k_y^2) \tag{2.55}$$

while

$$-\left(\left|\left\langle \Phi_1 \left| \frac{\hbar\boldsymbol{k}\cdot\boldsymbol{p}}{m} \right| \Gamma_{4c} \right\rangle\right|^2 \frac{1}{E_0'}\right)$$

is given by

$$-\left(\left|\left\langle \Phi_1 \left| \frac{\hbar\boldsymbol{k}\cdot\boldsymbol{p}}{m} \right| \Gamma_{4c} \right\rangle\right|^2 \frac{1}{E_0'}\right) = \frac{1}{2} M (k_x^2 + k_y^2 + 2k_z^2). \tag{2.56}$$

The result is

$$H_{11}' = \frac{\hbar^2 k^2}{2m} + \frac{1}{2} N (k_x^2 + k_y^2) + M k_z^2. \tag{2.57}$$

Similarly we can show that the remaining matrix elements are

$$
\begin{aligned}
H'_{12} &= -\frac{N}{\sqrt{3}}(k_x k_z - \mathrm{i}k_y k_z) \\
H'_{13} &= -\frac{1}{2\sqrt{3}}[(L-M)(k_x^2 - k_y^2) - 2\mathrm{i}Nk_x k_y] \\
H'_{14} &= 0 \\
H'_{15} &= \frac{1}{\sqrt{2}}H'_{12} \\
H'_{16} &= -\sqrt{2}H'_{13}
\end{aligned}
$$

$$
\begin{aligned}
H'_{22} &= \frac{\hbar^2 k^2}{2m} + \frac{1}{3}(M+2L)k^2 - \frac{1}{2}(L-M)(k_x^2 + k_y^2) \\
H'_{23} &= 0 \\
H'_{24} &= H'_{13} \\
H'_{25} &= \frac{1}{\sqrt{2}}(H'_{22} - H'_{11}) \\
H'_{26} &= \sqrt{\frac{3}{2}}H'_{12}
\end{aligned}
$$

$$
\begin{aligned}
H'_{33} &= H'_{22} \\
H'_{34} &= -H'_{12} \\
H'_{35} &= -(H'_{26})^* \\
H'_{36} &= H'_{25}
\end{aligned}
$$

$$
\begin{aligned}
H'_{44} &= H'_{11} \\
H'_{45} &= -\sqrt{2}(H'_{13})^* \\
H'_{46} &= -(H'_{15})^*
\end{aligned}
$$

$$
\begin{aligned}
H'_{55} &= \frac{\hbar^2 k^2}{2m} + \frac{1}{3}(2M' + L')k^2 - \Delta_0 \\
H'_{56} &= 0
\end{aligned}
$$

$$
H'_{66} = H'_{55}.
$$

The matrix $\{H'_{ij}\}$ is Hermitian, i. e., $H'_{ij} = [H'_{ji}]^*$. This 6×6 matrix can be diagonalized numerically without further simplification. Readers with access to a personal computer and a matrix diagonalization program are encouraged to

calculate the valence band structure of GaAs by diagonalizing this 6×6 matrix $\{H'_{ij}\}$ (Problem 2.14b).

The matrix $\{H'_{ij}\}$ can be diagonalized analytically with some approximations. We will now restrict k to values small enough that the matrix elements which couple the $J = 3/2$ and $J = 1/2$ bands, such as H'_{15}, H'_{16}, and H'_{25}, are negligible compared with the spin–orbit coupling. With this assumption, and limiting the expansion of the eigenvalue to terms of the order of k^2 only, the 6×6 matrix reduces to a 4×4 and a 2×2 matrix. The 2×2 matrix gives the energy of the doubly degenerate $j = 1/2$ Γ_7 band as

$$\begin{aligned} E_{\text{so}} &= H'_{55} = \frac{\hbar^2 k^2}{2m} + \frac{1}{3}(2M' + L')k^2 - \Delta_0 \\ &= -\Delta_0 + \frac{\hbar^2 k^2}{2m}\left[1 - \frac{2}{3}\left(\frac{P^2}{m(E_0 + \Delta_0)} + \frac{2Q^2}{m(E'_0 + \Delta_0)}\right)\right]. \end{aligned} \tag{2.58}$$

Thus, within the above approximation, the constant energy surface for the $j = 1/2$ split-off valence band is spherical and the band dispersion parabolic. In analogy with the conduction band we can define an effective mass $m_{v,\text{so}}$ for the split-off valence band given by

$$\frac{m}{m_{v,\,\text{so}}} = 1 - \frac{2}{3}\left(\frac{P^2}{m(E_0 + \Delta_0)} + \frac{2Q^2}{m(E'_0 + \Delta_0)}\right). \tag{2.59}$$

The dispersion of the $j = 3/2$ bands is obtained by diagonalizing the 4×4 matrix

$$\begin{vmatrix} H'_{11} & H'_{12} & H'_{13} & 0 \\ (H'_{12})^* & H'_{22} & 0 & H'_{13} \\ (H'_{13})^* & 0 & H'_{22} & -H'_{12} \\ 0 & (H'_{13})^* & -(H'_{12})^* & H'_{11} \end{vmatrix}.$$

The secular equation for this matrix reduces to two identical equations of the form

$$(H'_{11} - E)(H'_{22} - E) = |H'_{12}|^2 + |H'_{13}|^2 \tag{2.60}$$

and their solutions are

$$E_\pm = \tfrac{1}{2}(H'_{11} + H'_{22}) \pm \tfrac{1}{2}[(H'_{11} + H'_{22})^2 - 4(H'_{11}H'_{22} - |H'_{12}|^2 - |H'_{13}|^2)]^{\frac{1}{2}}. \tag{2.61}$$

Substituting the matrix elements H'_{ij} as defined earlier into (2.61) $E_\pm$ can be expressed as

$$E_\pm = Ak^2 \pm [B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)]^{\frac{1}{2}}, \tag{2.62}$$

an equation first derived by *Dresselhaus* et al. [2.21]. The constants A, B, and C in (2.62) are related to the electron momentum matrix elements and energy

gaps by

$$\frac{2m}{\hbar^2}A = 1 - \frac{2}{3}\left[\left(\frac{P^2}{mE_0}\right) + \left(\frac{2Q^2}{mE_0'}\right)\right] \tag{2.63}$$

$$\frac{2m}{\hbar^2}B = \frac{2}{3}\left[\left(\frac{-P^2}{mE_0}\right) + \left(\frac{Q^2}{mE_0'}\right)\right] \tag{2.64}$$

$$\left(\frac{2m}{\hbar^2}C\right)^2 = \frac{16P^2Q^2}{3mE_0mE_0'}. \tag{2.65}$$

Equations (2.63–65) show that it is more convenient to define the constants A, B, and C in units of $\hbar^2/2m$. Note that in the literature [2.17] the definitions of A, B, and C may contain a small additional term R, which is the matrix element of the electron momentum operator between the Γ_{4v} valence band and a higher energy Γ_{3c} conduction band. Inclusion of R is particularly important for large bandgap materials such as diamond [2.22]

The dispersion of the $\Gamma_8(J = 3/2)$ bands near the zone center is given by (2.62); this equation has been derived after much simplification and is valid only for energies small compared to the spin–orbit splitting. We note that both A and B are negative since the dominant term in both (2.63) and (2.64) is $2P^2/(3mE_0)$, which is $\gg 1$. As a result, the effective masses of these bands, and of the splitt-off valence band in (2.59), are negative. In many cases we have to consider the properties of a semiconductor in which a few electrons are missing from an otherwise filled valence band. Instead of working with electrons with negative masses, it is more convenient to introduce the idea of a **hole**. A filled valence band with one electron missing can be regarded as a band (known as a **hole band**) containing one hole. If the energy of the missing electron in the valence band is E (assuming that $E = 0$ is the top of the valence band) then the energy of the corresponding hole is $-E$ and is positive. With this definition the effective mass of a hole in the valence band is opposite to that of the corresponding missing electron and is positive also. The hole mass of the split-off valence band $m_{so} = -m_{v,so}$ is positive with this definition (Table 2.24). Since the valence band represented by E_+ has a smaller dispersion and hence larger mass, it is generally referred to as the **heavy hole** band, while the band represented by E_- is known as the **light hole** band. From now on the energies of these two hole bands will be written as E_{hh} and E_{lh} with the corresponding hole energies defined as

$$E_{hh} = -Ak^2 - [B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)]^{\frac{1}{2}}, \tag{2.66a}$$

$$E_{lh} = -Ak^2 + [B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)]^{\frac{1}{2}}. \tag{2.66b}$$

Constant energy surfaces represented by (2.66a) and (2.66b) are shown in Fig. 2.16. The shapes of these constant energy surfaces are referred to as "warped" spheres. The warping occurs along the [100] and [111] directions because of the cubic symmetry of the zinc-blende crystal. In fact one can argue that these warped spheres are the only possible shapes for constant energy surfaces described by

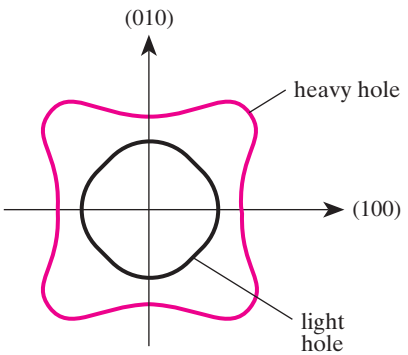


Fig. 2.16. Constant energy surfaces of the $J = 3/2(\Gamma_8)$ bands in diamond- and zinc-blende-type semiconductors

a second-order equation in cubic crystals. Assuming that odd-order terms in k are either zero or negligible, the lowest order terms even in k consistent with the cubic symmetry are k^2 and $[\alpha k^4 + \beta(k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2)]^{\frac{1}{2}}$. If we neglect higher order terms, the most general expression for the k dependence of the energy of the $\Gamma_8(j = 3/2)$ component of a Γ_4 state in a cubic crystal is of the form of (2.62), where A, B, and C are linearly independent parameters related to the electron momentum matrix elements. One may notice from the definitions of the coefficients A, B, and C in (2.63–65) that C can be expressed in terms of A and B. This is a result of neglecting in our model the coupling between the Γ_{4v} bands and higher conduction bands (such as Γ_{3c}), for the inclusion of the lowest Γ_{3c} state see Problem 2.15d.

The hole band dispersions along the [100] and [111] directions are parabolic, but the hole effective masses are different along the two directions:

$$k\|(100) \qquad \frac{1}{m_{\mathrm{hh}}} = \frac{2}{\hbar^2}(-A + B), \tag{2.67a}$$

$$\frac{1}{m_{\mathrm{lh}}} = \frac{2}{\hbar^2}(-A - B), \tag{2.67b}$$

$$k\|(111) \qquad \frac{1}{m_{\mathrm{hh}}} = \frac{2}{\hbar^2}\left[-A + B\left(1 + \frac{|C|^2}{3B^2}\right)^{\frac{1}{2}}\right], \tag{2.68a}$$

$$\frac{1}{m_{\mathrm{lh}}} = \frac{2}{\hbar^2}\left[-A - B\left(1 + \frac{|C|^2}{3B^2}\right)^{\frac{1}{2}}\right]. \tag{2.68b}$$

From the above expressions we see that the warping of the valence bands is caused by the term $|C|^2$, which is proportional to Q^2. If the term B^2 is much larger than $|C|^2/3$ warping can be neglected and we can obtain the approximate result that $m_{\mathrm{lh}} \approx 3m_{\mathrm{c}}^*/2$ and $m_{\mathrm{so}} \approx 3m_{\mathrm{c}}^*$. Note that Q^2 is crucial to m_{hh}. If we put $Q^2 = 0$ we obtain the incorrect result $m_{\mathrm{hh}} = -m_0$ (even the sign is wrong!). Often, for simplicity, it is expedient to assume that the valence band masses are isotropic. In such cases average heavy and light hole masses m_{hh}^* and m_{lh}^* can be obtained by averaging (2.67) and (2.68) over all possible directions of $\boldsymbol{k}$ (Problem 4.4):

$$\frac{1}{m^*_{\rm hh}} = \frac{1}{\hbar^2}\left[-2A + 2B\left(1 + \frac{2|C|^2}{15B^2}\right)\right], \tag{2.69a}$$

$$\frac{1}{m^*_{\rm lh}} = \frac{1}{\hbar^2}\left[-2A - 2B\left(1 + \frac{2|C|^2}{15B^2}\right)\right]. \tag{2.69b}$$

In Table 2.24 we have listed what we judge to be reliable values of the constants A, B, and $|C|^2$ for several semiconductors obtained from data in [2.18]. In this table the three valence band effective masses calculated from (2.67–69) using these values of A, B, and $|C|^2$ and experimental energy gaps are compared with the experimentally determined effective masses.

