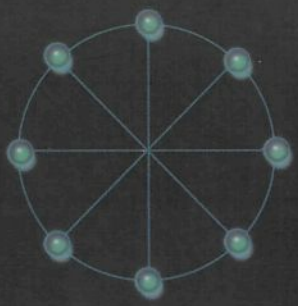


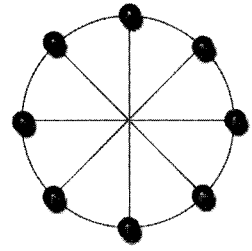
Introduction to Phonons and Electrons

Liang-fu Lou

World Scientific

Introduction to Phonons and Electrons

This page is intentionally left blank



Introduction to Phonons and Electrons

Liang-fu Lou
Former Member of Technical Staff
Northrop Grumman Corporation

World Scientific
New Jersey • London • Singapore • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: Suite 202, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

INTRODUCTION TO PHONONS AND ELECTRONS

ISBN 981-238-439-1
ISBN 981-238-461-8 (pbk)

Printed in Singapore by World Scientific Printers (S) Pte Ltd

In memory of my parents
and
to those
who are searching for
beauty and truth
in the world of condensed matters

This page is intentionally left blank

Preface

This book is based on my lecture notes developed for the undergraduate physics students. The lecture notes have two distinct parts. The first part goes into this book that focuses on achieving the basic understanding of phonons and electrons in crystalline solids. The second part is a selected set of classical and contemporary topics. It introduces to the students the diverse beauty of condensed matter physics. Since the selection is highly dependent on the personal interests of individual instructors, and some of the topics are still evolving, it is my judgment not to include the second part of my lecture notes here in this book for the time being. The title "Introduction to Phonons and Electrons" of the present book reflects this decision.

Both the phonons and electrons in crystalline solids form the two corner stones of the vast field of condensed matter physics. In order to embark on the study of these two topics, the students need to have a proper background in the quantum and statistical mechanics. However, many students who take the introductory solid state course are not yet comfortable with these tools. For this reason, students are frequently asked to accept the quantum and statistical results as axioms and just to learn how to apply them to solid state phenomena. The obvious drawback of this approach is from the shaky roots. The students can inadvertently end up with the misconceptions that would be difficult to rectify. In an attempt to minimize this undesirable outcome, this book makes an extra effort to bring together the fundamental building blocks. When required, the quantum theory and statistical physics are derived from scratch. The treatment is systematic in structure and tutorial in style. It contains the normally skipped details for the beginning students. This approach ensures a self-sufficient content for easier teaching and learning. The topics discussed are: crystal symmetries;

lattices; reciprocal lattices; x-ray diffraction; lattice vibrations; phonons; boson statistics; lattice thermal properties; free electrons; Fermion statistics; energy bands. The approach is sharply focused on building the concepts of phonons and electrons. The objective is to prepare the students with these essential concepts so that they do not have to relearn them in the more advanced solid state course.

This book can be used as the beginning module of a one-year introduction course to solid state physics, and the instructor will have a chance to choose the additional topics. Or, it can be taught as a stand-alone text for building the most needed foundation in just one semester. In the recent years, the long program was tried at Zhejiang University, and the short program was taught at Shanghai Jiao Tong University.

This book is suitable for the students who are learning solid state physics for the first time. Their field of study can range from physics to materials sciences and to solid state electronics. Both the teacher and students are encouraged to finish this book in its entirety over the course of one semester.

I would like to thank Zhejiang University for the visiting opportunity. It gave me a rare chance to explore in depth in the region around Hangzhou, where scenery, culture, and history are plentiful. The tea from the nearby Long-Jing Mountain hills is the best that China offers. It can provide the needed inspiration when one is writing something, even if it is a textbook. The regional rice wine is also the most famous, known throughout Chinese history. A couple of glasses could probably make any conscientious author feel comfortable with some of the overlooked typos and errors. For that, your comments and corrections will be appreciated. For my students who collectively helped with the original computer files, I wish them the best in their future endeavors. I would also like to thank Shanghai Jiao Tong University, where the teaching opportunity gave further tests and refinements to the manuscript. Special thanks go to Prof. R.O. Pohl of Cornell University and Prof. J. J. Chang of Wayne State University. Both of them spent their precious time to read the manuscript prior to the publication. The most appreciated help comes from Dr. K. J. Dunn of Cheveron Texaco Company and Dr. N. Wu from Raytheon Company. Without their involvement, the typesetting would have been much more difficult. I thank them both.

Liang-fu Lou
Palos Verdes Peninsula, California

Contents

Chapter 1

Crystal Structure

Introduction and Summary

Chapter 1 is a condensed but sufficient introduction to crystal structures. It is not intended to be a study of crystallography. This chapter is organized around the center theme of translation symmetry, which is highlighted at the beginning when introducing the definitions for lattice, basis, primitive lattice vectors, and primitive cells. The rotation symmetry is shown to be limited with a proof of the lack of 5, 7, and higher fold symmetry, if the translation symmetry is pre-required. This leads to the appreciation for why there are only 7 crystal types and 14 distinct lattices. The Bravais lattice is discussed in terms of both the qualitative and quantitative definitions. Whether a regular structure is a Bravais lattice or not must be examined against one or both of these definitions. As an example, it is shown that the two-dimensional honeycomb structure can not be viewed as a Bravais lattice unless the proper primitive cell and basis are selected. Along the same line, cubic lattices of sc, bcc, and fcc, close-packed lattices of fcc and hcp, and diamond structure are discussed. The advantage of visualizing the crystal symmetry makes the conventional cell the preferred one over the primitive cell in many cases. Precaution is given about the need to pick the correct basis in this conventional approach. The Wigner-Seitz primitive cell is introduced to prepare for future introduction of the first Brillouin zone. A final brief remark is given on structures beyond single crystals.

1.1 Lattice, Basis, Translation Symmetry, Primitive Lattice Vector, and Primitive Cell

The atom or group of atoms in crystalline solids appears to reproduce itself periodically through out the specimen, as if they were attached to an underlying set of points with the corresponding periodicity. These space points are not physical entities, but an analytical construction obtained from examining the crystalline geometry. These points are traditionally called lattice points, and the atom or group of atoms associated with each lattice point is the basis. Crystal structure is just lattice plus basis.

The atomic arrangement in a crystal looks identical regardless of being viewed from which lattice point, provided that it is far away from the surface. Translation invariance is how we formally describe this observation of equivalence. Specimens built with translation invariant structures are known as crystals or more precisely single crystals. Single crystals found in mineral deposits such as diamond, quartz, and some gemstones are formed inside the earth crest by high temperature and pressure. All elements and compounds can grow into their respective single crystals in the laboratory if the suitable thermodynamic environment can be created.

A lattice vector is a vector pointing to a lattice point from the origin, which can be an arbitrarily chosen lattice point. We define the lattice vector $\mathbf{R}$ by

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3. \tag{1.1}$$

The vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are unit vectors not all in the same plane. With n_1, n_2, and n_3 running through all integers, we can generate a complete set of lattice vectors or a complete set of lattice points. The lattice so generated will have translation symmetry. Namely, the lattice looks identical from point $\mathbf{r}$ and point $\mathbf{r}'$ if $\mathbf{r}' = \mathbf{r} + \mathbf{R}$.

The volume occupied by $\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3$ is the smallest building block that can fill the whole space by step-and-repeat processes along three directions of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. This cell is the primitive cell and $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are primitive lattice vectors. The primitive cell so defined has one lattice point on each of its eight corners, and each one belongs equally to eight neighboring cells. Therefore, each primitive cell contains only one lattice point. This is true regardless of how we choose the primitive cell.

1.2 Rotation Symmetry

A lattice can have more symmetry in addition to translation symmetry. For example, a cubic lattice looks the same after a $\pi/2$ rotation around one of its cubic edge. Translation operation moves the origin by a lattice vector $\mathbf{R}$, while keeping the coordinate axes parallel to the old system. Rotation operation rotates the coordinate by an angle, but leaves the origin unchanged. The symmetry operation that leaves the origin fixed belongs to a point group. Other symmetry operations in the point group include inversion and mirror reflection. The combination of point-group and translation operations belongs to the full space group.

The lattice with translation symmetry is identical when viewed from either point $\mathbf{r}$ or $\mathbf{r} + \mathbf{R}$, where $\mathbf{R}$ is any lattice vector including $-\mathbf{R}$. It implies that the paired points $\mathbf{r} + \mathbf{R}$ and $\mathbf{r}$ - $\mathbf{R}$ are equivalent. Therefore, the lattice with translation symmetry must be invariant under the inversion operation that changes $\mathbf{R}$ into $-\mathbf{R}$. Mirror reflection is an inversion after a π rotation around the axis perpendicular to the mirror. Both inversion and mirror reflection can be viewed as derived symmetry operations. Translation and rotation symmetries are more basic.

Because of the prerequisite translation symmetry, crystals can not have rotation symmetry for arbitrary angles. In fact, the allowed rotation symmetry is rather limited only for angles of $\pm\frac{2\pi}{6}$, $\pm\frac{2\pi}{4}$, $\pm\frac{2\pi}{3}$, $\pm\frac{2\pi}{2}$, and $\pm\frac{2\pi}{1}$. They are 1, 2, 3, 4, and 6 fold symmetries. There are no 5, 7, and higher fold symmetries. Let us prove it.

If the lattice is invariant under a rotation of θ as shown in Fig. 1.1(a), to repeat the rotation several times will have to return to the original configuration. It implies $\mathrm{m}\theta = 2\pi$. There must be lattice points on the ring corresponding to each consecutive rotation of θ. In other words, the lattice is invariant under rotations of $\mathrm{m}\theta$ for m = 0, 1, 2,...,m. Since a rotation of $-\theta$ is identical to the rotation of $(m-1)\theta$, the lattice must be also invariant under a rotation of $-\theta$.

Figure 1.1(b) shows a plane perpendicular to the rotation axis. A and B are two representative lattice points separated by the primitive lattice constant a. The lattice looks exactly the same from either point A or B as required by translation symmetry. If the lattice is invariant under a rotation of θ around point A, it must be also invariant under a rotation of $-\theta$ around the same axis. Consequently, the lattice must be invariant under a rotation of $-\theta$ around point B because the equivalence of point A and B. So, there must be a lattice point at B' with respect to θ rotation

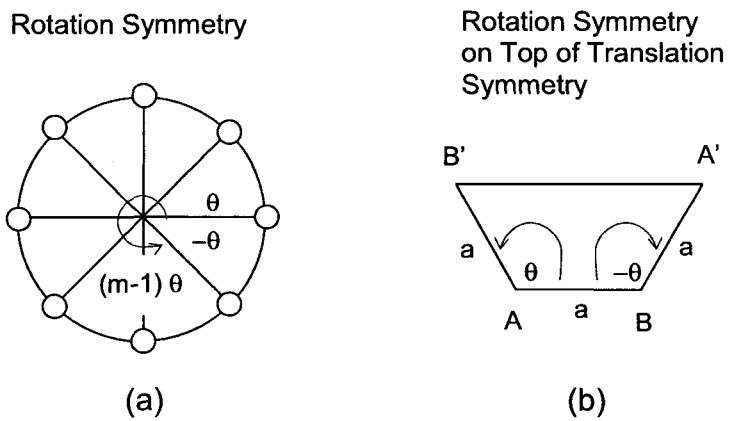


Fig. 1.1 (a) Necessary properties of rotation symmetry. Small circles are lattice points that are invariant under any rotation from θ to $m\theta$. We have $m\theta = 2\pi$. The rotation of $(m-1)\theta$ is identical to the rotation of $-\theta$. (b) Rotation symmetry on top of the required translation symmetry. If the lattice is invariant under a rotation of θ around A, it must be also invariant under the rotation of $-\theta$ around B. The two lattice points A and B are separated by the lattice constant a. The translation invariance requires that the distance A'B' equals na.

around point A, and a lattice point at A' for the rotation of $-\theta$ around point B. Translation symmetry requires that any two lattice-points must be connected by a lattice vector. In this case, the lattice vector between A' and B' has a length equal to $n \cdot a$ with the integer n to be determined. On the other hand, the geometry of Fig. 1.1(b) shows that the length A'B' also equals $a - 2a \cdot cos\theta$. Therefore, the condition for allowed angles of rotation symmetry in the lattice with prerequisite translation symmetry is

$$a - 2a \cdot cos\theta = n \cdot a. \tag{1.2}$$

The only possible integer solutions for n are 0, 1, 2, 3, and -1. They correspond to the allowed rotation angles of $\pm\frac{2\pi}{6}$, $\pm\frac{2\pi}{4}$, $\pm\frac{2\pi}{3}$, $\pm\frac{2\pi}{2}$, and $\pm\frac{2\pi}{1}$. There are only 1, 2, 3, 4, and 6 fold symmetries. There are no other rotation angles allowed, including 5, 7, and higher fold symmetries. The reason is simple, because these other angles do not satisfy the underlying translation symmetry. For example, fivefold symmetry is only possible if the translation symmetry is relaxed as in the AlMn alloy of 14% atomic Mn. It is not a crystal but a quasicrystal, because the lack of translation

symmetry.

1.3 Bravais Lattice

It is the extent of symmetry in rotation, inversion, and mirror reflection that gives rise to the different appearances of various crystals. There are total seven distinct crystal shapes: cubic, tetragonal, orthorhombic, monoclinic, triclinic, trigonal, and hexagonal. Each one has its own characteristic symmetry accountable by point-group operations. One crystal type may have several underlying lattice types characterized by linearly independent sets of primitive lattice vectors. For example, the cubic crystal may have the simple, body-centered, or face-centered cubic lattices. All share the characteristic cubic symmetry, but they are distinguished by different translation symmetry. There are a total of 14 distinct lattice types. They are simple, body-centered, and face-centered cubic; simple, and centered tetragonal; simple, base-centered, body-centered, and face-centered orthorhombic; simple, and centered monoclinic; triclinic; trigonal; hexagonal. A. Bravais first counted the number of lattice types correctly in 1845. Therefore, the 14 distinct lattice types are called Bravais lattices that exhaust all possible space lattices for building single crystals.

The advantage of categorizing space lattice into 14 Bravais lattices is for their reflection of actual crystal appearances, but the underlying translation symmetry may not be immediately clear. In fact, translation symmetry underlies all Bravais lattices and 14 of them exhaust all lattice types that have translation symmetry. Therefore, translation invariance is the necessary and sufficient condition for a set of space points to be one of the Bravais lattices. We give two equivalent definitions of Bravais lattice:

(a) The Bravais lattice is an infinite array of discrete points. From whichever points in it, the array looks exactly the same including arrangement and orientation of all points.

(b) The Bravais lattice consists of all points generated by the vector

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3,$$

where $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are three shortest vectors pointing to different directions that are not all in the same plane. The coefficient n_1, n_2, and n_3 range through all integers. Definition (a) is a qualitative one and (b) is a quantitative one. Two are totally equivalent. The vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$

in definition (b) are primitive lattice vectors. They generate or span the lattice.

1.4 Honeycomb Structure

Honeycomb is a waxy structure built by bees in their hives. Its basic building block is a tube with six sides of equal dimensions. Obviously, the bees can fill the hive space without any gaps by continuously adding more building blocks to the exposed sides. It is instructive to see whether the honeycomb structure is a Bravais lattice. A cross section of honeycomb is shown in Fig. 1.2(a), where we see the angle between any two sides is 120°. At the first glance, it looks like a Bravais lattice because the six-sided cell does fill the space without any gaps. In addition, the structure looks identical from either point P_2 or point b. However, closer examination shows that points b and c are not equivalent because the orientations around them are 180° apart. So, it is not completely symmetrical under translation operations. To illustrate this point quantitatively, a Cartesian coordinate is chosen to have its origin at point P_0, and the two adjacent sides of the honeycomb are chosen to be the possible primitive lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$. We have

$$\begin{aligned} \mathbf{a}_1 &= a\hat{\mathbf{x}}; \\ \mathbf{a}_2 &= -\frac{a}{2}\hat{\mathbf{x}} + \frac{\sqrt{3}a}{2}\hat{\mathbf{y}}. \end{aligned} \tag{1.3}$$

Indeed, the vector $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2$ can generate all points in the honeycomb structure, except that it also generates an extra point at the center of every honeycomb cell. We have

Table 1.1

(n_1, n_2)	*Point*
$(0, 0)$	P_0
$(1, 0)$	P_1
$(0, 1)$	P_5
$(2, 1)$	P_2
$(1, 2)$	P_4
$(2, 2)$	P_3
$(1, 1)$	P_c

The two-dimensional honeycomb structure with an additional point in the

cell center is just the hexagonal structure. A set of hexagonal lattice point is generated when attempting to generate the honeycomb structure. Results are shown in Fig. 1.2(b). Now, the points P_2, b, and c are all equivalent in terms of lattice arrangement and orientation around them. Therefore, the two-dimensional hexagonal structure is a Bravais lattice, and the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ defined above are its primitive unit vectors. The honeycomb

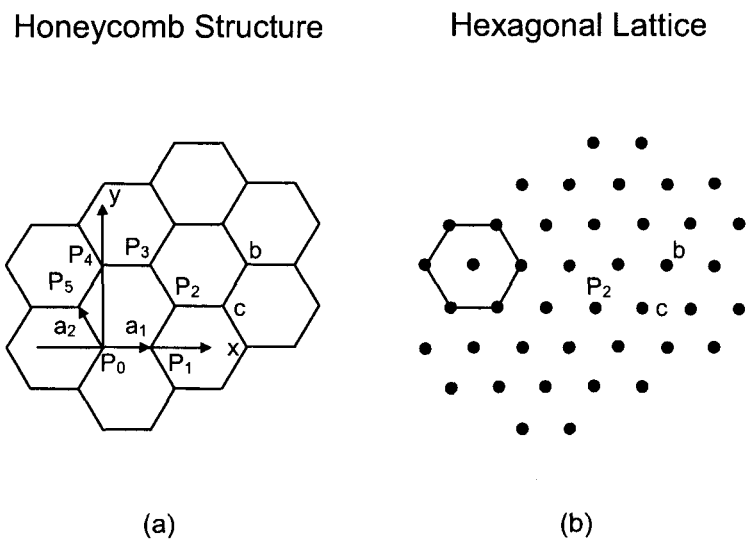


Fig. 1.2 (a) Honeycomb structure. The point P_2 and point b are equivalent, but b and c are not. Honeycomb structure is not a Bravais lattice. For quantitative verification, one may want to choose the two adjacent sides of a honeycomb cell as the basis vectors in the x-y coordinate as shown here. (b) Hexagonal Bravais lattice. All corner points of the honeycomb structure are indeed generated by the basis vector $\mathbf{a}_1$ and $\mathbf{a}_2$, but there is an extra point in the center of the honeycomb cell. The result is the hexagonal lattice. Now, points like P_2, b, and c are all equivalent.

structure is a subset of the hexagonal Bravais lattice. By eliminating the center point of the hexagonal cell in Fig. 1.2(b), it becomes the honeycomb structure in Fig. 1.2(a), where point b and c become asymmetrical. It is not possible to find a set of primitive unit vectors that can generate points like b and c without generating P_c. Therefore, the honeycomb is not a Bravais lattice. The problem stems from the fact that honeycomb has two types of corner sites, one half is b-like and the other half is c-like. Each b is

connected to three c sites and each c is connected to three b sites as their nearest neighbors. If the asymmetrical b and c sites are bundled into pairs, it will eliminate the asymmetry and the honeycomb structure becomes a set of symmetrical pairs. The symmetrical pairs may satisfy the definition of Bravais lattice. Let us try this idea.

In Fig. 1.3(a), we get rid of all the honeycomb cages except one to remind us about its unit cell. A different unit cell is obtained by connecting four adjacent centers of the honeycomb cells. The parallelogram is a hexagonal primitive cell because it has two equal adjacent sides that make an angle of 120°. The hexagonal primitive cell so obtained has a side-length equal to $\sqrt{3}$ times of that shown in Fig. 1.2(b). In this construction, all corners of the hexagonal primitive cell do not coincide with any honeycomb sites. Instead, there is a pair of honeycomb sites b and c inside it. Figure 1.3(b) shows the new view of the same honeycomb structure from the perspective of the new hexagonal primitive cells. The hexagonal primitive cell contains a b-c pair and is identical to each other. Therefore, the honeycomb structure can be viewed as having a hexagonal primitive cell with a basis of two points associated with the cell. There are many ways to choose the primitive cell. For instance, one may shift the cage in Fig. 1.3(b) such that all the c sites coincide with one corner of the primitive cell.

Another observation shows that the hexagonal primitive cell is made of two opposite triangles as seen in Fig. 1.3(c). One contains site b at its center and the opposite one contains site c. Each side of the triangle is the bisector of the line that connects b or c to one of its nearest neighbors, see Fig. 1.3(c). In fact, the triangle containing b or c has the smallest area enclosed by any set of bisectors that bisect site-pointing vectors originated from b or c. Cells obtained in this way in a Bravais lattice are called Wigner-Seitz cells. All Wigner-Seitz cells are identical in shape and orientation because of translation symmetry. Here in Fig. 1.3(c), the triangular cells containing sites b or c are identical only to its own kind, but not to each other. Triangles b and c are different in orientation because Honeycomb does not have complete translation invariance. However, if triangles b and c are considered as a jointed cell to house the paired basis of b and c, full translation symmetry is restored. The new primitive cells become identical as shown in Fig. 1.3(b). With the asymmetrical sites b and c combined into paired basis, we can view honeycomb as a structure with an underlying hexagonal Bravais lattice.

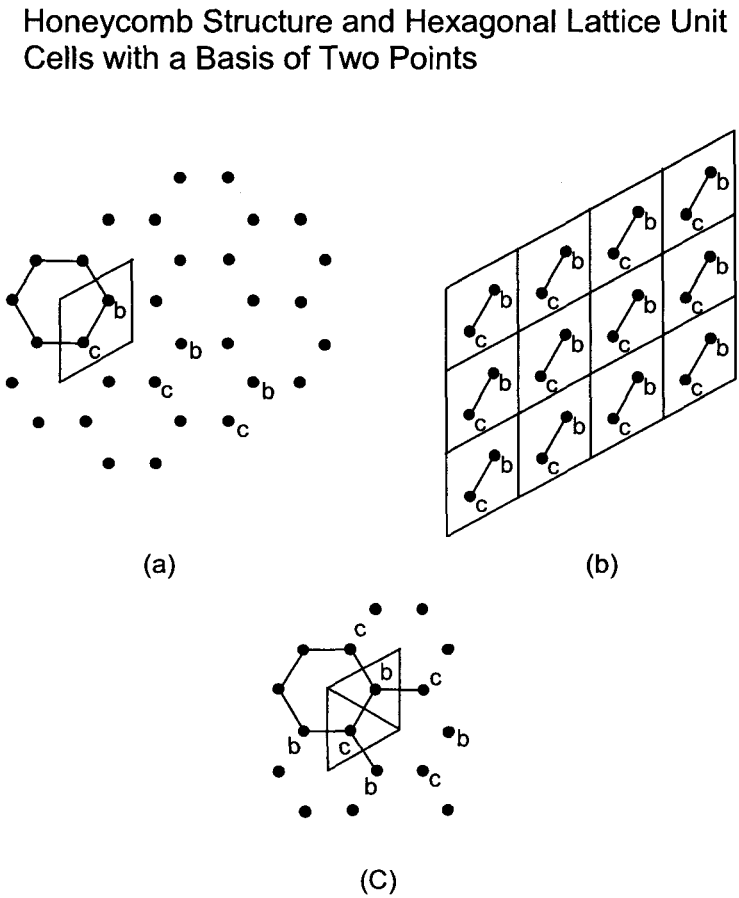


Fig. 1.3 (a) The 6-sided hexagonal polygon is a honeycomb cell, and the parallelogram is the primitive cell of the hexagonal lattice, which has a basis of two points b and c. (b) The same honeycomb structure can be viewed as the hexagonal lattice if points b and c are considered as one unit. (c) The point b and c are not equivalent in orientation, and translation symmetry does not hold. But, translation symmetry is regained if b and c are combined into one unit.

1.5 Cubic Crystals

Cubic crystals are more symmetrical than crystals of other shapes. The cubic symmetry is easier to visualize directly. There are three Bravais

lattice types for cubic crystals. We will discuss the cubic lattices in terms of the definition of Bravais lattice.

1.5.1 *Simple cubic (sc) lattice*

If we try to arrange billiard balls into simple cubicles, we find the balls touch at the highest point of each other, see Fig. 1.4(a). This configuration

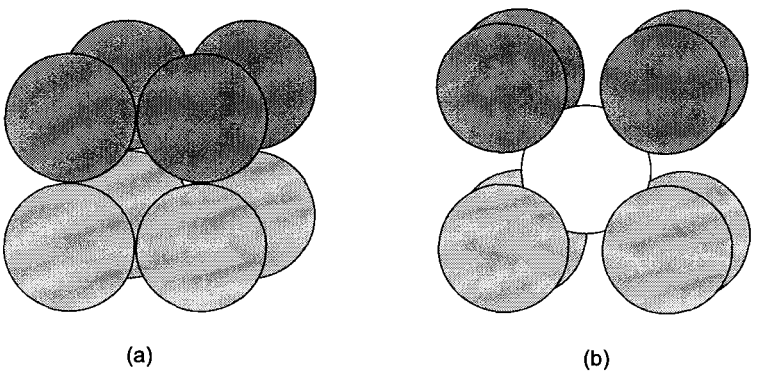


Fig. 1.4 (a)A hard-sphere model of the unit cell of simple cubic lattice. (b) The unit cell of bcc lattice.

is intuitively unstable because the top layer can slip easily. So, simple cubic crystals are not common and the alpha phase of polonium is the only element known to be simple cubic. Nevertheless, simple cubic lattice is still very important in the sense that two other popular cubic structures of body-entered and face-centered cubic lattices can be viewed as simple cubic cell with different basis. The translation invariance is rather easy to see in the simple cubic lattice, because any lattice point can be always located at one corner of a unit cell, and all corners of a unit cell are equivalent in terms of viewing the lattice arrangement and orientation. Quantitatively, we can choose an arbitrary lattice point as the origin and take three mutually perpendicular edges of the unit cell as coordinate axes. Assuming that $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ are unit vectors and a is the lattice constant which is defined as the length of the cell edge, we can generate the simple cubic lattice with $\mathbf{R} = n_1 a\hat{\mathbf{x}} + n_2 a\hat{\mathbf{y}} + n_3 a\hat{\mathbf{z}}$ by taking all integer combination of (n_1, n_2, n_3). Therefore, simple cubic lattice is a Bravais lattice, and the conventional cubic cell $\mathbf{a} \cdot \mathbf{a} \times \mathbf{a}$ is its primitive cell.

1.5.2 *Body-centered cubic (bcc) lattice*

Body-centered cubic (bcc) structure is obtained by adding a point to the center of the simple cubic cell. A billiard-ball model of bcc crystal is shown in Fig. 1.4(b). Comparing to the simple cubic model in Fig. 1.4(a), the center ball of Fig. 1.4(b) tends to lock up the bottom and top layers from slipping. It is intuitively more stable than the simple cubic structure. Many substances have bcc structures.

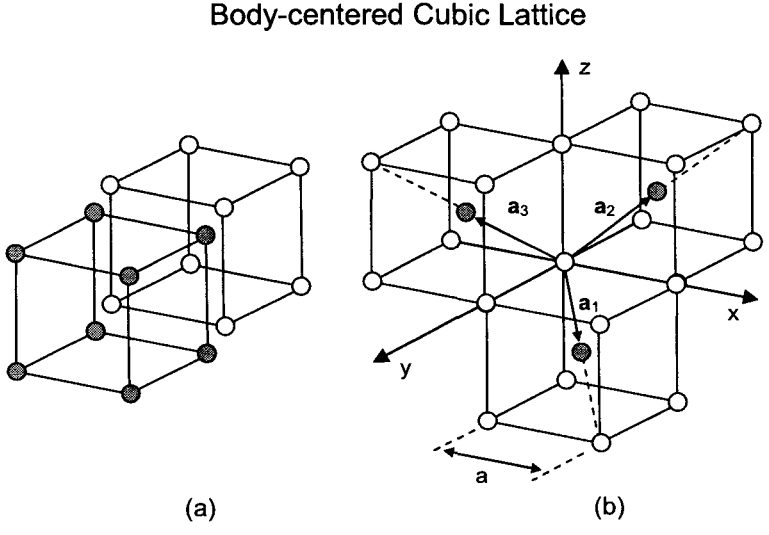


Fig. 1.5 (a) The cage of the bcc unit cell is shifted along the body diagonal line to show the equivalence between the corner and body-center points. (b) A choice of primitive unit vectors in the bcc lattice.

The conventional bcc unit cell is illustrated in Fig. 1.5(a). The wire cage is drawn to group certain atoms together for showing the crystal configuration, and it does not exist in real crystals. It is up to us to draw the wire cage to illustrate symmetry properties. If the cubic cage, not the atoms, is shifted by $\frac{1}{2}a$ in all x, y, and z direction as shown in Fig. 1.5(a), the body-center atoms in the old cage become the corner atoms in the new cage and vice versa. It shows the equivalence of all atomic positions in bcc

structure. Figure 1.5(b) shows a choice of primitive vectors:

$$\begin{aligned}\mathbf{a}_1 &= \frac{a}{2}(\hat{\mathbf{x}} + \hat{\mathbf{y}} - \hat{\mathbf{z}});\\ \mathbf{a}_2 &= \frac{a}{2}(\hat{\mathbf{z}} + \hat{\mathbf{x}} - \hat{\mathbf{y}});\\ \mathbf{a}_3 &= \frac{a}{2}(\hat{\mathbf{y}} + \hat{\mathbf{z}} - \hat{\mathbf{x}}).\end{aligned} \tag{1.4}$$

The vector $\mathbf{R} = n_1 a\hat{\mathbf{x}} + n_2 a\hat{\mathbf{y}} + n_3 a\hat{\mathbf{z}}$ can generate all the bcc lattice points. Therefore, bcc structure is a Bravais lattice. The cubic symmetry is immediately clear with the conventional bcc unit cell, but not so transparent with the rhombohedron primitive cell defined by $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. The conventional cubic cell $\mathbf{a} \cdot \mathbf{a} \times \mathbf{a}$ can be considered as the primitive cell if the body-center atom is bundled with a corner atom to form the basis. From this point of view, bcc lattice becomes a simple cubic lattice with a basis of two atoms, one at the corner and one at the body center.

1.5.3 *Face-centered cubic (fcc) lattice*

Face-centered cubic lattice is obtained by adding one lattice point to the center of every face of the simple cubic cell, see Fig. 1.6. The lattice looks identical from every corner points. So, all corner points are equivalent. By the same reason, all face-center points are also equivalent. However, it is not immediately clear whether the corner and face-center points are in fact equivalent. Let the cubic cage, not the atoms, be shifted by $\frac{1}{2}a$ in the x and y directions as shown in the Fig. 1.6(a). The face-center atoms in the top and bottom faces of the old cage become the corner atoms in the respective faces of the new cage and vice versa. This proofs that all corner and face-center points are equivalent, because we already know that all corner atoms are equivalent and all face-center atoms are also equivalent. To visualize that the face-center atoms of four side faces in Fig. 1.6(a) can be seen as corner atoms, one needs to move the cubic cage in the x-z and y-z planes in a similar manner, respectively. Figure 1.6(b) shows a choice of primitive lattice vectors for fcc lattice:

$$\begin{aligned}\mathbf{a}_1 &= \frac{a}{2}(\hat{\mathbf{x}} + \hat{\mathbf{y}});\\ \mathbf{a}_2 &= \frac{a}{2}(\hat{\mathbf{z}} + \hat{\mathbf{x}});\\ \mathbf{a}_3 &= \frac{a}{2}(\hat{\mathbf{y}} + \hat{\mathbf{z}}).\end{aligned} \tag{1.5}$$

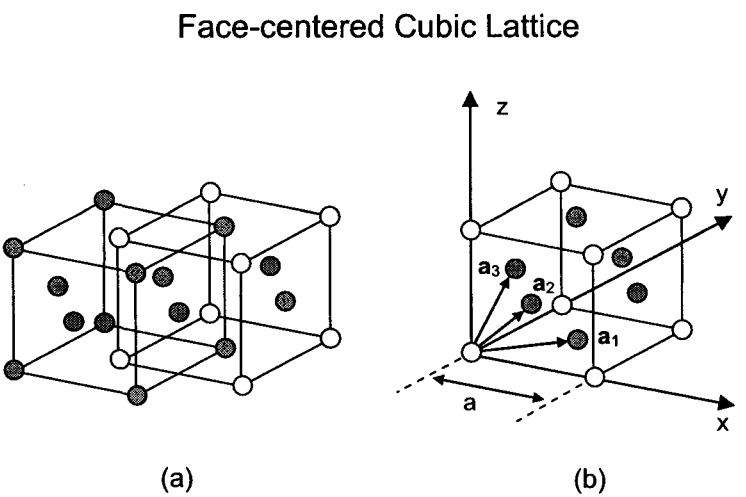


Fig. 1.6 (a) The cage of a fcc unit cell is shifted along the diagonal line on two parallel surfaces. It shows the equivalence between the face-center and corner points on these surfaces. (b) A choice of primitive unit vectors in the fcc lattice.

The vector space spanned by these primitive vectors is the fcc lattice. So, fcc is a Bravais lattice. Although the translation symmetry is mathematically described with the primitive vectors, the primitive cell does not offer a direct impression of cubic symmetry. In practice, the fcc conventional cell is almost always used because the clarity of cubic symmetry. The fcc structure can be also considered as the simple cubic lattice with a basis of 4 atoms, one corner and three centers of the adjacent faces.

1.6 Close-packed Crystals

Many crystals have an atomic configuration that resembles close packed billiard balls. It is the best packing scheme to save space. The close packed atomic layer forms a two-dimensional hexagonal lattice. We can quickly visualize this result by squeezing the billiard balls together on the pool table. Figure 1.7(a) shows a set of close packed billiard balls in one layer. The balls occupy the hexagonal lattice points labeled as A in light gray. The space in the middle of any three balls is indented. The indented positions

are labeled B or C. Both positions B and C are the center of a respective triangle formed by three A balls. However, the orientation of the triangle enclosing B is opposite to that of triangle C, see Fig. 1.7(a). The sites B and C are mirror images of each other. When we put another close-packed layer on top of this layer, the balls can either all sit on positions B or all on C. An example is shown in Fig. 1.7(b) where the second layer is on B positions. It leaves two possible ways to settle the third layer, either on A or on C.

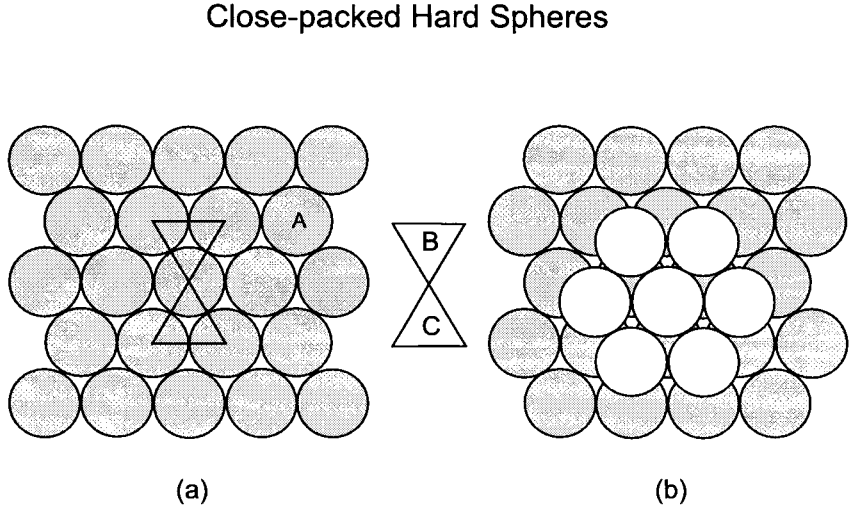


Fig. 1.7 (a) A layer of close-packed hard spheres. The centers of the sphere are A sites. Three adjacent spheres form a triangle, and the center of the triangle is either the B site or C site. (b) The second layer can sit either on B sites or on C sites. The second layer, shown here in white spheres, happens to sit on the B sites. In this case, the third layer can be on either A or C sites.

We can continue to stack layers. The simplest will be ABAB$\cdots$AB stacking where the AB sequence repeats itself. The periodicity is for every two layers. The next one will be ABCABC$\cdots$ABC stacking. Here, the repeating sequence is a three-layer of ABC. In fact, we can have a stacking of many layers before the sequence repeats itself. For example, we may have ABCBCBC as a sequence. A long stacking sequence is called polytypism. Zinc sulfide (ZnS) is known to have more than 150 different stacking

sequences (polytypes). Its long periodicity can go to 360 layers. Silicon carbide (SiC) has more than 45 polytypes. One SiC polytype known as 393R has a primitive cell with the c-axis as long as 989.6 Åin the stacking direction. The stacking can be of no specific repeating sequence. Units of different lengths are randomly stacked. There is no periodicity in the stacking direction, but there is a crystalline periodicity in the perpendicular planes. Substances with random stacking can be considered as two-dimensional crystals. Nevertheless, the most common close-packed crystals have either ABAB$\cdots$AB or ABCABC$\cdots$ABC stacking sequences. The former is the hexagonal close-packed (hcp) structure and the latter is the fcc structure.

1.6.1 *Hexagonal close-packed (hcp) structure*

Figure 1.8 shows a dissected model of the hcp conventional cell, where the top layer is removed to the side for viewing the stacking. More atoms are shown in the bottom layer to highlight the fact that it is a hexagonal close-packed layer. The top layer is an identical layer that is raised a distance of c directly above the bottom layer. The bottom and top layer are the base layers of the hcp structure. The middle layer is also a hexagonal close-packed layer, but it is shifted parallel to the base layer. All these layers are two-dimensional hexagonal Bravais lattices. However, the three-dimensional hcp structure of ABAB stacking is not a Bravais lattice.

The middle layer is identical to the base layer except being shifted to have its atoms located on B sites of the base layer. Relatively speaking, the atoms of base layer are on C sites of the middle layer, see Fig. 1.7(b). This geometry is further illustrated with the perpendicular projection of the hcp structure onto its base plane as shown in Fig. 1.9(a). In the projection, every atom of the middle layer is surrounded by a triangle of three atoms from the base layer. By comparison, every atom of the base layer is surrounded by an opposite triangle of three atoms from the middle layer. Two triangles are otherwise identical except that the orientation is different. The situation is similar to the b and c sites of the honeycomb in Fig. 1.3(c). In fact, Figure 1.9(a) is another representation of the honeycomb structure shown in Fig. 1.3(b). This observation demonstrates that hcp structure does not have complete translation symmetry in the c-direction.

A schematic of the hcp conventional cell is depicted in Fig. 1.9(b), where a choice of coordinates and the primitive vectors are shown. The primitive lattice vectors are chosen to be the three adjacent sides of the hcp cell. The

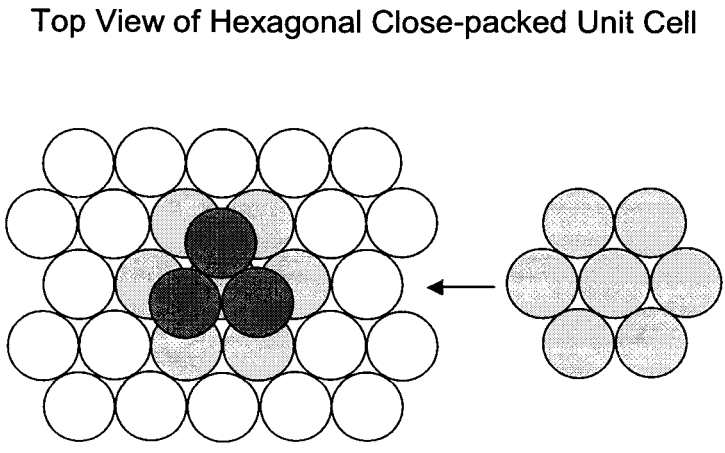


Fig. 1.8 Top view of the conventional hcp unit cell. Both the top and bottom of the hcp unit cell are shown in light gray. The middle layer is in darker gray. The top layer is moved to the side for clarity.

two vectors marked by their length a in the x-y base plane can generate the planar hexagonal lattice. The third vector marked by its length c will enable the generation of other base planes along the z-direction. We have

$$\begin{aligned} \mathbf{a}_1 &= a\hat{\mathbf{x}}; \\ \mathbf{a}_2 &= \frac{-a}{2}\hat{\mathbf{x}} + \frac{\sqrt{3}a}{2}\hat{\mathbf{y}}; \\ \mathbf{a}_3 &= c\hat{\mathbf{z}}. \end{aligned} \tag{1.6}$$

Indeed, the vector $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ can generate all the hexagonal sites on all base planes. The set of lattice points so generated is the hexagonal Bravais lattice. However, it fails to generate any of the hexagonal sites on the middle layer of the hcp structure. It is a matter of fact that we can not find any set of primitive vectors that can generate all the sites in the hcp structure. Therefore, hcp is not a Bravais lattice.

Since the hcp structure can be viewed as a hexagonal lattice with added middle layers, it may be possible to assign the mid-layer atoms to the base-layer atoms to form a basis. The two-dimensional projection in Fig. 1.9(a)

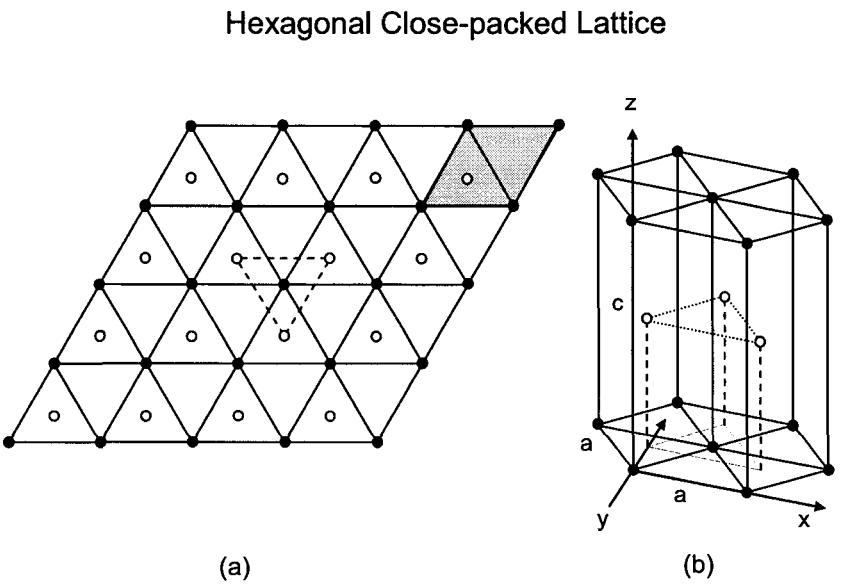


Fig. 1.9 (a) The perpendicular projection of hcp structure onto the base plane. The solid circles are A sites, and the open circles are B sites. The primitive unit cell is shown in gray with a basis of two points, a pair of solid and open circles. (b) A 3-dimensional view of a conventional hcp unit cell, and a choice of primitive unit vectors.

demonstrates exactly this idea, where a two-atom basis is clearly associated with each hexagonal primitive cell. The grayed region highlights one of the hexagonal primitive cells. The three-dimensional hexagonal primitive cell that contains this basis can be identified within the hcp conventional cell in Fig. 1.9(b). Therefore, the hcp structure can be viewed as a hexagonal lattice with a basis of two atoms.

1.6.2 *Cubic close-packed (fcc) structure*

It is not clear at the first glance that the fcc conventional cell of Fig. 1.6 is actually a close packed structure. In fact it is if we look along the cubic diagonal direction. Figure 1.10(a) shows a perspective view of this approach. A fcc cell is highlighted with its 14 atoms distributed in four consecutive close packed layers. One corner atom of the fcc unit cell is from the bottom base layer (A site). Six fcc atoms are from the second

layer occupying B sites and another six are from the third layer occupying C sites. The opposite corner one is from the repeating base layer on the very top. The stacking sequence is ABC along the cubic diagonal direction. Figure 1.10(b) shows the bottom view of the stacking. By matching the color in Figs. 1.10(a) and 1.10(b), it is clear that the second layer in darker gray is on B sites and the third layer in lighter gray is on C sites. By contrast, only B or C sites are occupied in the hcp structure.

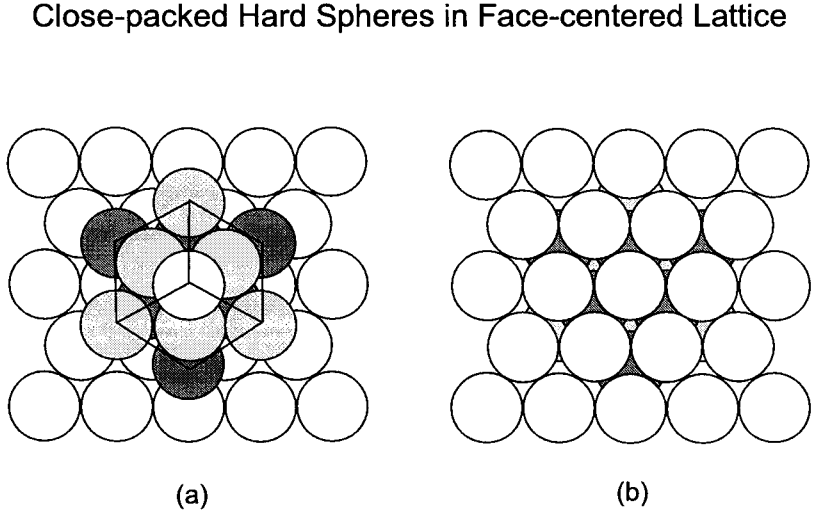


Fig. 1.10 (a) Top view along the diagonal direction of a fcc unit cell. Layer A is in white. Layer B is in darker gray, and layer C is in lighter gray. (b) Bottom view of the same fcc unit cell. The corresponding gray shades show through both B and C sites.

1.7 Diamond Structure

Diamond structure is another crystal structure that is not a Bravais lattice if a basis of only one atom is required, but it can be viewed as a fcc lattice with a basis of two atoms in the primitive cell. It is the crystal structure of diamond, a single crystal of carbon. It is also the crystal structure of the technologically important semiconductors such as Si, Ge, GaAs, GaAlAs,

InSb, HgCdTe, and many others. The diamond structure has a conventional cell of a fcc cell with 4 additional atoms. These 4 atoms are added on the diagonal axes separately at $\frac{1}{4}$ length from the non-neighboring corners. Each atom is connected to four nearest neighbors in a characteristic way called tetrahedral bonding. Figure 1.11(a) shows a diamond conventional cell. The atoms of the fcc cell are shown in light and darker gray, and the four added atoms on the diagonal axes are shown as clear spheres. The diamond structure can be viewed as two fcc cell shifted $\frac{1}{4}$ length along one diagonal axis. Figure 1.11(a) shows that the added four atoms are actually atoms from the shifted fcc cell.

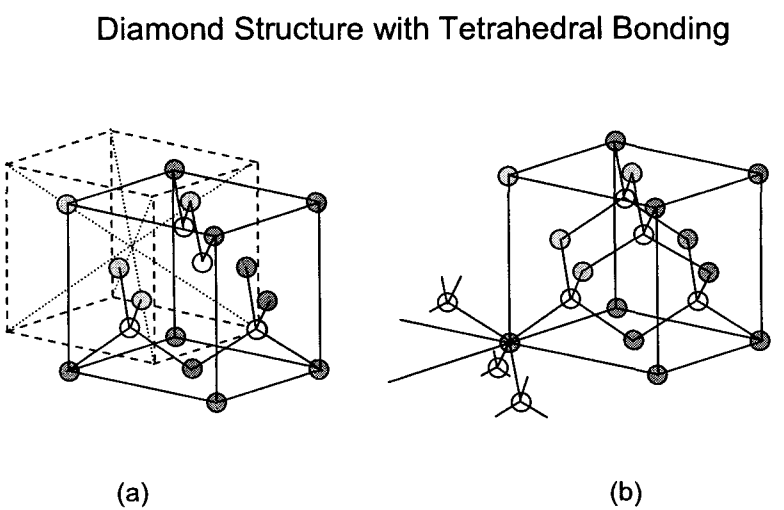


Fig. 1.11 Diamond unit cell is a fcc unit cell with four additional atoms. They are the four clear atoms on the diagonal axes, located at $\frac{1}{4}$ of the axis length from four non-adjacent corners. The four light gray and other darker gray atoms are fcc atoms. (a) shows a shifted cell cage from the diamond unit cell along the diagonal axis by $\frac{1}{4}$ of its length. The four clear atoms become the fcc atoms, and the four light gray fcc atoms become the diagonal atoms in the shifted cell. (b) shows the different orientation of the tetrahedral bonding when viewed from the corner and diagonal atoms, respectively.

The diamond structure has the same primitive cell of fcc lattice, except each cell contains a basis of two atoms. Employing the same coordinates as in Fig. 1.6(b), a typical basis involves atoms at $(0,0,0)$ and $(\frac{1}{4},\frac{1}{4},\frac{1}{4})$. Without forming a basis of two atoms, we can not find a set of primitive

vectors that can generate all the fcc sites together with the four extra sites. Figure 1.11(b) shows the tetrahedral bonds seen by a corner fcc atom. The bond orientation is a mirror image after a π rotation of that seen by the adjacent atom on the diagonal axis. Therefore, diamond structure lacks complete translation symmetry and is not a Bravais lattice. When the diagonal atom is bundled with the corner atom to form a paired basis, the translation symmetry is restored and the underlying lattice becomes fcc.

1.8 Wigner-Seitz Cell

There are many ways to choose the primitive cell in a Bravais lattice. We can draw lattice vectors from one lattice point to all the others, and then draw the bisecting plane of each vector. The smallest volume enclosed by such a set of bisecting plane is a primitive cell called Wigner-Seitz cell. It contains only this lattice point. In fact every lattice point is enclosed inside an identical Wigner-Seitz cell. Any two of them are replica of each other connected by a lattice vector because of translation symmetry. Wigner-Seitz cell in reciprocal lattice space is of great value for understanding the wave propagation in crystals.

1.9 Beyond Single Crystals

To the other extreme from crystalline solids, atoms may arrange themselves in a completely disordered way in the specimen. No identical atomic arrangement and orientation can be found when viewed from atom to atom. In other words, there is no translation invariance, both in the short and long ranges. These are amorphous materials. Window glasses, eyeglasses, and drinking glasses are common examples. For this reason, the amorphous or disordered materials are sometimes called glasses. If the thermodynamic environment during growth does not favor crystallization, the material may end up in the amorphous form. Some metallic alloys are actually glasses.

Most stuff we are familiar with has atomic configurations somewhere in between. These are polycrystalline materials. They consist of randomly oriented single-crystalline grains with their grain boundaries in somewhat disordered states. Gold, silver, copper, aluminum, jade, ceramics, salt, sugar, and many others are usually in polycrystalline forms. The mixture of polycrystals and glasses also exists. Examples are certain china and some lava rocks, etc. Some metallic alloys are called solid solutions with

one species randomly distributed in the orderly host species. The host can be either a single crystal or a polycrystal.

Some crystals can be classified as low-dimensional crystals. Two-dimensional crystals have perfectly ordered layer structures, but the layers stack up in a random way. One-dimensional crystals have periodicity in a chain, but not in the plane of other two dimensions. Furthermore, no single crystals are so pure that they do not contain any defects and impurities. Important technological achievements are based on controlling the impurities and defects. No crystals are of infinitive size. Surfaces and interfaces are important on their own merits. When the crystal size shrinks down to microns or even smaller to nanometers, the importance of surfaces is further underscored. There are a great many varieties of solids. Regardless of what our research encounters, an understanding of crystal structures is always needed for it is the very foundation.

Problems

1. In a given two-dimensional lattice, we have $\mathbf{a}_1 \neq \mathbf{a}_2$ and the angle between $\mathbf{a}_1$ and $\mathbf{a}_2$ is not 90°.
   (a) Draw the lattice points, and show the choice of primitive unit vectors is not unique. Give 3 examples.
   (b) What is the area of primitive cell?
   (c) Draw a Wigner-Seitz cell.

2. Find the integer coordinates (n_1, n_2, n_3) for every atom in the conventional bcc unit cell in the coordinate system of the primitive unit vectors defined in Eq. (1.4).

3. Find the integer coordinates (n_1, n_2, n_3) for every atom in the conventional fcc unit cell in the coordinate system of the primitive unit vectors defined in Eq. (1.5).

4. Fill Table 1.2. Use hard sphere model to calculate the packing fraction.

5. Find the ratio c/a for the hcp structure. See Fig. 1.9(b).

6. Find the angles between two adjacent tetrahedral bonds in the diamond structure. What is the relationship between the four tetrahedral bonds

seen by the corner atom and those seen by the adjacent atom on the diagonal axis? See Fig. 1.11(b).

Table 1.2

	sc	bcc	fcc
Volume of conventional cell			
Lattice points per conventional cell			
Volume of primitive cell			
Lattice points per unit volume			
Number of nearest neighbors			
Nearest neighbor distance			
Number of second neighbors			
Second neighbor distance			
Packing fraction			

Chapter 2

Reciprocal Lattice and X-ray Diffraction

Introduction and Summary

Chapter 2 has two parts. the first part discusses the reciprocal lattice by considering wave propagation in crystal lattices. The objective is to establish the concept of the reciprocal lattice, and then to introduce the scheme for specifying the crystal orientation. The lattice translation symmetry, when imposed on the propagating wave, is shown to lead to a special set of wave vectors that defines the reciprocal lattice in k space. The lattice in real space can be dissected into parallel lattice planes. For each such set, there is always a set of reciprocal lattice vectors perpendicular to it. The shortest of such vectors is proportional to the inverse of the lattice-plane spacing. Proof is given to these statements and to their converse statements, leading to the scheme of Miller indices for identifying the lattice plane and direction in crystals. Precaution is given for using Miller indices in the conventional-cell coordinate.

The second part is about x-ray diffraction in crystals. The objective is to show the experiment that firmly established the concept of lattice and reciprocal lattice. The explanation of x-ray diffraction leads to the formulation of Bragg law and von Laue diffraction law. Under the assumption of specular reflection from lattice planes, Bragg law predicts the location of diffraction peaks. The law relates the x-ray wavelength to the lattice plane spacing, and it offers insights about why lights of longer wavelengths do not produce diffraction effects. A closer look reveals a more realistic picture. The electrons must be distributed with the same translation symmetry of the lattice. If the x-ray scattering off the electron clouds is elastic, von Laue diffraction law can be derived. It maps the reciprocal lattice directly into the bright spots in diffraction. The diffraction condition gives a physical

significance to the Brillouin zone, because the law selects those wave vectors pointing onto the zone boundary only. The diffraction beam intensity is shown to be calculable from knowing the lattice type and the number of valence electrons in the basis atoms. This exercise serves as a practical example to demonstrate how one can apply the knowledge of crystal structures.

Reciprocal Lattice

2.1 A Special Set of Wave Vectors in the k Space

The most important property of crystal structures is their periodicity. Physical observables such as electron density, local potential, geometric arrangement of atoms, and many others are identical when viewed from points that are separated by lattice vectors. The physics is the same either at point $\mathbf{r}$ or at $\mathbf{r} + \mathbf{R}$, provided that $\mathbf{R}$ is a lattice vector. We call this property translation invariance.

Let us consider an electromagnetic plane wave traveling in a given Bravais lattice. The wave amplitude is proportional to the phase factor $\exp(i\mathbf{k} \cdot \mathbf{r})$, where the value of wave vector $\mathbf{k}$ is $2\pi/\lambda$, and λ is the wavelength. On a plane perpendicular to wave vector $\mathbf{k}$, all phase angles $(\mathbf{k} \cdot \mathbf{r})$ are equal. Therefore, this plane is a wave front of equal amplitude. For a given $\mathbf{k}$ vector of arbitrary direction and magnitude, the phase factor may not necessarily be the same after the wave front travels a distance of $(\mathbf{k} \cdot \mathbf{R})/\mathrm{k}$, where $\mathbf{R}$ is any lattice vector. However, there must be a special subset of $\mathbf{k}$ that does have this periodicity. For these special vectors in k space, we will label them as $\mathbf{G}$ vectors. Both the direction and the magnitude of $\mathbf{G}$ vectors are restricted, because the periodical condition requires that the phase factor of any two wave fronts separated by a distance of $(\mathbf{G} \cdot \mathbf{R})/\mathrm{G}$ remains equal to each other. So, we have

$$e^{i\mathbf{G}\cdot(\mathbf{r}+\mathbf{R})} = e^{i\mathbf{G}\cdot\mathbf{r}}. \tag{2.1}$$

It leads to the condition that G must satisfy, i.e.,

$$e^{i\mathbf{G}\cdot\mathbf{R}} = 1. \tag{2.2}$$

This identity requires the phase angle $(\mathbf{G} \cdot \mathbf{R})$ to be multiples of 2π. Hence, it defines the set of $\mathbf{G}$ vectors for a given Bravais lattice. Recall that a Bravais lattice is a set of vectors generated by three primitive vectors in

all their possible linear combinations with integer coefficients, i.e., $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 = \Sigma_i n_i \mathbf{a}_i$. In a similar manner, the $\mathbf{G}$ vector may also be written as a linear combination of three $\mathbf{b}$ vectors, which are not all in the same plane. We have $\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3 = \Sigma_j m_j \mathbf{b}_j$. If we further require $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$, we will have the inner product $\mathbf{G} \cdot \mathbf{R} = 2\pi \cdot (m_1 n_1 + m_2 n_2 + m_3 n_3)$. Since n_1, n_2, and n_3 are arbitrary integers, the sum $(m_1 n_1 + m_2 n_2 + m_3 n_3)$ can be guaranteed to be an integer if and only if all m_1, m_2, and m_3 are integers. Then, the condition $\exp(i\mathbf{G} \cdot \mathbf{R}) = 1$ is satisfied, i.e., the same amplitude or phase occurs on all lattice points defined by $\mathbf{R}$ if the wave vector is one of the $\mathbf{G}$ vectors. Since the set of $\mathbf{G}$ vectors is generated by all possible linear combinations of $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$ with integer coefficients, $\mathbf{G}$ vectors also form a Bravais lattice. The set of points generated by $\mathbf{G}$ is the reciprocal lattice in k space. Its counter part $\mathbf{R}$ is the direct lattice in real space. Since the choice of primitive vectors for a Bravais lattice is not unique, let the $\mathbf{b}$ vectors just be written in the following form:

$$\begin{aligned}
\mathbf{b}_1 &= 2\pi \cdot \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3}; \\
\mathbf{b}_2 &= 2\pi \cdot \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3}; \\
\mathbf{b}_3 &= 2\pi \cdot \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3}.
\end{aligned} \tag{2.3}$$

These vectors satisfy the necessary orthogonal condition $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$. We can easily observe, for example, that the vector $\mathbf{b}_1$ is perpendicular to the plane defined by $\mathbf{a}_2$ and $\mathbf{a}_3$, and its value is proportional to the reciprocal value of $\mathbf{a}_1$, i.e., $b_1 \propto 2\pi/a_1$. This is why the lattice generated by the $\mathbf{b}$ vectors is named reciprocal lattice. More importantly, we should remember that the reciprocal lattice is a set of wave vectors $\mathbf{G}$ defined by requiring $\exp(i\mathbf{G} \cdot \mathbf{R}) = 1$, and the set of $\mathbf{G}$ vectors also forms itself a Bravais lattice.

2.1.1 *Examples*

For cubic lattices that have the lattice constant equal to a, the origin may be set at one corner of the conventional cubic cell. Three joining cubical edges can then be chosen as the coordinate axes. The unit vectors are denoted as $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ along the respective axes. In this coordinate system, the primitive vectors of the simple cubic lattice are

$$\mathbf{a}_1 = a\hat{\mathbf{x}};$$

$$\begin{aligned} \mathbf{a}_2 &= a\hat{\mathbf{y}}; \\ \mathbf{a}_3 &= a\hat{\mathbf{z}}. \end{aligned} \tag{2.4}$$

We have the $\mathbf{b}$ vectors equal to

$$\begin{aligned} \mathbf{b}_1 &= \frac{2\pi}{a}\hat{\mathbf{x}}; \\ \mathbf{b}_2 &= \frac{2\pi}{a}\hat{\mathbf{y}}; \\ \mathbf{b}_3 &= \frac{2\pi}{a}\hat{\mathbf{z}}. \end{aligned} \tag{2.5}$$

This example shows that the reciprocal lattice constant equals $2\pi/a$. Vectors in the real lattice space are measured in units of length such as cm, and reciprocal lattice vectors are measured in units of cm^{-1}, or wave numbers. Also noticed here is that the primitive cell volume in the reciprocal lattice space equals $(2\pi)^3/v$, where v equals a^3 and is the primitive cell volume in the real space.

For the bcc lattice in the real space, using the same coordinate system, the primitive vectors pointing to three adjacent body-center points are

$$\begin{aligned} \mathbf{a}_1 &= \frac{a}{2}(\hat{\mathbf{x}} + \hat{\mathbf{y}} - \hat{\mathbf{z}}); \\ \mathbf{a}_2 &= \frac{a}{2}(\hat{\mathbf{z}} + \hat{\mathbf{x}} - \hat{\mathbf{y}}); \\ \mathbf{a}_3 &= \frac{a}{2}(\hat{\mathbf{y}} + \hat{\mathbf{z}} - \hat{\mathbf{x}}). \end{aligned} \tag{2.6}$$

From the definition of $\mathbf{b}$ vectors, we obtain

$$\begin{aligned} \mathbf{b}_1 &= (\frac{4\pi}{a})(\frac{1}{2})(\hat{\mathbf{x}} + \hat{\mathbf{y}}); \\ \mathbf{b}_2 &= (\frac{4\pi}{a})(\frac{1}{2})(\hat{\mathbf{z}} + \hat{\mathbf{x}}); \\ \mathbf{b}_3 &= (\frac{4\pi}{a})(\frac{1}{2})(\hat{\mathbf{y}} + \hat{\mathbf{z}}). \end{aligned} \tag{2.7}$$

These turn out to be the primitive vectors of fcc lattice, and the lattice constant equals $4\pi/a$. Therefore, the reciprocal lattice of bcc lattice is the fcc lattice. It is also true that the reciprocal lattice of fcc lattice is the bcc lattice.

2.2 Lattice Planes, Crystal Directions, and Miller Indices

Lattice points in the three-dimensional space can always be grouped into parallel lattice planes. Obviously, there are many possible choices. For a particular set of parallel lattice planes, we can always send plane waves traveling perpendicular to these planes. Among all these waves there must be a subset that is periodical as going from plane to plane. If s is the perpendicular distance from the coordinate origin to the first plane, the phase angle will be $k \cdot s$ when the wave front arrives at this plane. The phase angles on consecutive lattice planes, separated by an equal spacing of d, will be $(k \cdot s + k \cdot d)$, $(k \cdot s + k \cdot 2d)$, $(k \cdot s + k \cdot 3d)$, and so forth. In order to have the same amplitude on all lattice planes, the wave vector must satisfy the condition $k \cdot d = 2n\pi$, so that the phase factor becomes all equal to $\exp(ik \cdot s)$. The $\mathbf{k}$ vectors selected by this condition are actually reciprocal lattice vectors, because they satisfy the requirement, $\mathbf{k} \cdot \mathbf{R} = k \cdot m \cdot d = 2mn\pi$, that defines reciprocal lattice vectors. Here, we have made use of the fact that all lattice points are on these planes. So, we find the reciprocal lattice vectors perpendicular to this set of lattice planes to be $2\pi/d$, $4\pi/d$, $6\pi/d$, and so on. The shortest one is $G_s = 2\pi/d$. The shortest G vector is also related to the maximum wavelength λ_m by $G_s = 2\pi/\lambda_m$. Therefore, the wavelength can not be larger than d and still maintain periodicity over all lattice planes.