We note that the constant energy surfaces for the valence bands as described by (2.62) have inversion symmetry: $E(\boldsymbol{k}) = E(-\boldsymbol{k})$, even though the crystal may not have such symmetry. This is a consequence of the electron Hamiltonian we have used being invariant under time reversal (**time-reversal symmetry**). A Bloch wave traveling with wave vector $\boldsymbol{k}$ is transformed into a Bloch wave with wave vector $-\boldsymbol{k}$ under time reversal. If the Hamiltonian is invariant under time reversal, these two Bloch waves will have the same energy.

Finally we point out that there is an alternate equivalent approach often used in the literature to represent the valence band dispersion in diamond- and zinc-blende-type semiconductors. Using group theory it is possible to derive an effective $\boldsymbol{k}\cdot\boldsymbol{p}$ Hamiltonian which is appropriate for the Γ_4 valence bands. An example of such a Hamiltonian was proposed by *Luttinger* [2.23]:

$$H_{\rm L} = \frac{\hbar^2}{2m}\left[\left(\gamma_1 + \frac{5}{2}\gamma_2\right)\nabla^2 - 2\gamma_3(\nabla\cdot\boldsymbol{J})^2 + 2(\gamma_3 - \gamma_2)(\nabla_x^2 J_x^2 + \text{c. p.})\right], \tag{2.70}$$

where the parameters γ_1, γ_2, and γ_3 are known as the **Kohn–Luttinger parameters**; $\boldsymbol{J} = (J_x, J_y, J_z)$ is an operator whose effects on the Γ_8 valence bands are identical to those of the angular momentum operator on the $J = 3/2$ atomic states, and c. p. stands for cyclic permutations. This approach facilitates the diagonalization of $H_{\rm L}$ together with additional perturbations applied to the crystal. In Chap. 4 we will see an application of this Hamiltonian to calculate the energies of acceptor states. The first two terms in (2.70) have spherical symmetry while the last represents the effect of the lower, cubic symmetry. It is thus clear that the warping of the valence band is directly proportional to the difference between γ_2 and γ_3. The Kohn–Luttinger parameters can be shown to be related to the coefficients A, B, and C in (2.62) by

$$(\hbar^2/2m)\gamma_1 = -A \tag{2.71a}$$

$$(\hbar^2/2m)\gamma_2 = -B/2 \tag{2.71b}$$

$$(\hbar^2/2m)\gamma_3 = \left[(B^2/4) + (C^2/12)\right]^{1/2} \tag{2.71c}$$

The proof of these results is left as an exercise (Problem 2.15).

2.7 Tight-Binding or LCAO Approach to the Band Structure of Semiconductors

The pseudopotential approach to calculating the band structure of semiconductors discussed in Sect. 2.5 starts with the assumption that electrons are nearly free and their wave functions can be approximated by plane waves. In this section we will approach the problem from the other extreme. We will assume that the electrons are tightly bound to their nuclei as in the atoms. Next we will bring the atoms together. When their separations become comparable to the lattice constants in solids, their wave functions will overlap. We will approximate the electronic wave functions in the solid by linear combinations of the atomic wave functions. This approach is known as the **tight-binding** approximation or **Linear Combination of Atomic Orbitals** (LCAO) approach. One may ask: how can two completely opposite approaches such as the pseudopotential method and the tight-binding method both be good starting points for understanding the electronic properties of the same solid? The answer is that in a covalently bonded semiconductor there are really two kinds of electronic states. Electrons in the conduction bands are delocalized and so can be approximated well by nearly free electrons. The valence electrons are concentrated mainly in the bonds and so they retain more of their atomic character. The valence electron wave functions should be very similar to **bonding orbitals** found in molecules. In addition to being a good approximation for calculating the valence band structure, the LCAO method has the advantage that the band structure can be defined in terms of a small number of **overlap parameters**. Unlike the pseudopotentials, these overlap parameters have a simple physical interpretation as representing interactions between electrons on adjacent atoms.

2.7.1 Molecular Orbitals and Overlap Parameters

To illustrate the tight-binding approach for calculating band structures, we will restrict ourselves again to the case of tetrahedrally bonded semiconductors. The valence electrons in the atoms of these semiconductors are in s and p orbitals. These orbitals in two identical and isolated atoms are shown schematically in Figs. 2.17a, 2.18a, and 2.19a. The p_z orbitals are not shown since

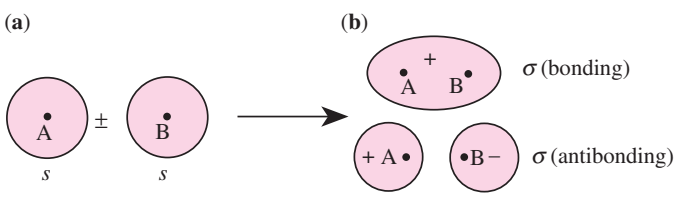


Fig. 2.17a,b. Overlap of two s orbitals to form bonding and antibonding σ orbitals

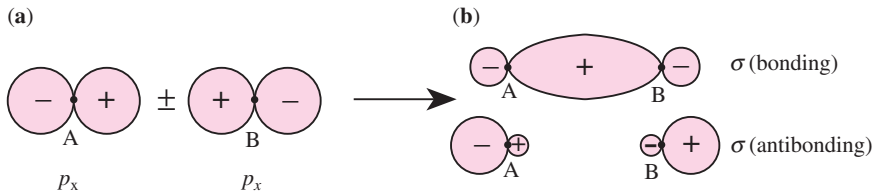


Fig. 2.18a,b. Overlap of two p_x orbitals along the x axis to form bonding and antibonding σ orbitals

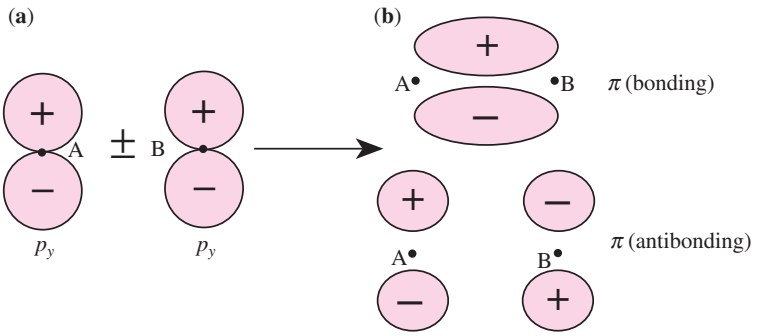


Fig. 2.19a,b. Overlap of two p_y orbitals to form bonding and antibonding π orbitals

their properties are similar to those of the p_y orbitals. Figures 2.17b, 2.18b, and 2.19b show schematically what happens to the atomic orbitals when the two atoms are brought together along the x direction until the atomic orbitals overlap to form a diatomic molecule. The interaction between the two atomic orbitals produces two new orbitals. One of the resultant orbitals is symmetric with respect to the interchange of the two atoms and is known as the **bonding orbital** while the other orbital, which is antisymmetric, is known as the **antibonding orbital**. In the case of p orbitals there are two ways for them to overlap. When they overlap along the direction of the p orbitals, as shown in Fig. 2.18b, they are said to form σ **bonds**. When they overlap in a direction perpendicular to the p orbitals they are said to form π **bonds**, as shown in Fig. 2.19b.

The interaction between the atomic orbitals changes their energies. Typically the antibonding orbital energy is *raised* by an amount determined by the interaction Hamiltonian H. The energy of the bonding orbital is *decreased* by the same amount. The changes in orbital energies are shown schematically in Fig. 2.20a for a homopolar molecule and in Fig. 2.20b for a heteropolar one. In both cases V is the matrix element of the interaction Hamiltonian between the atomic orbitals and is usually referred to as the **overlap parameter**. For a homopolar molecule containing only s and p valence electrons, there are four nonzero overlap parameters. To derive this result we will denote the atomic orbital on one of the atoms as $|\alpha\rangle$ and that on the second atom as $|\beta\rangle$. These

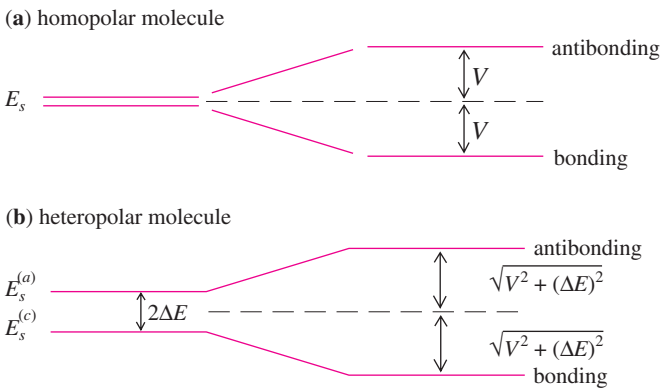


Fig. 2.20. Effect of orbital overlap on the energy levels in (**a**) a diatomic homopolar molecule and (**b**) a diatomic heteropolar molecule. V represents the matrix element of the interaction Hamiltonian

orbitals can be expressed as products of a radial wave function and a spherical harmonic $Y_{lm}(\theta, \phi)$ with the atom chosen as the origin. We will denote the vector going from the first atom (designated as A in Fig. 2.21) to the second atom (B) as $\boldsymbol{d}$. For both orbitals $|\alpha\rangle$ and $|\beta\rangle$ we will choose the coordinate axes such that the z axes are parallel to $\boldsymbol{d}$ and the azimuthal angles ϕ are the same (see Fig. 2.21). In these coordinate systems the spherical harmonic wave functions of the two atoms A and B are $Y_{lm}(\theta, \phi)$ and $Y_{l'm'}(\theta', \phi)$, respectively. The Hamiltonian H has cylindrical symmetry with respect to $\boldsymbol{d}$ and therefore cannot depend on ϕ. Thus the matrix element $\langle\alpha | H | \beta\rangle$ is proportional to the integral of the azimuthal wave functions $\exp[\mathrm{i}(m' - m)\phi]$. This integral vanishes except when $m = m'$. As a result of this selection rule we conclude that

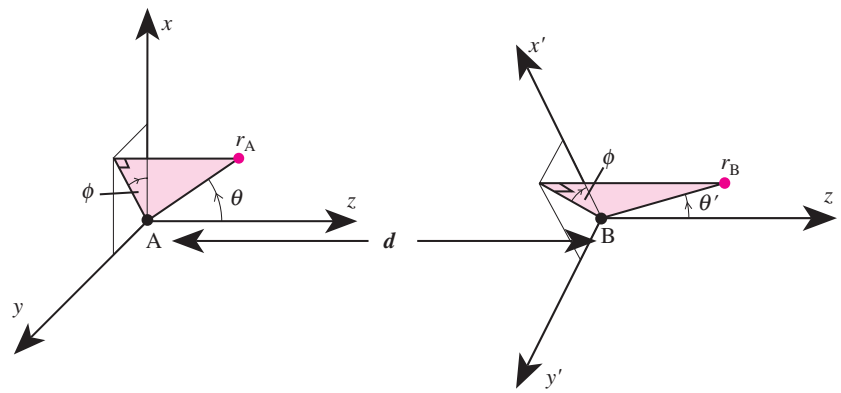


Fig. 2.21. Choice of the polar coordinate systems for the two atoms A and B in a diatomic molecule in order that the z axis be parallel to the vector joining the two atoms A and B and the azimuthal angle ϕ be identical for both atoms

there are four nonzero and linearly independent overlap parameters between the s and p electrons:

$$\langle s\,|\,H\,|\,s\rangle = V_{ss\sigma};\quad \langle s\,|\,H\,|\,p_z\rangle = V_{sp\sigma};\quad \langle p_z\,|\,H\,|\,p_z\rangle = V_{pp\sigma};$$
$$\text{and}\quad \langle p_x\,|\,H\,|\,p_x\rangle = V_{pp\pi}.$$

We notice that $\langle p_x\,|\,H\,|\,p_y\rangle = 0$ and $\langle p_y\,|\,H\,|\,p_y\rangle = \langle p_x\,|\,H\,|\,p_x\rangle$ as a result of symmetry. The overlap parameters are usually labeled σ, π and δ for ($l = 2$ wave functions), depending on whether m is 0, 1, or 2 (in analogy with the s, p, and d atomic wave functions).