This argument proves that there are always reciprocal lattice vectors perpendicular to a complete set of parallel lattice planes. The shortest G vector is determined by $G_s = 2\pi/d$, where d is the spacing between adjacent lattice planes. The converse of this statement is also true. Namely, there is always a complete set of lattice planes perpendicular to a given set of parallel reciprocal lattice vectors G. The spacing d between two adjacent planes is determined by the shortest G_s, i.e., $d = 2\pi/G_s$.

To see why the converse statement is true, we will first construct a set of parallel planes perpendicular to G and have the smallest intra-plane spacing d equal to $2\pi/G_s$. Using the same convention to choose the origin such that the perpendicular distance to the first plane is s, the phase angle on the first plane is $(G \cdot s)$ and the phase factor is $\exp(iG \cdot s)$. The phase factor on the p^{th} plane is

$$e^{(iG \cdot s + iG \cdot pd)} = e^{(iG \cdot s + i2p\pi \cdot G/G_s)} = e^{(iG \cdot s + i2pn\pi)} = e^{(iG \cdot s)}. \tag{2.8}$$

Therefore, all space points on any of these planes will have the same phase factor of $\exp(iG \cdot s)$. For any space points in between two adjacent planes,

say between the p and $p+1$ plane, the additional phase angle would be $G \cdot (p+\delta) \cdot d = 2(p+\delta)n\pi$, which is not a multiple of 2π because δ is a fraction number. These points in between planes do not have the same phase factor of $\exp(iG \cdot s)$. Therefore, the set of planes we constructed exhausts all space positions where the phase factor is $\exp(iG \cdot s)$. It does not loose generality to assume that there is at least one lattice point on the first plane, because we can always move all the planes together in parallel until the first plane hits one lattice point. Since any other lattice point is related to this one by a lattice vector $\mathbf{R}$ and $\exp(i\mathbf{G} \cdot \mathbf{R}) = 1$, the phase factor on any lattice point must be

$$e^{(iG \cdot s + i\mathbf{G} \cdot \mathbf{R})} = e^{(iG \cdot s)}. \tag{2.9}$$

Since the set of planes we constructed exhausts all space positions where the phase factor is $\exp(iG \cdot s)$, all the lattice points must lie on these planes. For the set of parallel planes having d larger than $2\pi/G_s$, some lattice points will lie outside these planes. For planes having d smaller than $2\pi/G_s$, some planes will contain no lattice points. These arguments prove the converse statement.

Planes are normally specified using their normal vectors. For specifying a set of parallel lattice planes, the shortest of the perpendicular reciprocal lattice vectors becomes handy. It not only provides indices for specification purposes, but also measures the intra-plane spacing. For lattice plane indexing purposes, the reciprocal lattice vector is traditionally written as

$$\mathbf{G} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3. \tag{2.10}$$

The coefficient h, k, and l are reduced integers that do not have common factors. This is equivalent to say that $\mathbf{G}$ is the shortest in its direction. The indices (h,k,l) are Miller indices for planes perpendicular to $\mathbf{G}$. There is a geometric meaning to the Miller indices. If a plane perpendicular to $\mathbf{G} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3$ intercepts with the three primitive-vector axes at $x_1\mathbf{a}_1$, $x_2\mathbf{a}_2$, and $x_3\mathbf{a}_3$, we have $\mathbf{G} \cdot x_1\mathbf{a}_1 = \mathbf{G} \cdot x_2\mathbf{a}_2 = \mathbf{G} \cdot x_3\mathbf{a}_3 = C$, where C is a constant. From the definition of $\mathbf{b}$ we also have $\mathbf{G} \cdot \mathbf{a}_1 = 2\pi h$, $\mathbf{G} \cdot \mathbf{a}_2 = 2\pi k$, and $\mathbf{G} \cdot \mathbf{a}_3 = 2\pi l$. Combining the results, we have $x_1 = C/(2\pi h)$, $x_2 = C/(2\pi k)$, and $x_3 = C/(2\pi l)$. The ratio of Miller indices is therefore given by

$$h : k : l = \frac{1}{x_1} : \frac{1}{x_2} : \frac{1}{x_3}. \tag{2.11}$$

This result offers a conventional way to obtain the Miller indices for parallel

lattice planes. The first step is to find the intercepts of a plane on three axes defined by the primitive lattice vectors. Secondly, we take the ratio of their reciprocals and reduce the ratio to the smallest integers.

Rigorously speaking, we must use the primitive vectors for obtaining Miller indices. However, for instance, the conventional cubic cell with orthogonal unit vectors is always preferred for indexing cubic crystals. The reason is simple, because the cubic symmetry is much easier to visualize. For the simple cubic lattice, there is no confusion because the conventional cubic cell is just the primitive cell. For fcc and bcc lattices, the conventional cubic cell contains more than one lattice point and is not the primitive cell. If the three joining edges of the cubic cell are taken to be the unit vectors, the lattice becomes the simple cubic lattice that counts the face-center or body-center positions as part of the basis. In certain direction, the parallel lattice planes obtained from the conventional-cell scheme will have an intra-plane spacing larger than the actual spacing defined by the primitive lattice vectors. The reason is quite simple, because planes containing only the face-center or body-center points are not considered as lattice planes in the simple cubic convention. They are part of the basis. This effect is quantitatively reflected in the G vectors. The shortest G, say $G_s{}'$, obtained from using the fcc or bcc conventional unit cell could be shorter than the actual G_s from the primitive cell. Physical quantities that involve wave vectors shorter than G_s are not periodic over all lattice points. But, it would be confusing if the wave vector falls inside the range of $G_s{}' < k < G_s$. In this case, a clear statement about the basis is needed if one uses the conventional unit cell. With this precaution in mind, we may proceed to use the cubic convention to specify the fcc and bcc lattice planes.

For cubic lattices that have the lattice constant equal to a, the lattice plane parallel to the y-z plane and intercept the x-axis at $1 \cdot a$ will have three intercepts at $(x_1, x_2, x_3) = (1, \infty, \infty)$. Taking their inverse value, we obtain the Miller indices $(h, k, l) = (1, 0, 0)$, or in short (100). All planes parallel to it with intercepts at $n \cdot a$ on the x-axis will have the same indices (100). Other parallel planes with intercepts at $-n \cdot a$ on the negative x-axis will have indices $(\bar{1}, 0, 0)$. With cubic symmetry, the six planes (100), (010), (001), $(\bar{1}00)$, $(0\bar{1}0)$, and $(00\bar{1})$ are the same in terms of physical probing. They are collectively labeled as {100} plane. A similar convention is adapted to describe the crystal direction: [100], [010], [001], $[\bar{1}00]$, $[0\bar{1}0]$, and $[00\bar{1}]$ are six directions perpendicular to the above mentioned planes. They are collective written as $< 100 >$ direction.

X-ray Diffraction

2.3 Bragg Law

When an incident x-ray beam of sharply defined wavelengths bounces off a crystal, intense reflection peaks are observed in certain directions. The resulted pattern of discrete bright spots is not obtainable with liquids. It must be a characteristic of crystallinity. In 1913, W. H. and W. L. Bragg discovered this monumental effect. W. L. Bragg presented a simple explanation. His theory assumes that the incident x-ray is reflected specularly off the array of atoms in the crystal. Atoms on parallel planes can reflect x-ray into either constructive or destructive interference, depending upon the path length difference similar to the optical diffraction from a grating.

In Fig. 2.1, a set of lattice planes is shown in parallel to the x-y plane. The incident x-ray is parallel to the y-z plane and enters the crystal at an angle θ with respect to the lattice plane. The reflected beam will also lie in the same y-z plane and exit at the same angle because of the assumed specular reflection. This mirror-like reflection also ensures that the wave vector only changes its direction but not the magnitude. The path length difference is $2d \cdot \sin\theta$ for the x-ray reflected from two adjacent planes. If this path length difference equals at least one or multiple wavelengths, two reflected wave fronts will be in phase and constructive interference occurs, where an intensity peak will be detected. This condition is the Bragg law of diffraction in crystals, and we have

$$2d \cdot \sin\theta = n\lambda, \tag{2.12}$$

where n is an integer and λ is the wavelength.When Bragg condition is satisfied at an incident angle of θ, it is actually satisfied by all parallel lattice planes associated with this direction. In other words, the beam reflected from every parallel lattice plane is in phase. This is the key to the formation of bright diffraction peaks, and the necessary condition is $\lambda \leq 2d$. The lattice spacing d is of the order of a few angstroms (Å) that matches or is greater than typical wavelengths of x-ray. Therefore, x-ray can satisfy the Bragg condition and is suitable for diffraction experiments. Not all planes can contribute in phase if $\lambda > 2d$. For visible lights that have wavelengths around 5000 Å, the Bragg condition can only be satisfied about every 5000 layers, and the random-phase reflection from layers in between will wash out any brightness contrast. So, Bragg diffraction will not happen for visible lights. Bragg law can predict the angle of diffraction for a given d

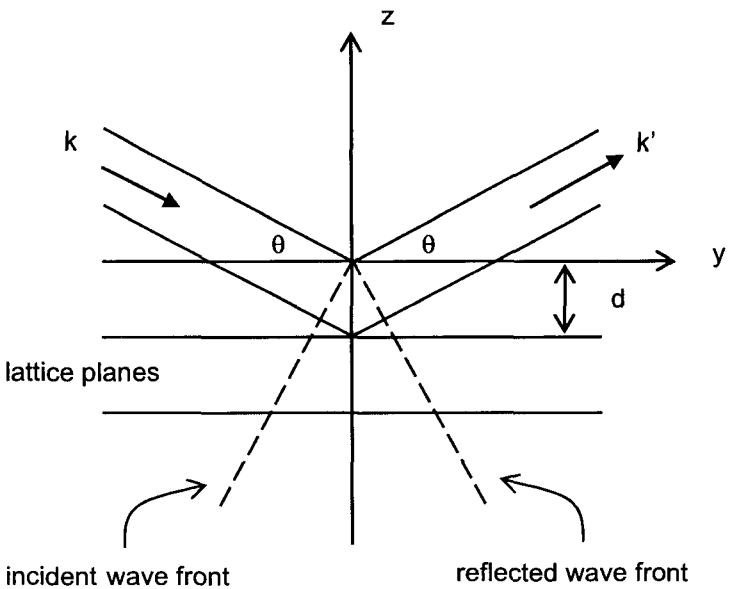


Fig. 2.1 Schematic drawing to show the Bragg law of x-ray diffraction. The x-axis is perpendicular to the paper and the wave front is perpendicular to the beam. The lattice plane is assumed to reflect like a mirror at the atomic site and is transparent elsewhere.

and λ value, but it does not say anything about the intensity and width of the diffracted beam. This deficiency stems from the assumption of mirror-like lattice planes, and it can be rectified if we take the actual scattering centers into account. The x-ray, with the wavelength in the vicinity of 1Å and the energy around 12.3 KeV, can not penetrate the electron cloud surrounding the nucleus. Photons in this energy range will either bounce off the electron cloud or be absorbed by the electron and re-emitted later. The electron cloud is a distribution of electrons described by the electron density function $n(\mathbf{r})$.

2.4 Electron Density in Crystal Lattice

Since electron density is a physical observable, the density function $n(\mathbf{r})$ must look identical as we move from $\mathbf{r}$ to $\mathbf{r} + \mathbf{R}$ by a lattice vector $\mathbf{R}$. We

have the translation invariance condition expressed by

$$n(\mathbf{r}) = \mathbf{n}(\mathbf{r} + \mathbf{R}). \tag{2.13}$$

We also know that we can always expand a periodic function into a Fourier series.

$$\begin{aligned} n(\mathbf{r}) &= \sum_{\mathbf{k}} n_{\mathbf{k}} \cdot e^{i\mathbf{k}\cdot\mathbf{r}} \\ n(\mathbf{r}+\mathbf{R}) &= \sum_{\mathbf{k}} n_{\mathbf{k}} \cdot e^{i\mathbf{k}\cdot(\mathbf{r}+\mathbf{R})} \\ &= \sum_{\mathbf{k}} n_{\mathbf{k}} \cdot e^{i\mathbf{k}\cdot\mathbf{r}} \cdot e^{i\mathbf{k}\cdot\mathbf{R}} \\ &= n(\mathbf{r}). \end{aligned} \tag{2.14}$$

In order to guarantee a translation invariant $n(\mathbf{r})$, i.e., $n(\mathbf{r}) = \mathbf{n}(\mathbf{r} + \mathbf{R})$, we must require $\exp(i\mathbf{k} \cdot \mathbf{R}) = 1$ for every $\mathbf{k}$ in the Fourier series. Any $\mathbf{k}$ satisfies $\exp(i\mathbf{k} \cdot \mathbf{R}) = 1$ is a reciprocal lattice vector. Therefore, only the $\mathbf{k}$ vectors that are reciprocal lattice vectors are allowed in the Fourier expansion of $n(\mathbf{r})$. To reflect this fact, we replace $\mathbf{k}$ with $\mathbf{G}$ in the Fourier expansion of the electron density function.

$$n(\mathbf{r}) = \sum_{\mathbf{G}} \mathbf{n}_{\mathbf{G}} \cdot \mathbf{e}^{\mathbf{i}\mathbf{G}\cdot\mathbf{r}}. \tag{2.15}$$

2.5 Diffraction Beam Intensity

When a beam of x-ray enters a crystal, the elastically reflected beam will contain rays that have traveled different path lengths, because the finite dimension of the beam. This situation is illustrated in Fig. 2.2. If one edge of the beam gets reflected at point O that we take as the coordinate origin, and if the opposite edge of the beam gets reflected at point $\mathbf{r}$, there is a path-length difference between the rays reflected from these two points. The path-length difference in the incident rays is $(\mathbf{k}/k) \cdot \mathbf{r}$, and the path-length difference in the reflected rays is $-(\mathbf{k}'/k') \cdot \mathbf{r}$. The total path-length difference is

$$S = (\mathbf{k}/k) \cdot \mathbf{r} - (\mathbf{k}'/k') \cdot \mathbf{r}. \tag{2.16}$$

Since $k = k' = 2\pi/\lambda$, the total lag in phase angle is

$$(S/\lambda) \cdot 2\pi = \mathbf{k} \cdot \mathbf{r} - \mathbf{k}' \cdot \mathbf{r}. \tag{2.17}$$

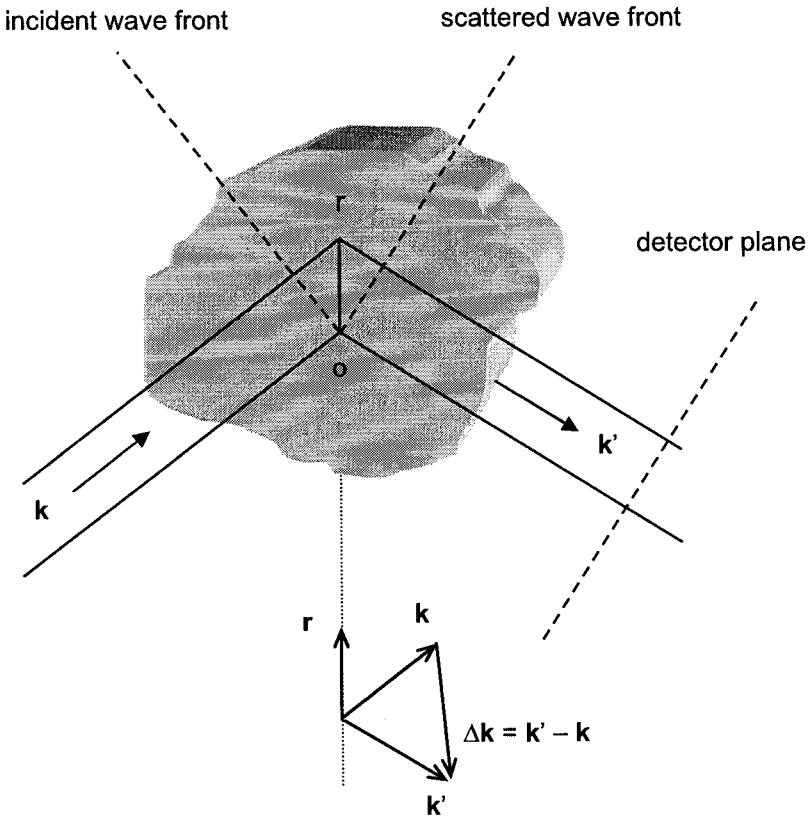


Fig. 2.2 The schematic of path-length difference in elastic scattering of x-ray from two representative points, o and r. The incident and scattered wave vectors are equal in value, but altered in direction. The geometry of $\mathbf{r}$, $\mathbf{k}$, $\mathbf{k}'$, and $\Delta\mathbf{k}$ is highlighted. At diffraction peaks, where the von Laue's law of $\Delta\mathbf{k} = \mathbf{G}$ is obeyed, $\Delta\mathbf{k}$ is by definition perpendicular to a set of parallel lattice planes.

Therefore, the rays reflected by a small volume ΔV around point $\mathbf{r}$ will have an intensity that is proportional to $n(\mathbf{r})$ and the phase factor $\exp i(\mathbf{k} \cdot \mathbf{r} - \mathbf{k}' \cdot \mathbf{r})$. The overall intensity detected at the detector is proportional to the superposition of all rays scattered by all the illuminated crystal volume. Let us define this integral as F, which will represent the

scattering amplitude,

$$\begin{aligned} F &= \int n(\mathbf{r}) e^{i(\mathbf{k}\cdot\mathbf{r}-\mathbf{k}'\cdot\mathbf{r})} dV \\ &= \int \sum_{\mathbf{G}} n_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \cdot e^{-i\Delta\mathbf{k}\cdot\mathbf{r}} dV \\ &= \sum_{\mathbf{G}} \int n_{\mathbf{G}} e^{i(\mathbf{G}-\Delta\mathbf{k})\cdot\mathbf{r}} dV. \end{aligned} \tag{2.18}$$

If we know the electron density function $n(\mathbf{r})$, the Fourier coefficient $n_{\mathbf{G}}$ can be determined. Usually, a few dominant terms in the Fourier series are sufficient and the integration of F can be carried out. Even if we do not know anything about $n_{\mathbf{G}}$, we can still learn some important properties about the F integral by examining its phase angle. First, let us express $\mathbf{r}$ in terms of the primitive vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ in the real space. Both $\mathbf{G}$ and $\Delta\mathbf{k}$ are vectors in the k space spanned by $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$. We have

$$\mathbf{r} = r_1\mathbf{a}_1 + r_2\mathbf{a}_2 + r_3\mathbf{a}_3; \tag{2.19}$$

$$\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3; \tag{2.20}$$

$$\Delta\mathbf{k} = \Delta_1\mathbf{b}_1 + \Delta_2\mathbf{b}_2 + \Delta_3\mathbf{b}_3; \tag{2.21}$$

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi \cdot \delta_{ij}. \tag{2.22}$$

The indices i and j range from 1 to 3. The electron density $n(\mathbf{r})$ is a continuous function of r_1, r_2, and r_3. These position variables vary continuously, not like the discrete integers when we count for the lattice points. For any given reciprocal lattice vector $\mathbf{G}$, m_1, m_2, and m_3 are always integers. But, the wave vector difference $\Delta\mathbf{k}$, i.e.,Δ_1, Δ_2, and Δ_3 are continuously changing due to the rotation of the crystal in the experiment. A complete scan of all scattering angles will exhaust all phase angles in the F integral. When F goes through peaks and valley at the corresponding phase angles, we will be able to determine the relationship between (Δ_1, Δ_2, Δ_3) and (m_1, m_2, m_3).

Since $n_{\mathbf{G}}$ is not a function of $\mathbf{r}$, we will put it outside the F integral, and decompose the volume integral along three primitive lattice vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. We have

$$F = \sum_{m_1,m_2,m_3} n_{\mathbf{G}} v \int e^{i2\pi(m_1-\Delta_1)r_1} e^{i2\pi(m_2-\Delta_2)r_2} e^{i2\pi(m_3-\Delta_3)r_3} dr_1 dr_2 dr_3, \tag{2.23}$$

where the summation is over all m_i, and v is the primitive cell volume in real space.

$$F = \sum_{m_1,m_2,m_3} n_{\mathbf{G}} v F_1 F_2 F_3 \tag{2.24}$$

$$\propto \int_{-N_1}^{N_1}\int_{-N_2}^{N_2}\int_{-N_3}^{N_3} e^{i2\pi(m_1-\Delta_1)r_1} e^{i2\pi(m_2-\Delta_2)r_2} e^{i2\pi(m_3-\Delta_3)r_3} dr_1 dr_2 dr_3.$$

The integration limits go from $-N_1$ to N_1, $-N_2$ to N_2, and $-N_3$ to N_3. Those are the boundary of the crystal volume penetrated by the x-ray beam. Since this volume is quite large comparing to the lattice spacing, we do not lose generality to set the volume boundary on lattice points. Therefore, N_1, N_2, and N_3 are assumed to be integers. Let us examine the r_1 component first.

$$\begin{aligned} F_1 &= \int_{-N_1}^{N_1} e^{i2\pi(m_1-\Delta_1)r_1} dr_1 \\ &= \frac{1}{i2\pi(m_1-\Delta_1)} \cdot [e^{i2\pi(m_1-\Delta_1)N_1} - e^{-i2\pi(m_1-\Delta_1)N_1}] \\ &= 2N_1 \frac{\sin[2\pi(m_1-\Delta_1)N_1]}{2\pi(m_1-\Delta_1)N_1}. \end{aligned} \tag{2.25}$$

In x-ray diffraction experiment using a diffractometer, the crystal is rotated with respect to the incident beam. This is in effect equivalent to scanning $\Delta\mathbf{k}$ through the k space. When $\Delta_1 = m_1$, $F_1 \propto 0/0 = 1$ which gives the maximum value for $F_1 = 2N_1$. If Δ_1 deviates from m_1, F_1 decreases quickly to zero. To see this, let us assume that $\Delta_1 = m_1 + 1/2N_1$, and evaluate F_1. We found $F_1 \propto 0/(-\pi/N_1) = 0$. Since N_1 is a large number, $1/2N_1$ is very small and the width of diffraction peak is narrow. The diffraction peak intensity is proportional to F_1^2. We plot the normalized F_1 square, $(F_1/2N_1)^2$, versus $(\Delta_1 - m_1)\cdot N_1$ to illustrate the point in Fig. 2.3. The diffraction peak has fine structures, namely, a very pronounced center peak at $\Delta_1 = m_1$ and a series of decaying peaks as Δ_1 moves away from m_1. The zero intensity occurs at $\Delta_1 = m_1 + j/2N_1$, where $j = \pm$ integers. By the same token, we can verify the same for the other two directions. In order to have $F \propto F_1F_2F_3 = 1$, we must have $\Delta_1 = m_1$, $\Delta_2 = m_2$, and $\Delta_3 = m_3$. If any one of Δ_i deviates from their corresponding m_i slightly by as little as $1/2N_i$, F goes to zero. Except the term $\mathbf{G} = \Delta\mathbf{k}$, other $\mathbf{G}$ terms in the summation $\sum_{\mathbf{G}}$ drop out. We have at the center peak

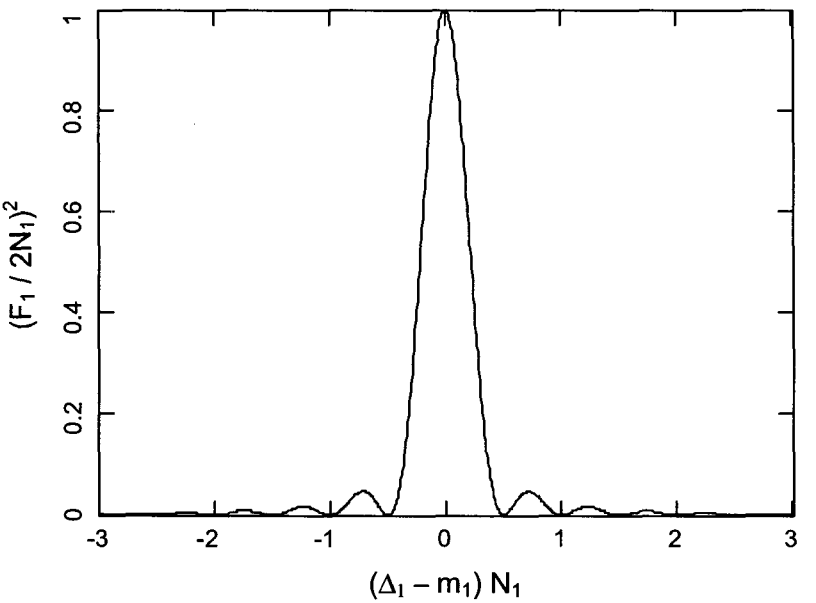


Fig. 2.3 Normalized diffraction intensity as a function of normalized phase angle. Only one component of the F integral is shown. In the normalized plot, the other two components are identical to this one. Just change the subscript to 2 and 3, respectively.

$F = n_{\mathbf{G}} \cdot (2N_1 \cdot 2N_2 \cdot 2N_3) \cdot v = n_{\mathbf{G}} \cdot V$, where V is the volume sampled by the x-ray.

We come to a very important conclusion: The scattering vector $\Delta\mathbf{k}$ must equal to a reciprocal lattice vector $\mathbf{G}$ in order to have a maximum in the scattering amplitude F. Once $\Delta\mathbf{k}$ deviates from $\mathbf{G}$, F decreases to zero quickly. This effect repeats itself when $\Delta\mathbf{k}$ is scanned to match another $\mathbf{G}$ vector.

2.6 von Laue's Diffraction Law

We have just derived the diffraction condition based on von Laue's approach. Translation invariance in the Bravais lattice is assumed to describe the electron density. Also assumed is the elastic scattering between photons

and electrons. The von Laue diffraction law is simply expressed as

$$\Delta \mathbf{k} = \mathbf{G}. \tag{2.26}$$

If we have an experimental way to scan all possible $\Delta \mathbf{k}$ in k-space, we will have a map of bright spots whenever the condition $\Delta \mathbf{k} = \mathbf{G}$ is met. Therefore, x-ray diffraction is a direct way to image the reciprocal lattice defined by $\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3$. Neutron and electron diffraction will achieve similar goals. By comparison, scanning tunneling electron microscope (STEM) maps the lattice in real space. The experiment and von Laue theory of x-ray diffraction was first carried out around 1912. The STEM was not invented until the early 1980s.

We may rearrange the von Laue diffraction condition from $\mathbf{k}' - \mathbf{k} = \mathbf{G}$ to the square form of $(\mathbf{k} + \mathbf{G})^2 = \mathbf{k}'^2$. Since only the beam direction is changed in elastic scattering, we can make use of the fact of $k' = k$ to reduce the square equation to a simpler form. We have

$$|\mathbf{k} \cdot \mathbf{G}| = \frac{1}{2}G^2. \tag{2.27}$$

This form of von Laue diffraction condition says explicitly that the projection of $\mathbf{k}$ on $\mathbf{G}$ must equal $G/2$ for diffraction to happen. The geometry is shown in Fig. 2.4. The bisecting plane of $\mathbf{G}$ contains the end points of $\mathbf{k}$ and $\mathbf{k}'$, which can be scattered into each other by involving the $\mathbf{G}$ vector. At this juncture, we may want to recall that there is a set of parallel lattice planes perpendicular to a set of parallel $\mathbf{G}$ vectors. The intra-plane spacing is related to the shortest $\mathbf{G}$ vector by $G_s = 2\pi/d$, and the $\mathbf{G}$ vector in the set is a multiple of $\mathbf{G}_s$, i.e., $\mathbf{G} = n\mathbf{G}_s$. With these properties in mind and with the help of Fig. 2.4, we can verify that the von Laue diffraction law and Bragg's law are actually equivalent. Therefore, the elastic scattering involving a reciprocal lattice vector as shown in Fig. 2.4 is usually referred as Bragg scattering. The condition for Bragg scattering and von Laue diffraction is the same. Bragg scattering is a rather severe scattering process, because the wave moving in the direction of $\mathbf{G}$ will be reflected totally into the reverse direction.

2.7 Brillouin Zones

From a reciprocal lattice point O in k-space as shown in Fig. 2.4, we can draw a reciprocal lattice vector $\mathbf{G}$ pointing to another reciprocal lattice

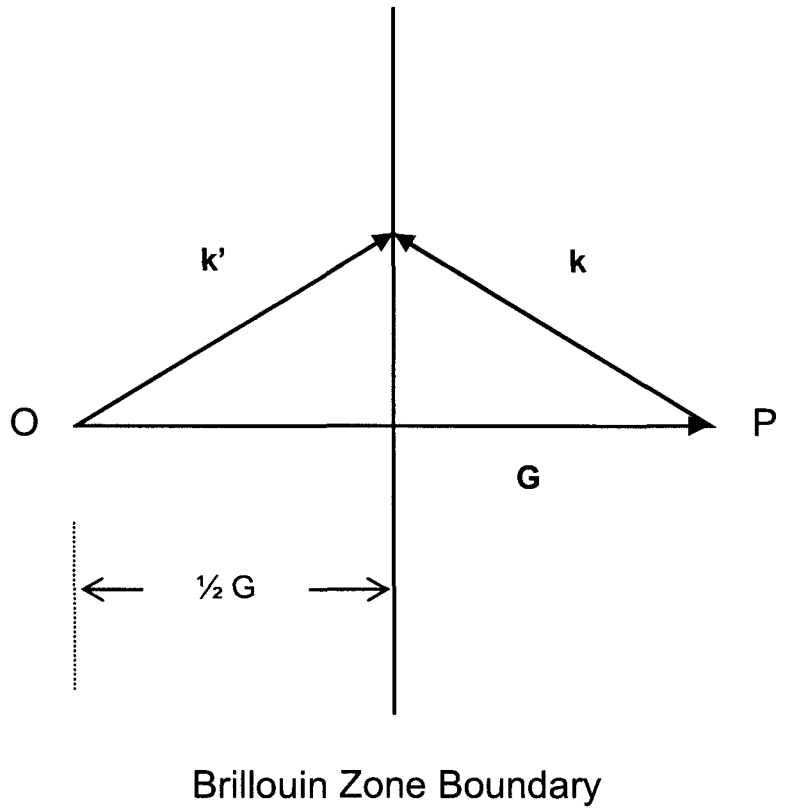


Fig. 2.4 Brillouin zone boundary is a perpendicular plane that bisects the reciprocal lattice vector **G**. Here, the wave vectors **k** and **k'** that satisfy the von Laue diffraction law are undergone Bragg scattering. When **G** lies in the plane of the paper as shown here, both **k** and **k'** are not necessarily so in general.

point P. We can then draw a bisecting plane perpendicular to **G** at midpoint of OP. Any k vector starting from either O or P and ending on the bisecting plane will satisfy the diffraction condition $\mathbf{k} \cdot \mathbf{G} = \frac{1}{2}G^2$ or $\mathbf{k}' = \mathbf{k} + \mathbf{G}$.

By the same procedure, we can keep drawing perpendicular bisecting planes to all **G** vectors originated from point O. These planes will form closed cells of various sizes. The smallest one enclosing the reciprocal lattice point O at its center is by definition the Wigner-Seitz cell in the reciprocal lattice space. Customarily, we call it the 1^{st} Brillouin zone. The perpen-

dicular planes bisecting all **G** vectors divide the k space into fragments of n^{th} order of Brillouin zones. Any **k** vector originated from the center of the 1^{st} Brillouin zone will not cross any zone boundaries if **k** stays inside the 1^{st} Brillouin zone. If **k** crosses the zone boundary once, **k** is in the 2^{nd} Brillouin zone. If it crosses $n-1$ zone boundaries, it is in the n^{th} Brillouin zone. The zone boundaries are bisecting planes that extend as far as the reciprocal lattice goes. The zone boundaries of a lower order zone will participate to define higher order zones. When we talk about Brillouin zones, we always refer to the k-space or the reciprocal lattice space.

Now, let us go back to the 1^{st} Brillouin zone that is the Wigner-Seitz cell in the reciprocal lattice space. Any **k** vector originates from the center and ends up on the zone boundary will satisfy the diffraction condition. The resulted Bragg scattering causes a large directional change for the traveling wave by involving a reciprocal lattice vector **G**, but the energy is conserved. The Bragg scattering at the zone boundary characteristically affects wave propagation, and is of great importance in solid state physics. In addition to x-ray diffraction, thermal conductivity in insulators and electron energy band structures are other examples that are greatly influenced by the Bragg scattering. Here, phonons and electrons are scattered at the zone boundary similar to electromagnetic waves.

In Fig. 2.5, a 5×5 array is shown to schematically represent a square reciprocal lattice. The dash lines are reciprocal lattice vectors. They are **G** vectors originated from the center point to all the other 24 points. The solid lines are part of their perpendicular bisecting planes. The center cell in off-white is the 1^{st} Brillouin zone. The 2^{nd}, 3^{rd}, and 4^{th} zones are shaded with dark gray, light gray, and gray, respectively.

2.8 Effects of Electron Distribution around Atoms in the Unit Cell

When atoms are brought together to form the crystal, the electron distribution of the free atom must be altered by the proximity of other atoms. Since x-ray diffraction originates from photon-electron scattering, we may think that the detailed knowledge about the electron distribution in crystals is important if we attempt to analyze the diffraction data quantitatively. Surprisingly, this consideration does not seem to matter very much in x-ray diffraction. Experiments show that the x-ray diffraction peak intensities of metallic iron, copper, and aluminum agree with their theoretical free-atom

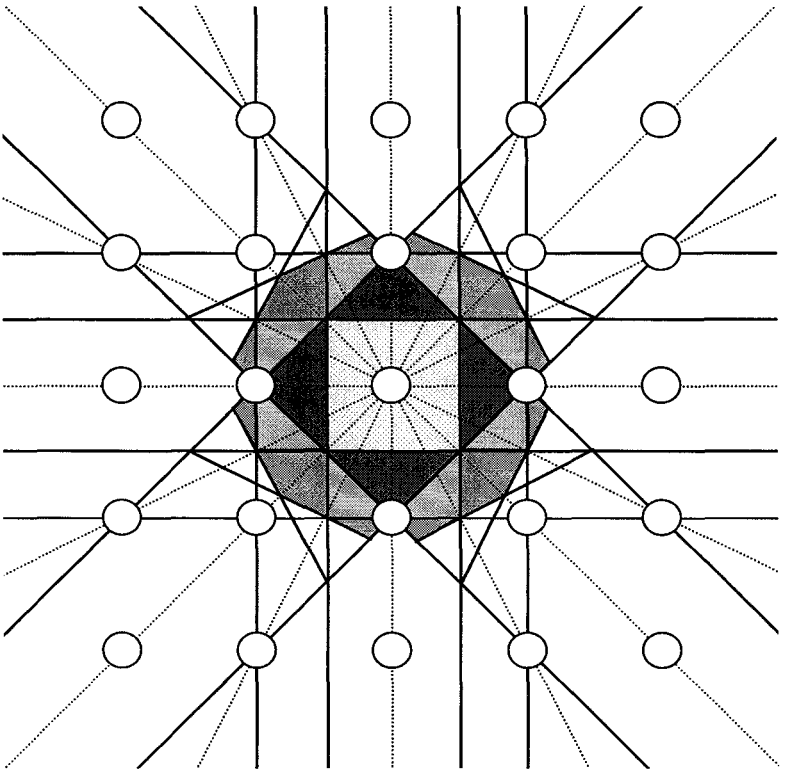


Fig. 2.5 Some lower-order Brillouin zones in the two-dimensional reciprocal lattice of a 5×5 square array. The reciprocal lattice vectors are shown in dash lines, connecting the center points to the other 24 points. The solid lines are some representatives of the Brillouin zone boundaries. Each one is a bisector of one particular reciprocal lattice vector. The smallest area enclosing the center reciprocal lattice point is the first Brillouin zone shown in off-white gray. The second zone is in darker gray, the third in light gray, and the fourth is in gray.

values within 1%. Therefore, we do not need to know every details of the electron density function $n(\mathbf{r})$ in order to make theoretical calculations for the x-ray diffraction experiments.

Let us recall that the scattering amplitude that satisfies diffraction condition $\Delta\mathbf{k} = \mathbf{G}$ is

$$F = \int n(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} dV. \tag{2.28}$$

The integration over the whole scattering volume is equivalent to integrat-

ing over one unit-cell first, and then multiplying it with the number of cells. Hence, we rewrite F as

$$F = N \int_{cell} n(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} dV = N \cdot S. \tag{2.29}$$

S is called the *structure factor*, and N is the total number of unit cells in the scattering volume. Sometimes it is more convenient to write $n(\mathbf{r})$ as a superposition of contributions from all atomic sites. Let the j^{th} atom be located at $\mathbf{r}_j$, we have

$$n(\mathbf{r}) = \sum_j n_j(\mathbf{r} - \mathbf{r}_j), \tag{2.30}$$

where the index j goes through all basis atoms in the cell. The index j is equal to 1 in a primitive cell. For a conventional cell, j only goes through the basis atoms. The structure factor S may be written as a summation of integrals over all basis atoms in a unit cell:

$$\begin{aligned} S &= \sum_j \int_{cell} n_j(\mathbf{r} - \mathbf{r}_j) e^{-i\mathbf{G}\cdot\mathbf{r}} dV \\ &= \sum_j \int_{cell} n_j(\mathbf{r} - \mathbf{r}_j) e^{-i\mathbf{G}\cdot(\mathbf{r}-\mathbf{r}_j)} \cdot e^{-i\mathbf{G}\cdot\mathbf{r}_j} dV \\ &= \sum_j e^{-i\mathbf{G}\cdot\mathbf{r}_j} \int_{cell} n_j(\mathbf{r} - \mathbf{r}_j) e^{-i\mathbf{G}\cdot(\mathbf{r}-\mathbf{r}_j)} dV \\ &= \sum_j f_j e^{-i\mathbf{G}\cdot\mathbf{r}_j}, \end{aligned} \tag{2.31}$$

where f_j is called the *atomic form factor*. We have

$$f_j = \int_{cell} n_j(\mathbf{r} - \mathbf{r}_j) e^{-i\mathbf{G}\cdot(\mathbf{r}-\mathbf{r}_j)} dV. \tag{2.32}$$

The atomic form factor describes the diffraction of x-ray by a single atom in the unit cell. As talked about earlier, the electron distribution of a freely standing atom can be used to calculate the atomic form factor. It agrees with experiment within 1%.

2.8.1 *Atomic form factor*

When we calculate f_j for the j^{th} atom, we can set the coordinate zero at $\mathbf{r}_j$ and let $\mathbf{G}$ coincide with the z-axis. Since the electrons surrounding this

atomic site is closely resemble that of a free atom, we can safely assume a spherical symmetry for the electron density function.

$$\begin{aligned} f_j &= \int_{cell} n_j(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} dV \\ &= \int_{cell} n_j(\mathbf{r}) e^{-iG\cdot r\cdot\cos\theta} \cdot r^2 \sin\theta d\theta d\phi dr \\ &= 2\pi \int_{r,\theta} n_j(\mathbf{r}) e^{-iG\cdot r\cdot\cos\theta} \cdot r^2 d(\cos\theta) dr \\ &= 4\pi \int n_j(\mathbf{r}) \cdot r^2 \cdot \frac{\sin(G\cdot r)}{G\cdot r} dr. \end{aligned} \tag{2.33}$$

We know that the electron density integration must be equal to the total number of atomic electrons Z. We have

$$Z = 4\pi \int n_j(\mathbf{r}) \cdot r^2 \cdot dr. \tag{2.34}$$

In a hypothetical situation if all Z electrons are concentrated at $r = 0$, we have $(\sin(G\cdot r))/(G\cdot r) = 1$ and $f_j = Z$. Take this hypothetical case further by assuming $n(r) = Zn_0(r)$ where $n_0(r)$ is the one-electron probability distribution, we can say that the atomic form factor f is equal to Z times the scattering amplitude of a single electron. In other words, the atomic scattering amplitudes should be the same if two atoms have identical number of electrons. Experimentally, this is verified by K^+ and Cl^- ions in KCl crystals.

2.8.2 *Structure factor for the bcc lattice*

Now we will examine the structure factor in more detail. We recall the formula for the structure factor:

$$\begin{aligned} S &= \sum_j f_j \cdot e^{-i\mathbf{G}\cdot\mathbf{r}_j} \\ &= \sum_j f_j \cdot e^{-i2\pi(m_1x_j + m_2y_j + m_3z_j)}. \end{aligned} \tag{2.35}$$

For a conventional unit cell in bcc lattice, we have a basis of two identical atoms located at sites (0,0,0) and $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$. The structure factor becomes

$$S = f \cdot \{1 + e^{-i\pi(m_1+m_2+m_3)}\}. \tag{2.36}$$

Therefore,we have

$$\begin{aligned} S &= 0, \qquad \text{if } m_1 + m_2 + m_3 = \text{odd integers;} \\ S &= 2f, \qquad \text{if } m_1 + m_2 + m_3 = \text{even integers.} \end{aligned} \tag{2.37}$$

For example, metallic sodium has bcc lattice structure. Its diffraction pattern does not contain lines like (100), (111), (221), and so on, as long as the Miller indices add up to an odd number. The lines like (200), (110), and (222) etc. are present.

For a properly selected x-ray wavelength, one can always find an angle of incidence, such that the Bragg law is satisfied for the (100) plans in the simple cubic lattice. In other words, the x-ray reflected by the consecutive (100) planes differs by a phase of 2π, and constructive interference results. There will be a (100) line in the diffraction pattern in the sc lattice. We may consider the bcc structure as a sc lattice with an intervening plane added in between two adjacent (100) planes. Now, the x-ray reflected from consecutive planes will differ by a phase of π, and it causes destructive interference. Therefore, the x-ray gets canceled out and there is no (100) line for the bcc lattice.

2.8.3 *Structure factor for the fcc lattice*

The basis of the fcc structure in a conventional unit cell has 4 identical atoms at $(0,0,0)$, $(0,\frac{1}{2},\frac{1}{2})$, $(\frac{1}{2},0,\frac{1}{2})$, and $(\frac{1}{2},\frac{1}{2},0)$. We have the structure factor equal to

$$S = f \cdot \{1 + e^{-i\pi(m_1+m_2)} + e^{-i\pi(m_2+m_3)} + e^{-i\pi(m_1+m_3)}\}. \tag{2.38}$$

It is not hard to see that

$$\begin{aligned} S &= 0, \qquad \text{if one of } (m_1, m_2, m_3) \text{ is even and two are odd,} \\ &\qquad\quad \text{or, one odd and two even;} \\ S &= 4f, \qquad \text{if } m_1, m_2, \text{ and } m_3 \text{ are all even, or all odd.} \end{aligned} \tag{2.39}$$

KBr has the fcc structure. Its x-ray powder pattern has lines like (200), (220), (222), and (400) etc. for all even indices, and (111), (311), and (331) etc. for all odd indices. It does not have lines like (122) and (211).

Problems

1. Verify that the reciprocal lattice of bcc lattice is a fcc lattice, and the reciprocal lattice of fcc lattice is a bcc lattice. If the lattice constant of the conventional unit cell is $a = 2\mathring{A}$ in real space, find the corresponding reciprocal lattice constants for both bcc and fcc lattices.

2. Consider the planes with indices (100) and (001) obtained from the conventional fcc unit cell, what are the indices of these planes in the system of primitive unit vectors, referring to Fig. 1.6(b) in Chapter 1.

3. Consider a set of lattice planes (hkl).
   (a) Prove the reciprocal lattice vector $\mathbf{G} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3$ is perpendicular to these planes.
   (b) Prove that two adjacent planes are separated by $d = \frac{2\pi}{G}$.
   (c) Show $d = \frac{a}{\sqrt{h^2+k^2+l^2}}$ for simple cubic lattice.

4. Verify that von Laue diffraction law and Bragg's law are equivalent.

5. Design an experiment to detect all possible lattice planes in a given crystal type.

6. Make 3-dimensional cardboard models of the first Brillouin zone for both fcc and bcc lattices.

Chapter 3

Lattice Vibrations and Phonons

Introduction and Summary

Chapter 3 has two parts. The objective of the first part is to learn about the wave nature of lattice vibration. The seemly random lattice vibration can always be analyzed in terms of plane waves. For the special case where all constituting plane waves are propagating along the direction of a reciprocal lattice vector, the entire lattice plane perpendicular to the wave vector will move in phase as either longitudinal or transverse waves. The analysis reduces to a problem that can be represented with a one-dimensional lattice. The basic concepts of lattice vibration learned here are as general as what can be learned in the 3-dimensional case. The equation of motion for a particular atom in the linear lattice is derived from Hooke's law and Newton's second law. The discreteness of the lattice is shown to diminish in the long wavelength limit. The equation of motion reduces to the familiar wave equation that has an obvious solution of the traveling plane wave. When the same solution is extended to the discrete lattice for all the wavelengths, both the sufficient and necessary conditions turn out to be the dispersion relation $\omega(\mathbf{k})$. The force constant is shown to be calculable from the experimental dispersion relation. The long wavelength approximation, applied either to the dispersion relation or to the equation of motion, is shown to give the identical result. The periodic function of $\omega(\mathbf{k})$ is plotted versus the wave vector by using the nearest neighbor approximation. An elaborated explanation is given to show why the phase velocity is ω/k and the group velocity is $d\omega/dk$. Why the k and $k+2n\pi/a$ waves are equivalent is explained. It reaches the conclusion that all independent modes should be located inside the first Brillouin zone. These facts give additional physical significance to the first Brillouin zone. The concepts of wave packet,

lattice diffraction, standing wave, and the first Brillouin zone are all pulled together. So, it is possible to explain why the group velocity $d\omega/dk$ is zero at the first Brillouin zone boundary, and why it increases slowly and approaches ω/k as k moves toward the center of the zone.

The second part takes on the task to build the concept of phonons. The difficulties with the classical approach to the understanding of lattice vibrations are pointed out, and the need for quantum considerations becomes clear. The quantization procedure begins with setting up the classical Hamiltonian of all atoms. In the limit of small oscillations, the potential energy reduces to a quadratic function of position variables, where the atoms are coupled in all possible pairs. The cross term requires simultaneous solution to the dynamic state of all atoms. To get around this challenging hurdle, the need for a transformation that can diagonalize the potential matrix becomes clear. At this juncture, periodic boundary condition is introduced to construct the N normal modes for the linear lattice of N atoms. The atomic displacement is obviously made of the contributions from all N normal modes. This observation leads to the appreciation of Fourier relationships between (q_s, p_s) and (Q_k, P_k). The normal modes are proved to be mutually orthogonal, and the vector interpretation of Fourier series is discussed. The normal-mode variables (Q_k, P_k) are shown to be equivalent to (q_s, p_s) in describing the same dynamic state of N atoms. Quantum operators, commutation rules, and the uncertainty principle are then introduced. The Hamiltonian under consideration is a Hermitian matrix that can be diagonalized by a unitary matrix. It is shown that the plane-wave phase factor is actually the matrix element of such a transformation matrix. When the coordinate is transformed from (q_s, p_s) to (Q_k, P_k), the Hamiltonian is diagonalized from N coupled atoms into separated pairs of $(k, -k)$ plane waves. At this point, Dirac's treatment of the simple harmonic oscillator is introduced. In this Hilbert vector space, the creation, annihilation, and number operators are defined. The analysis based on the axioms of Hilbert space leads to the construction of the basis eigenvectors and the eigenvalue spectrum. The result reveals the other side of the quantum simple harmonic oscillator. It turns out to be a set of elementary energy excitations. Back to the field of lattice vibrations, it is shown that one can construct the creation and annihilation operators if the proper combinations of the normal-mode operators are made. The Hamiltonian is then transformed to the trace of the diagonal matrix of number operators. Each diagonal term of a specific k value represents the simple harmonic oscillator that is already solved in the Dirac's treatment. Hence,

the solution can be carried over without any more modifications. Emerged naturally is the interpretation of the energy spectrum in terms of phonons. Final remarks are given to show the connection between phonons and the classical picture of lattice vibrations in the limit of large quantum numbers.

Lattice Vibration

3.1 Equation of Motion in the One Dimensional Lattice

Atoms in crystals vibrate naturally around their equilibrium lattice positions because of temperature. Both sound and light, for example, can induce additional atomic vibrations. We refer all these to lattice vibrations. The lattice vibration is not a collection of isolated atomic events, but a correlated motion governed by the coupling forces among all atoms. The net force on each atom primarily results from the Coulomb interaction with other nuclei and their surrounding electron charges. Since electrons are much lighter than nuclei, the electron redistribution in response to any nucleus movement is considered instantaneous. If all atoms are at their equilibrium lattice sites, the net force will be zero. As soon as one of the atoms moves, it feels a restoring force and also initiates the pushing as well as the pulling forces on its neighbors in the same time. The positional disturbance of atoms gets transmitted in the manner as if the atoms were connected to each other by massless springs. Taking any pair of atoms, we can expand their coupling potential in terms of the Taylor series around their equilibrium positions. For small atomic displacements, much smaller than the lattice spacing, the only important term in the pair coupling potential is the quadratic term, and the higher-order terms are negligible unless the atomic displacement gets very large. Because the difference in their governing potential, the small oscillation is harmonic and the large oscillation is anharmonic. The former is simpler to analyze, but yet very important fundamentally. Hence, we will concern ourselves only with the harmonic lattice vibration for the time being. Let us draw analogy from the familiar harmonic oscillator of a mass suspended by a single spring. When the mass is put into small oscillations, a transverse sinusoidal wave will propagate perpendicular to the oscillatory direction, for instance, in an attached string. Meanwhile, a longitudinal sinusoidal wave will appear in the air compressed by the oscillatory mass. We may intuitively guess that similar wave motions will appear in the lattice if atoms vibrate around

their equilibrium positions, although it may not be a simple sinusoidal wave because the coupling of multiple springs.

With the help of Fourier analysis, we will be able to decompose the complex lattice vibration into a series of sinusoidal waves. These are plane waves of various wavelengths and frequencies. The understanding of lattice vibrations boils down to the understanding of plane lattice waves, which are usually given the name of normal modes borrowed from the elastic wave theory. Each normal mode is characterized by a particular harmonic oscillation going through all atoms, and consists of three mutually orthogonal branches. They are one longitudinal and two transverse branches, specified by the amplitude polarization with respect to the wave vector. The polarization comes from the atomic degree-of-freedom in the primitive cell, and is not too hard to visualize in isotropic cubic crystals. In anisotropic crystals, the polarization vectors are not so obvious to see. They depend upon the actual symmetry of the crystal. Nevertheless, as proved in Chapter 2, there is always a set of lattice planes perpendicular to any given reciprocal lattice vector. If we choose the plane lattice wave to propagate along a symmetry axis in the direction of a reciprocal lattice vector, the entire lattice plane perpendicular to this axis will move in phase as the wave front passing through, either in parallel or perpendicular to the wave vector. The former is the longitudinal wave, and the latter is the transverse wave. In either case, the analysis reduces down to the simpler problem of a one-dimensional lattice.

In the spring model of a linear lattice where Hooke's law applies, the force acted on the s^{th} atom by the n^{th} atom depends linearly upon the distance between them. This distance is determined by the displacements of both atoms measured from their respective equilibrium lattice sites. When they are at the lattice sites, the individual displacements are zero and so is their relative displacement. The imaginary spring is relaxed. Hence, the force exerts on each other is zero. When their relative displacement is not zero, the imaginary spring is either compressed or stretched to produce the pushing or pulling force, respectively. In either case, both atoms will feel the same force exerted by the spring except in opposite directions.

Figure 3.1 shows that the force F_s exerted on the s^{th} atom by the $(s+1)^{th}$ atom is always proportional to their relative atomic displacements $q_{s+1} - q_s$. The total force exerted on the s^{th} atom by $s \pm 1$, $s \pm 2$, and up to $s \pm n$ atoms is the summation of forces between all individual pairs. We obtain the equation of motion according to Newton's second law, $F = ma$, where the acceleration is the second derivative of the displacement. We

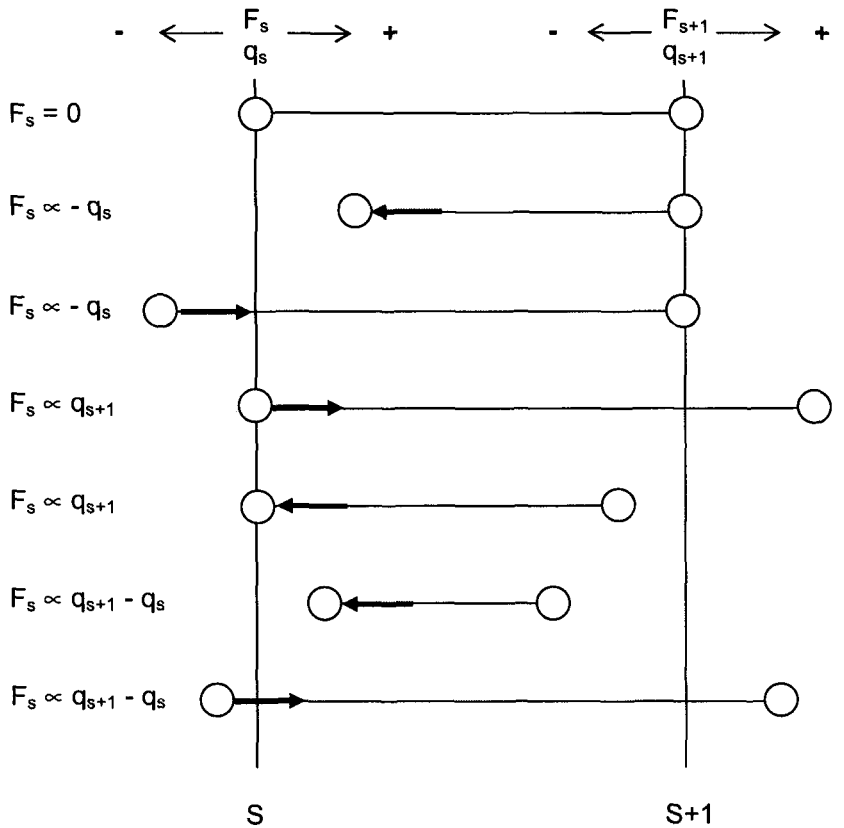


Fig. 3.1 When the spring is compressed, there is a pushing force on the s^{th} atom toward the left as shown. When the spring is stretched, the force is pulling toward the right. The state of the spring depends on the relative atomic displacement.

have

$$F_s = \sum_n C_n(q_{s+n} - q_s), \text{ and}$$

$$m\frac{d^2q_s}{dt^2} = \sum_n C_n(q_{s+n} - q_s), \tag{3.1}$$

where C_n is the force constant; and both q_s and q_{s+n} are atomic displacements. This differential equation describes a correlated motion of the lin-

ear lattice. The atomic displacement is a continuous function of time, but depends on lattice sites discretely in space. Before solving the discrete differential equation, it is instructive to discuss the more familiar case where the lattice space becomes continuous.

3.2 Elastic Waves in Continuous Media

If a wave is introduced into the linear lattice with a very long wavelength, the atomic displacement will only vary by an infinitesimal amount over a great number of lattice points. On the scale of the wavelength, the lattice points are so densely packed that it can be considered as a continuous medium. Therefore, the length of a few lattice points can be considered infinitesimal. We may approximate $\Delta q_s = q_{s+n} - q_s$ by dq_s, and na by dx, for n running up to the site where the pair coupling becomes negligible. Noticing that $C_n = C_{-n}$ from translation symmetry, the governing differential equation, Eq. (3.1), can be rewritten as

$$\begin{aligned} m\frac{d^2q_s}{dt^2} = {} & C_1[(q_{s+1} - q_s) - (q_s - q_{s-1})] \cdot \left(\frac{a}{a}\right)^2 + \\ & C_2[(q_{s+2} - q_s) - (q_s - q_{s-2})] \cdot \left(\frac{2a}{2a}\right)^2 + \\ & C_3[(q_{s+3} - q_s) - (q_s - q_{s-3})] \cdot \left(\frac{3a}{3a}\right)^2 + \cdots\cdots . \end{aligned} \tag{3.2}$$

Applying the approximation $q_{s+n} - q_s = dq_s$ and $na = dx$, we have

$$\frac{d^2q_s}{dt^2} = \frac{d^2q_s}{dx^2}\left[\frac{C_1}{m}a^2 + \frac{C_2}{m}(2a)^2 + \frac{C_3}{m}(3a)^2 + \cdots\right], \quad \text{and}$$

$$\frac{d^2q}{dt^2} = v^2\frac{d^2q}{dx^2}. \tag{3.3}$$

The subscript is omitted in the final step of the derivation of Eq. (3.3). Since the lattice looks as if it were continuous on the scale of long wavelengths, we can safely ignore the lattice indexing subscript and consider the atomic displacement as a continuous function of x and t. It turns out that Eq. (3.3) is the familiar wave equation in continuous media, such as the wave in a rope or in the violin string. The wave velocity is given by

$$v = \left[\frac{C_1}{m}a^2 + \frac{C_2}{m}(2a)^2 + \frac{C_3}{m}(3a)^2 + \cdots\right]^{\frac{1}{2}}. \tag{3.4}$$

The solution of the wave equation is a traveling plane wave of the form

$$q = A \cdot e^{-i\omega\left(t-\frac{x}{v}\right)}, \tag{3.5}$$

and we can plug Eq. (3.5) into Eq. (3.3) to verify it. The pre-factor A in Eq. (3.5) is obviously the amplitude. Knowing that the wave must be periodic over the time of one period T, we obtain from Eq. (3.5) that ωT must equal 2π , which defines ω as the angular frequency related to the frequency ν by $\omega = 2\pi/T = 2\pi\nu$. The wave is also periodic over the length of one wavelength λ, and we have

$$\frac{\omega\lambda}{v} = 2\pi, \quad \lambda = \frac{v}{\nu} = vT, \quad \text{and } \frac{\omega}{v} = \frac{2\pi}{\lambda} = k. \tag{3.6}$$

The parameter k is the wave vector that scales linearly with the angular frequency ω, and the proportion constant is the wave velocity v. This relationship shows that the long-wavelength wave propagating in continuous media does not suffer from dispersion. The wave function described in Eq. (3.5) is usually expressed explicitly in terms of parameters ω and k, i.e.,

$$q = A \cdot e^{-i\omega t+ikx}. \tag{3.7}$$

By requiring $\lambda \gg a$, the equation of motion of the discrete lattice could be transformed into the wave equation as if the lattice space were continuous. The solution becomes nothing but a continuous elastic wave. This exercise leads us to believe intuitively that Eq. (3.1) will probably have a wave solution in general, including the case of shorter wavelengths. We will pursue this idea in the following.

3.3 Waves of Lattice Vibration and the Dispersion Relation $\omega(k)$

Assuming the wave function in Eq. (3.7) is the correct solution to Eq. (3.1) of the discrete lattice, we need first to reinstall the indexing subscript s, and substitute $s \cdot a$ for x to count only for the lattice points. This is the right thing to do because Eq. (3.1) only governs the vibration of lattice points, and does not care what happens about the empty space in between lattice points. Eq. (3.7) reduces to

$$q_s = A \cdot e^{-i(\omega t-ksa)}. \tag{3.8}$$

Plugging Eq. (3.8) into Eq. (3.1), we obtain

$$-m\omega^2 \cdot A \cdot e^{iska} = \sum_n C_n \cdot \left(e^{i(s+n)ka} - e^{iska}\right) \cdot A, \quad \text{and}$$

$$\begin{aligned}
\omega^2 m &= -\sum_n C_n \left(e^{inka} - 1\right) \\
&= -\sum_{n>0} \left[C_n \left(e^{inka} - 1\right) + C_{-n} \left(e^{-inka} - 1\right)\right] \\
&= -\sum_{n>0} C_n \left(e^{inka} + e^{-inka} - 2\right) \\
&= -\sum_{n>0} 2C_n(\cos(nka) - 1). \qquad (3.9)
\end{aligned}$$

Here in Eq. (3.9), we have used $C_{-n} = C_n$ from translation symmetry. We have

$$\omega^2 = \frac{2}{m} \sum_{n>0} C_n(1 - \cos(nka)). \qquad (3.10)$$

The function $\omega(k)$ such as Eq. (3.10) is the dispersion relation, which is no longer proportional to k linearly. It is the necessary and sufficient condition for the equation of motion, described by Eq. (3.1), to have the traveling plane wave solution of the form of Eq. (3.8). The frequency has a cut-off maximum and can not increase indefinitely as wavelength decreases. However, for an allowed frequency, there are indefinite wave solutions. Their k values are separated by multiples of $2\pi/a$, and all satisfy the dispersion relation. The periodical dispersion is the property of lattice waves in the discrete lattice, and elastic waves do not suffer dispersion in continuous media.

3.4 Force Constants

Multiplying the dispersion relation in Eq. (3.10) by $\cos(\gamma ka)$ and integrating over all k values, we obtain

$$\begin{aligned}
m \int_{-\pi/a}^{\pi/a} \omega^2 \cos(\gamma ka) \cdot dk &= 2 \sum_{n>0} C_n \int_{-\pi/a}^{\pi/a} (1 - \cos(nka)) \cos(\gamma ka) \cdot dk \\
&= -\frac{2\pi C_\gamma}{a}. \qquad (3.11)
\end{aligned}$$

In order to see this result, we substrate θ for ka in the integral of the right hand side of Eq. (3.11), and proceed with the integration. We have

$$\begin{aligned}
&\frac{2}{a}\sum_{n>0} C_n \int_{-\pi}^{\pi} (1-\cos n\theta)\cos\gamma\theta \cdot d\theta \\
&= \frac{2}{a}\sum_{n>0} C_n \left[\frac{1}{\gamma}\sin\gamma\theta - \frac{\sin(n-\gamma)\theta}{2(n-\gamma)} - \frac{\sin(n+\gamma)\theta}{2(n+\gamma)}\right]_{-\pi}^{\pi} \\
&= \begin{cases} 0 & \text{for } n \neq \gamma \\ -\frac{2}{a}\cdot C_n \left[\lim\limits_{n\to\gamma} \dfrac{\sin(n-\gamma)\theta}{2(n-\gamma)}\right]_{-\pi}^{\pi} & \text{for } n \to \gamma \end{cases} \\
&= -\frac{2}{a}\cdot C_\gamma \left[\frac{\theta\cos(n-\gamma)\theta}{2}\right]_{-\pi}^{\pi} \\
&= -\frac{2}{a}\cdot C_\gamma \left[\frac{\pi}{2}+\frac{\pi}{2}\right] = \frac{-2\pi C_\gamma}{a}. \qquad (3.12)
\end{aligned}$$

The relation,

$$C_\gamma = -\frac{ma}{2\pi}\int_{-\pi/a}^{\pi/a} \omega^2 \cos(\gamma ka)\cdot dk, \qquad (3.13)$$

allows the calculation of the force constant between any two atoms in the linear lattice from an experimentally determined dispersion relation $\omega(k)$.

3.5 Long Wavelength Approximation

The wave solution represented by Eq. (3.8) is very general without any prior assumption made on wavelengths. The dispersion relation, that is as general as the original equation of motion, plays the role to set the range of wavelength and frequency. It would be interesting to take the long wavelength approximation to see whether the result agrees with the elastic wave in the continuous medium.

When the wavelength is much longer than the lattice spacing, i.e., $\lambda \gg na$, we have $nka \ll 1$, and $\cos(nka) \approx 1 - (1/2)(nka)^2$. The dispersion relation, Eq. (3.10), reduces to

$$\omega^2 = \frac{k^2a^2}{m}\sum_{n>0} n^2 C_n, \quad \text{and}$$

$$\omega = k\sqrt{\frac{a^2}{m}\sum_{n>0} n^2 C_n}. \tag{3.14}$$

The frequency becomes linearly proportional to the wave vector and the proportion constant is the wave velocity defined in Eq. (3.4), an identical result obtained previously for the elastic wave in continuous media. Indeed, the discrete lattice behaves as if it were continuous in the long wavelength limit. Since sound waves can propagate in crystalline solids and their wavelengths are much longer than the lattice spacing, the wave velocities described in Eqs. (3.4) and (3.14) are identifiable with the sound velocity. We can find an easy explanation here for the commonly observed fact that the sound travels faster in stiffer materials, because stiffer materials have larger force constants. We also notice that the sound should go faster in materials with less mass density.