The concept of bonding and antibonding orbitals introduced for molecules can be easily extended to crystals if one assumes that the orbitals of each atom in the crystal overlap with those of its nearest neighbors only. This is a reasonable approximation for most solids. The results of orbital overlap in a solid is that the bonding and antibonding orbitals are broadened into bands. Those occupied by electrons form valence bands while the empty ones form conduction bands. Figure 2.22 shows schematically how the s and p orbitals evolve into bands in a tetrahedral semiconductor. In this case the bonding orbitals are filled with electrons and become the valence bands while the antibonding orbitals become the conduction bands. As may be expected, the crystal structure affects the overlap between atomic orbitals. For example, in a tetrahedrally coordinated solid each atom is surrounded by four nearest neighbors. The vectors $\boldsymbol{d}$ linking the central atom to each of its nearest neighbors are different, so it is not convenient to choose the z axis parallel to $\boldsymbol{d}$. Instead it is more convenient to choose the crystallographic axes as the coordinate axes. The spherical harmonics $Y_{lm}(\theta,\phi)$ of the atomic orbitals are then defined with respect to this fixed coordinate system. In calculating the overlap parameter for any pair of neighboring atoms, one expands the spherical harmonics defined with respect to $\boldsymbol{d}$ in terms of $Y_{lm}(\theta,\phi)$. An example of this expansion is shown schematically in Fig. 2.23.

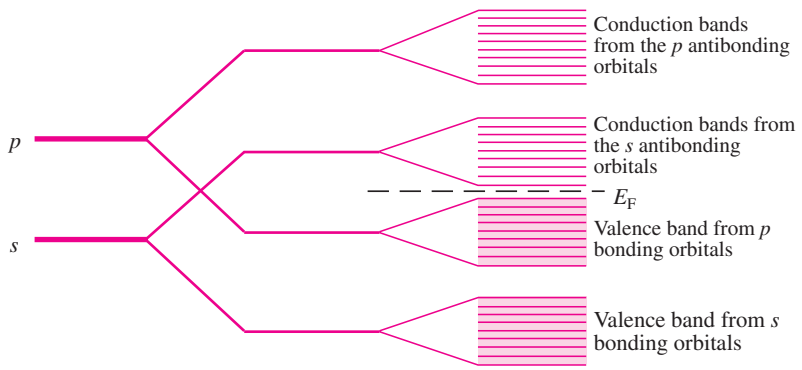


Fig. 2.22. Evolution of the atomic s and p orbitals into valence and conduction bands in a semiconductor. E_F is the Fermi energy

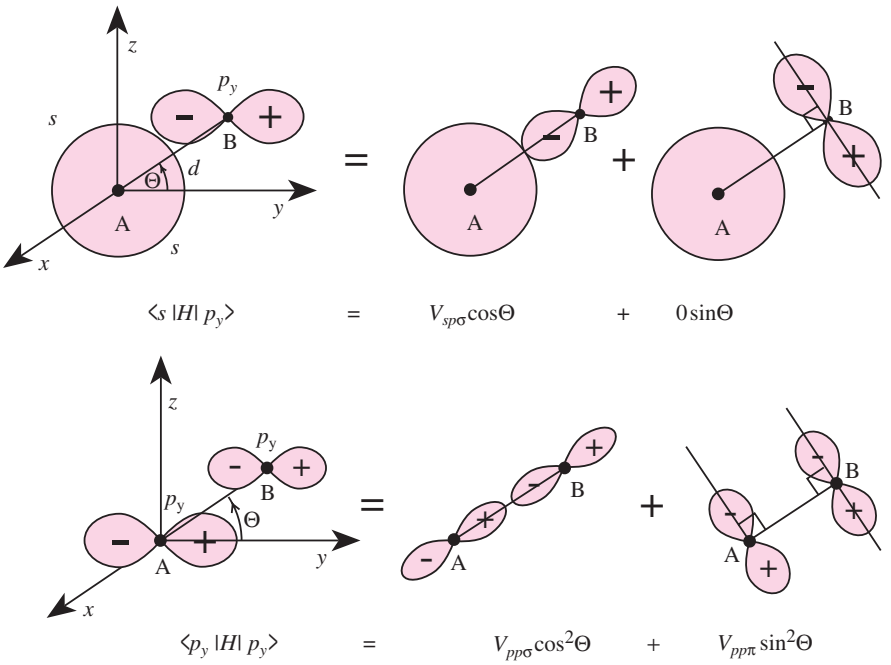


Fig. 2.23. Projection of the overlap parameter between an s and a p_y orbital, and between p_y orbitals, along the vector $\boldsymbol{d}$ joining the two atoms and perpendicular to $\boldsymbol{d}$

2.7.2 Band Structure of Group-IV Elements by the Tight-Binding Method

After this introduction to the interaction between atomic orbitals we are ready to perform a quantitative calculation of the electronic band structure using the method of Linear Combination of Atomic Orbitals (LCAO). While the method has been utilized by many authors [Ref. 2.25, p. 75], the approach we will describe follows that of *Chadi* and *Cohen* [2.25].

The position of an atom in the primitive cell denoted by j will be decomposed into $\boldsymbol{r}_{jl} = \boldsymbol{R}_j + \boldsymbol{r}_l$, where $\boldsymbol{R}_j$ denotes the position of the jth primitive cell of the Bravais lattice and $\boldsymbol{r}_l$ is the position of the atom l within the primitive cell. For the diamond and zinc-blende crystals $l = 1$ and 2 only. Let $h_l(\boldsymbol{r})$ denote the Hamiltonian for the isolated atom l with its nucleus chosen as the origin. The Hamiltonian for the atom located at $\boldsymbol{r}_{jl}$ will be denoted by $h_l(\boldsymbol{r} - \boldsymbol{r}_{jl})$. The wave equation for h_l is given by

$$h_l\phi_{ml}(\boldsymbol{r} - \boldsymbol{r}_{jl}) = E_{ml}\phi_{ml}(\boldsymbol{r} - \boldsymbol{r}_{jl}), \tag{2.72}$$

where E_{ml} and ϕ_{ml} are the eigenvalues and eigenfunctions of the state indexed by m. The atomic orbitals $\phi_{ml}(\boldsymbol{r} - \boldsymbol{r}_{jl})$ are known as **Löwdin orbitals**. They are different from the usual atomic wave functions in that they have been constructed in such a way that wave functions centered at different atoms are

orthogonal to each other. Next we assume that the Hamiltonian for the crystal $\mathcal{H}$ is equal to the sum of the atomic Hamiltonians and a term $\mathcal{H}_{\text{int}}$ which describes the interaction between the different atoms. We further assume the interaction between the atoms to be weak so that $\mathcal{H}$ can be diagonalized by perturbation theory. In this approximation the unperturbed Hamiltonian $\mathcal{H}_0$ is simply

$$\mathcal{H}_0 = \sum_{j,l} h_l(\boldsymbol{r} - \boldsymbol{r}_{jl}) \tag{2.73}$$

and we can construct the unperturbed wave functions as linear combinations of the atomic wave functions. Because of the translational symmetry of the crystal, these unperturbed wave functions can be expressed in the form of Bloch functions:

$$\Phi_{ml\boldsymbol{k}} = \frac{1}{\sqrt{N}} \sum_{j} \exp(\mathrm{i}\boldsymbol{r}_{jl} \cdot \boldsymbol{k}) \phi_{ml}(\boldsymbol{r} - \boldsymbol{r}_{jl}), \tag{2.74}$$

where N is the number of primitive unit cells in the crystal. The eigenfunctions $\Psi_{\boldsymbol{k}}$ of $\mathcal{H}$ can then be written as linear combinations of $\Phi_{ml\boldsymbol{k}}$:

$$\Psi_{\boldsymbol{k}} = \sum_{m,l} C_{ml} \Phi_{ml\boldsymbol{k}}. \tag{2.75}$$

To calculate the eigenfunctions and eigenvalues of $\mathcal{H}$, we operate on $\Psi_{\boldsymbol{k}}$ with the Hamiltonian $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{\text{int}}$. From the orthogonality of the Bloch functions we obtain a set of linear equations in C_{ml}:

$$\sum_{m,l} \left(H_{ml,m'l'} - E_{\boldsymbol{k}} \delta_{mm'} \delta_{ll'} \right) C_{m'l'}(\boldsymbol{k}) = 0, \tag{2.76}$$

where $H_{ml,m'l'}$ stands for the matrix element $\langle \Phi_{ml\boldsymbol{k}} \,|\, \mathcal{H} \,|\, \Phi_{m'l'\boldsymbol{k}} \rangle$ and $E_{\boldsymbol{k}}$ are the eigenvalues of H. To simplify the solution of (2.76) we introduce the following approximations.

- We include only the s^2 and p^6 electrons in the outermost partially filled atomic shells. We neglect spin–orbit coupling (although it can be included easily). The two atomic orbitals of s symmetry for the two atoms in the unit cell will be denoted by $S1$ and $S2$, respectively. Correspondingly, the atomic orbitals with p symmetry will be denoted by: $X1$, $X2$, $Y1$, $Y2$, $Z1$ and $Z2$, respectively. In the following equations the index m will represent the s, p_x, p_y, and p_z orbitals.
- When we substitute the wave functions $\Phi_{ml\boldsymbol{k}}$ defined in (2.74) into (2.76) we obtain

$$\begin{aligned} H_{ml,m'l'}(\boldsymbol{k}) = {} & \sum_{j}^{N} \sum_{j'}^{N} \frac{\exp[\mathrm{i}(\boldsymbol{r}_{j'l'} - \boldsymbol{r}_{jl}) \cdot \boldsymbol{k}]}{N} \\ & \times \langle \phi_{ml} \,|\, (\boldsymbol{r} - \boldsymbol{r}_{jl}) \,|\, H \,|\, \phi_{m'l'}(\boldsymbol{r} - \boldsymbol{r}_{j'l'}) \rangle \end{aligned} \tag{2.77}$$

$$\begin{aligned} = {} & \sum_{j}^{N} \exp[\mathrm{i}(\boldsymbol{R}_j + \boldsymbol{r}_{l'} - \boldsymbol{r}_l) \cdot \boldsymbol{k}] \\ & \times \langle \phi_{ml}(\boldsymbol{r} - \boldsymbol{r}_{jl}) \,|\, H \,|\, \phi_{m'l'}(\boldsymbol{r} - \boldsymbol{r}_{jl'}) \rangle. \end{aligned} \tag{2.78}$$

Instead of summing j over all the unit cells in the crystal, we will sum over the nearest neighbors only. In the diamond and zinc-blende crystals this means j will be summed over the atom itself plus four nearest neighbors. These atoms will be denoted as $j = 1,2,3,4,5$. If needed, one can easily include second neighbor or even further interactions.

Within the above approximation the collection of matrix elements of the form in (2.77) constitutes an 8×8 matrix (note that the dimensions of the matrix depend only on the number of basis functions, not the number of neighbors included). Applying symmetry arguments allows the number of nonzero and linearly independent matrix elements of $\mathscr{H}_{\text{int}}$ to be greatly reduced. As an example, we will consider the matrix element $H_{s1,\,s2}$. From (2.78) this matrix element is given by

$$H_{s1,\,s2} = [\exp(i\boldsymbol{k}\cdot\boldsymbol{d}_1) + \exp(i\boldsymbol{k}\cdot\boldsymbol{d}_2) + \exp(i\boldsymbol{k}\cdot\boldsymbol{d}_3) + \exp(i\boldsymbol{k}\cdot\boldsymbol{d}_4)] \times \langle S1 \,|\, \mathscr{H}_{\text{int}} \,|\, S2\rangle, \tag{2.79}$$

where we have assumed that atom 1 is located at the origin and $\boldsymbol{d}_\alpha$ ($\alpha = 1$ to 4) are the positions of its four nearest neighbors, with

$$\boldsymbol{d}_1 = (1,1,1)(a/4);$$
$$\boldsymbol{d}_2 = (1,-1,-1)(a/4);$$
$$\boldsymbol{d}_3 = (-1,1,-1)(a/4);$$

and

$$\boldsymbol{d}_4 = (-1,-1,1)(a/4).$$

The matrix element $\langle S1 \,|\, \mathscr{H}_{\text{int}} \,|\, S_2\rangle$ is basically the same overlap parameter $V_{ss\sigma}$ as we have defined for molecules. The other matrix elements $H_{s1,\,p_x2}$, and $H_{p_x1,\,p_x2}$, etc., can also be expressed in terms of the overlap parameters $V_{sp\sigma}$, $V_{pp\sigma}$, and $V_{pp\pi}$. For example $H_{s1,\,p_x2}$ can be shown to contain four terms involving the four phase factors $\exp(i\boldsymbol{k}\cdot\boldsymbol{d}_\alpha)$ and the matrix element $\langle S1 \,|\, \mathscr{H}_{\text{int}} \,|\, X2\rangle$. However, for each nearest neighbor $\langle S1 \,|\, \mathscr{H}_{\text{int}} \,|\, X2\rangle$ has to be decomposed into σ and π components as shown in Fig. 2.23. This decomposition introduces a factor of $\cos\Theta = \pm(1/\sqrt{3})$. The $-$ or $+$ sign depends on whether the s orbital lies in the direction of the positive or negative lobe of the p_x orbital. As a result, it is convenient to introduce a new set of four overlap parameters appropriate for the diamond lattice:

$$V_{ss} = 4V_{ss\sigma}, \tag{2.80a}$$

$$V_{sp} = 4V_{sp\sigma}/\sqrt{3}, \tag{2.80b}$$

$$V_{xx} = (4V_{pp\sigma}/3) + (8V_{pp\pi}/3), \tag{2.80c}$$

$$V_{xy} = (4V_{pp\sigma}/3) - (4V_{pp\pi}/3), \tag{2.80d}$$

With this notation the matrix element $\langle S1(\boldsymbol{r}) \,|\, \mathscr{H}_{\text{int}} \,|\, X2(\boldsymbol{r} - \boldsymbol{d}_1)\rangle$ is given by $(V_{sp\sigma})/\sqrt{3} = V_{sp}/4$. The remaining three matrix elements are related to $\langle S1(\boldsymbol{r}) \,|\, \mathscr{H}_{\text{int}} \,|\, X2(\boldsymbol{r} - \boldsymbol{d}_1)\rangle$ by symmetry. For example, a two-fold rotation about

the y axis will transform (x, y, z) into $(-x, y, -z)$, so $\boldsymbol{d}_1$ is transformed into $\boldsymbol{d}_3$. The s-symmetry wave function $|S1\rangle$ is unchanged while the p-symmetry wave function $|X2\rangle$ is transformed into $-|X2\rangle$ under this rotation. As a result, $\langle S1(\boldsymbol{r})\,|\,\mathcal{H}_{\text{int}}\,|\,X2(\boldsymbol{r}-\boldsymbol{d}_3)\rangle = -\langle S1(\boldsymbol{r})\,|\,\mathcal{H}_{\text{int}}\,|\,X2(\boldsymbol{r}-\boldsymbol{d}_1)\rangle$. By applying similar symmetry operations we can show that

$$\sum_{\alpha} \exp[\mathrm{i}(\boldsymbol{d}_\alpha \cdot \boldsymbol{k})]\langle S_1(\boldsymbol{r})\,|\,\mathcal{H}_{\text{int}}\,|\,X2(\boldsymbol{r}-\boldsymbol{d}_\alpha)\rangle = \tfrac{1}{4}V_{sp}\{\exp[\mathrm{i}(\boldsymbol{d}_1 \cdot \boldsymbol{k})] + \exp[\mathrm{i}(\boldsymbol{d}_2 \cdot \boldsymbol{k})] - \exp[\mathrm{i}\boldsymbol{d}_3 \cdot \boldsymbol{k})] - \exp[\mathrm{i}(\boldsymbol{d}_4 \cdot \boldsymbol{k})]\} \tag{2.81}$$

In the zinc-blende structure, because the atoms 1 and 2 are different, $\langle S1\,|\,\mathcal{H}_{\text{int}}\,|\,X2\rangle$ is, in principle, different from $\langle S2\,|\,\mathcal{H}_{\text{int}}\,|\,X1\rangle$. They are, however, often assumed to be equal [Ref. 2.24, p. 77]. The case of the zinc-blende crystal is left as an exercise in Problem 2.16. Here we will restrict ourselves to the case of the diamond structure.

The 8×8 matrix for the eight s and p bands can be expressed as in Table 2.25. E_s and E_p represent the energies $\langle S1\,|\,\mathcal{H}_0\,|\,S1\rangle$ and $\langle X1\,|\,H_0\,|\,X1\rangle$, respectively. The four parameters g_1 to g_4 arise from summing over the factor $\exp[\mathrm{i}(\boldsymbol{k} \cdot \boldsymbol{d}_\alpha)]$ as in (2.81). They are defined by

$$\begin{aligned}
g_1 &= (1/4)\{\exp[\mathrm{i}(\boldsymbol{d}_1 \cdot \boldsymbol{k})] + \exp[\mathrm{i}(\boldsymbol{d}_2 \cdot \boldsymbol{k})] + \exp[\mathrm{i}(\boldsymbol{d}_3 \cdot \boldsymbol{k})] + \exp[\mathrm{i}(\boldsymbol{d}_4 \cdot \boldsymbol{k})]\},\\
g_2 &= (1/4)\{\exp[\mathrm{i}(\boldsymbol{d}_1 \cdot \boldsymbol{k})] + \exp[\mathrm{i}(\boldsymbol{d}_2 \cdot \boldsymbol{k})] - \exp[\mathrm{i}(\boldsymbol{d}_3 \cdot \boldsymbol{k})] - \exp[\mathrm{i}(\boldsymbol{d}_4 \cdot \boldsymbol{k})]\},\\
g_3 &= (1/4)\{\exp[\mathrm{i}(\boldsymbol{d}_1 \cdot \boldsymbol{k})] - \exp[\mathrm{i}(\boldsymbol{d}_2 \cdot \boldsymbol{k})] + \exp[\mathrm{i}(\boldsymbol{d}_3 \cdot \boldsymbol{k})] - \exp[\mathrm{i}(\boldsymbol{d}_4 \cdot \boldsymbol{k})]\},\\
g_4 &= (1/4)\{\exp[\mathrm{i}(\boldsymbol{d}_1 \cdot \boldsymbol{k})] - \exp[\mathrm{i}(\boldsymbol{d}_2 \cdot \boldsymbol{k})] - \exp[\mathrm{i}(\boldsymbol{d}_3 \cdot \boldsymbol{k})] + \exp[\mathrm{i}(\boldsymbol{d}_4 \cdot \boldsymbol{k})]\}.
\end{aligned}$$

If $\boldsymbol{k} = (2\pi/a)(k_1, k_2, k_3)$ the g_j's can also be expressed as

$$g_1 = \cos(k_1\pi/2)\cos(k_2\pi/2)\cos(k_3\pi/2) - \mathrm{i}\sin(k_1\pi/2)\sin(k_2\pi/2)\sin(k_3\pi/2), \tag{2.82a}$$

$$g_2 = -\cos(k_1\pi/2)\sin(k_2\pi/2)\sin(k_3\pi/2) + \mathrm{i}\sin(k_1\pi/2)\cos(k_2\pi/2)\cos(k_3\pi/2), \tag{2.82b}$$

Table 2.25. Matrix for the eight s and p bands in the diamond structure within the tight binding approximation

	$S1$	$S2$	$X1$	$Y1$	$Z1$	$X2$	$Y2$	$Z2$
$S1$	$E_s - E_k$	$V_{ss}g_1$	0	0	0	$V_{sp}g_2$	$V_{sp}g_3$	$V_{sp}g_4$
$S2$	$V_{ss}g_1^*$	$E_s - E_k$	$-V_{sp}g_2^*$	$-V_{sp}g_3^*$	$-V_{sp}g_4^*$	0	0	0
$X1$	0	$-V_{sp}g_2$	$E_p - E_k$	0	0	$V_{xx}g_1$	$V_{xy}g_4$	$V_{xy}g_3$
$Y1$	0	$-V_{sp}g_3$	0	$E_p - E_k$	0	$V_{xy}g_4$	$V_{xx}g_1$	$V_{xy}g_2$
$Z1$	0	$-V_{sp}g_4$	0	0	$E_p - E_k$	$V_{xy}g_3$	$V_{xy}g_2$	$V_{xx}g_1$
$X2$	$V_{sp}g_2^*$	0	$V_{xx}g_1^*$	$V_{xy}g_4^*$	$V_{xy}g_3^*$	$E_p - E_k$	0	0
$Y2$	$V_{sp}g_3^*$	0	$V_{xy}g_4^*$	$V_{xx}g_1^*$	$V_{xy}g_2^*$	0	$E_p - E_k$	0
$Z2$	$V_{sp}g_4^*$	0	$V_{xy}g_3^*$	$V_{xy}g_2^*$	$V_{xx}g_1^*$	0	0	$E_p - E_k$

$$g_3 = -\sin(k_1\pi/2)\cos(k_2\pi/2)\sin(k_3\pi/2) + \mathrm{i}\cos(k_1\pi/2)\sin(k_2\pi/2)\cos(k_3\pi/2), \tag{2.82c}$$

$$g_4 = -\sin(k_1\pi/2)\sin(k_2\pi/2)\cos(k_3\pi/2) + \mathrm{i}\cos(k_1\pi/2)\cos(k_2\pi/2)\sin(k_3\pi/2), \tag{2.82d}$$

The valence and lowest conduction band energies of the diamond-type crystals can be obtained by diagonalizing the 8×8 matrix of Table 2.25, provided the four parameters V_{ss}, V_{sp}, V_{xx}, and V_{xy} are known. These four parameters can be determined by comparing the calculated band structure with a first principles or empirical band structure calculation. For example *Chadi and Cohen* [2.25] obtained the tight-binding parameters for C, Si, and Ge by comparison with empirical pseudopotential calculations. Their results are shown in Table 2.26. Note that the signs of V_{ss} etc. are, in part, arbitrary and are determined by the choice of the relative phases of the two overlaping atomic orbitals. The signs in Table 2.26 correspond to the choices shown in Figs. 2.17a and 2.23. The magnitudes of the interaction parameters decrease in the sequence C to Ge. We will show later that this trend can be understood from the increase in the lattice constant along this sequence. When the second-nearest neighbor interactions are included, only V_{xx} decreases somewhat. Since V_{xx} is the smallest interaction, the overall band structure is not significantly affected.