3.6 Nearest Neighbor Approximation

The force coupling strength naturally decreases when two atoms are located further apart. The significant pair coupling occurs usually just within a few near neighbors, and the nearest neighbor coupling is no doubt the strongest. For most practical purposes, the nearest neighbor coupling alone would provide a good approximation for estimations. Taking this approximation, we have $n = 1$ and the dispersion relation reduces to

$$\omega^2 = \frac{2C_1}{m}(1 - \cos ka) = \frac{4C_1}{m}\sin^2\left(\frac{1}{2}ka\right), \quad \text{and}$$

$$\omega = \sqrt{\frac{4C_1}{m}}\left|\sin\frac{1}{2}ka\right|. \tag{3.15}$$

The normalized frequency $\omega/(4C_1/m)^{1/2}$ is plotted against k in Fig. 3.2. The period of the sinusoidal $\omega(k)$ is $2\pi/a$. There are multiple solutions of $k + 2n\pi/a$ for a given ω . The range of k between $-\pi/a$ and π/a is the first Brillouin zone. Beyond this range the function $\omega(k)$ will repeat periodically into the extended zone region. In the region where $|k| \ll \pi/a$ or $\lambda \gg a$, the continuum or long wavelength approximation applies, the dispersion relation of Eq. (3.15) further reduces to

$$\omega = \sqrt{\frac{4C_1}{m}}\frac{1}{2}ka, \quad \omega = \sqrt{\frac{C_1 a^2}{m}}k, \quad \text{or} \quad \omega = vk. \tag{3.16}$$

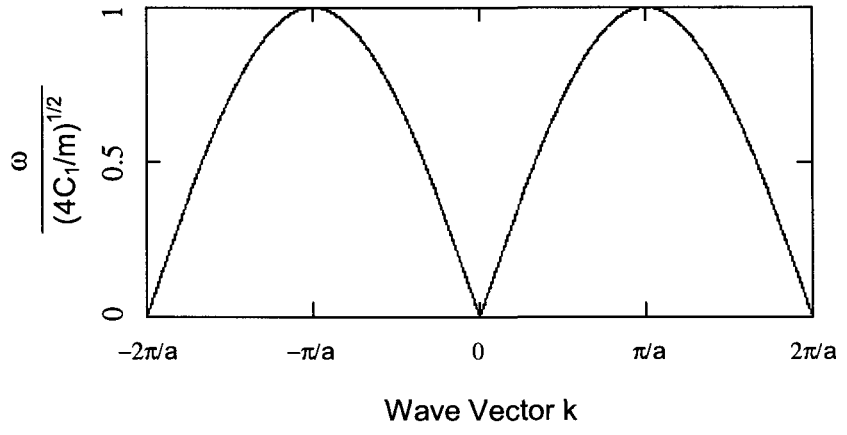


Fig. 3.2 Normalized dispersion relation of Eq. (3.15) is plotted as a function of wave vector k in the nearest neighbor approximation. The slope is the group velocity, which is a constant in the long wavelength region and approaches zero near the Brillouin zone boundary.

As seen before for the more general case, the frequency ω becomes proportional to k, and the proportion constant is the speed of sound that equals $(C_1 \cdot a^2/m)^{1/2}$. When the wave vector increases to approach $\pm\pi/a$, the wavelength is getting comparable to the lattice spacing, and the long wavelength approximation breaks down. The dispersion relation of Eq. (3.15) becomes increasingly non-linear with decreasing slope $d\omega/dk$. Eventually, we have $d\omega/dk = 0$ when $k = \pm\pi/a$, see Fig. 3.2.

3.7 Phase Velocity and Group Velocity

The wave function such as the one represented by Eq. (3.7) is a moving monochromatic plane wave. The wave peak, for example, appears to move toward one direction at a constant speed. Following the peak and record its location x_1 at time t_1 and x_2 at t_2, we can obtain the wave velocity, i.e., $(x_2 - x_1)/(t_2 - t_1)$. For the wave to move with this velocity, the wave phase at (x_1, t_1) must equal the phase at (x_2, t_2). The wave velocity turns out to be the phase velocity that measures how fast the same phase on the wave travels. The equal phase requirement, $(\omega t_1 - kx_1) = (\omega t_2 - kx_2)$, leads to $(x_2 - x_1)/(t_2 - t_1) = \omega/k$, and we have the phase velocity $v_{ph} = \omega/k$.

However, the phase velocity is not related to energy flow. A good proof can be found by examining the monochromatic plane lattice wave of Eq. (3.8). Here, atoms in the lattice wave oscillate around their lattice point in the same simple harmonic motion described by $Ae^{-i\omega t}$. Since their phases are different according to e^{isak}, the kinetic energy and potential energy will vary along the lattice chain. However, the total energy remains constant for every atom at all times. There is no net energy transfer from atom to atom, although the wave travels at the phase velocity. In general, the total energy associated with the monochromatic plane wave is uniform over space and time, and therefore, it does not transport the energy.

In contrary to the monochromatic plane wave that spreads energy uniformly over space, the wave packet confines energy to a limited spatial region. We can follow the movement of the wave packet in the course of time, and learn about the flow of energy. Therefore, it is important to learn about how the wave packet is actually formed and what is its traveling velocity. Since the wave packet is limited in its spatial extension, we may guess that it could result from some sort of interference effects. For waves of different wavelengths running simultaneously in space, it is plausible to anticipate that they interfere constructively inside the wave-packet region and destructively anywhere else, although certain conditions may have to be met. To examine whether this intuitive thought has any merit, we begin with a group of monochromatic plane waves. Suppose a large sample of N measurements on the wave vector is collected, and the number of counts $n(k)$ at each k value is sorted out. It will not be surprising to find, as frequently found in experiments, that the $n(k)$-versus-k plot converges into the Gauss distribution. Dividing $n(k)$ by N will give us the probability for finding each k, i.e., $A(k) = n(k)/N$ as a function of k. We have

$$A(k) = A_0 \cdot e^{-\frac{\sigma^2 \cdot (k-k_0)^2}{2}}. \tag{3.17}$$

Integrating $A(k)$ over the entire k-space gives the total probability for finding all the wave vectors, and naturally it should be unity. This normalization condition determines the height of the distribution, and we have $A_0 = \sigma/(2\pi)^{1/2}$. The center of the distribution is the mean value of k located at k_0. The width is measured in terms of $(1/\sigma)^2$, the variant of the distribution. The parameter $1/\sigma$ is the root-mean-square (rms) value of $k - k_0$, also called the standard deviation. It makes statistical sense to define the uncertainty in measuring k by its rms deviation from the mean,

i.e.,

$$\Delta k = (< (k - k_0)^2 >)^{1/2} = \frac{1}{\sigma}. \tag{3.18}$$

Since a bigger probability is expected when detecting the wave of larger amplitude, the waves in our experiment can therefore be described categorically by the wave function

$$\varphi(k) = \sqrt{A(k)} \cdot e^{-i(\omega(k)\cdot t + k\cdot x)}. \tag{3.19}$$

The square of the amplitude, obtainable from the product $\varphi\varphi^*$, will be the probability distribution $A(k)$ in agreement with our experimental measurements. To add up all the waves of different wavelengths in the group, we will have the net wave function $\phi(x,t)$ in the real space,

$$\phi(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{A_0} \cdot e^{\frac{-\sigma^2\cdot(k-k_0)^2}{4}} \cdot e^{-i(\omega(k)\cdot t - k\cdot x)} dk. \tag{3.20}$$

The pre-factor of $1/(2\pi)^{1/2}$ is for proper normalization of $\phi(x,t)$. Equation (3.20) turns out to be the Fourier transform of $A(k)^{1/2}$. The function $A(k)^{1/2}$ diminishes rapidly for wave vectors going away from k_0, and becomes insignificant outside a range of $k - k_0$. Therefore, numerical integration of Eq. (3.20) can be carried out rather easily for $t = 0$ when the $\omega(k) \cdot t$ term drops out, see Fig. 3.3. When t is not zero, the knowledge of the dispersion relation $\omega(k)$ is needed for the integration. If $A(k)^{1/2}$ distribution is narrowly peaked, the wavelength spread will be small and $k - k_0$ is a small number. We may approximate $\omega(k)$ with the first two terms in its Taylor series expansion. We have $\omega(k) \approx \omega_0 + (d\omega/dk)_0 \cdot (k - k_0)$, where the subscript 0 means that the function value is evaluated at k_0. This approximation converts Eq. (3.20) into the following form,

$$\phi(x,t) = e^{-i(\omega_0\cdot t - k_0\cdot x)} \cdot \sqrt{\frac{A_0}{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-\sigma^2\cdot(k-k_0)^2}{4}} \cdot e^{i(k-k_0)\cdot(x-\left(\frac{d\omega}{dk}\right)_0\cdot t)} dk. \tag{3.21}$$

The integral in Eq. (3.21) will become more transparent for calculation, if we substitute p for $(k - k_0)$ and q for $(x - (d\omega/dk)_0 \cdot t)$,

$$\int_{-\infty}^{\infty} e^{\frac{-\sigma^2\cdot p^2}{4}} \cdot e^{ip\cdot q} dp$$
$$= \int_{-\infty}^{\infty} e^{\frac{-\sigma^2}{4}\cdot(p - \frac{2iq}{\sigma^2})^2} \cdot e^{\frac{-q^2}{\sigma^2}} dp$$

$$= \frac{2\sqrt{\pi}}{\sigma} \cdot e^{\frac{-q^2}{\sigma^2}}. \tag{3.22}$$

Equation (3.22) begins with the Fourier transform of a Gauss function in the k-space, and interestingly, also ends up with a Gauss function in the real space. Their variance is the inverse of each other, $2/\sigma^2$ in the k-space and $\sigma^2/2$ in the real space. The narrower the k distribution is, and the wider the x distribution will be, or vise versa. Substituting $(x - (d\omega/dk)_0 \cdot t)$ back for q, we have

$$\phi(x,t) = \frac{\sqrt{2A_0}}{\sigma} \cdot e^{\frac{-(x-\left(\frac{d\omega}{dk}\right)_0 \cdot t)^2}{\sigma^2}} \cdot e^{-i(\omega_0 \cdot t - k_0 \cdot x)}. \tag{3.23}$$

This is a wave packet in the real space, consisting of the k_0 plane wave with its amplitude bounded within the Gaussian envelope. To find the position x of any physical entity associated with the wave packet, or just to find the wave packet itself somewhere in space, the probability must be proportional to the square of the wave-packet amplitude, obtainable from $B(x) = \phi\phi^*$. The integration of $\phi\phi^*$ over the whole space is the total probability that equals unity logically, leading to $A_0 = \sigma/(2\pi)^{1/2}$, the same value obtained in the k-space earlier. For comparing with the probability distribution in the k-space described by Eq. (3.17), the probability distribution in the real space is given here. We have

$$B(x) = \frac{2A_0}{\sigma^2} \cdot e^{\frac{-2(x-x_0)^2}{\sigma^2}}. \tag{3.24}$$

The center of the distribution is the mean value of x located at $x_0 = (d\omega/dk)_0 \cdot t$. The width is measured in terms of $(\sigma/2)^2$, the variant of the distribution. The parameter $\sigma/2$ is the root-mean-square (rms) value of $x - x_0$. Defining the uncertainty in measuring x as the rms deviation from the mean, we have $\Delta x = (< (x - x_0)^2 >)^{1/2}$, or simply $\Delta x = \sigma/2$. Combining with $\Delta k = 1/\sigma$, we have $\Delta x \cdot \Delta k = 1/2$. The wider the wave packet is in the real space, the narrower it will be in the k-space, or vise versa. The compounded spread of the wave packet is 1/2. The rather detailed analysis of the wave packet should help the understanding of the uncertainty principle. It states that the compounded accuracy in measuring both x and k can not be smaller than 1/2.

The amplitude envelope of the wave packet is proportional to the Gauss function $F(x,t)$,

$$F(x,t) = e^{\frac{-(x-\left(\frac{d\omega}{dk}\right)_0 \cdot t)^2}{\sigma^2}}. \tag{3.25}$$

The peak is at $x = (d\omega/dk)_0 \cdot t$, that moves obviously from $x = 0$ to $x = (d\omega/dk)_0 \cdot t$ in t seconds. The velocity of the envelope, that is of the wave packet, is therefore equal to $(d\omega/dk)_0$. Since the wave packet is the interference manifestation of a group of waves, the wave-packet velocity is also the group velocity, $v_g = (d\omega/dk)_0$. Physical insights can be gained, if we examine the phase of any two contributing waves in the wave packet. In order to interfere constructively, their phase must equal to each other at the point (x_i, t_i) where the wave packet peaks, i.e., $\omega_1 t_i - k_1 x_i = \omega_2 t_i - k_2 x_i$. When the peak moves to (x_f, t_f), these two waves must be in phase again at this point, i.e., $\omega_1 t_f - k_1 x_f = \omega_2 t_f - k_2 x_f$. So, we have $(\omega_2 - \omega_1) \cdot t_i = (k_2 - k_1) \cdot x_i$, and $(\omega_2 - \omega_1) \cdot t_f = (k_2 - k_1) \cdot x_f$. Subtracting one from the other and rearranging the factors, we obtain

$$\frac{x_f - x_i}{t_f - t_i} = \frac{\omega_2 - \omega_1}{k_2 - k_1}. \tag{3.26}$$

The left hand side is by definition the group velocity, and the right hand side approaches $(d\omega/dk)_0$ in the limit of small range of k. Finally, an illustration of the wave packet is shown in Fig. 3.3. It can be obtained from the numerical integration of Eq. (3.20). Because of the small range of wave vectors, the plot is literally identical to what described by Eq. (3.23). The oscillatory wave inside the wave-packet envelope follows the phase factor $\exp(-i(\omega_0 t + k_0 x))$ very closely. The phase velocity ω_0/k_0 can be faster or slower than the group velocity $(d\omega/dk)_0$, depending on whether the dispersion curve $\omega(k)$ is concave downward or upward. In the long wave limit where there is no dispersion, the phase velocity equals the group velocity.

Given the dispersion relation in the linear lattice described by Eq. (3.10), the group velocity is

$$\begin{aligned} v_g = \frac{d\omega}{dk} &= \frac{2}{m}\frac{1}{2\omega}\sum_{n>0} C_n na \cdot \sin(nka) \\ &= \frac{\sum_{n>0} C_n na \cdot \sin(nka)}{\sqrt{2m\sum_{n>0} C_n \cdot (\sin(\frac{nka}{2}))^2}}. \end{aligned} \tag{3.27}$$

The group velocity v_g is 0 at the zone boundary because $\sin(nka) = 0$ and $\sin(nka/2) = 1$ when k equals the multiple of $\pm\pi/a$. Since we still have $\sin(nka) \approx 0$ and $\sin(nka/2) \approx 1$ near the zone boundary, v_g must remain to be small around this region. When we move away from the zone boundary and into the long wavelength realm, where the approximation

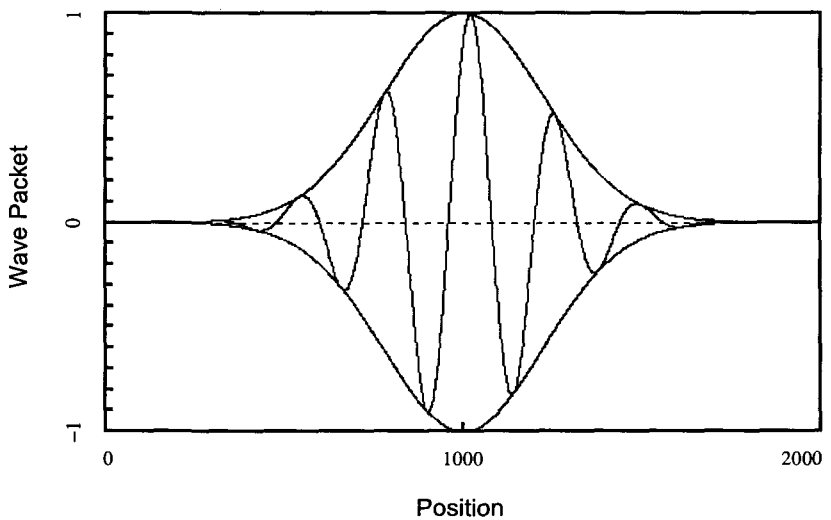


Fig. 3.3 The wave packet is obtained from a sinusoidal wave when we multiply it with a pulse of limited spatial extension. The wave-packet is described by the function $F(x,t_1)\cdot sin(\omega_0 t_1 - k_0 x)$, where the sinusoidal function represents the plane traveling wave of well defined ω_0 and k_0. The pulse function $F(x,t_1)$ peaks at location x_1 when time is at t_1, and it forces the sinusoidal wave into an envelope. The plot is generated for the fixed time t_1, and x_1 happens to be 1000. The position on the x-axis may be considered as the lattice site index. This wave packet extends over roughly 1500 lattice sites.

$\sin(nka) \approx nka$ becomes valid, the group velocity in Eq. (3.27) approaches the phase velocity of Eq. (3.14). For practical purposes of getting estimations, the group velocity in the nearest neighbor approximation is usually sufficient, and is listed here for references.

$$v_g = \sqrt{\frac{C_1 a^2}{m}} \cos\frac{1}{2}ka,$$

$$v_g = 0 \quad \text{for } k = \pm\frac{\pi}{a},$$

$$v_g = \sqrt{\frac{C_1 a^2}{m}} \quad \text{for } 0 \approx ka \ll \pi. \tag{3.28}$$

3.8 First Brillouin Zone

In the wave representation of lattice vibrations, the wave function evaluated at the lattice point is the atomic displacement. The wave function value in between lattice points relates to no physical quantities and is therefore irrelevant. The irrelevant part of the wave function needs not to be defined uniquely. As a consequence, the solution of wave vector for a given frequency is multi-valued in the dispersion relation of Eq. (3.10). Multiple waves of different wavelengths can be constructed to describe the same lattice vibration. These waves, when viewed as continuous waves, are obviously different because their different wavelengths, but they are actually identical when their function values are compared at the lattice sites. In the lattice wave, each atom oscillates according to $\exp(-i\omega t)$, and is related to its neighbor by a phase-factor shift of $\exp(ika)$. The mathematical description of these correlated oscillations is a discrete wave function proportional to $\exp(-i\omega t) \cdot \exp(iksa)$, where $s = 0, \pm 1, \pm 2, \cdots$ for numbering the lattice sites. This is a subset of the continuous wave function that is proportional to $\exp(-i\omega t) \cdot \exp(ikx)$. To examine whether two lattice waves are equivalent for the same frequency, we only need to compare the respective phase factors at each lattice point and ignore what happens in between lattice points. Two lattice waves, characterized by (ω, k) and (ω, k'), are considered to be equivalent if and only if their phase factors are equal at every lattice point. That is

$$e^{iksa} = e^{ik'sa} \quad \text{for } s \text{ that runs through all lattice points.} \tag{3.29}$$

For wave vectors inside the first Brillouin zone,

$$-\frac{\pi}{a} \leq k \leq \frac{\pi}{a}, \tag{3.30}$$

Equation (3.29) can not be satisfied, except for the pair of $k = \pm\pi/a$ at the zone boundary. Therefore, all lattice waves inside the first Brillouin zone are independent modes, and the pair $k = \pm\pi/a$ are identical modes. For any wave vector k' outside the first Brillouin zone, we can always subtract a reciprocal lattice vector $G(= 2n\pi/a)$ from it, and obtain the corresponding k vector to be located inside the first Brillouin zone. This procedure looks like a smart shift of coordinate, but there is in fact a physical reason for it. Namely, the k' wave gets scattered from the lattice and will remerge constructively into the k wave provided $k' - k = G$. This is just the von Laue diffraction law. The relationship $k' = k + 2n\pi/a$ also ensures that the

phase factor of the k' wave will reduce to that of the k wave at all lattice points, i.e.,

$$e^{ik'sa} = e^{iksa} \cdot e^{i\frac{2ns\pi}{a}\cdot a}, \quad \text{where n and s are integers.} \tag{3.31}$$

The result shows that the k' wave is not independent from its corresponding k wave. In fact, for a wave vector k inside the first Brillouin zone, there are unlimited equivalent waves that are related by $k' = k + 2n\pi/a$. All these wave vectors satisfy Eq. (3.29) as shown by Eq. (3.31), and they are solutions to the dispersion relation of Eq. (3.10) for the same ω. Since the points $(\omega, k+2n\pi/a)$ periodically appear on the dispersion curve in the extended zone scheme, see Fig. 3.2, their group velocities $d\omega/dk$ are obviously equal to one another. The modes $(\omega, k + 2n\pi/a)$ are therefore equivalent to the mode (ω, k). It proves that any wave vector outside the first Brillouin zone can always be reduced to an equivalent one located inside the zone without loosing the power to describe the same lattice vibration. The multiple wavelengths of the equivalent waves do not have any independent physical meaning. One may argue that these waves do have different phase velocity, i.e., $\omega/(k + 2n\pi/a)$. It turns out that these phase velocities do not affect the phase relationship among lattice points at any time, neither they approach the sound velocity in the long wavelength limit. Therefore, the phase velocity of waves outside the first Brillouin zone is physically meaningless and immaterial in dealing with lattice waves. To label all the independent modes of lattice vibrations, it is sufficient, and arguably necessary, to resort to the k vectors in the first Brillouin zone. There is no reason to take k vectors beyond the first Brillouin zone for describing the lattice vibration, because it adds no new physics but confusion. Fig. 3.4 shows an example to illustrate this point further.

The shorter wave with the wave vector outside the first Brillouin zone, represented by curve (b) in Fig. 3.4, appears to oscillate more cycles through empty space with slower phase velocity. Curve (a) with its wave vector located inside the first Brillouin zone oscillates less with faster phase velocity. If we erase the traces of both curves, nothing changes about the atomic displacements and their relative phases. Therefore, the different phase velocities and wavelengths related by $k' = k + 2n\pi/a$ are immaterial. The confusion, if any, comes from the illusive continuous waves drawn through both the lattice and irrelevant empty space. To avoid confusion, the wavelength of lattice waves is always understood to be located inside the range of $-\pi/a \leq k \leq \pi/a$, that is the first Brillouin zone. Any shorter wavelength

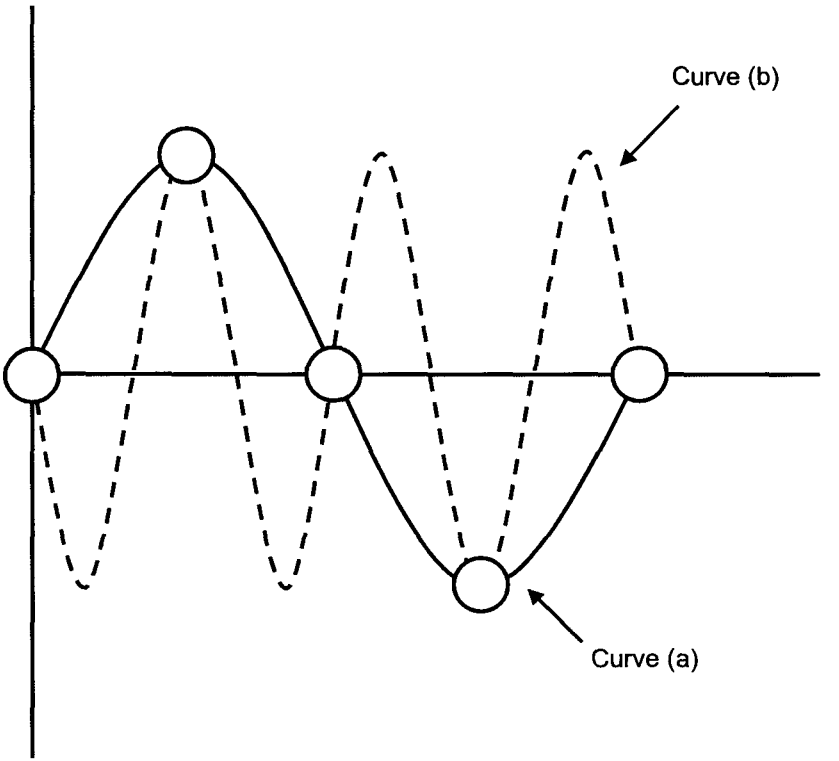


Fig. 3.4 Curve (a) is the lattice wave with its wavelength equal to 4 times of lattice spacing, i.e., $\lambda = 4a$ or $k = \pm\pi/2a$. The wave vector is inside the first Brillouin zone. Curve (b) is the lattice wave describes exactly the same atomic displacement, but its wavelength is shorter and its wave vector is outside the first Brillouin zone. There are unlimited waves, with the wave vector residing outside the first Brillouin zone, describe the same lattice displacement. They are not physically independent from curve (a).

outside the first Brillouin zone should be reduced to the corresponding one inside the zone. Wave (b) in Fig. 3.4 should be reduced to that of wave (a) by either adding or subtracting a reciprocal lattice vector, i.e.,

$$\begin{aligned}
&\text{curve(a):} \quad \lambda = 4a \quad \longrightarrow \quad k = \pm\frac{\pi}{2a};\\
&\text{curve(b):} \quad \lambda = \frac{4}{3}a \quad \longrightarrow\\
&\qquad\qquad k = \frac{6\pi}{4a} \quad \longrightarrow \quad k = \frac{6\pi}{4a} - \frac{2\pi}{a} = -\frac{\pi}{2a},
\end{aligned}$$

or

$$k = -\frac{6\pi}{4a} \quad \longrightarrow \quad k = -\frac{6\pi}{4a} + \frac{2\pi}{a} = \frac{\pi}{2a}. \tag{3.32}$$

When we have $k = \pi/2a$, curve (a) represents a positively going wave because the group velocity $d\omega/dk$ is positive. The same curve can represent a negatively going wave when $k = -\pi/2a$, where $d\omega/dk$ is negative. Both are independent mode in the first Brillouin zone. The wave vector $k = -6\pi/4a$ of curve (b) is not an independent wave and should be reduced back into the first Brillouin zone. It is considered to be a positively going wave because it is physically equivalent to $k = \pi/2a$ including the same group velocity. If we need to describe the atomic displacements in Fig. 3.4 with a negatively going wave, we should take $k = -\pi/2a$, not $k = -6\pi/4a$. The group velocity $d\omega/dk$ is not pointing along the k vector in general, but they are for the one dimension lattice in the first Brillouin zone.

It is interesting to note that the wave vector at the first Brillouin zone boundary satisfies von Laue's diffraction condition to the first order, i.e., $k = \pi/a = \frac{1}{2}G_s$. Here, we have $\lambda = 2a$ and the incident angle $\theta = 90°$. The Bragg condition, $2a \cdot \sin(\theta) = n\lambda$, becomes $2a = 2na$ and is satisfied for $n = 1$. For higher-order n values, the condition will be satisfied with k vectors located on the higher Brillouin zone boundaries. These higher-order Bragg scattering of lattice waves are not independent from the first-order one, because these lattice wave vectors should be always reduced to the equivalent one on the first Brillouin zone boundary. However, electromagnetic waves such as x-ray and particle waves such as electrons do have physical meaning throughout the solid including the empty space between atoms. They can certainly have very short wavelengths, so that the wave vectors will extend way beyond the first Brillouin zone. These wave vectors can not be reduced to locate inside the first Brillouin zone. Therefore, higher order diffraction peaks do exist for x-ray and electrons.

The lattice wave of $k = \pi/a$ at the zone boundary will be reflected into the wave of $k = -\pi/a$, because the Bragg condition is satisfied. Assuming the reflection is elastic with no losses, the reflected wave will maintain the same amplitude of the incoming wave. Two waves of equal strength traveling opposite to each other with the same phase speed will naturally form a standing wave. The same argument applies to the lattice wave of $k = -\pi/a$ at the opposite boundary of the first Brillouin zone. Although the incoming wave becomes $k = -\pi/a$ and the reflected wave is $k = \pi/a$ in this case, the standing wave formed is not distinguishable from the previous one. Therefore, lattice waves of $k = \pm\pi/a$ are not independent waves, but the same standing wave. The wave function is simply the superposition

of $A_0 \cdot \exp(-i\omega t + ix\pi/a)$ and $A_0 \cdot \exp(-i\omega t - ix\pi/a)$, which reduces to $2A_0 \cdot \exp(-i\omega t) \cdot \cos(x\pi/a)$. When evaluated at lattice sites with $x = sa$, the atomic displacements are found to be $(-1)^s \cdot 2A_0 \cdot \exp(-i\omega t)$. It is interesting to observe that two adjacent lattice sites are always out of phase, but this fact alone is not sufficient to claim that the lattice wave of either $k = \pi/a$ or $k = -\pi/a$ is a standing wave. The real underlying reason is the Bragg reflection from $k = \pi/a$ to $k = -\pi/a$, or vice versa.

The group velocity $d\omega/dk$ is zero at $k = \pm\pi/a$. It is rather straightforward to see mathematically from Eq. (3.27), but it is not quite so obvious if one asks for a physical explanation. The concepts of wave packet, lattice diffraction, and first Brillouin zone all play implicitly to manifest the show of zero group velocity at the Brillouin zone boundary. We begin with a wave packet formed around $k = \pi/a$, containing n plane waves symmetrically distributed in the small range of $(\pi/a - \Delta k) \leq k \leq (\pi/a + \Delta k)$. Since the upper half of the range is outside the first Brillouin zone, the lattice diffraction will map it into the physically equivalent range of $-(\pi/a - \Delta k) \geq k \geq -\pi/a$ located inside the first Brillouin zone. Now, for each positively going wave k in the range of $(\pi/a - \Delta k) \leq k \leq \pi/a$, we can see a negatively going wave $-k$ in the range of $-(\pi/a - \Delta k) \geq k \geq -\pi/a$, so that they will form a standing wave. Therefore, the original n constituting waves in the wave packet combine into $n/2$ standing waves. Somewhere all standing waves will be in phase and add up to form the wave-packet peak. Since the standing wave does not travel in time, the wave packet can not move either, and the group velocity is zero. In the limit of small Δk as seen in Eq. (3.26), zero group velocity leads to $d\omega/dk = 0$. This is the proof for $d\omega/dk = 0$ from picturing the physics at the zone boundary, much more insightful than just taking the derivative of $\omega(k)$ mathematically.

When the wave packet is formed near the zone boundary inside the first Brillouin zone, the distribution of constituting plane waves will be slightly asymmetrical around $k = \pi/a$ in the range of $(\pi/a - \epsilon - \Delta k) \leq k \leq (\pi/a - \epsilon + \Delta k)$. The distribution peaks at $k = \pi/a - \epsilon$. The portion outside the zone, $\Delta k - \epsilon$, is smaller than the portion $\Delta k + \epsilon$ inside the zone. After reducing the wave vectors back into the first Brillouin zone, it is easier to see that the positively going waves out weigh the negatively going waves both in numbers and in amplitudes. The imbalance gets severer as ϵ increases, and there will be no negatively going waves in the wave packet when $\epsilon = \Delta k$. Accordingly, the group velocity $d\omega/dk$ will increase steadily from zero as we move away from the zone boundary. When we enter the region where the long wavelength limit applies, all constituting waves in the

wave packet will have the same phase velocity. It follows from Eq. (3.26) that $d\omega/dk$ finally approaches $v_{ph} = \omega/k$. Similar argument applies to the negative branch of the dispersion curve.

Phonons

3.9 Difficulties with the Classical Approach

We have learned some basic wave properties of lattice vibrations from the point of view of classical mechanics. Knowing the constituent plane waves, we can presumably calculate the position and momentum of each atom, and therefore its energy. For example, the total energy in a plane wave will be the sum of the kinetic and potential energy of all atoms involved. To sum them up and average over time, we should be able to find the energy of this mode, which is proportional to the square of the amplitude. The total energy of the system is obviously the sum of the energy of all modes. This procedure appears to be workable, but two difficulties arise if we examine it in more detail. The uncertainty principle tells us that the position and momentum of an atom can not be measured precisely in the same time. The classical approach obviously does not take this quantum consideration into account. Secondly, we do not know the plane-wave amplitudes, or more importantly, their statistical distribution at a given temperature. The energy of $3NkT$ borrowed from the statistics of kinetic gas theory is at least partially false because it does not explain the low-temperature specific heat of crystals. The quantum mechanical approach will remove these two difficulties and construct a new picture for us to view lattice vibrations. We normally say that the lattice wave is quantized. Namely, a mode of lattice wave of frequency $\omega_{\mathbf{k}}$ will be mapped into a collection of quanta called phonons, and each phonon has the energy of $\hbar\omega_{\mathbf{k}}$. The number of phonons in the mode will depend on temperature obeying the Bose-Einstein statistics. It is a matter of fact that we need the quantum picture of phonons in all areas of solid state physics. We will now pursue the quantum theory of a simpler one-dimensional monatomic lattice in order to develop the concept of phonons. Except the simpler mathematics, the underlying phonon picture is no narrower than what we can learn from the real crystal in three dimensions.