Table 2.26. Tight-binding interaction parameters (in eV) for C, Si, and Ge obtained by *Chadi* and *Cohen* [2.25] when only nearest-neighbor interactions are included

	$E_p - E_s$	V_{ss}	V_{sp}	V_{xx}	V_{xy}
C	7.40	−15.2	10.25	3.0	8.3
Si	7.20	−8.13	5.88	3.17	7.51
Ge	8.41	−6.78	5.31	2.62	6.82

To gain some insight into the band structure obtained with the tight-binding approach, we will calculate the band energies at the $k = 0$ point. From (2.82a–d) we find $g_2 = g_3 = g_4 = 0$ and $g_1 = 1$ at $k = 0$. Thus the 8×8 matrix simplifies into a 2×2 matrix for the s electrons and three identical 2×2 matrices for the p levels:

$$\begin{vmatrix} E_s - E(0) & V_{ss} \\ V_{ss} & E_s - E(0) \end{vmatrix} \tag{2.83a}$$

and

$$\begin{vmatrix} E_p - E(0) & V_{xx} \\ V_{xx} & E_p - E(0) \end{vmatrix}. \tag{2.83b}$$

These two matrices can be easily diagonalized to yield four energies:

$$E_{s\pm}(0) = E_s \pm |V_{ss}| \tag{2.84a}$$

and

$$E_{p\pm}(0) = E_p \pm |V_{xx}| \tag{2.84b}$$

As a result of the overlap of the atomic orbitals the two s and p levels of the two atoms inside the primitive cell are split by an amount equal to $2|V_{ss}|$ and $2|V_{xx}|$, respectively. The level E_{s+} is raised in energy and its wave function is antisymmetric with respect to the interchange of the two atoms. This state corresponds to the antibonding s state in a diatomic molecule. The level E_{s-} corresponds to the bonding s state. From Table 2.5 we expect the antisymmetric antibonding state to have $\Gamma_{2'}$ symmetry and the symmetric bonding state to have Γ_1 symmetry. Using a similar analogy, the triply degenerate antisymmetric Γ_{15} conduction band states correspond to the antibonding p orbitals while the symmetric $\Gamma_{25'}$ valence band states are identified with the bonding p orbitals.

In Fig. 2.24 the valence band structure of Si calculated by the tight-binding method is compared with that obtained by the empirical pseudopotential method. Figure 2.24 also compares the valence band density of states obtained by the two methods (We will define density of states of a band in Sect. 4.3.1 and also in Chap. 8, where this concept will be utilized). In this tight-binding calculation one second-nearest-neighbor interaction has been included in addition to the nearest-neighbor interactions. The agreement between the two methods is quite good for the valence bands. Figure 2.25 shows a comparison between the band structure of Ge calculated by the tight-binding method, the empirical pseudopotential method, and the nearly free electron model. While the valence bands are well reproduced by the tight-binding method with the simple sp^3 base used here, this is not true for the conduction bands since the

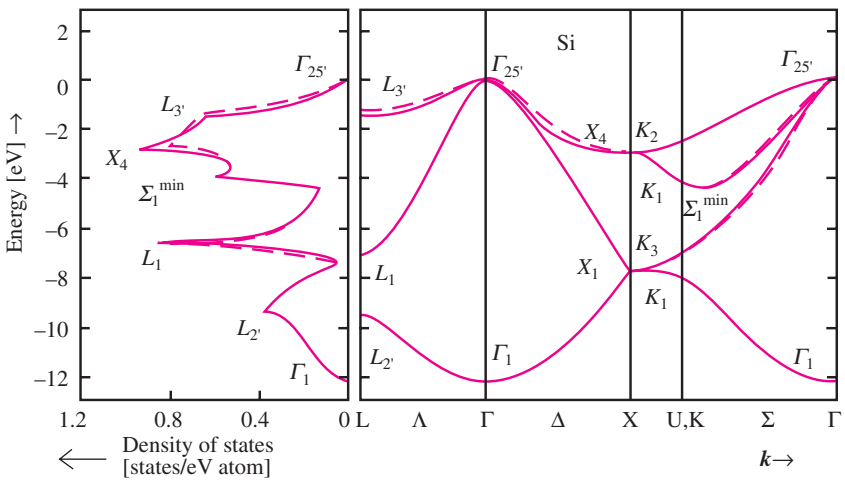
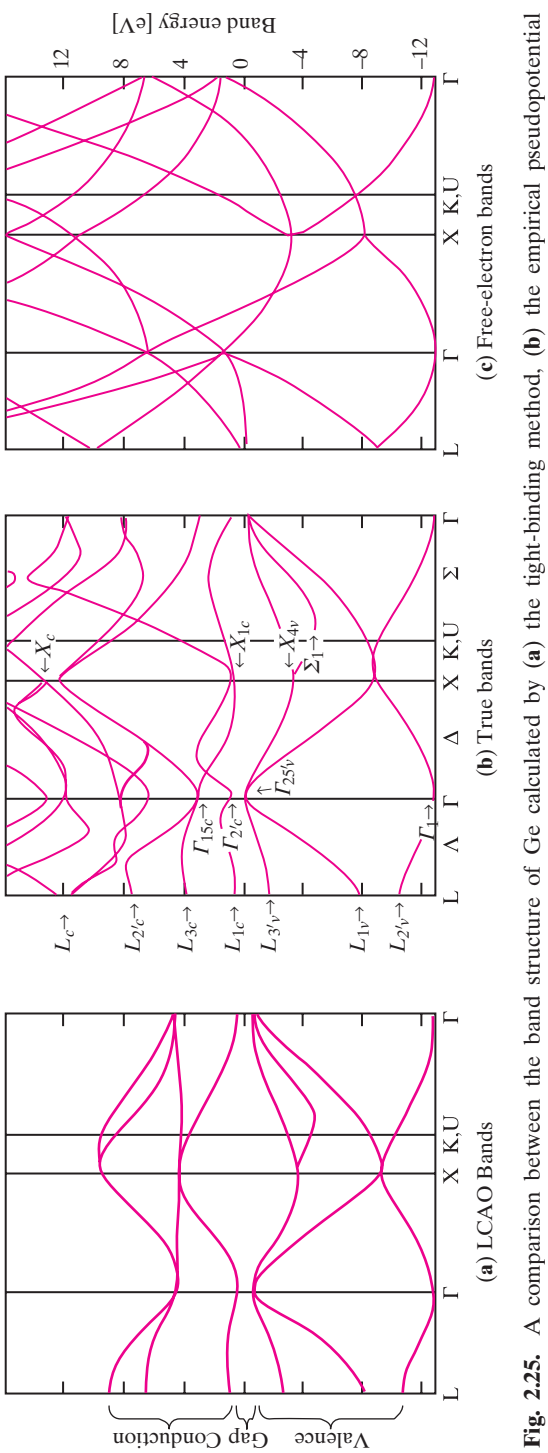


Fig. 2.24. The valence band structure and density of states (see Sect. 4.3.1 for definition) of Si calculated by the tight-binding method (*broken curves*) and by the empirical pseudopotential method (*solid lines*) [2.25]



(a) LCAO Bands (b) True bands (c) Free-electron bands

Fig. 2.25. A comparison between the band structure of Ge calculated by (**a**) the tight-binding method, (**b**) the empirical pseudopotential method, and (**c**) the nearly free electron model [Ref. 2.24, p. 79]

conduction band electrons are more delocalized. The accuracy of the conduction bands in the tight-binding calculations can be improved by introducting additional overlap parameters. However, there is another shortcoming in the tight-binding model presented here. There are only four conduction bands in this model because we have included only four s and p orbitals. To correct this problem additional orbitals and overlap parameters are required; unfortunately they destroy the simplicity of this model.

2.7.3 Overlap Parameters and Nearest-Neighbor Distances

So far we have shown that the advantage of the tight-binding approach is that the valence band structures of semiconductors can be calculated in terms of a small number of atomic energies and overlap parameters. Now we will demonstrate that these overlap parameters in different semiconductors can be expressed as a simple function of the nearest-neighbor distance multiplied by a geometric factor. These results combined make the tight-binding method very powerful for predicting the properties of many compounds (not just semiconductors) with only a small number of parameters [Ref. 2.24, p. 49].

One may expect some relationship between the overlap parameters and the interatomic distance based on the following simple argument. Figures 2.20 and 2.22 show that the atomic energy levels broaden into bands due to overlap of the atomic orbitals. The width of the band is essentially $2V$, where V is the relevant overlap parameter. At the same time the electron wave functions become delocalized over a distance given by the nearest-neighbor separation (i. e., the bond length) d as a result of this overlap. Using the uncertainty principle the momentum of the delocalized electron is estimated to be $(\hbar\pi/d)$, so the electron kinetic energy is given by $\hbar^2\pi^2/(2md^2)$. This result suggests that the overlap parameters depend on d as d^{-2}. This simple heuristic argument can be made more rigorous by comparing the band structures calculated by the tight-binding method and by the nearly free electron model. As an example, we will consider the lowest energy valence band in a crystal with the simple cubic structure. This band can be identified with the bonding s orbitals and its dispersion along the [100] direction can be shown to be given by $E_s - 4V_{ss\sigma} - 2V_{ss\sigma}\cos kx$ (Problem 2.16). Thus the width of this band is equal to $4V_{ss\sigma}$. On the other hand the nearly free electron model gives the band width as $\hbar^2\pi^2/(2md^2)$. Equating the band widths obtained by these two different methods we get

$$4V_{ss\sigma} = \frac{\hbar^2\pi^2}{2md^2}. \tag{2.85}$$

In general, all four overlap parameters for the s and p orbitals can be expressed in the form

$$V_{ll'm} = \tau_{ll'm}\frac{\hbar^2}{md^2} \tag{2.86}$$

where $\tau_{ll'm}$ is a factor which depends on the crystal symmetry. From (2.85) we see that $\tau_{ss\sigma} = \pi^2/8$ in crystals with the simple cubic structure. Table 2.27 lists the values of $\tau_{ll'm}$ for the simple cubic and diamond lattices.

For the diamond and zinc-blende crystals, *Harrison* [2.26] has treated the factors $\tau_{ll'm}$ as adjustable parameters in fitting the energy bands of Si and Ge. He found excellent agreement between the calculated values and the adjusted values for three of the parameters. The only exception is $\tau_{pp\pi}$, where the fitted value of -0.81 is somewhat lower than the calculated one.