3.10 Two-body Potential

Suppose we take a segment of N atoms in a linear monatomic lattice of infinite length, what will be the total energy in this segment? Rather obviously, we need to find both the kinetic energy and potential energy of each atom first, and then sum them up over the whole segment. The kinetic energy of the s^{th} atom is simply equal to ${p_s}^2/2m$, where the momentum p_s is implicitly determinable pending upon what form of potential the atom is in. The potential is in general a function of the position of all atoms. In the model of coupled atoms with springs, the imaginary spring between any two atoms is either compressed, or stretched, or relaxed, depending on their relative displacement $q_s - q_{s'}$. Here, q_s is the displacement of s^{th} atom from its equilibrium lattice position. When $q_s - q_{s'}$ equals zero for all pairs, all springs are relaxed and the total potential energy will be at its lowest value. We may set this ground level equal to zero for convenience. The pair potential increases as $(q_s - q_{s'})$ deviates from zero. Since the spring does not care from which end it is compressed or stretched, the pair potential will be a function of $|q_s - q_{s'}|$, and the total potential energy is just the sum of all the pair potentials, i.e.,

$$V(q_1, q_2, \cdots, q_N) = \frac{1}{2}\sum_{s,s'} v(|q_s - q_{s'}|). \tag{3.33}$$

The summation over s and s' are both from 1 to N. Since the pair potential between s and s' atoms is counted twice as $v(|q_s - q_{s'}|)$ and $v(|q_{s'} - q_s|)$, two equal quantities, the summation must be divided by 2 in order to count for the total potential energy correctly. Denoting $u_{ss'} = |q_s - q_{s'}|$ and $v_{ss'} = v(|q_s - q_{s'}|)$, we will expand V in terms of the Taylor series around $u_{ss'} = 0$.

$$\begin{aligned} V &= \frac{1}{2}\sum_{s,s'} v_{ss'} \\ &= V_0 + \sum_{s,s'} \left(\frac{\partial V}{\partial u_{ss'}}\right)_0 \cdot u_{ss'} + \\ &\quad \frac{1}{2}\sum_{s,s'} \left[\sum_{i,j} \left(\frac{\partial}{\partial u_{ij}}\frac{\partial V}{\partial u_{ss'}}\right)_0 \cdot u_{ss'} \cdot u_{ij}\right] + \text{higher order terms.} \end{aligned} \tag{3.34}$$

Since the reference potential is arbitrary, we have already set $V_0 = 0$ for convenience. When the displacements of any two atoms are equal, the force exerted on each other between them is zero, i.e., $F = -\left(\frac{\partial v_{ss'}}{\partial u_{ss'}}\right)_0 = 0$. With this in mind we have

$$\begin{aligned}
&\sum_{s,s'} \left(\frac{\partial V}{\partial u_{ss'}}\right)_0 \cdot u_{ss'} \\
&= \sum_{s,s'} \left(\frac{\partial}{\partial u_{ss'}} \left(\frac{1}{2}\sum_{i,j} v_{ij}\right)\right)_0 \cdot u_{ss'} = \sum_{s,s'} \left(\frac{1}{2}\frac{\partial v_{ss'}}{\partial u_{ss'}}\right)_0 \cdot u_{ss'} = 0.
\end{aligned} \tag{3.35}$$

Neglecting higher order terms, we find that the only non-zero term in the potential is the quadratic term.

$$\begin{aligned}
&\frac{1}{2}\sum_{s,s'} \left[\sum_{i,j} \left(\frac{\partial}{\partial u_{ij}}\frac{\partial V}{\partial u_{ss'}}\right)_0 \cdot u_{ss'} \cdot u_{ij}\right] \\
&= \frac{1}{2}\sum_{s,s'} \left(\frac{1}{2}\frac{\partial^2 v_{ss'}}{\partial u_{ss'}^2}\right)_0 \cdot u_{ss'}^2 \\
&= \frac{1}{2}\sum_{s} \left[\frac{1}{2}\sum_{s'} \left(\frac{\partial^2 v_{ss'}}{\partial u_{ss'}^2}\right)_0 \cdot u_{ss'}^2\right].
\end{aligned} \tag{3.36}$$

The second derivative of $v_{ss'}$ with respect to $u_{ss'}$ is a constant when evaluated at $u_{ss'} = 0$. It turns out that $\frac{\partial^2 v_{ss'}}{\partial u_{ss'}^2}$ is always equal to this constant during the oscillation in the quadratic approximation. The reason becomes clear when we examine the physical meaning of the second derivative of the potential. Since $\frac{\partial^2 v_{ss'}}{\partial u_{ss'}^2}$ measures how much the force changes in a small change of $u_{ss'}$, it gauges the stiffness of the imaginary spring and is the force constant $C_{ss'}$ between these two atoms. The force constant of a spring does not change in small oscillations. The total potential takes the following form

$$V(q_1, q_2, \cdots, q_N) = \frac{1}{2}\sum_{s} \left[\frac{1}{2}\sum_{s'} C_{ss'}(q_s - q_{s'})^2\right]. \tag{3.37}$$

3.11 Hamiltonian of Coupled Harmonic Oscillators

We can now write down the system's total energy, or the Hamiltonian, as

$$H = \frac{1}{2}\sum_s \left(\frac{p_s^2}{m} + \frac{1}{2}\sum_{s'} C_{ss'}(q_s - q_{s'})^2 \right). \tag{3.38}$$

In the approximation of nearest neighbor coupling, the Hamiltonian reduces to

$$\begin{aligned} H &= \frac{1}{2}\sum_s \left(\frac{p_s^2}{m} + \frac{1}{2}\left[C_{s(s+1)}(q_s - q_{s+1})^2 + C_{s(s-1)}(q_s - q_{s-1})^2 \right] \right) \\ &= \frac{1}{2}\sum_s \left(\frac{p_s^2}{m} + C_1(q_s - q_{s+1})^2 \right). \end{aligned} \tag{3.39}$$

If we assume the coupling is isotropic, we have $C_{s(s+1)} = C_{s(s-1)} = C_1$, where C_1 is the spring constant between any two adjacent atoms. By the same token, the Hamiltonian for n-neighbor coupling is given by

$$H = \frac{1}{2}\sum_s \left(\frac{p_s^2}{m} + \sum_{n>0} C_n(q_s - q_{s+n})^2 \right). \tag{3.40}$$

The summation over n extends to the end of diminishing coupling strength. The range is usually much smaller than N. The spring constant C_n is for the coupling between the s^{th} and $(s \pm n)^{th}$ atoms. The Hamiltonian in Eq. (3.40) is obtained in the approximation of small oscillations where only the quadratic term in the potential is retained. It clearly consists of N coupled harmonically oscillating atoms, and each has the individual Hamiltonian equal to

$$H_s = \frac{p_s^2}{2m} + \frac{1}{2}\sum_{n>0} C_n(q_s - q_{s+n})^2. \tag{3.41}$$

As in the case of a simple harmonic oscillator, the total of kinetic and potential energy is constant regardless of the position of the oscillator during the oscillation. This is also true for the s^{th} atom governed by Eq. (3.41), if the lattice wave going through the chain is a plane wave. The condition that H_s is a constant when q_s changes, i.e., $\frac{dH_s}{dq_s} = 0$, will lead to the equation of motion that was already derived in Eq. (3.1) from Newton's second law and Hooke's law.

3.12 Periodic Boundary Condition

For the function $f(q_s)$ to be translation invariant in the linear lattice, we must have $f(q_s) = f(q_{s+j})$ with j running through all lattice sites. This is a very stringent condition that requires the function to be periodic over the distance as short as just one lattice spacing. For example, among all plane lattice waves where $f(q_s) = q_s$, only the uniform mode is fully invariant under translation. For the more relaxed requirement of periodicity over a segment of N lattice sites, the reduced translation symmetry becomes $f(q_s) = f(q_{s+N})$. Applying the periodic condition $q_s = q_{s+N}$ to plane lattice waves, we have

$$e^{ik'sa} = e^{ik'(s+N)a} \tag{3.42}$$

for the arbitrarily chosen wave vector k'. It does not loose generality to let $k' = k + \frac{\pi}{a}$, and the periodic condition becomes

$$e^{i(k+\frac{\pi}{a})Na} = 1. \tag{3.43}$$

It leads to

$$(k + \frac{\pi}{a}) \cdot Na = 2n\pi, \quad \text{or}$$

$$k = (\frac{2n}{N} - 1) \cdot \frac{\pi}{a} \qquad \text{for} \quad n = 0, 1, 2, \cdots, N. \tag{3.44}$$

Since the modes $k = \pm\frac{\pi}{a}$ obtained respectively with $n = N$ and 0 are the same standing wave, there are N independent modes in the first Brillouin zone, i.e., in the range of $-\frac{\pi}{a} \leq k \leq \frac{\pi}{a}$. Other k values outside the zone may satisfy Eq. (3.44), but are not independent. They can always be reduced to become those located inside the zone. The restriction of the wave vector k to N discrete values is an important consequence of the periodic condition over a segment of N lattice sites. Since the positional configuration $(q_1, q_2, \cdots, q_N)$ is resulted from the statistical superposition of these allowed plane lattice waves, instead of using the atomic positions directly, the dynamic state of the segment of N atoms can be equally specified in terms of the constituting plane waves.

3.13 Normal Mode Coordinate — Q_k and P_k in the k Space

The Hamiltonian in Eq. (3.40) is derived using the atomic position and momentum as variables. The advantage goes with our familiarity with the 3-dimensional world, but the disadvantage is in the cross term of the pair potential, which couples all the other atoms to the one under consideration. It posts an immediate challenge for any attempt to solve for the dynamic state of the system because there are N coupled variables involved. Fortunately, the atomic coordinates are not random but correlated. The lattice plane wave provides that correlation. The positional configuration $(q_{1,k}, q_{2,k}, \cdots, q_{N,k})$, corresponding to the plane wave of wave vector k, is fully specified by $q_{s,k} \propto Q_k \cdot e^{iksa}$. Since the position q_s is in general resulted from the statistical superposition of all plane waves allowed by Eq. (3.44), we have

$$q_s = \sqrt{\frac{1}{N}} \sum_k Q_k e^{iksa}. \tag{3.45}$$

The pre-factor $\sqrt{1/N}$ is introduced for getting the proper normalization, and the summation is over N allowed k values in Eq. (3.44). It turns out that Eq. (3.45) is just the Fourier-series expansion of q_s in terms of the plane waves allowed by the periodic boundary condition. In reverse, we can expand Q_k in terms of q_s, also as a Fourier series. We first change the summation index from k to k' in Eq. (3.45), multiply both sides of the equation with $\sqrt{1/N} \cdot e^{-iksa}$, and sum over s. Then, we have

$$\begin{aligned} \sqrt{\frac{1}{N}} \sum_s q_s e^{-iksa} &= \frac{1}{N} \sum_{k',s} Q_{k'} e^{i(k'-k)sa} \\ &= \sum_{k'} Q_{k'} \delta_{k'k} = Q_k. \end{aligned} \tag{3.46}$$

The delta function is either 1 or 0 depending on whether $k' = k$ or not, i.e., $\delta_{k'=k} = 1$ or $\delta_{k' \neq k} = 0$. It is easy to see why the term $k' = k$ reduces to Q_k, because the summation over s in the right hand side of Eq (3.46) simply equals N in this case. For the other $N - 1$ terms where $k' \neq k$, the summation over s needs to be zero. To prove this orthogonal property, we need to substitute k and k' with values defined in Eq. (3.44), and evaluate

the right hand side of Eq. (3.46). We have

$$k = \left(\frac{2m}{N} - 1\right) \cdot \frac{\pi}{a} \quad \text{where } m \text{ is a fixed integer between 1 and } N,$$

$$k' = \left(\frac{2n}{N} - 1\right) \cdot \frac{\pi}{a} \quad \text{where } n = 1, 2, \cdots, N, \text{ and}$$

$$\frac{1}{N} \sum_{k'} \sum_{s} Q_{k'} \cdot e^{i(k'-k)sa} = \frac{1}{N} \sum_{n} Q_n \sum_{s} e^{(i2\pi \frac{(n-m)}{N})s}. \tag{3.47}$$

The summation over s from 1 to N in the right hand side is a geometric series with $r = e^{i2\pi \frac{(n-m)}{N}}$ and the result is

$$\frac{1}{N} \sum_{s} e^{i2\pi \frac{(n-m)}{N} s} = \frac{1}{N} \sum_{s} r^s = \frac{1}{N} r \frac{1 - r^N}{1 - r}$$

$$= \frac{1}{N} e^{i2\pi \frac{(n-m)}{N}} \cdot \frac{[1 - e^{i2\pi(n-m)}]}{[1 - e^{i2\pi \frac{(n-m)}{N}}]}$$

$$= \begin{cases} 0 & \text{for } n \neq m \\ 1 & \text{for } n = m. \end{cases} \tag{3.48}$$

This completes the proof of Eq. (3.46) with the orthogonal statement between two k waves

$$\frac{1}{N} \sum_{s} e^{i(k'-k)sa} = \delta_{kk'}. \tag{3.49}$$

If we take the general matrix element at k^{th} row and s^{th} column in a matrix U to be $\sqrt{1/N} \cdot e^{-iksa}$, the corresponding element at s^{th} row and k^{th} column in U^+ would be $\sqrt{1/N} \cdot e^{iska}$. Note the Hermitian conjugate U^+ is the transpose of the complex conjugate of U. Equation (3.49) turns out to be the matrix element, $(UU^+)_{kk'}$, resulted from the multiplication of U with its Hermitian conjugate U^+. The orthogonality of Eq. (3.49) ensures that all the diagonal elements are equal to 1 and all the off-diagonal elements are 0, i.e., $UU^+ = I$, a unity matrix. The result reveals that U is unitary. Here, we have

$$(UU^+)_{kk'} = \sum_{s} U_{k,s} U^*_{s,k'} = \sum_{s} \left(\sqrt{\frac{1}{N}} e^{-iksa}\right) \left(\sqrt{\frac{1}{N}} e^{isk'a}\right) = \delta_{kk'}. \tag{3.50}$$

There is another geometric meaning to the orthogonal property of Eq. (3.49). Namely, the plane waves of different k vectors are mutually orthogonal. The plane wave k going through all lattice points can be viewed as a vector in the N dimensional space spanned by a set of orthogonal unit vectors $\{\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_N\}$. Let us denote this vector as $|k>$ and the transpose of its complex conjugate as $<k|$. We have

$$|k> = \sqrt{\frac{1}{N}}\left(e^{-ika}\hat{x}_1 + e^{-ik2a}\hat{x}_2 + \cdots + e^{-ikNa}\hat{x}_n\right), \quad \text{and}$$
$$<k| = \sqrt{\frac{1}{N}}\left(e^{ika}\hat{x}_1 + e^{ik2a}\hat{x}_2 + \cdots + e^{ikNa}\hat{x}_n\right). \tag{3.51}$$

In this picture, the matrix element in Eq. (3.50) can be viewed as the inner product of $<k'|k>$, and we have $<k'|k> = \delta_{kk'}$. We come to the conclusion that all N normal-mode plane waves can be viewed as mutually orthogonal vectors and the norm of each vector is 1. Therefore, the plane waves of N normal modes form a complete orthogonal basis to span the space of lattice vibrations of N atoms. With similar procedures, we can prove

$$\begin{aligned}\sqrt{\frac{1}{N}}\sum_k Q_k e^{iksa} &= \frac{1}{N}\sum_{k,s'} q_{s'} e^{ik\cdot(s-s')a} \\ &= \sum_{s'} q_{s'}\delta_{ss'} \\ &= q_s. \end{aligned} \tag{3.52}$$

The corresponding orthogonal statement is

$$\frac{1}{N}\sum_k e^{ik\cdot(s-s')a} = \delta_{ss'}. \tag{3.53}$$

The analysis made for Eq. (3.49) applies just as well to Eq.(3.53) if we switch the role between k and s. For example, the matrix element in Eq. (3.53) becomes $(U^+U)_{ss'}$. Regarding the plane wave e^{iksa}, we need to see it as an "s" wave going through all the k points defined by N normal modes in the k space. All the "s" waves are mutually orthogonal, and the norm of each of them is 1.

By the same mathematics, we have a similar corresponding pair of p_s and P_k related by the Fourier series. The two sets of Fourier series are

listed here for future references. We have

$$Q_k = \sqrt{\frac{1}{N}} \sum_s q_s e^{-iksa}; \qquad P_k = \sqrt{\frac{1}{N}} \sum_s p_s e^{iksa}. \tag{3.54}$$

Or,

$$q_s = \sqrt{\frac{1}{N}} \sum_k Q_k e^{iksa}; \qquad p_s = \sqrt{\frac{1}{N}} \sum_k P_k e^{-iksa}. \tag{3.55}$$

The selection of opposite signs in the exponential argument of the Fourier series for Q_k and P_k in Eq. (3.54), or for q_s and p_s in Eq. (3.55), seems to be arbitrary at the first glance, but it is needed for the commutator $[Q_k, P_k]$ to work properly when all q_s, p_s, Q_k, and P_k are considered as quantum mechanical operators. The normal mode variables Q_k and P_k are obviously associated with the position and momentum, respectively. In quantum mechanics, they are the "position" and "momentum" operators in the normal mode coordinate associated with the k wave. The periodic boundary condition restricts the basic lattice vibration to N independent plane waves, or normal modes, which satisfy the orthogonal property of Eqs. (3.49) and (3.53). These plane waves form a complete basis for the Fourier-series expansion of any lattice vibration. This allows the one-to-one mapping of the N dimensional real space of (q_s, p_s) into the N dimensional k space of (Q_k, P_k). There are N pairs of (q_s, p_s) in the real space and N pairs of (Q_k, P_k) in the k space. The dynamic state is fully specified with the knowledge of either (q_s, p_s) or (Q_k, P_k).

3.14 Quantum Mechanical Operators, Commutation Relations, and the Uncertainty Principle

The variables of p_s, q_s, Q_k, and P_k considered in the framework of classical mechanics correspond to operators in quantum mechanic. The operator represents the tool for making observations, and the wave function contains the dynamic state to be measured. For example, when we operate the momentum operator on the wave function, we are making an observation on the momentum. Similarly, when we operate the Hamiltonian on the wave function, we are measuring the energy. The measurement involves probing into the system in some ways that will have to disturb the system's dynamic state. The intrusive probing can not measure the value of dynamic variables with absolute certainty, instead, only the statistical distribution

is experimentally meaningful. Along the same thinking path, we wonder whether the order of measurements may make a difference. Indeed, it does in the microscopic world. Namely, to measure the position and momentum sequentially is different from measuring them in the reversed order. The commutative law in ordinary algebra does not always apply in the operator algebra. When both q_s and p_s are considered as quantum mechanical operators, their commutation relation is described by

$$[q_s, p_{s'}] = i\hbar\delta_{ss'}. \tag{3.56}$$

The order of measuring the position and momentum of the same atom does make a difference. The difference is $i\hbar$. However, the order does not matter when the position and momentum are measured on two different atoms, respectively. The following example proves it. By definition we have

$$[q_s, p_{s'}] = q_s p_{s'} - p_{s'} q_s = q_s(-i\hbar\frac{\partial}{\partial q_{s'}}) - (-i\hbar\frac{\partial}{\partial q_{s'}})q_s. \tag{3.57}$$

Let us operate the commutator on the wave function.

$$[q_s, p_{s'}]\phi = q_s(-i\hbar\frac{\partial\phi}{\partial q_{s'}}) + i\hbar\delta_{ss'}\phi - q_s(-i\hbar\frac{\partial\phi}{\partial q_{s'}})$$

$$[q_s, p_{s'}]\phi = i\hbar\delta_{ss'}\phi. \tag{3.58}$$

The $\delta_{ss'}$ function equals 1 when $s = s'$, but it equals 0 when $s \neq s'$. Eq. (3.58) shows that the wave function ϕ is the eigenfunction of the commutator and the eigenvalue equals $i\hbar\delta_{ss'}$. Since this is true for any ϕ, the commutator $[q_s, p_{s'}]$ must equal $i\hbar\delta_{ss'}$. The accuracy of measuring both q and p on the same atom simultaneously is limited by the Heisenberg uncertainty principle, i.e., $\Delta q \cdot \Delta p \geq \frac{1}{2}\hbar$. Any pair of operators that satisfies the commutation relation of $[a, b] = i\hbar$ is the canonical pair, and will have the similar uncertainty of $\Delta a \cdot \Delta b \geq \frac{1}{2}\hbar$.

As seen in Eq. (3.54), if we consider Q_k as a reverse Fourier series of q_s, P_k is then the Fourier series of p_s. The asymmetry so selected is for Q_k and P_k of the same normal mode to satisfy the commutation rule, in analogy to q_s and p_s of the same atom. Substituting Eq. (3.54) into the commutator $[Q_k, P_{k'}]$, we have

$$[Q_k, P_{k'}] = \frac{1}{N}\sum_{ss'}[q_s, p_{s'}] \cdot e^{i(k's'-ks)a}$$

$$= \frac{1}{N}\sum_{ss'} i\hbar\delta_{ss'} \cdot e^{i(k's'-ks)a}$$

$$= \frac{i\hbar}{N}\sum_{s} e^{i(k'-k)sa}, \quad \text{for } s = 1 \text{ to } N. \tag{3.59}$$

From the proof discussed for the orthogonality in Eq. (3.49), we already know that $\frac{1}{N}\sum_s e^{i(k'-k)sa} = \delta_{kk'}$. The canonical commutation relation between operators Q_k and $P_{k'}$ is therefore preserved, and we have

$$[Q_k, P_{k'}] = i\hbar\delta_{kk'}. \tag{3.60}$$

In analogy to q_s and $p_{s'}$, the normal-mode operators Q_k and $P_{k'}$ do not commute for the same mode, but they commute with each other for different modes.

3.15 Diagonalized Hamiltonian in the Normal Mode Coordinate

The Hamiltonian derived in Eq. (3.38) is a function of q_s and p_s. The cross-term in the potential couples the atoms in all possible pairs. If we want to solve for one particular atom, we need to know the position of all the others. This is obviously not the easiest situation. Fortunately, the cross term can be eliminated through a proper coordinate transformation. The clearest way to see how this can be accomplished is to see the Hamiltonian through its matrix form. Let us assume that the characteristic matrices T and V can produce the kinetic and potential energy, respectively, through the following matrix operation.

$$H = \frac{1}{2}\sum_s \left(\frac{p_s^2}{m} + \sum_{s'} \frac{1}{2} C_{ss'}(q_s - q_{s'})^2 \right)$$

$$= \frac{1}{2m} \cdot (p_1, p_2, \cdots, p_N) \cdot T \cdot \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{pmatrix} + \frac{1}{2}(q_1, q_2, \cdots, q_N) \cdot V \cdot \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix}. \tag{3.61}$$

It is rather obvious that the matrix T is a diagonal unit matrix of N^{th} order. So, we have

$$T = \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & 0 \\ 0 & 0 & \cdot & 0 & 0 \\ 0 & \cdot & 0 & 1 & 0 \\ 0 & \cdot & \cdot & 0 & 1 \end{pmatrix}. \tag{3.62}$$

The matrix V is not as transparent, but after counting and rearranging all the terms in $\sum_{ss'} C_{ss'}(q_s - q_{s'})^2$ we can verify that V is of the following form.

$$V = \begin{pmatrix} \frac{1}{2}\sum_s (C_{1s} + C_{s1}) - C_{11} & -C_{12} & \cdot & \cdot \\ -C_{21} & \frac{1}{2}\sum_s (C_{2s} + C_{s2}) - C_{22} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ -C_{N1} & -C_{N2} & \cdot & \cdot \\ \cdot & \cdot & -C_{1N} & \\ \cdot & \cdot & -C_{2N} & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \frac{1}{2}\sum_s (C_{Ns} + C_{sN}) - C_{NN} & \end{pmatrix}. \tag{3.63}$$

Since the force constants $C_{ss'}$ are physically observable quantities and the lattice is isotropic, the matrix V must be real and symmetrical ($C_{ss'} = C_{s's}$). Therefore, V is equal to the complex conjugate of its own transpose, i.e., $V = V^+$, and by definition V is a Hermitian matrix that can always be diagonalized by a unitary transformation. The Hermitian conjugate of a unitary matrix is equal to its own inverse, i.e., $U^+ = U^{-1}$, implying $UU^+ = U^+U = I$, which is a unit diagonal matrix. The unitary matrix transforms the matrix V into a diagonal matrix D with the operation $D = UVU^+$. Let us introduce a shorthand notation $(q|$ to represent the row vector $(q_1, q_2, \cdots, q_N)$ and $|q)$ to represent the corresponding column vector in Eq. (3.61). The coordinate transformation on the potential term in Eq. (3.61) is simply given by

$$(q|V|q) = (q|U^+UVU^+U|q)$$

$$= (Q|UVU^+|Q)$$

$$= (Q|D|Q). \tag{3.64}$$

While V is diagonalized by the unitary transformation UVU^+, the $(q|$ and $|q)$ vectors are transformed according to $(Q| = (q|U^+$ and $|Q) = U|q)$, respectively. Both vectors $|q)$ and $|Q)$ represent the same N-dimensional vector in different coordinate systems. The matrix V contains non-zero off-diagonal elements in the $\{q\}$ coordinate, but the diagonal matrix D in the $\{Q\}$ coordinate does not. The diagonal matrix elements of D are the allowed characteristic eigenvalues. The corresponding eigenvectors are vectors along coordinate axes in the $\{Q\}$ representation, i.e., along the component vectors of $|Q)$. The pair coupling mixes position variables in N^2 pairs that appear explicitly in the $\{q\}$ representation, but are hidden in the $\{Q\}$ representation where the new variables $(Q_1, \cdots, Q_N)$ are separated. Without tangling the variables in the new coordinate system, the benefit is obvious because we will be able to solve for the potential energy by each variable individually.

Since $U^+IU = UIU^+ = I$, the matrix T is not altered by the same unitary transformation, and remains the same diagonal unit matrix. Therefore, the Hamiltonian H is diagonalized when V is diagonalized. The standard method to diagonalize V is to solve the secular equation resulted from requiring the determinant $|V - \lambda I| = 0$. The solution of this N^{th} order equation is the set of eigenvalues equal to λ_1, λ_2,$\cdots$,and λ_N. For each eigenvalue λ_s, the equation $\sum_{s'} V_{ss'} q_{s'} - \lambda_s q_s = 0$ determines the direction of the corresponding eigenvector in the $\{q\}$ coordinate. There are N unit eigenvectors that forms the columns of the unitary matrix U. The obvious challenge lies in how to find the solution to the N^{th} order secular equation. To bypass this hurdle we may take an intuitive approach by trying the normal-mode coordinate to see whether V and H can be diagonalized in this reference framework.

We recall that the same dynamic state of lattice vibration can be represented by the superposition of normal modes. Knowing the N pairs of (Q_k, P_k) is equivalent to knowing the N pairs of (q_s, p_s), or vise versa, because they are related by the Fourier transformation in Eqs. (3.54) and (3.55). Comparing $Q_k = \sqrt{\frac{1}{N}} \sum_s q_s e^{-iksa}$ of Eq. (3.54) and the k^{th} component of the column vector $|Q)$, i.e., $Q_k = \sum_s U_{ks} q_s$, we notice that the matrix element $U_{ks} = \sqrt{\frac{1}{N}} e^{-iksa}$. If the matrix U is unitary, it would be able to diagonalize V and H, and Q_k would be the eigenvectors of the diagonalized potential matrix. To see whether U is unitary, we need to

determine whether the matrix product UU^+ is a diagonal unity matrix or not. With the help of Eq. (3.49), we have already learned from Eq. (3.50) that

$$(UU^+)_{kk'} = \sum_s U_{ks}U^+_{sk'} = \sum_s U_{ks}U^*_{k's} = \frac{1}{N}\sum_s e^{i(k'-k)sa} = \delta_{kk'}. \quad (3.65)$$

Indeed, the matrix U is unitary and we should be able to diagonalize V by carrying out the matrix operation of UVU^+. Let us calculate the matrix element $(UVU^+)_{kk'}$ and see what happens.

$$\begin{aligned}(UVU^+)_{kk'} &= \sum_s \sum_{s'} U_{ks}V_{ss'}U^+_{s'k'} \\ &= \frac{1}{N}\sum_s e^{-iksa} \cdot \sum_{s'} V_{ss'}e^{ik's'a} \\ &= \frac{1}{N}\sum_s \left(\sum_{s'} V_{ss'}e^{ik'(s'-s)a}\right) \cdot e^{i(k'-k)sa}. \end{aligned} \quad (3.66)$$

For a given atom at the s^{th} site in the segment of N atoms, the summation over s' in Eq. (3.66) is to count the coupling of the s^{th} atom with atoms located at $s' = s, s \pm 1, s \pm 2, \cdots, s \pm n$, and so on until the coupling strength diminishes. Because the linear lattice is isotropic and uniform, the coupling constant between the s^{th} and $(s+n)^{th}$ atoms is identical to that between the s^{th} and $(s-n)^{th}$ atoms, and it is independent of the index s. For instance, the coupling constant to the n^{th} neighbor from both sides of an atom is all the same regardless from which atom we are looking. This symmetry can be expressed explicitly by making use of the identity $C_{s(s+n)} = C_{s(s-n)} = C_{sn} = C_n$, where C_n is the n^{th} neighbor coupling constant. From the matrix V given in Eq. (3.63), we have the off-diagonal element $V_{ss'} = -C_{ss'}$, and the diagonal element equals to

$$\begin{aligned}V_{ss} &= \frac{1}{2}\sum_{s'} \cdot (C_{s's} + C_{ss'}) - C_{ss} = \sum_{s' \neq s} C_{ss'} \\ &= \sum_{n>0} (C_{s(s+n)} + C_{s(s-n)}) \\ &= 2\sum_{n>0} C_n. \end{aligned} \quad (3.67)$$

In fact, we can regroup the entire summation of $\sum_{s'}$ in Eq. (3.66), including

both the diagonal V_{ss} and the off diagonal $V_{ss'}$, into $(n, -n)$ pairs. So, we have

$$
\begin{aligned}
(UVU^{+})_{kk'} &= \frac{1}{N}\sum_{s}\left(\sum_{s'} V_{ss'}e^{ik'(s'-s)a}\right)\cdot e^{i(k'-k)sa} \\
&= \frac{1}{N}\sum_{s}\left(2\sum_{n>0} C_n - \sum_{n>0} C_n(e^{-ik'na}+e^{ik'na})\right)\cdot e^{i(k'-k)sa} \\
&= \frac{1}{N}\sum_{s}\left(2\sum_{n>0} C_n - 2\sum_{n>0} C_n\cos(k'na)\right)\cdot e^{i(k'-k)sa} \\
&= \frac{2}{N}\sum_{n>0} C_n\left(1-\cos(k'na)\right)\cdot\sum_{s} e^{i(k'-k)sa} \\
&= 2\sum_{n>0} C_n[1-\cos(k'na)]\cdot\delta_{kk'}. \qquad (3.68)
\end{aligned}
$$

This result shows that all off-diagonal matrix elements $(UVU^{+})_{kk'} = 0$. The obvious reason why we can diagonalize V in the normal-mode coordinate is the orthogonal property between two plane waves, i.e., $\sum_s e^{i(k'-k)sa} = N\cdot\delta_{kk'}$. The underlying reason is the symmetry of force constants and the lattice periodicity. Recalling the normal-mode frequency defined in the dispersion relation in Eq. (3.10), we come to realize that the diagonal matrix elements are actually nothing but a measure of all possible vibration frequencies, i.e.,

$$
\begin{aligned}
(UVU^{+})_{kk} &= 2\cdot\sum_{n>0} C_n(1-\cos(kna)) \\
&= m\cdot\omega_k^2. \qquad (3.69)
\end{aligned}
$$

Since the seemly random vibrations of N atoms in the lattice can always be decomposed into normal modes that belong to the complete set of N plane waves, it would be natural to expect that the system's total energy will be obtainable from the involved normal modes. This idea is quantified when we transform the coordinate from the $\{q\}$ to $\{Q\}$ representation. Let us recall that the k^{th} component of the column vector $|Q)$ is obtained from

$$
|Q)_k = \sum_s U_{ks}q_s = \sqrt{\frac{1}{N}}\sum_s q_s e^{-iksa} = Q_k, \qquad (3.70)
$$

and the row vector is transformed according to

$$(Q|_k = \sum_s q_s U_{sk}^+ = \sum_s q_s U_{ks}^* = \sqrt{\frac{1}{N}} \sum_s q_s e^{iksa} = Q_{-k}. \tag{3.71}$$

Therefore, the total potential energy is equal to the sum of that of all participating normal modes as shown here by the following equation

$$\frac{1}{2}(q|V|q) = \frac{1}{2}(Q|UVU^+|Q) = \frac{1}{2}\sum_k m \cdot \omega_k^2 Q_{-k} Q_k. \tag{3.72}$$

The left-hand side of Eq. (3.72) is the potential energy calculated in the $\{q\}$ representation, where atoms are counted individually and added up. The right-hand side is the same potential energy calculated in the $\{Q\}$ representation in terms of normal modes. Each k value involves a pair of oppositely going waves vibrating at the same frequency $\omega_k (= \omega_{-k})$, and the potential energies of $\pm k$ modes add up to $\frac{1}{2} m \cdot \omega_k^2 \cdot (Q_{-k} Q_k + Q_k Q_{-k})$.