Table 2.27. The geometric factor $\tau_{ll'm}$ relating the overlap parameters for the s and p bands to the free electron band width $\hbar^2/(md^2)$ as shown in (2.86). The last column represents the adjusted values obtained by fitting the energy bands of Si and Ge [Ref. 2.24, p. 49]

	Simple cubic	Diamond and zinc–blende	Adjusted values
$\tau_{ss\sigma}$	$-\pi^2/8 = -1.23$	$-9\pi^2/64 = -1.39$	-1.40
$\tau_{sp\sigma}$	$(\pi/2)[(\pi^2/4) - 1]^{1/2} = 1.90$	$(9\pi^2/32)[1 - (16/3\pi^2)]^{1/2} = 1.88$	1.84
$\tau_{pp\sigma}$	$3\pi^2/8 = 3.70$	$21\pi^2/64 = 3.24$	3.24
$\tau_{pp\pi}$	$-\pi^2/8 = -1.23$	$-3\pi^2/32 = -0.93$	-0.81

Table 2.27 together with (2.86) and the lattice constants are all that is needed to calculate the overlap parameters for computing the valence bands and the lowest conduction bands in many zinc-blende- and diamond-type semiconductors. Even without any detailed calculations, we can understand qualitatively the symmetries of the conduction and valence bands at the Brillouin zone center of the three group-IV elements Si, Ge, and gray tin (or α-Sn). The lattice constant increases from Si to α-Sn. This results in a decrease in the overlap parameters $|V_{ss}|$ and $|V_{xx}|$ (the variation from C to Ge is shown in Table 2.26). The decrease is larger for $|V_{ss}|$ than for $|V_{xx}|$. As a result, the ordering of the s and p orbitals changes from Si to α-Sn in the manner shown in Fig. 2.26. The Fermi level is located by filling the bands with the eight valence electrons available. In this way it is easily seen that the lowest conduction band at zone center in Si is p-like while the corresponding band in Ge is s-like. In this scheme α-Sn turns out to be a semi-metal because of the lower energies of the bands derived from the s orbitals. It was first shown by *Herman* [2.27] that relativistic effects are responsible for this in gray tin (and also in HgTe and HgSe. Note, however, that the s-p reversal for HgSe has recently been the object of controversy; see [2.28]).

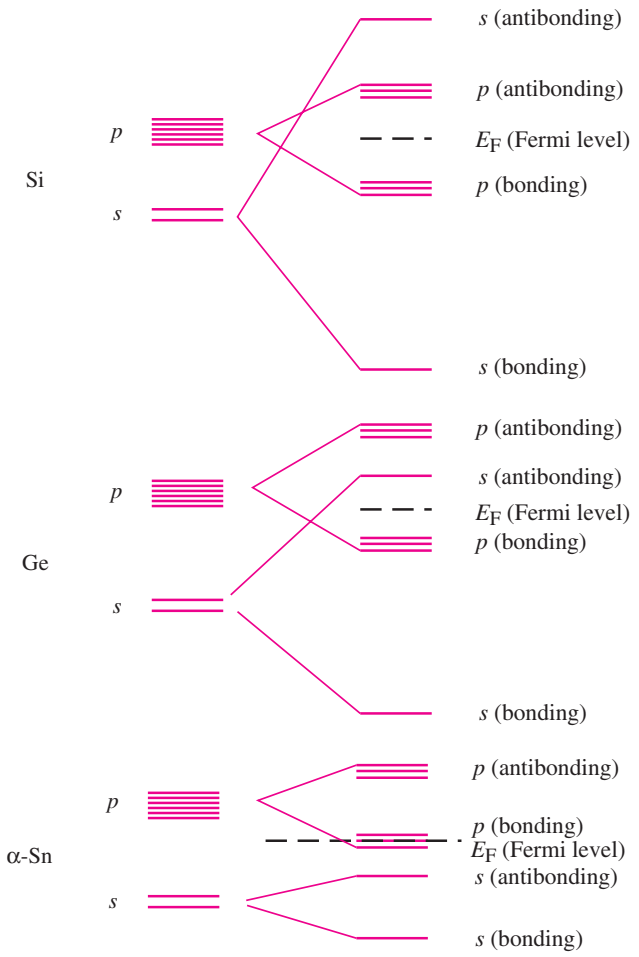


Fig. 2.26. Evolution of s and p atomic orbitals into the conduction and valence bands at zone center within the tight-binding approximation for Si, Ge, and α-Sn. The band ordering for diamond is similar to that of Si.

PROBLEMS

2.1 *Template of an fcc Brillouin Zone*

Construct a model of the Brillouin zone of the fcc lattice by pasting a copy of the template shown in Fig. 2.27 on cardboard and cutting along the solid lines. Score along the broken lines. Tape the edges together.

2.2 *Group Theory Exercises*

a) Verify the character table of the T_d point group as given in Table 2.3.

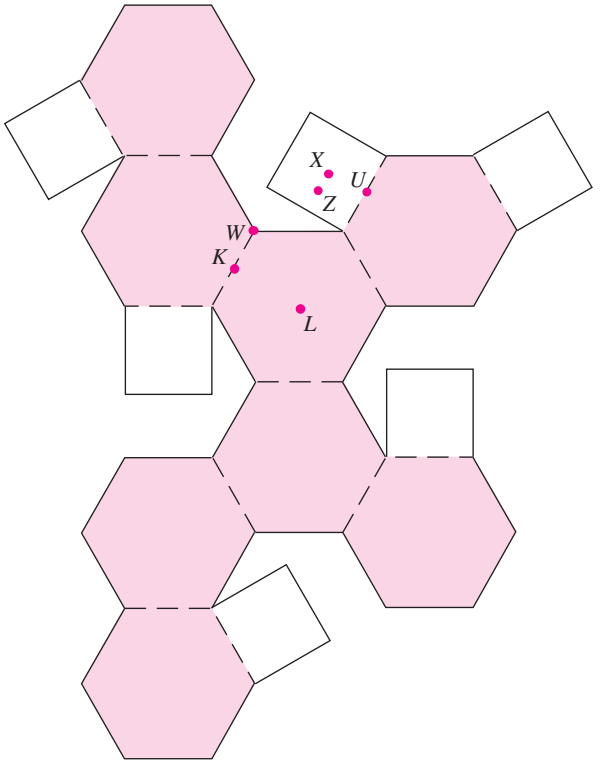


Fig. 2.27. Template for constructing a model of the Brillouin zone of the fcc lattice. Paste this sheet on thin cardboard and cut along *solid lines*. Score along *broken lines* and tape the joints

b) By applying the symmetry operations of T_d to the basis functions in Table 2.3 show that the functions transform according to their respective irreducible representations.

2.3 *Group Theory Exercises*

a) By using the character table of T_d show that $T_2 \otimes T_2 = T_1 \oplus T_2 \oplus E \oplus A_1$.

b) Verify that the symmetrized linear combinations of the matrix elements of a second-rank tensor given in Sect. 2.3.4 transform according to the irreducible representations T_1, T_2, E, and A_1.

2.4 *Symmetrized Wave Functions: Transformation Properties*

Verify that the symmetrized wave functions in Tables 2.9 and 2.10 transform according to their respective irreducible representations.

2.5 *Characters of C_{3v} and C_{2v} Point Groups*

Deduce by inspection the characters for the C_{3v} and C_{2v} point groups in Tables 2.12 and 2.14, respectively.

2.6 *Compatibility Relations*

Use Tables 2.3, 2.4, 2.12 and 2.14 to verify the following compatibility relations:

Γ_1	Δ_1	Λ_1
Γ_2	Δ_2	Λ_2
Γ_3	$\Delta_1 \oplus \Delta_2$	Λ_3
Γ_4	$\Delta_1 \oplus \Delta_3 \oplus \Delta_4$	$\Lambda_1 \oplus \Lambda_3$
Γ_5	$\Delta_2 \oplus \Delta_3 \oplus \Delta_4$	$\Lambda_2 \oplus \Lambda_3$.

2.7 *Representations of Nonsymmorphic Groups*

a) Using Tables 2.15 and 2.19 show that the doubly degenerate X_1 and X_2 states in the diamond crystal split into the $X_1 \oplus X_3$ and $X_2 \oplus X_4$ states, respectively, when the diamond crystal (nonsymmorphic) is transformed into a zinc-blende crystal (symmorphic) by making the two atoms in the primitive cell different. Under the same transformation the X_3 and X_4 states remain doubly degenerate and become the X_5 state in the zinc-blende crystal.

b) *Some Insight into the Doubly Degenerate Wavefunctions at the X point of the Brillouin Zone in the Diamond Structure.*

Within the free electron approximation, the wave functions at the X point of the Brillouin Zone can be written as: $\exp[\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}]$ where $\boldsymbol{k} = (2\pi/a)(\pm 1, 0, 0)$, $(2\pi/a)(0, \pm 1, 0)$, or $(2\pi/a)(0, 0, \pm 1)$. Let us consider the wave functions $\psi_1 = \sin[(2\pi/a)x]$ and $\psi_2 = \cos[(2\pi/a)x]$.

Assume that the crystal structure of diamond simply has inversion symmetry I: $(x, y, z) \rightarrow (-x, -y, -z)$; then by applying this symmetry operation to ψ_1 we obtain $-\psi_1$. Since the crystal is invariant under I we expect ψ_1 and $I\psi_1$ to have the same energy. We find that this is trivially satisfied since ψ_1 and $I\psi_1$ are linearly dependent. Thus we cannot conclude that the states ψ_1 and ψ_2 should be degenerate.

Now we take into account that the inversion operation in the diamond lattice is not simply I but rather I': $(x, y, z) \rightarrow (-x + (a/4), -y + (a/4), -z + (a/4))$. Applying I' to ψ_1 we find that: $I'\psi_1 = \sin[(2\pi/a)(-x + (a/4)] = \sin[(2\pi/a)(-x) + (\pi/2)] = -\cos[(2\pi/a)(-x)] = -\cos[(2\pi/a)x] = -\psi_2$. Since ψ_1 and ψ_2 are not linearly independent we have to conclude that ψ_1 and ψ_2 are degenerate from the fact that the crystal is invariant under I'. Similarly one can show that all the other plane wave states at the X point are doubly degenerate because of this symmetry operation I'.

2.8 *Pseudopotential Band Structure Calculation by Hand*

The purpose of this exercise is to show how pseudopotentials lift degeneracies in the nearly-free-electron band structure and open up energy gaps. Since the pseudopotentials are weak enough to be treated by perturbation theory, rather

accurate band energies can be evaluated with a pocket calculator without resorting to a large computer.

We will consider only the six lowest energy wave functions at the X point of a zinc-blende-type semiconductor. In the nearly-free-electron model, the electron wave functions are given by $\exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r})$, where $\boldsymbol{k} = (2\pi/a)(\pm 1, 0, 0)$ and $(2\pi/a)(0, \pm 1, \pm 1)$. For brevity these six wave functions will be denoted by $|100\rangle$, $|\bar{1}00\rangle$, $|011\rangle$, $|0\bar{1}1\rangle$, $|01\bar{1}\rangle$, and $|0\bar{1}\,\bar{1}\rangle$.

a) Show that these six wave functions can be symmetrized according to the following irreducible representations:

$$\psi_1 = (1/\sqrt{2})[|011\rangle + |0\bar{1}\,\bar{1}\rangle] \text{ and } \psi_2 = (1/\sqrt{2})[|011\rangle - |0\bar{1}\,\bar{1}\rangle] \leftrightarrow X_5$$
$$\psi_3 = (1/2)\{[|011\rangle - |0\bar{1}\,\bar{1}\rangle] + \mathrm{i}[|01\bar{1}\rangle + |0\bar{1}1\rangle]\} \leftrightarrow X_3$$
$$\psi_4 = (1/2)\{[|011\rangle - |0\bar{1}\,\bar{1}\rangle] - \mathrm{i}[|01\bar{1}\rangle + |0\bar{1}1\rangle]\} \leftrightarrow X_1$$
$$\psi_5 = (1/2)\{[|100\rangle + |\bar{1}00\rangle] + \mathrm{i}[|100\rangle - |\bar{1}00\rangle]\} \leftrightarrow X_1$$
$$\psi_6 = (1/2)\{[|100\rangle + |\bar{1}00\rangle] - \mathrm{i}[|100\rangle - |\bar{1}00\rangle]\} \leftrightarrow X_3$$

It should be noted that the pseudopotential form factors in Table 2.21 have been defined with the origin chosen to be the midpoint between the two atoms in the primitive cell. In order to conform with this coordinate system, the symmetry operations for the group of X have to be defined differently from those in Sect. 2.3.2. Taking the axes and planes of the point group to intersect at the midpoint some of the symmetry operations must involve a translation.

b) Calculate the matrix elements of the pseudopotential between these wave functions. This task can be greatly simplified by using the matrix element theorem. Since the pseudopotential V has the full symmetry of the crystal, it has Γ_1 symmetry. The only states that are coupled by V are then the X_3 states ψ_3 and ψ_6 and the X_1 states ψ_4 and ψ_5. Show that the resulting 6×6 matrix $\{V_{ij}\}$ is

$$\begin{vmatrix}
-v_8^s & 0 & 0 & 0 & 0 & 0 \\
0 & -v_8^s & 0 & 0 & 0 & 0 \\
0 & 0 & v_8^s + 2v_4^a & 0 & 0 & \mathrm{i}\sqrt{2}(-v_3^a - v_3^s) \\
0 & 0 & 0 & v_8^s - 2v_4^a & \mathrm{i}\sqrt{2}(-v_3^a + v_3^s) & 0 \\
0 & 0 & 0 & -\mathrm{i}\sqrt{2}(-v_3^a + v_3^s) & -v_4^a & 0 \\
0 & 0 & -\mathrm{i}\sqrt{2}(-v_3^a - v_3^s) & 0 & 0 & +v_4^a
\end{vmatrix}$$

c) Diagonalize the secular determinant

$$\left| \left(\frac{\hbar^2 k^2}{2m} - E \right) \delta_{ij} + V_{ij} \right| = 0$$

by reducing it to three 2×2 determinants. Show that the resultant energy levels are

$$E(X_5) = \frac{4\pi^2\hbar^2}{ma^2} - v_8^s,$$

$$E(X_1) = \frac{1}{2}\left(\frac{6\pi^2\hbar^2}{ma^2} + v_8^s - 3v_4^a\right)$$
$$\pm \frac{1}{2}\left[\left(\frac{2\pi^2\hbar^2}{ma^2} + v_8^s - v_4^a\right)^2 + 8(v_3^a - v_3^s)^2\right]^{1/2},$$
$$E(X_3) = \frac{1}{2}\left(\frac{6\pi^2\hbar^2}{ma^2} + v_8^s + 3v_4^a\right)$$
$$\pm \frac{1}{2}\left[\left(\frac{2\pi^2\hbar^2}{ma^2} + v_8^s + v_4^a\right)^2 + 8(v_3^a + v_3^s)^2\right]^{1/2}.$$

d) Calculate the energies of the X_1, X_3, and X_5 levels in GaAs by substituting into the expression in (c) the pseudopotential form factors for GaAs. Take the lattice parameter a to be 5.642 Å. In Fig. 2.28 these results are compared with the nearly free electron energies and with the energies obtained by the EPM.

e) If you want to improve on the present calculation, what are the plane wave states and pseudopotential form factors you should include?

Note: Often in the literature, the origin of the coordinates adopted by the authors is not specified.[4] The symmetry of the band structure at the X point of the zinc-blende-type crystal depends on the choice of origin and this has

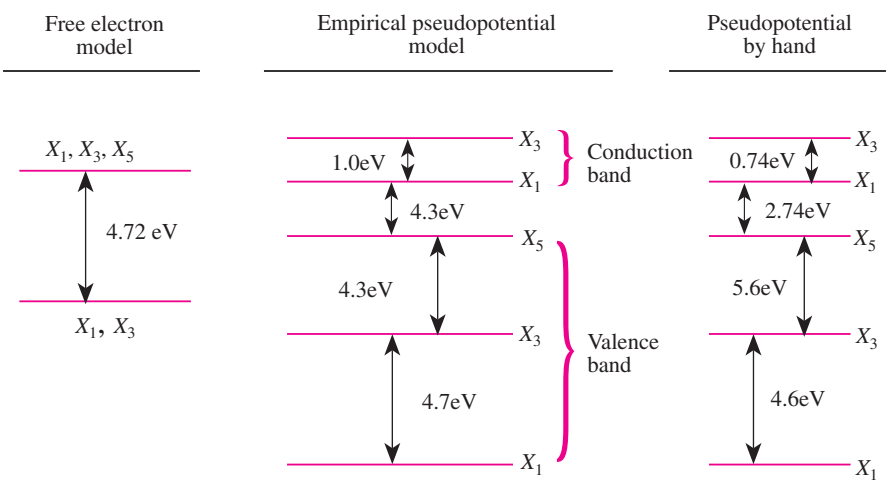


Fig. 2.28. The lowest energy bands at the X point of GaAs computed by the nearly free electron model, the EPM, and the perturbation approach of Problem 2.8. The X_1–X_3 notation corresponds to $v_j^a > 0$, i. e., to placing the cation at $(a/4)(111)$ and the anion at the origin.

[4] We assume, implicitly, that the origin is also the common point of the point group axes which specify the symmetry.

caused considerable confusion, see [2.29]. For example, if the origin is chosen at the anion the conduction band with the X_1 symmetry is mainly composed of the anion s wave function and cation p wave function. On the other hand, the X_3 conduction band state is made up of the cation s wave function and the anion p wave function. In all zinc-blende-type semiconductors with the exception of GaSb [2.30, 31] the X_1 state has lower energy than the X_3 state. If the origin is chosen at the cation, the signs of v_j^a and, correspondingly, the roles of X_1 and X_3 are reversed.

2.9 *Wave Functions of the L-Point of Zinc-Blende*
Using the symmetrized $\boldsymbol{k} = (2\pi/a)(1,1,1)$ wave functions in the nearly free electron model for zinc-blende-type crystals:

Γ_1: $\sqrt{8}\cos(2\pi x/a)\cos(2\pi y/a)\cos(2\pi z/a)$;
$\Gamma_4(x)$: $\sqrt{8}\sin(2\pi x/a)\cos(2\pi y/a)\cos(2\pi z/a)$,
and similar wave functions for $\Gamma_4(y)$ and $\Gamma_4(z)$ in Table 2.9,

show that the matrix elements of the momentum operator $\boldsymbol{p}$ between the Γ_1 and Γ_4 functions are given by

$$|\langle\Gamma_1\,|\,p_x\,|\,\Gamma_4(x)\rangle|^2 = |\langle\Gamma_1\,|\,p_y\,|\,\Gamma_4(y)\rangle|^2 = |\langle\Gamma_1\,|\,p_z\,|\,\Gamma_4(z)\rangle|^2 = (2\hbar\pi/a)^2$$

while all the other matrix elements of p_i such as $|\langle\Gamma_1\,|\,p_x\,|\,\Gamma_4(y)\rangle|^2$ are equal to 0.

2.10 *Double Group Representations*
In many quantum mechanics textbooks one can find the following result. The effect of a rotation by an infinitesimal amount $\delta\theta$ with respect to an axis defined by the unit vector $\hat{\boldsymbol{n}}$ on an orbital wave function $f(\boldsymbol{r})$ can be obtained by applying the operator $\exp[-\mathrm{i}\delta\Theta\boldsymbol{n}\cdot\boldsymbol{l}/\hbar]$ to $f(\boldsymbol{r})$. For a spin $s = 1/2$ particle the corresponding operator on the spin wave functions due to rotation by an angle Θ is given by $\exp[-\mathrm{i}\Theta\hat{\boldsymbol{n}}\cdot\boldsymbol{\sigma}/2]$. Using this operator, show that:

a) The effect of a 2π rotation on the wave functions α and β of a spin 1/2 particle is to change the sign of α and β, and hence the corresponding trace of $\hat{E}$ is -2;

b) the traces corresponding to the symmetry operations in Table 2.23 within the basis α and β are

$\{E\}$	$\{3C_2\}$ $\{3\hat{E}C_2\}$	$\{6S_4\}$	$\{6\sigma\}$ $\{6\hat{E}\sigma\}$	$\{8C_3\}$	$\{\hat{E}\}$	$\{6\hat{E}S_4\}$	$\{8\hat{E}C_3\}$
2	0	$\sqrt{2}$	0	1	-2	$-\sqrt{2}$	-1

c) *The Double Group at the X Point of the Zinc-Blende Structure*
As an additional exercise on the calculation of double group character table, we shall consider the X point of the zinc-blende structure.

The first step is to decide what are the classes in the double group. In this case we need only to compare the single group and double group classes at the zone center since the classes of X form a subset of these classes. It should not be difficult to see that there are now 7 classes:

$$\{E\},\ \{C_4^2(x), \hat{E}C_4^2(x)\},\ \{2C_4^2(y,z), 2\hat{E}C_4^2(y,z)\},\ \{2S_4\},\ \{2m_d\},\ \{\hat{E}\} \text{ and } \{\hat{E}S_4\}.$$

Using the results of Problem 2.10 one can show that the characters of these operations on the two spin wavefunctions are:

E	$C_4^2(x), \hat{E}C_4^2(x)$	$2C_4^2(y,z), 2\hat{E}C_4^2(y,z)$	$2S_4$	$2m_d$	$\hat{E}$	$\hat{E}S_4$
2	0	0	$\sqrt{2}$	0	-2	$-\sqrt{2}$

Using this result we can show that the character table for the double group of the X point in the zinc-blende crystal is:

	E	$C_4^2(x), \hat{E}C_4^2(x)$	$2C_4^2(y,z), 2\hat{E}C_4^2(y,z)$	$2S_4$	$2m_d$	$\hat{E}$	$\hat{E}S_4$
X_1	1	1	1	1	1	1	1
X_2	1	1	1	-1	-1	1	-1
X_3	1	1	-1	-1	1	1	-1
X_4	1	1	-1	1	-1	1	1
X_5	2	-2	0	0	0	2	0
X_6	2	0	0	$\sqrt{2}$	0	-2	$-\sqrt{2}$
X_7	2	0	0	$-\sqrt{2}$	0	-2	$\sqrt{2}$

Using these characters the reader should show that the X_1, X_3, and X_5 representations in the zinc-blende structure (see Problem 2.10) go over to the $X_6 \otimes X_1 = X_6$, $X_6 \otimes X_3 = X_7$ and $X_6 \otimes X_5 = X_6 \oplus X_7$ representations in the double group (see, for example, the band structure of GaAs in Fig. 2.14).

2.11 *The Structure Factor of Bond Charges in Si*

The intensity of x-ray scattering peaks from a crystal depends on the structure factor S of the crystal.