While the unitary matrix U transforms q_s into Q_k according to $|Q) = U|q)$, the Fourier transform defined in Eqs. (3.54) and (3.55) dictates that it takes the complex conjugate of U to transform p_s into P_k, i.e., $|P) = U^*|p)$. Taking the shorthand notations of $(p|$ and $|p)$ to replace the respective row and column momentum vectors in Eq. (3.61), we can perform similar matrix and vector manipulation as we did for the position variable. Accordingly, we have

$$\begin{aligned}
\frac{1}{2m}(p|T|p) &= \frac{1}{2m}(p|(U^*)^+ U^* I (U^*)^+ U^*|p) \\
&= \frac{1}{2m}(P|U^* I (U^*)^+|P) \\
&= \frac{1}{2m}(P|(UIU^+)^*|P) \\
&= \frac{1}{2m}(P|I|P) \\
&= \frac{1}{2m}\sum_k P_{-k} P_k.
\end{aligned} \tag{3.73}$$

The left-hand side of Eq. (3.73) is the kinetic energy calculated by counting atoms individually. The right-hand side is the same kinetic energy calculated in terms of normal modes. The pair of $\pm k$ modes, both vi-

brating at the same frequency ω_k, will have the kinetic energy add up to $\frac{1}{2m} \cdot (P_{-k}P_k + P_kP_{-k})$.

Summing up the kinetic and potential energy, we have the Hamiltonian equal to

$$H = \frac{1}{2m}(p|T|p) + \frac{1}{2}(q|V|q)$$

$$= \frac{1}{2m}\sum_k P_{-k}P_k + \frac{1}{2}\sum_k m \cdot \omega_k^2 Q_{-k}Q_k. \tag{3.74}$$

The Hamiltonian is indeed diagonalized and de-coupled into $\frac{N}{2}$ pairs of simple harmonic oscillators. Both $\pm k$ modes in the pair oscillate at the frequency ω_k and can be treated separately from other frequencies. This is a pleasing outcome for our trial initiative to change the coordinate from (p_s, q_s) to (P_k, Q_k). A special case of only two coupled harmonic oscillators is depicted in Fig. 3.5 in the problem section.

Up to this point, we have implicitly treated (p_s, q_s) and (P_k, Q_k) as classical variables. The Hamiltonian in Eq. (3.74) is derived as the classical Hamiltonian in the normal mode coordinate, but it also corresponds to the Hamiltonian operator in quantum mechanics. When p_s and q_s are considered as quantum operators, they are Hermitian operators having the property of ${p_s}^+ = p_s$ and ${q_s}^+ = q_s$. Consequently from Eqs. (3.54) and (3.55), also refer to Eqs. (3.70), (3.71), and (3.73), we notice that ${P_k}^+ = P_{-k} \neq P_k$ and ${Q_k}^+ = Q_{-k} \neq Q_k$, and hence neither P_k nor Q_k is Hermitian. The mixing of $\pm k$ modes in the Hamiltonian prevents us from taking analogous solutions directly from simple harmonic oscillators without further transformation of the Hamiltonian. There are a couple of methods that can solve the energy eigenvalue problem governed by Eq. (3.74). The method we will follow involves Dirac's creation and annihilation operators in the Hilbert space of energy eigenvectors. This method is the natural way to introduce the concept of phonons as quantized excitations of lattice waves.

3.16 Creation, Annihilation, and Number Operators — The N Representation

Dirac's method to treat the quantum simple harmonic oscillator involves a pair of operators that satisfy the commutation relation of $[a, a^+] = 1$. When the Hamiltonian, $H = \frac{1}{2m}P^2 + \frac{m\omega^2}{2}Q^2$, is transformed into $H = (a^+a + \frac{1}{2})\hbar\omega$, the problem to solve for the energy spectrum and wave func-

tions becomes the one to find the eigenvalues and eigenvectors of the Hermitian operator a^+a. This can be accomplished, basing the treatment solely upon the axioms of Hilbert vector space. Assuming $|\nu>$is a normalized eigenvector of a^+a, we have by definition that

$$a^+a|\nu>= \nu|\nu>, \tag{3.75}$$

and

$$<\nu|\nu>= 1. \tag{3.76}$$

The corresponding eigenvalue, ν, is equal to the norm of $a|\nu>$, i.e.,

$$\nu =<\nu|a^+a|\nu>=\| a|\nu>\|^2 . \tag{3.77}$$

Since the norm of any vector in Hilbert space is real and not negative, the eigenvalue ν must be real and $\nu \geq 0$.

If $|\nu>$ is an eigenvector of a^+a and the commutation relation $[a, a^+] = 1$ is satisfied, the vector $a|\nu>$ will be also an eigenvector. To see why, let us examine the commutator $[a^+a, a]$ first and we have

$$[a^+a, a] = a^+aa - aa^+a = [a^+, a]a = -a. \tag{3.78}$$

Since $a^+aa = aa^+a-a$, the vector $a|\nu>$ must satisfy the following equation,

$$(a^+a)a|\nu>= a(a^+a - 1)|\nu>= a(\nu - 1)|\nu>= (\nu - 1)a|\nu> . \tag{3.79}$$

Here, we see that $a|\nu>$ is indeed an eigenvector of a^+a with eigenvalue equal to $\nu - 1$. We can therefore write $a|\nu>= |\nu - 1>$. Because the eigenvalue is always real and positive, we must have $\nu \geq 1$. For the smallest value of $\nu = 1$, the vector resulted from operating a on $|1>$ is a null eigenvector, i.e., $|0>= a|1>$. The corresponding eigenvalue is 0 and the norm $<0|0>$ equals 1. If we continue to operate a on $|0>$, the resulted vector will have vanishing norm because $\| a|0>\|^2= 0$. The vector with zero norm does not exist by definition, i.e., $a|0>= 0$. The vanishing vector $a|0>$ can not be the eigenvector of a^+a, because the eigenvalue would be equal to -1 that violates the required positive eigenvalue in Hilbert space. We come to an important conclusion. Namely, the lowest eigenvalue is 0, the corresponding ground-state eigenvector is $|0>$, and the vector $a|0>$ does not exist. Similarly, we have

$$[a^+a, a^+] = a^+aa^+ - a^+a^+a = a^+[a, a^+] = a^+ \tag{3.80}$$

and

$$(a^+a)a^+|\nu>= a^+(a^+a+1)|\nu>= a^+(\nu+1)|\nu>= (\nu+1)a^+|\nu> \ . \quad (3.81)$$

Clearly $a^+|\nu>$ is also an eigenvector and the eigenvalue equals $\nu+1$. So, we have $a^+|\nu>= |\nu+1>$. Since we must have $\nu \geq 0$, the eigenvalue $\nu+1$ will never be negative. In other words, the vector $a^+|\nu>$ will never vanish because

$$\| \ a^+|\nu>\|^2 = <\nu|aa^+|\nu> = <\nu|(a^+a+1)|\nu> = \nu+1. \quad (3.82)$$

Starting from the ground state $|0>$, we can operate a^+ on $|0>$ repeatedly to obtain unlimited number of related eigenvectors. They are $|0>, a^+|0>, (a^+)^2|0>, \cdots, (a^+)^n|0>$, and so on. The corresponding eigenvalues are $0, 1, 2, \cdots, n$, and higher integers. The eigenvectors so generated have their norm given by

$$\begin{aligned}
<0|0> &= 1 = 0!, \\
<0|aa^+|0> &= <0|(a^+a+1)|0> = 1 = 1!, \\
<0|a^2(a^+)^2|0> &= <0|a(aa^+)a^+|0> = <0|a(a^+a+1)a^+|0> = 2 = 2!, \\
<0|a^3(a^+)^3|0> &= <0|a^2(aa^+)(a^+)^2|0> \\
&= <0|a^2(a^+a+1)(a^+)^2|0> = 2\cdot 2! + 2! = 3!, \\
<0|a^4(a^+)^4|0> &= <0|a^3(aa^+)(a^+)^3|0> \\
&= <0|a^3(a^+a+1)(a^+)^3|0> = 3\cdot 3! + 3! = 4!, \\
<0|a^5(a^+)^5|0> &= <0|a^4(aa^+)(a^+)^4|0> \\
&= <0|a^4(a^+a+1)(a^+)^4|0> = 4\cdot 4! + 4! = 5!, \\
&\ \ \vdots \\
<0|a^n(a^+)^n|0> &= <0|a^{n-1}(a^+a+1)(a^+)^{n-1}|0> \\
&= (n-1)\cdot(n-1)! + (n-1)! = n!.
\end{aligned} \quad (3.83)$$

The normalized eigenvector is the eigenvector divided by its own length and multiplied by an arbitrary phase factor. The phase factor can be fixed to be unity for simplicity, and we have

$$|n>= \sqrt{\frac{1}{n!}}(a^+)^n|0> \ . \quad (3.84)$$

Because a^+a is a Hermitian operator, the eigenvalue obtained from the ket $|n>$ is identical to that obtained from its conjugate bra $< n|$. For two

eigenvalues n and n', we can write one eigenequation in the ket space and the other in the bra space, i.e.,

$$a^+a|n'>= n'|n'>, \tag{3.85}$$

and

$$<n|a^+a =<n|n. \tag{3.86}$$

We will get identical result of $<n|a^+a|n'>$ if we multiply the ket equation with $<n|$ from the left side, and the bra equation with $|n'>$ from the right. Subtracting two results, we have

$$(n - n') <n|n'>= 0. \tag{3.87}$$

The inner product $< n|n'>$ must equal 0 if $n' \neq n$. On the other hand, $< n|n'>=<n|n>= 1$ if $n' = n$. Therefore, the eigenvectors in the set $\{|0>, |1>, |2>, \cdots, |n>, \cdots\}$ are pair orthonormal, and we have

$$<n|n'>= \delta_{nn'}. \tag{3.88}$$

It can be shown that the operator $N = a^+a$ forms a complete set of commuting observables, implying that the orthonormal eigenvectors $\{|0>, |1>, |2>, \cdots, |n>, \cdots\}$ is a complete set of basis vectors. Together with the following characteristic relations, the space where N, a, and a^+ are defined is completely constructed. We have

$$[a, a^+] = 1; \tag{3.89}$$

$$N = a^+a; \tag{3.90}$$

$$N|n> = n|n>; \tag{3.91}$$

$$\begin{aligned} a^+|n> &= \frac{1}{\sqrt{n!}} (a^+)^{n+1}|0> \\ &= \sqrt{n+1}\ |n+1>; \end{aligned} \tag{3.92}$$

$$\begin{aligned} a|n> &= aa^+ \frac{1}{\sqrt{n!}} (a^+)^{n-1}|0> \\ &= \frac{1}{\sqrt{n}} (a^+a + 1)|n - 1>= \sqrt{n}\ |n - 1>; \end{aligned} \tag{3.93}$$

$$a|0> = 0; \tag{3.94}$$

$$<n|n'> = \delta_{nn'}. \tag{3.95}$$

Now, we can go back to the simple harmonic oscillator. It is a simple harmonic motion with internal structure of the Hilbert space that we just

discussed. From the Hamiltonian $H = (N + \frac{1}{2})\hbar\omega$, we know its energy eigenvalue already, i.e., $E_n = (n + \frac{1}{2})\hbar\omega$. The corresponding eigenvector is $|n>$. Operating N on $|n>$, we will get the quantum number n characterizing the n^{th} energy level. The operator a^+ raises the energy by one level, while the operator a lowers the energy level by one notch. In a different but equivalent view, we can consider the energy level E_n contains n unit excitations and each one possesses the energy of $\hbar\omega$. In this picture, excitations can appear and disappear in units of $\hbar\omega$. N is the number operator that measures the number of excitations. The operator a^+ will add one excitation to the system, and the operator a will destroy one. Hence, a^+ is named the creation operator and a is the annihilation operator.

3.17 Quantized Hamiltonian of Creation and Annihilation Operators

The Hamiltonian in Eq. (3.74) has not been decomposed into simple harmonic oscillators because the mixing of $\pm k$ modes. The question is whether we can transform it to the form of $H = \sum_k (a_k{}^+ a_k + \frac{1}{2})\hbar\omega_k$. The key is whether we can find the correct creation and annihilation operators. It turns out these operators do exist. To simplify the notation, we let $A = \frac{1}{(2m\hbar\omega_k)^{1/2}}$ and $B = (\frac{m\omega_k}{2\hbar})^{1/2}$, and define the operator a_k and its Hermitian conjugate $a_k{}^+$ as

$$a_k = BQ_k + iAP_k{}^+ = BQ_k + iAP_{-k}, \tag{3.96}$$

and

$$a_k{}^+ = BQ_k{}^+ - iAP_k = BQ_{-k} - iAP_k. \tag{3.97}$$

The first step is to verify whether they satisfy the commutation relation of $[a_k, a_k{}^+] = 1$. The verification goes as follows.

$$a_k a_k{}^+ = \left(A^2 P_k{}^+ P_k + iAB\left(P_k{}^+ Q_k{}^+ - Q_k P_k\right) + B^2 Q_k Q_k{}^+\right), \tag{3.98}$$

and

$$a_k{}^+ a_k = \left(A^2 P_k P_k{}^+ + iAB\left(Q_k{}^+ P_k{}^+ - P_k Q_k\right) + B^2 Q_k{}^+ Q_k\right). \tag{3.99}$$

Noting that

$$\left[P_k, P_k{}^+\right] = 0, \qquad \left[Q_k, Q_k{}^+\right] = 0, \qquad \text{and} \qquad A \cdot B = \frac{1}{2\hbar}, \tag{3.100}$$

we have

$$\begin{aligned}\left[a_k, a_k{}^+\right] &= a_k a_k{}^+ - a_k{}^+ a_k \\ &= -iAB\left(Q_k{}^+ P_k{}^+ - P_k{}^+ Q_k{}^+\right) - iAB\left(Q_k P_k - P_k Q_k\right) \\ &= -iAB \cdot \left[Q_k{}^+, P_k{}^+\right] - iAB \cdot \left[Q_k, P_k\right] \\ &= \frac{-i}{2\hbar} \cdot (i\hbar + i\hbar) \\ &= 1. \end{aligned} \tag{3.101}$$

Indeed, the operator a_k and $a_k{}^+$ as defined in Eqs. (3.96) and (3.97) do satisfy the basic commutation relation required for the creation and annihilation operators. Next, we want to examine whether the Hamiltonian can be transformed into the correct form of $\sum_k \left(a_k{}^+ a_k + \frac{1}{2}\right)\hbar\omega_k$. Since the Hamiltonian in Eq. (3.74) is invariant under the exchange of index between k and $-k$, it hints that we must involve all four terms of $a_k a_k{}^+, a_k{}^+ a_k, a_{-k} a_{-k}{}^+$, and $a_{-k}{}^+ a_{-k}$, so that the transformed Hamiltonian can preserve the same symmetry. Proceeding along this line of thinking, we have

$$\begin{aligned}&a_k a_k{}^+ + a_k{}^+ a_k \\ &= 2A^2 P_k P_k{}^+ + 2B^2 Q_k{}^+ Q_k + iAB\left(Q_k{}^+ P_k{}^+ - P_k Q_k + P_k{}^+ Q_k{}^+ - Q_k P_k\right) \\ &= 2A^2 P_k P_{-k} + 2B^2 Q_{-k} Q_k + iAB\left(Q_{-k} P_{-k} - P_k Q_k + P_{-k} Q_{-k} - Q_k P_k\right),\end{aligned} \tag{3.102}$$

$$\begin{aligned}&a_{-k} a_{-k}{}^+ + a_{-k}{}^+ a_{-k} \\ &= 2A^2 P_{-k} P_k + 2B^2 Q_k Q_{-k} + iAB\left(Q_k P_k - P_{-k} Q_{-k} + P_k Q_k - Q_{-k} P_{-k}\right),\end{aligned} \tag{3.103}$$

and

$$\begin{aligned}&\sum_k \frac{\hbar\omega_k}{4} \cdot \left\{\left(a_k a_k{}^+ + a_k{}^+ a_k\right) + \left(a_{-k} a_{-k}{}^+ + a_{-k}{}^+ a_{-k}\right)\right\} \\ &= \sum_k \hbar\omega_k \left(A^2 P_{-k} P_k + B^2 Q_{-k} Q_k\right) \\ &= \sum_k \left(\frac{1}{2m} \cdot P_{-k} P_k + \frac{m\omega_k{}^2}{2} \cdot Q_{-k} Q_k\right) \\ &= \sum_k H_k \\ &= H.\end{aligned} \tag{3.104}$$

At this juncture, we have transformed the Hamiltonian into a linear function of four operators of $a_k a_k^+, a_k^+ a_k, a_{-k} a_{-k}^+$, and $a_{-k}^+ a_{-k}$. Since the summation is over the k values inside the first Brillouin zone, it is symmetrical about $k = 0$. Therefore, we have

$$\sum_k \frac{\hbar\omega_k}{4} \cdot \left(a_k a_k^+ + a_k^+ a_k\right) = \sum_k \frac{\hbar\omega_k}{4} \cdot \left(a_{-k} a_{-k}^+ + a_{-k}^+ a_{-k}\right). \quad (3.105)$$

This leads to our final result, i.e.,

$$\begin{aligned} H &= \sum_k H_k \\ &= \sum_k \frac{\hbar\omega_k}{2} \cdot \left(a_k a_k^+ + a_k^+ a_k\right) \\ &= \sum_k \left(a_k^+ a_k + \frac{1}{2}\right) \hbar\omega_k \\ &= \sum_k \left(N_k + \frac{1}{2}\right) \hbar\omega_k. \end{aligned} \quad (3.106)$$

Our effort really pays off, because the seemly impossible problem to solve for N vibrating atoms has been transformed into the rather easy task of handling N de-coupled simple harmonic oscillators. All we need to do now is to apply what we have learned in the last section and write down the eigenvalue and eigenvector solutions. If we want, we can follow the established prescription to calculate the wave function of each energy level, the expectation values for p_s and q_s, and other observable quantities. But, the most important gain out of our long exercise is the concept of phonons that we will discuss in the rest of this Chapter.

3.18 Phonon Interpretation — Quantization of Lattice Vibration Waves

Since the Hamiltonian in Eq. (3.106) consists of N independent simple harmonic oscillators specified by the normal-mode wave vectors k, they can be treated separately. The energy eigenvalue for the k^{th} mode at the ${n_k}^{th}$ quantum level is given by

$$E_{n_k,k} = \left(n_k + \frac{1}{2}\right) \hbar\omega_k, \quad (3.107)$$

In the realm of quantum mechanics, the normal mode of wave vector k and frequency ω_k is a dynamic state that possesses a discrete spectrum of energy levels. The ground state energy is $\frac{1}{2}\hbar\omega_k$, the first excited level is at $\left(1+\frac{1}{2}\right)\hbar\omega_k$, and so forth. The energy described by Eq. (3.107) can be also viewed naturally as a collection of basic energy excitations that we call phonons. Each phonon possesses a fixed energy of $\hbar\omega_k$ and its value solely depends upon how fast the parent normal mode vibrates. There is no limit on how many phonons we can have in a normal mode, as long as they can add up to count for the total energy of the mode. In this sense, the energy-level quantum number n_k has the meaning of occupation number for phonons. It measures how many phonons there are in the k mode. The adjacent energy levels such as $E_{n_k-1,k}, E_{n_k,k}$, and $E_{n_k+1,k}$ differ only by the energy of a single phonon. The energy eigenvector $|n_k>$ represents the dynamic state that contains n phonons in the k mode. When ${a_k}^+$ operates on $|n_k>$, the dynamic state goes from $|n_k>$ to $|n_k+1>$. In essence, the operator ${a_k}^+$ adds one phonon to the k mode. When a_k operates on $|n_k>$, it destroys one phonon as the dynamic state goes from $|n_k>$ to $|n_k-1>$. Therefore, ${a_k}^+$ is called the creation operator and a_k is the annihilation operator. Phonons can be created and destroyed as long as the energy and k vector are both conserved. When N_k operates on $|n_k>$, it does not change the number of phonons, but it reads the phonon occupation number by getting the eigenvalue n_k. So, N_k is the number operator. When the Hamiltonian H_k operates on $|n_k>$, the energy eigenvalue $E_{n_k,k} = \left(n_k+\frac{1}{2}\right)\hbar\omega_k$ results. For the ground state $|0>$ where no phonon exists, the energy is not zero but equal to $\frac{1}{2}\hbar\omega_k$, which is a consequence of the uncertainty principle. The state $|0>$ is a vacuum state for phonons, where we have no phonons to destroy and can only create phonons. With the phonon interpretation, we have quantized the lattice vibration wave in terms of particle-like phonons.

3.19 Lattice Vibrations and Phonons — Connection Remarks

We have learned from the example of a one-dimensional lattice on how to model the lattice vibration. From monochromatic plane waves to wave packets, everything discussed up to that point are within the realm of classical mechanics. The mode of lattice vibration is fully specified in terms of its wave vector k, frequency ω_k, and amplitude. The energy carried by the

vibration mode is presumably obtainable from summing up the kinetic and potential energy of all participating atoms. The resulted total energy will be proportional to the square of the amplitude. Traditionally in the classical point of view, we take both the energy and amplitude as continuously varying quantities. They can presumably decrease continuously to zero.

When quantum mechanics is applied to the same problem, we find that the classical mode of k and ω_k is replaced with a quantum harmonic oscillator. Its energy is counted for by counting phonons, the corpuscle-like energy excitations that appear naturally in the quantum picture of lattice vibrations. The phonon occupation number n_k obeys Bose-Einstein statistics. At extremely low temperatures, there are only a few or even zero phonons. The discreteness of the zero-point energy $\frac{1}{2}\hbar\omega_k$ and other low-lying excited states are apparent. This is where the classical picture of continuous amplitudes fails and we must use the quantum picture of phonons. For large phonon occupation numbers, usually occurs except at fairly low temperatures, the energy level spacing becomes relatively small in comparison to the energy level itself. In the limit where we can consider $\hbar \approx 0$, the discrete energy levels approach a continuous spectrum that converges to what the classical lattice wave describes. In other words, the averaged energy in a classical vibration mode should equal to the energy of its corresponding quantum harmonic oscillator, i.e.,

$$\varepsilon = \left(n + \frac{1}{2}\right) \cdot \hbar\omega. \tag{3.108}$$

The quantum number k is omitted for simplicity since we only deal with one mode here as an example. In order to find the averaged energy of the vibration mode sustained in the volume V of N atoms, we need to calculate the kinetic and potential energy of each atom at a given time, add them up, and average over time.

For a standing wave with k vector in the x direction, the vibrating lattice plane perpendicular to k is described in the continuum limit by $u = u_0 \cdot \cos(kx) \cdot \cos(\omega t)$. The average energy of this vibration mode is half kinetic energy and half potential energy as all harmonic oscillator does. We can just find the average kinetic energy if it is easier to do, and by simply doubling it we will have the total energy. The kinetic energy of the vibration mode is $\frac{1}{2}\rho\left(\frac{du}{dt}\right)^2$, where ρ is the mass density. To apply the boundary condition, let us require $u(L) = u(0)$. When we sum up the whole volume of L^3, y and z directions are trivial. In the x direction, we do the following integral from 0 to L first. Note that the periodic condition

requires $\cos(k \cdot L) = \cos(k \cdot 0)$, and we have

$$\int_0^L \cos^2(kx)dx = \int_0^L \frac{1}{2}\left(1 + \cos(2kx)\right) dx = \frac{L}{2}. \tag{3.109}$$

The total kinetic energy in the volume $V(= L^3)$ at a given time t is

$$\frac{\rho L^3 u_0^2 \omega^2}{4} \sin^2(\omega t). \tag{3.110}$$

Now we must do the time average. An average over one period T is sufficient. We have the time averaged kinetic energy equal to

$$\begin{aligned} &\frac{1}{T}\frac{\rho L^3 u_0^2 \omega^2}{4} \int_0^T \sin^2(\omega t)dt \\ &= \frac{\rho L^3 u_0^2 \omega^2}{4T} \int_0^T \frac{1}{2}\left(1 - \cos(2\omega t)\right) dt \\ &= \frac{\rho L^3 u_0^2 \omega^2}{8}. \end{aligned} \tag{3.111}$$

The total energy will be twice the value of the average kinetic energy, and it will equal that of the quantum oscillator. We have

$$\begin{aligned} \frac{1}{4}\rho L^3 \omega^2 u_0^2 &= \left(n + \frac{1}{2}\right)\hbar\omega \\ u_0^2 &= 4\left(n + \frac{1}{2}\right)\frac{\hbar}{\rho L^3 \omega}. \end{aligned} \tag{3.112}$$

In this example, we have derived the relation between the classical vibration amplitude and phonon occupation number for a given normal-mode frequency, or the phonon frequency. By forcing the averaged energy of the classical vibration mode to equal $\left(n + \frac{1}{2}\right)\hbar\omega$, we literally replace a volume of vibrating atoms with a bag of phonons.

Problems

1. Show the plane wave in Eq. (3.5) is the solution to the wave equation in Eq. (3.3).

2. Consider the dispersion relation in the nearest neighbor approximation, see Eq. (3.15). For three cases of $k = 0.5(\frac{\pi}{a}), 0.8(\frac{\pi}{a})$, and $(\frac{\pi}{a})$, plot the wave packet described by Eq. (3.25) for two cases of $t = 0$ and $t = 10$ sec. Show from the plots why $\frac{d\omega}{dk}$ is the group velocity. What are the

group velocities for all three cases?

3. Show by indicating on a copy of Fig. 3.2.
   (a) why the Bragg scattering near the zone boundary will form standing waves.
   (b) why the scattering that transfers a G vector will not produce standing waves for k vectors away from the zone boundary.

4. (a) Why the first Brillouin zone is important when dealing with the lattice vibration?
   (b) Where is the first Brillouin zone boundary in a continuous medium?
   (c) Do we have Brillouin zones in amorphous materials and why?

5. What is the difference between the translation symmetry and the periodic boundary condition in a given lattice?

6. Consider a linear lattice where each atom oscillates around its own lattice site with small amplitudes. Show that you understand Eq. (3.35) and Eq. (3.36). That is to show why the following two equalities are true.

$$\frac{\partial \sum_{i,j} v_{i,j}}{\partial u_{s,s'}} = \frac{\partial v_{s,s'}}{\partial u_{s,s'}}, \quad \text{and}$$

$$\sum_{i,j} \frac{\partial}{\partial u_{i,j}} \frac{\partial V}{\partial u_{s,s'}} = \frac{\partial^2 v_{s,s'}}{\partial u_{s,s'}^2}. \tag{3.113}$$

7. Derive equation of motion in Eq. (3.1) from the Hamiltonian in Eq. (3.41).

8. (a) Verify that Eq. (3.62) and Eq. (3.63) are true.
   (b) Write the unitary matrix U defined in Eq. (3.64) in its explicit matrix form.

9. (a) Show that ${P_k}^+ = P_{-k} \neq P_k$ and ${Q_k}^+ = Q_{-k} \neq Q_k$.
   (b) Show that the Hamiltonian in Eq. (3.74) is a Hermitian operator.

10. Consider the following setup of two equal masses connected by three identical springs in between two walls.
    (a) Find the Hamiltonian as a function of (p_1, q_1, p_2, q_2).

(b) Express the Hamiltonian in the matrix form explicitly.
(c) Find the matrix that can diagonalize the Hamiltonian.
(d) Find the eigenvalues and eigenvectors.
(e) Find the Hamiltonian in the normal mode coordinate.
(f) Find the normal mode frequencies

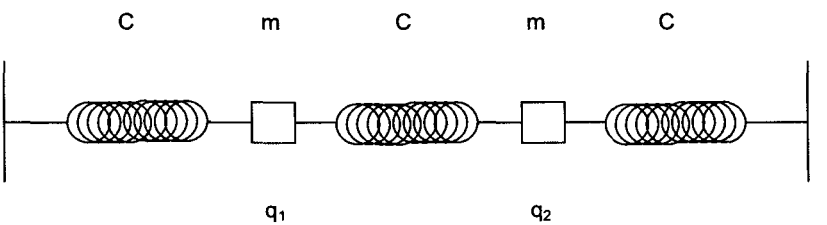


Fig. 3.5 Two coupled simple harmonic oscillators.

This page is intentionally left blank

Chapter 4

Thermal Properties of Insulators

Introduction and Summary

Chapter 4 has two parts. The first part is a brief introduction to the statistical physics. The fundamental concept of statistical equilibrium is explained by examining the case when a small quantum system is immersed in a large heat bath. As the initial energy flow dies out and the combined system reaches the statistical equilibrium, the Gibbs distribution over the quantum levels of the small system is derived from basic probability considerations. The temperature enters in the picture as the constant of statistical equilibrium, while the Boltzmann constant relates the temperature to energy. The partition function is defined by normalizing the total probability to one. The thermodynamic parameters are then derived from it. This extra effort is worthwhile, especially for those readers who are not familiar with the fundamental statistical mechanics. The objective is to offer an explanation about the origin of the basic Gibbs distribution. So, the flow of its application to phonons would be smoother with no undesirable misconceptions. It also sets the stage for introducing electron statistics in Chapter 5.

When statistics is applied to the quantum system of phonons, each phonon mode can be considered independently. It is shown as a demonstration that the Helmholtz free energy of the whole system of phonons is equal to the summation of the Helmholtz free energy of each phonon mode. Hence, the Gibbs distribution, the Helmholtz free energy, the partition function, and the averaged total energy are all derived for one phonon mode only. By definition, the averaged phonon occupation number when multiplied by the phonon energy must be the averaged total energy. This very definition offers an easy way to obtain the averaged phonon occupa-

tion number as a function of phonon energy and temperature. This function turns out to be nothing but the Planck distribution.