The structure factor of the Si crystal (face centered cubic or fcc lattice), in particular, is discussed in many standard textbooks on solid state physics, such as Kittel's "Introduction to Solid State Physics" (Chap. 2 in 6th Edition). The basis of the fcc structure is usually taken to be the cubic unit cell with four atoms per unit cube. These four atoms can be chosen to have the locations at $(0,0,0)$; $(0,1/2,1/2)$; $(1/2,0,1/2)$ and $(1/2,1/2,0)$ [in units of the size of the cube: a]. The structure factor $S_{\text{fcc}}(hkl)$ for a wave vector (h,k,l) in reciprocal space then vanishes if the integers h, k and l contain a mixture of even and odd numbers. In the case of the Si crystal there are now 8 atoms per unit cube since there are two interpenetrating fcc sublattices displaced from each other by the distance $(1/4,1/4,1/4)$. As a result, the structure factor of the Si crystal

$S_{Si}(hkl)$ is given by:

$$S_{Si}(hkl) = S_{fcc}(hkl)[1 + \exp(i\pi/2)(k + k + l)].$$

This implies that $S_{Si}(hkl)$ will be zero if the sum $(h + k + l)$ is equal to 2 times an odd integer. When combining the above two conditions one obtains the result that $S_{Si}(hkl)$ will be non-zero only if (1) (k, k, l) contains only even numbers and (2) the sum $(h + k + l)$ is equal to 4 times an integer. See, for example, Kittel's "Introduction to Solid State Physics" (Chap. 2 in 6th Edition), Problem 5 at the end of Chap. 2. Based on this result one expects that the diffraction spot corresponding to $(2, 2, 2)$ in the x-ray diffraction pattern of Si will have zero intensity since $h + k + l = 6$.

It has been known since 1959 that the so-called forbidden $(2, 2, 2)$ diffraction spot in diamond has non-zero intensity (see Ref. [3.23] or Kittel's "Introduction to Solid State Physics", p. 73 in 3rd Edition). It is now well established that the presence of this forbidden $(2, 2, 2)$ diffraction spot can be explained by the existence of bond charges located approximately mid-way between the atoms in diamond or silicon. What is the structure factor of the bond charges in the Si crystal if one assumes that they are located exactly mid-way between two Si atoms?

2.12 *Matrix Elements of* $\boldsymbol{p}$

a) Show that all matrix elements of $\boldsymbol{p}$ between the Γ_4 valence bands and the Γ_4 conduction bands of zinc-blende-type semiconductors of the form $\langle X \,|\, p_x \,|\, \Gamma_{4c}(z)\rangle$, $\langle Z \,|\, p_y \,|\, \Gamma_{4c}(z)\rangle$, or $\langle X \,|\, p_y \,|\, \Gamma_{4c}(y)\rangle$, where at least two of the labels x, y, or z are identical, vanish as a result of the requirement that the crystal is invariant under rotation by 180° with respect to one of the three equivalent [100] axes.

b) As a result of (a), the only nonzero matrix elements of $\boldsymbol{p}$ are of the form $\langle X \,|\, p_y \,|\, \Gamma_{4c}(z)\rangle$. Using the three-fold rotational symmetries of the zinc-blende crystal, show that

$$\langle X \,|\, p_y \,|\, \Gamma_{4c}(z)\rangle = \langle Y \,|\, p_z \,|\, \Gamma_{4c}(x)\rangle = \langle Z \,|\, p_x \,|\, \Gamma_{4c}(y)\rangle$$

and

$$\langle X \,|\, p_z \,|\, \Gamma_{4c}(y)\rangle = \langle Y \,|\, p_x \,|\, \Gamma_{4c}(z)\rangle = \langle Z \,|\, p_y \,|\, \Gamma_{4c}(x)\rangle.$$

c) Finally, use the reflection symmetry with respect to the (110) planes to show that

$$\langle X \,|\, p_y \,|\, \Gamma_{4c}(z)\rangle = \langle Y \,|\, p_x \,|\, \Gamma_{4c}(z)\rangle.$$

2.13 *Linear Terms in* $\boldsymbol{k}$

Show that the k linear term due to the $\boldsymbol{k} \cdot \boldsymbol{p}$ interaction is zero in the zinc-blende crystal at the Γ-point.

2.14 *$\boldsymbol{k}\cdot\boldsymbol{p}$ Method*

a) Use (2.52) to calculate the elements of the 6×6 matrix $\{H'_{ij}\}$.

b) Use a computer and a matrix diagonalization program to calculate the valence band structure of GaAs from these parameters for GaAs: $P^2/(m_0) = 13$ eV; $Q^2/(m_0) = 6$ eV; $E_0 = 1.519$ eV; $E'_0 = 4.488$ eV; $\Delta = 0.34$ eV and $\Delta'_0 = 0.171$ eV.

2.15 *Valence Bands; $\boldsymbol{k}\cdot\boldsymbol{p}$ Hamiltonian*

a) Calculate the 4×4 matrix obtained by taking matrix elements of the Luttinger Hamiltonian in (2.70) between the $J_z = \pm 3/2$ and $\pm 1/2$ states of the $J = 3/2$ manifold.

b) Diagonalize this 4×4 matrix to obtain two sets of doubly degenerate levels with energies

$$E_{\pm} = \frac{\hbar^2}{2m}\{\gamma_1 k^2 \pm [4\gamma_2^2 k^4 + 12(\gamma_3^2 - \gamma_2^2)(k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2)]^{1/2}\}.$$

c) By comparing the results in (b) with (2.66) derive (2.71).

d) Calculate the contributions of the lowest Γ_3^- conduction band term to γ_1, γ_2, and γ_3. Show that it is not negligible for silicon and diamond [2.22].

2.16 *Energy Bands of a Semiconductor in the Tight-Binding Model*

a) Derive the 8×8 matrix for the s and p band energies in a zinc-blende-type semiconductor using the tight-binding model.

b) Show that at $k = 0$ the energies of the s and p bands are given by

$$E_{s\pm}(0) = \tfrac{1}{2}(E_{s1} + E_{s2}) \pm \tfrac{1}{2}[(E_{s1} - E_{s2})^2 + 4\,|\,V_{ss}\,|^2]^{1/2}$$

and

$$E_{p\pm}(0) = \tfrac{1}{2}(E_{p1} + E_{p2}) \pm \tfrac{1}{2}[(E_{p1} - E_{p2})^2 + 4\,|\,V_{xx}\,|^2]^{1/2}$$

instead of (2.84a) and (2.84b). E_{s1} and E_{s2} are the atomic s level energies $\langle S1\,|\,\mathcal{H}_0\,|\,S1\rangle$ and $\langle S2\,|\,\mathcal{H}_0\,|\,S2\rangle$, respectively, while E_{p1} and E_{p2} are the corresponding energies for the atomic p levels.

2.17 *Tight Binding Overlap Integrals*

Evaluate the geometric factors $\tau_{ll'm}$ in Table 2.27.

2.18 *Tight Binding Hamiltonian*

Given two p orbitals, one located at the origin and the other at the point $d(\cos\Theta_x,\ \cos\Theta_y,\ \cos\Theta_z)$, where d is the distance between the two p orbitals and $\cos\Theta_x$, $\cos\Theta_y$, and $\cos\Theta_z$ are the directional cosines of the second p orbital, show that the overlap parameters V_{xx} and V_{xy} are given by

$$V_{xx} = V_{pp\sigma}\cos^2\Theta_x + V_{pp\pi}\sin^2\Theta_x,$$

$$V_{xy} = [V_{pp\sigma} - V_{pp\pi}]\cos\Theta_x\cos\Theta_y.$$

2.19 *Conduction and Light Hole Bands in Small Band Gap Semiconductors*

Write down the 2×2 Hamiltonian matrix which describes the conduction and

the light hole band of a narrow gap semiconductor such as InSb. Diagonalize it and discuss the similarity of the resulting expression with the relativistic energy of free electrons and positrons [4.28]. Use that expression to estimate non-parabolicity effects on the conduction band mass.

Comments on the Equivalence of the Brillouin Zone Edge Points U and K 3
In these figures the zone-edge points U and K are indicated to be equivalent. This is true strictly only for Bloch waves indexed by these points which lie on the Brillouin Zone (BZ) surface but not for for waves with wave vector along the *lines* joining them to X nor to Γ. This is easily seen from the fcc lattice template in Fig. 2.27. Both U and K lie on the hexagonal Brillouin zone face along the [111] direction. However, the line X-U is in general not equivalent to the line X-K. For the endpoints U and K which lie on the BZ boundary their Bloch waves become equivalent in the diamond lattice as a result of the three-fold rotation and the "inversion symmetry" operations of the lattice. Also only the line Γ-K is along the [110] or Σ direction so the label for the horizontal axis Γ-Σ-U,K in these figures refers to K only. In the zincblende structure without the "inversion symmetry" U and K are no longer equivalent as can be seen from Fig. 2.14 and 2.15. In Figures 3.2, 3.10 and 3.11 this problem is avoided by not labeling the symmetry points indicated by a broken vertical line in these figures. This symmetry point is really K as we can conclude from the above discussion. We are grateful to Dr. L.C. Andreani of the University of Pavia, Italy for pointing out this ambiguity to us.

2.20 *Parity of the s-like Electron Wavefunctions at the **L** point of Si and Ge*
In discussing the valence and conduction band structures of diamond-type semiconductors based on the tight-binding approximation, we point out that the valence bands are formed from the bonding orbitals while the conduction bands are formed from the anti-bonding orbitals. Furthermore, we note that the bonding orbitals have even parity under inversion in order that the electron charge density be higher at the center of the bond. On the other hand, the antibonding orbitals have odd parity under inversion. This result explains why the valence bands at the zone center of Si (see **Fig. 2.10**) and Ge (see **Fig. 2.13**) have even parity while the conduction bands at the zone center have odd parity. This is not true at the L point of the Brillouin zone. [It should be noted that along many high symmetry directions, such as Δ and Λ, inversion is not a symmetry operation since k and $-k$ are not identical. However, at the L point k and $-k$ are the same in the reduced zone scheme and hence parity is a good quantum number.].

The lowest energy valence band state at the L-point (which arises from the bonding s-orbital) has odd parity (L_6^- in Ge and $L_{2'}$ in Si) while the lowest energy conduction band (which arises mainly from the anti-bonding s-orbital) has even parity (L_6^+ in Ge and L_1 in Si) under the special inversion operator I' of the diamond lattice.

(a) Calculate the energy and wave functions of these bands at the L point directly by using the nearly-free electron model (use reasonable approximations to simplify the problem as far as possible) and check the parity of the resulting wave functions under the "inversion" operation I'.

(b) Repeat your calculation by using the tight-binding model and check again the parity of the resulting wave functions under the operation I'.

(c) Why does the odd parity state have lower energy rather than the even parity state at the L point? Does this result somehow violate the "compatibility relation" in changing a bonding state at $k = 0$ into an anti-bonding state at the L point?

SUMMARY

A semiconductor sample contains a very large number of atoms. Hence a quantitative quantum mechanical calculation of its physical properties constitutes a rather formidable task. This task can be enormously simplified by bringing into play the symmetry properties of the crystal lattice, i. e., by using group theory. We have shown how wave functions of electrons and vibrational modes (phonons) can be classified according to their behavior under symmetry operations. These classifications involve *irreducible representations* of the group of symmetry operations. The translational symmetry of crystals led us to *Bloch's theorem* and the introduction of *Bloch functions* for the electrons. We have learnt that their eigenfunctions can be indexed by wave vectors (Bloch vectors) which can be confined to a portion of the reciprocal space called the first *Brillouin zone*. Similarly, their energy eigenvalues can be represented as functions of wave vectors inside the first Brillouin zone, the so-called electron *energy bands*. We have reviewed the following main methods for calculating energy bands of semiconductors: the *empirical pseudopotential* method, the *tight-binding* or *linear combination of atomic orbitals* (LCAO) method and the $\boldsymbol{k}\cdot\boldsymbol{p}$ method. We have performed simplified versions of these calculations in order to illustrate the main features of the energy bands in diamond- and zinc-blende-type semiconductors.

http://www.springer.com/978-3-642-00709-5

Fundamentals of Semiconductors
Physics and Materials Properties
YU, P.; Cardona, M.
2010, XX, 775 p., Hardcover
ISBN: 978-3-642-00709-5