The second part begins with a revisit to the averaged phonon occupation number from a different derivation. All the possible expectation values of the phonon occupation number are calculated from the Gibbs distribution, and then summed up to give the averaged phonon occupation number. Einstein model of lattice heat capacity, despite of its over simplified assumption of only one phonon frequency, offers a quick demonstration to show the power of the phonon concept. It also shows the need for taking all phonon frequencies into account. We begin with the enumeration of all allowed k states in the first Brillouin zone, which leads to the definition of density of states in k space. From there, the density of states in the frequency domain is obtained through the use of the dispersion relation. The results, which are inversely proportional to the phonon group velocity, are derived for both the one and three dimensional cases. In the long wavelength limit, the density of states can be integrated analytically and the result is the Debye density of states. The Debye cutoff frequency and wave vector are introduced in order to keep the total number of states equal to the number of normal modes. Using three sound velocities to model for the polarization differences, we can plot the Debye density of states to simulate two branches of the transverse mode and one in longitudinal. This exercise is simple but very insightful. The Debye approximation is no longer valid for wave vectors near the first Brillouin zone boundary. The actual density of states will be larger than the Debye value because the group velocity slows down. For wave vectors at the zone boundary where the group velocity vanishes, Van Hove singularities result. The phonon spectrum in lanthanum, deduced from the superconducting tunneling data, is reproduced here to show that all the discussed features are actually measurable experimentally. The Debye model of lattice heat capacity is then formulated. Both the energy and heat capacity are expressed in their respective integral form, and the Debye temperature is defined. The integral can be evaluated analytically, if the temperature is either much higher or much lower than the Debye temperature. The former is the Dulong-Petit law, and the latter is the Debye T^3 law. The concept of dominant phonons is introduced for gaining deeper insights. As an example, a quick but insightful derivation of the Debye T^3 law is offered.

The phonon crystal momentum, Umklapp and normal processes, and the phonon mean free path are the essential concepts in any attempt to understand the phenomena that involve the scattering of phonons. The

lattice thermal conductivity is discussed as a vehicle to introduce these concepts. Why phonons do not have real momentum is explained in terms of the zero net atomic momentum in the normal mode. The conservation of wave vector is always satisfied in the scattering process. It resembles the law of conservation of momentum within a factor of the Planck constant. Since phonons do not exist outside the first Brillouin zone, any extra momentum must be transferred to the entire lattice as a whole. The reciprocal lattice vector involved in the conservation law actually counts for the real recoil momentum of the crystal. Therefore, the quantity $\hbar\mathbf{G}$ is called the crystal momentum, and so is called for the momentum-like quantity $\hbar\mathbf{k}$. When a reciprocal lattice vector is needed to conserve the crystal momentum, the scattering is of the Umklapp process that can cause thermal resistance. When the reciprocal lattice vector is not needed to conserve the crystal momentum, the scattering is a normal process. The concepts of phonon mean free path and relaxation time are introduced to hide the details of phonon scattering. The lattice thermal conductivity is calculated within the scheme of relaxation time. The result shows a well-known relationship that connects the thermal conductivity to the specific heat, sound velocity, and the phonon mean free path. The phonon mean free path is analyzed in two regions of both the high and low temperatures. It offers a feel for the temperature dependence in the thermal conductivity.

Phonon Statistics

4.1 Equilibrium, Probability, Gibbs Distribution, Partition Function, and Temperature

When a small object drops into a large volume of water, regardless whether the object is originally hotter or colder, it will soon reach the same temperature of the bath and the bath temperature will not change noticeably at all. This ordinary phenomenon frequently seen in daily life is an example of thermal equilibrium. The key here is the size of the bath that is large enough to actually maintain a constant temperature environment. The energy exchanged may be huge to the tiny object, but it is negligible when compared with the total energy of the bath. Taking the analysis further, we can learn more about the statistical equilibrium reachable by a quantum system in a large heat bath. Especially important is the probability distributed to the likelihood of each quantum level.

Consider a small quantum system S, such as the one-dimensional lattice we have studied, that is brought into contact with a large heat bath H. There will be an one-way energy flow initially, either from S to H or vice versa, depending upon which one is hotter. After the initial energy redistribution dies down, the system S enters into a long lasting state with no net energy flow either into or out of its boundary. However, there are always energy exchanges between S and H microscopically, and each such exchange is characterized by a relaxation time. On the time scale much longer than the relaxation time, the random energy exchanges are averaged out and result in zero net flow. It is in this macroscopic sense that the system S is considered in statistical equilibrium. The exact solution of a steady-state wave function for S does not exist because of the continuous random interactions with H. It takes an ensemble of wave functions $\{\Psi\}$ to describe the physical state of S at statistical equilibrium. None of $\{\Psi\}$ is a true steady-state wave function. On the other hand, if S is completely isolated, we can in principle solve for its steady-state wave functions $\{\phi_n\}$ and the corresponding energy spectrum $\{E_n\}$. The wave functions $\{\phi_n\}$ form a complete set of orthonormal basis. Now each Ψ can be expressed as a linear combination of ϕ_n, i.e., $\Psi = \sum c_n \psi_n$. The average of ${c_n}^* c_n$ over the ensemble $\{\Psi\}$ is the statistical weight or probability density $P(E_n)$ for system S to be at energy E_n when it is in statistical equilibrium.

Just like S can not be in a steady state even at statistical equilibrium, the equilibrium energy of the perfectly isolated H is also smeared somewhat due to the fluctuating exchange between S and H. Normally, the larger the system the tighter its energy levels are packed together. Take the one-dimensional lattice as an example. When the length increases, the spacing between adjacent k states decreases, and there will be more k states squeezed into a given energy interval of ΔE. Suppose H is so large that its energy spectrum becomes continuous. Then the exchange between S and H, no matter how small, is capable of scattering the bath from E to $E+\delta E$. In addition to the statistical smearing, there is also quantum uncertainty about the energy. In order to count for all the uncertainties, it makes more sense to specify the bath with a small range of $\pm\frac{1}{2}\Delta E$ around the averaged value E. The probability to find the bath within the range is peaked at E and decreases quickly outside the range of $E \pm \frac{1}{2}\Delta E$.

While the system S and heat bath H exchange energy and settle down at the equilibrium energy of E_s and E, respectively, their combined energy E_t remains unchanged. So, we have the condition $E_t = E + E_s$. Examining the energy E_s microscopically, we find that the entire energy spectrum $\{E_n\}$

contributes according to the probability distribution $P(E_n)$. Namely, the system's equilibrium energy E_s is just the sum of the expectation value of each E_n, i.e., $E_s = \sum E_n P(E_n)$. We are obviously interested in the explicit function form of $P(E_n)$, that is the probability to find the entire system S at the energy level E_n. For S to have the probability $P(E_n)$ to fall on E_n, it requires H to be in the range of $E \pm \frac{1}{2}\Delta E$ such that

$$E_t = E + E_n. \tag{4.1}$$

The width ΔE of the bath determines how wide the sample's wave functions $\{\Psi\}$ will spread. Consequently, the width of $\{c_n{}^* c_n\}$ is proportional to ΔE. Since the peak $P(E_n)$ decreases when the spread of $\{c_n{}^* c_n\}$ widens, $P(E_n)$ must be inversely proportional to ΔE. For the reason of $\Delta E/E \ll 1$, the density-of-states $D(E)$ in the bath is unlikely to change abruptly within the small range of ΔE around E. The number of states inside ΔE can therefore be approximated by $\Delta\Gamma(E) = D(E)\Delta E$. On the contrary, the small system S with discrete energy levels will have only a single state at E_n if there is no degeneracy. The product $\Delta\Gamma(E) \cdot 1$ counts for all the possible combinations of available states for both S and H to fall into their respective energy levels. This is just the number of ways for the said scenario to occur. The larger this number is, the higher the probability $P(E_n)$ will be. $P(E_n)$ must be proportional to $\Delta\Gamma(E) \cdot 1$. So, we have

$$P(E_n) \propto \frac{\Delta\Gamma(E)}{\Delta E} \propto D(E) = D(E_t - E_n). \tag{4.2}$$

Then, we can take the ratio between $P(E_n)$ and $P(E_{n+1})$ from two adjacent levels to eliminate the proportion constants, and we have

$$\frac{P(E_n)}{P(E_{n+1})} = \frac{D(E_t - E_n)}{D(E_t - E_{n+1})} = e^{\ln D(E_t - E_n) - \ln D(E_t - E_{n+1})}. \tag{4.3}$$

Because the sheer size difference between H and S, we will always have $E = E_t - E_n \gg E_n$, and the approximation $E \approx E_t$ holds well for all E_n. So, we have the density-of-states function $D(E) = D(E_t - E_n) \approx D(E_t)$, i.e., $D(E)$ does not deviate $D(E_t)$ much at all in the range of $E \pm \frac{1}{2}\Delta E$. We can replace $D(E)$ with $D(E_t)$ and set $dD(E)/dE = (dD(E)/dE)_{E_t}$ whenever it is appropriate. This consideration leads to

$$\frac{d\ln D(E_t - E_n)}{d(E_t - E_n)} = \frac{d\ln D(E)}{dE} = \frac{1}{D(E)}\frac{dD(E)}{dE} = \frac{1}{D(E_t)}\left(\frac{dD(E)}{dE}\right)_{E_t}$$

$$= \beta(E_t) = \beta. \tag{4.4}$$

The resulted parameter β is a function of E_t only. As long as E_t remains constant, β will be a constant. The logarithm of $\Delta\Gamma(E)$ is defined as the entropy, i.e., $S = k_B \ln(\Delta\Gamma(E))$, where k_B is the Boltzmann constant. The corresponding temperature of the system is defined by the derivative $dS/dE = 1/T$. Since $d\Delta E/dE \approx 0$ is true, we also have $dS/dE = k_B d\ln(\Delta\Gamma(E))/dE = k_B \cdot d\ln(D(E))/dE = k_B \cdot \beta$. We notice here that the parameter β must be $1/(k_B T)$. The temperature defined here makes no difference whether it is for the heat bath H or for the combined system of H + S at equilibrium. Rearranging the exponential argument in Eq. (4.3), we have

$$\frac{P(E_n)}{P(E_{n+1})} = e^{\frac{\ln D(E_t - E_n) - \ln D(E_t - E_{n+1})}{(E_n - E_{n+1})} \cdot (E_n - E_{n+1})} = e^{-\beta(E_n - E_{n+1})}. \quad (4.5)$$

Therefore, the probability $P(E_n)$ for S to be in the energy state E_n is proportional to $e^{-\beta E_n}$. Introducing the proportion constant A, we have

$$P(E_n) = Ae^{-\beta E_n}. \quad (4.6)$$

This is the Gibbs distribution for the quantum system characterized by the energy spectrum of $\{E_n\}$ at temperature T. It measures the probability for the system to be at the energy level E_n. The proportional constant will be determined by normalizing the total probability to unity. Summing up over all allowed levels, we have

$$\sum_i P(E_i) = 1 = A \sum_i e^{-\beta E_i}, \quad (4.7)$$

leading to

$$Q = \frac{1}{A} = \sum_i e^{-\beta E_i}, \quad (4.8)$$

and

$$P(E_i) = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}} = \frac{1}{Q} \cdot e^{-\beta E_i}. \quad (4.9)$$

The function $Q = \sum_i e^{-\beta E_i}$ is the partition function, resulted from requiring the total probability to equal to unity.

4.2 Thermodynamic Parameters

The Gibbs distribution is fundamental to statistical mechanics and thermodynamics. Once the partition function is constructed from the Gibbs distribution, all thermodynamic parameters and their relationships can be deduced from there. Let us define the Helmholtz free energy F as

$$F = -k_B T \ln Q = -k_B T \ln \left(\sum_n e^{\frac{-E_n}{k_B T}} \right), \tag{4.10}$$

where the partition function Q is from the Gibbs distribution defined in Eq. (4.8), and also related to the Helmholtz free energy through Eq. (4.10), i.e.,

$$Q = \sum_n e^{\frac{-E_n}{k_B T}} = e^{\frac{-F}{k_B T}}. \tag{4.11}$$

The entropy S is defined as

$$S = -k_B \sum_n P_n \cdot \ln P_n, \tag{4.12}$$

where the probability to find the system at energy level E_n is

$$P_n = \frac{1}{Q} \cdot e^{\frac{-E_n}{k_B T}}. \tag{4.13}$$

The expectation value of energy E_n is $< E_n >= E_n \cdot P_n$, and the averaged energy of the system is the sum of all energy expectation values, i.e.,

$$\begin{aligned} U &= \frac{1}{Q} \cdot \sum_n E_n \cdot e^{\frac{-E_n}{k_B T}} \\ &= \frac{k_B T^2}{Q} \cdot \frac{1}{k_B T^2} \cdot \sum_n E_n \cdot e^{\frac{-E_n}{k_B T}} \\ &= \frac{k_B T^2}{Q} \cdot \left(\frac{\partial Q}{\partial T} \right) \\ &= -T^2 \cdot \frac{\partial}{\partial T} \left(\frac{F}{T} \right) \\ &= F - T \cdot \frac{\partial F}{\partial T} \\ &= \frac{\partial}{\partial \frac{1}{T}} \left(\frac{F}{T} \right). \end{aligned} \tag{4.14}$$

The entropy is

$$\begin{aligned} S &= -k_B \sum_n P_n \cdot \ln P_n \\ &= -k_B \cdot \sum_n \frac{1}{Q} \cdot e^{\frac{-E_n}{k_B T}} \left(\frac{-E_n}{k_B T} - \ln Q \right) \\ &= -\frac{\partial F}{\partial T}. \end{aligned} \tag{4.15}$$

We have the averaged total energy, or the internal energy, given by

$$U = F + TS. \tag{4.16}$$

The pressure at constant temperature is defined as

$$P = -(\frac{\partial F}{\partial V})_T = -\frac{1}{Q} \cdot \sum_n e^{\frac{-E_n}{k_B T}} \cdot \frac{\partial E_n}{\partial V}. \tag{4.17}$$

4.3 Statistical Quantities for Phonons — Free Energy, Partition Function, Average Energy, Gibbs Distribution, and the Average Occupation Number

A crystalline solid at a given temperature can be considered as a collection of phonons in thermal equilibrium. For example, there are $3N$ acoustic phonon modes in a sample of N atoms. Each phonon mode is characterized by the wave vector $\mathbf{k}$, frequency $\omega_{\mathbf{k}}$, and a discrete energy spectrum of $E_{\mathbf{k},n} = (n + \frac{1}{2})\hbar\omega_{\mathbf{k}}$. The quantum number n, when specified with the quantum number $\mathbf{k}$, is interpreted as the number of phonons in the $\mathbf{k}$ mode. Each phonon has an energy of $\hbar\omega_{\mathbf{k}}$, and there could be $0, 1, 2, \cdots$,or n such phonons in the mode. Each of the phonon energy levels has a finite chance to be occupied according to the Gibbs distribution $P(E_{\mathbf{k},n})$ depending on the temperature T and energy $n\hbar\omega_{\mathbf{k}}$. Since the phonon modes are independent from each other in the harmonic approximation, they can be considered as independent sub-systems in thermal equilibrium with one another. Therefore, each mode can be treated individually. For instance, we can calculate the energy of each mode first, and add them up to get the energy of the entire sample. An equivalent view is to treat all phonon modes together as one quantum system with an energy spectrum of $\{E_n\}$. The n^{th} quantum level E_n must be the summation of the n^{th} level energies from all modes, and we have $E_n = \sum_{\mathbf{k}} E_{\mathbf{k},n}$. To see how this integrated view can

be decomposed into individual modes, let us deduce from the Helmholtz free energy of the whole system, i.e.,

$$\begin{aligned} F &= -k_B T \ln Q \\ &= -k_B T \ln \left(\sum_n e^{-\beta E_n} \right) \\ &= -k_B T \ln \left(\sum_n e^{-\beta \sum_{\mathbf{k}} E_{\mathbf{k},n}} \right) \\ &= -k_B T \ln \prod_{\mathbf{k}} \left(\sum_n e^{-\beta E_{\mathbf{k},n}} \right) \\ &= \sum_{\mathbf{k}} -k_B T \ln \left(\sum_n e^{-\beta E_{\mathbf{k},n}} \right) \\ &= \sum_{\mathbf{k}} F_{\mathbf{k}}. \end{aligned} \tag{4.18}$$

We come to the conclusion we have already known intuitively. The free energy of the whole system is indeed the sum of free energies of all modes. Now, let us concentrate on one particular mode $\mathbf{k}$. The partition function is

$$\begin{aligned} Q_{\mathbf{k}} &= \sum_n e^{\frac{-E_{\mathbf{k},n}}{k_B T}} \\ &= \sum_n e^{-(n+\frac{1}{2})\frac{\hbar\omega_{\mathbf{k}}}{k_B T}} \\ &= e^{\frac{-\hbar\omega_{\mathbf{k}}}{2k_B T}} \cdot \left(1 + e^{\frac{-\hbar\omega_{\mathbf{k}}}{k_B T}} + e^{\frac{-2\hbar\omega_{\mathbf{k}}}{k_B T}} + \cdots \right) \\ &= \frac{e^{\frac{-\hbar\omega_{\mathbf{k}}}{2k_B T}}}{1 - e^{\frac{-\hbar\omega_{\mathbf{k}}}{k_B T}}}. \end{aligned} \tag{4.19}$$

The free energy is

$$\begin{aligned} F_{\mathbf{k}} &= -k_B T \ln Q_{\mathbf{k}} \\ &= \frac{\hbar\omega_{\mathbf{k}}}{2} + k_B T \cdot \ln \left(1 - e^{\frac{-\hbar\omega_{\mathbf{k}}}{k_B T}} \right). \end{aligned} \tag{4.20}$$

The average energy of mode $\mathbf{k}$ in thermal equilibrium is

$$\begin{aligned} U_{\mathbf{k}} &= \frac{1}{Q_{\mathbf{k}}} \cdot \sum_n \left(E_{\mathbf{k},n} \cdot e^{\frac{-E_{\mathbf{k},n}}{k_B T}} \right) \\ &= \frac{\partial}{\partial(\frac{1}{T})}(\frac{F_{\mathbf{k}}}{T}) \\ &= \frac{\hbar\omega_{\mathbf{k}}}{2} + \hbar\omega_{\mathbf{k}} \cdot \frac{e^{\frac{-\hbar\omega_{\mathbf{k}}}{k_B T}}}{1 - e^{\frac{-\hbar\omega_{\mathbf{k}}}{k_B T}}} \\ &= \frac{\hbar\omega_{\mathbf{k}}}{2} + \frac{\hbar\omega_{\mathbf{k}}}{e^{\frac{\hbar\omega_{\mathbf{k}}}{k_B T}} - 1}. \end{aligned} \tag{4.21}$$

The probability to find the mode $\mathbf{k}$ having n phonons is

$$\begin{aligned} P_{\mathbf{k},n} &= \frac{1}{Q_{\mathbf{k}}} \cdot e^{\frac{-E_{\mathbf{k},n}}{k_B T}} \\ &= (\frac{1}{Q_{\mathbf{k}}}) \cdot e^{-(n+\frac{1}{2})\frac{\hbar\omega_{\mathbf{k}}}{k_B T}}. \end{aligned} \tag{4.22}$$

This is just the Gibbs Distribution for this mode. Because there is a finite probability for the phonon mode to have $0, 1, 2, \cdots$,or n phonons, it is useful to introduce the statistical expectation value of phonon occupation number, i.e., $n \cdot P_{\mathbf{k},n}$, the number multiplied by the probability. Let us define an average phonon occupation number $< n_{\mathbf{k}} >$ for mode $\mathbf{k}$ by requiring

$$(< n_{\mathbf{k}} > + \frac{1}{2}) \cdot \hbar\omega_{\mathbf{k}} = U_{\mathbf{k}}. \tag{4.23}$$

We have

$$< n_{\mathbf{k}} > = \frac{1}{e^{\frac{\hbar\omega_{\mathbf{k}}}{k_B T}} - 1}. \tag{4.24}$$

With temperature as a parameter, we can generate a set of curves where the average phonon occupation number $< n_{\mathbf{k}} >$ is plotted as a function of $\hbar\omega_{\mathbf{k}}$. The functional dependence is nothing but the same as Planck distribution for blackbody radiation. This is not a surprising outcome, since phonons are particle-like energy excitations of the quantized lattice wave, just like photons to the electromagnetic wave. Both are in the harmonic realm. The total free energy of all phonon modes is

$$F = \sum_{\mathbf{k}} F_{\mathbf{k}}$$

$$= \sum_{\mathbf{k}} \left(\frac{\hbar\omega_{\mathbf{k}}}{2} + k_B T \cdot \ln\left(1 - e^{\frac{-\hbar\omega_{\mathbf{k}}}{k_B T}}\right)\right), \tag{4.25}$$

and the total averaged internal energy is

$$\begin{aligned} U &= \sum_{\mathbf{k}} U_{\mathbf{k}} \\ &= \sum_{\mathbf{k}} \left(\frac{\hbar\omega_{\mathbf{k}}}{2} + \frac{\hbar\omega_{\mathbf{k}}}{e^{\frac{\hbar\omega_{\mathbf{k}}}{k_B T}} - 1} \right). \end{aligned} \tag{4.26}$$

Lattice Thermal Properties

4.4 Phonon Occupation Number — A Revisit

Lattice vibrations and electron movements are both responsible for thermal properties in conductors. In insulators, the only important contribution is from the lattice because there are no free electrons. We will discuss some lattice thermal properties common to all crystalline solids. Starting with a single phonon mode of wave vector $\mathbf{k}$ and energy $\hbar\omega_{\mathbf{k}}$, we have already learned that the probability for having n phonons is

$$\frac{e^{\frac{-n\hbar\omega_{\mathbf{k}}}{k_B T}}}{\sum_{s=0}^{\infty} e^{\frac{-s\hbar\omega_{\mathbf{k}}}{k_B T}}}. \tag{4.27}$$

To multiply the number n by its corresponding probability gives the expectation value of n. This is similar to gambling with dice. If the chance to win 12 dollars is $1/6$, what we can expect to win, after a large number of trials, approaches 2 dollars per game in average. Therefore, the expectation value of winning 12 dollars is just 2 dollars. Since it is possible to have $0, 1, 2, \cdots$, or n phonons in the phonon mode, the expectation value of each phonon number will add up to the averaged phonon occupation number, i.e.,

$$< n_{\mathbf{k}} > = \frac{\sum_{s=0}^{\infty} s \cdot e^{-\frac{s\hbar\omega_{\mathbf{k}}}{k_B T}}}{\sum_{s=0}^{\infty} e^{\frac{-s\hbar\omega_{\mathbf{k}}}{k_B T}}}.$$

Let $x = e^{-\frac{\hbar\omega_{\mathbf{k}}}{k_B T}}$, and we have

$$< n_{\mathbf{k}} > = \frac{\sum_{s=0}^{\infty} s \cdot x^s}{\sum_{s=0}^{\infty} x^s} = \frac{x\frac{d}{dx}\sum_{s=0}^{\infty} x^s}{\frac{1}{1-x}} = \frac{\frac{x}{(1-x)^2}}{\frac{1}{1-x}} = \frac{x}{1-x};$$

$$< n_{\mathbf{k}} > = \frac{1}{e^{\frac{\hbar\omega_{\mathbf{k}}}{k_B T}} - 1}. \tag{4.28}$$

At temperature T, we can expect to have $< n_{\mathbf{k}} >$ phonons with energy $\hbar\omega_{\mathbf{k}}$. It is an averaged number. Namely, if we take many measurements to determine the number of phonons, the average would turn out to be $< n_{\mathbf{k}} >$. For the phonon mode of energy $\hbar\omega_{\mathbf{k}}$ at temperature T, $< n_{\mathbf{k}} >$ is the averaged phonon occupation number, and $(< n_{\mathbf{k}} > + \frac{1}{2})\hbar\omega_{\mathbf{k}}$ is the averaged energy. This result is identical to what we deduced previously from the energy point of view.

4.5 Einstein Model for Lattice Heat Capacity

The heat capacity of a sample is the amount of heat energy it takes to raise the sample temperature by one degree. The heat capacity is usually meant to be the heat capacity at constant volume, which is more fundamental than the heat capacity at constant pressure. The difference is negligible in solids, so the constant-volume condition is usually not enforced in experimental measurements. The specific heat is the heat capacity of one unit mass or one mole of the substance.

Einstein model of lattice heat capacity assumes that all lattice vibration modes are at the same frequency. The assumption is an obvious oversimplification. However, the idea is to demonstrate that the experimental heat capacity of insulators can be counted for qualitatively in both high and low temperature limits, if the lattice vibration is represented by phonons. In Einstein model, there are $3N$ phonon modes counting for 2 transverse and 1 longitudinal lattice waves of N atoms. All modes are assumed to vibrate at the frequency ω. The corresponding phonon energy is $\hbar\omega$, and the average phonon occupation number is $< n >$. The total energy in the phonon representation is

$$U = 3N < n > \hbar\omega = \frac{3N\hbar\omega}{e^{\frac{\hbar\omega}{k_B T}} - 1}. \tag{4.29}$$

The heat capacity at constant volume is

$$C_v = \left(\frac{\partial U}{\partial T}\right)_v = 3Nk_B \left(\frac{\hbar\omega}{k_B T}\right)^2 \frac{e^{\frac{\hbar\omega}{k_B T}}}{\left(e^{\frac{\hbar\omega}{k_B T}} - 1\right)^2}. \tag{4.30}$$

Now, we take two limiting cases at high and low temperatures.
For high T:

$$\frac{\hbar\omega}{k_B T} \ll 1, \qquad C_v \approx 3Nk_B \left(\frac{\hbar\omega}{k_B T}\right)^2 \cdot \frac{1}{\left(\frac{\hbar\omega}{k_B T}\right)^2} \approx 3Nk_B. \tag{4.31}$$

For low T:

$$\frac{\hbar\omega}{k_B T} \gg 1, \qquad C_v \approx 3Nk_B \left(\frac{\hbar\omega}{k_B T}\right)^2 e^{-\frac{\hbar\omega}{k_B T}} \approx e^{-\frac{\hbar\omega}{k_B T}}. \tag{4.32}$$

The Einstein heat capacity approaches a constant equal to $3Nk_B$ at high temperatures, known as the Dulong-Petit law. It goes to zero as the temperature goes to zero. This is what experiments measure on all solids. Both high and low temperature limits are accounted for simply by assuming that all thermal vibration modes are represented by phonons of a single energy $\hbar\omega_k$. However, Einstein model predicted the wrong temperature dependence for the heat capacity at low temperatures. The measured temperature dependence is $C_v \sim T^3$, but Einstein model predicts $C_v \sim \exp\left(\frac{-\hbar\omega}{k_B T}\right)$. This discrepancy comes from the oversimplified assumption of a single phonon frequency for all vibration modes. The improved Debye model corrects for it.

4.6 Density of States

In the one-dimensional lattice with N mobile atoms, there are N allowed k states. For each k state, the vibration can always be decomposed into 3 polarizations, one longitudinal and two transverse ones. So, there are $3N$ acoustic phonon modes in thermal equilibrium. Each mode is characterized by a well-defined wave vector k, polarization p, and frequency $\omega_{k,p}$. Considering one polarization at a time, we may omit the polarization index, and add the averaged energies from all N k-states to get the thermal energy , i.e.,

$$U = \sum_k < n_k > \hbar\omega_k. \tag{4.33}$$

The average number of phonons in the k mode, $< n_k >$, is a function of ω_k, and ω_k is a function of k. When N is very large, which is usually true for the sample size we deal with, the spacing between adjacent k states becomes very small. There will be numerous k states within an incremental Δk. So, both ω_k and $< n_k >$ can be considered as continuous functions of k. We may replace ω_k with the dispersion function $\omega(k)$, and $< n_k >$ with the phonon distribution function $n(\omega(k), T)$. If we know the density of phonon modes in k space, i.e., the number of k states per unit range of k, we can calculate the number of k states in Δk from $\rho(k) \cdot \Delta k$, where $\rho(k)$ is the said density of states. In the limit as Δk shrinks to the infinitesimal dk, we have the differential thermal energy $dU = \hbar\omega(k) \cdot n(\omega(k), T) \cdot \rho(k)dk$, and the summation in Eq. (4.33) can be converted into an integration over k. The mathematics would be pretty messy if we actually substitute the dispersion function $\omega(k)$ into the integrand. It is much easier if we can hide the explicit dependence on k by changing the integration variable from k to ω. Replacing dk with $(dk/d\omega) \cdot d\omega$, we have $dU = \hbar\omega \cdot n(\omega, T) \cdot \rho(k) \cdot dk/d\omega \cdot d\omega$. The product $\rho(k) \cdot dk/d\omega$ measures how many k states per unit frequency, and by definition it is the density of states $D(\omega)$ at the frequency ω. So, the key here is to find $D(\omega)$, i.e., to find both $\rho(k)$ and $d\omega/dk$. The former can be obtained by counting the k states, and the latter is just the group velocity. It turns out that we can indeed calculate U through the much easier integration over ω , i.e.,

$$U = \int \hbar\omega D(\omega) n(\omega, T) d\omega. \tag{4.34}$$

The meaning of this integral is quite clear. In the phonon frequency range from ω to $\omega + d\omega$, there are $D(\omega)d\omega$ phonon modes and $n(\omega, T)$ phonons per mode. Since all these phonons have the same energy $\hbar\omega$, the differential thermal energy is simply $dU = \hbar\omega \cdot n(\omega, T) \cdot D(\omega)d\omega$. Obviously, the integration will yield the total thermal energy for one polarization. In general, the dispersion function $\omega(k)$ depends somewhat on the polarization. We need first to calculate U from Eq. (4.34) individually for each polarization and add them up in the final step. For most estimate purposes, to multiply U of Eq. (4.34) by 3 will be approximately justified. The only temperature dependence in U is from the phonon distribution $n(\omega, T)$. Hence, the specific heat $(\frac{dU}{dT})_v$ will be determined by $\frac{dn(\omega,T)}{dT}$. In this approach to calculate the total thermal energy and specific heat, we need to find the density-of-states function $D(\omega)$ first, i.e., how many phonon modes per unit energy range for each polarization.

Let us begin with the one-dimensional lattice that satisfies the periodic boundary condition over a segment of N lattice points. We have already learned previously that there are N independent k values, i.e.,

$$k = \frac{2n\pi}{L} \qquad \text{with } n = 0, \pm 1, \pm 2, \cdots, \pm\frac{N}{2}. \tag{4.35}$$

The length of the lattice vector connecting the s^{th} and $(s+N)^{th}$ sites is $L = Na$, where a is the lattice spacing. The periodic boundary condition $q(sa) = q(sa+L)$, that holds for $k = \pm\frac{\pi}{a}$, implies that N is an even number, because for an odd N we would have $q(sa) = -q(sa+L)$ for $k = \pm\frac{\pi}{a}$. The total number of modes appears to be $2 \cdot (\frac{N}{2}) + 1$ by counting all the possible n values. However, $k = \frac{\pi}{a}$ and $k = -\frac{\pi}{a}$ are identical standing-wave modes because $\exp(is\pi) = \exp(-is\pi) = \cos(\pm s\pi)$. Therefore, we have N independent modes including the uniform mode at $k = 0$, and a standing wave at $k = \pm\frac{\pi}{a}$. At the first glance, we might count from s to $s+N$ a total of $N+1$ mobile atoms, but the periodic condition implies that the s^{th} and $(s+N)^{th}$ atoms are identical. Therefore, for N mobile atoms we have N independent modes per polarization, all in the first Brillouin zone $(-\frac{\pi}{a} \leq k \leq \frac{\pi}{a})$ with equal spacing of $\Delta k = \frac{2\pi}{L}$. The density of states is simply equal to N divided by the width of the zone, i.e., $\rho(k) = N/(\frac{2\pi}{a}) = \frac{L}{2\pi}$. An alternative method is to count one mode per interval $\Delta k = \frac{2\pi}{L}$, which leads to the same result $\rho(k) = \frac{1}{k} = \frac{L}{2\pi}$. The number of modes in an infinitesimal range of dk is $(\frac{L}{2\pi})dk$. Replacing dk with $(\frac{dk}{d\omega})d\omega$, we have

$$\frac{L}{2\pi} \cdot dk = \frac{L}{2\pi}\frac{dk}{d\omega}d\omega. \tag{4.36}$$

The quantity $(\frac{L}{2\pi})(\frac{dk}{d\omega}) = \rho(k)(\frac{dk}{d\omega})$ obviously measures how many k states in a unit frequency range. By definition, it is the density of states $D(\omega)$. Noting that the derivative $d\omega/dk$ of the dispersion function is the group velocity v_g, we have

$$\begin{aligned} D(\omega) &= \rho(k)\frac{dk}{d\omega} \\ &= \frac{L}{2\pi}\frac{dk}{d\omega} = \frac{L}{2\pi} \cdot \frac{1}{v_g}. \end{aligned} \tag{4.37}$$

From the dispersion curve and group velocity discussed in this chapter earlier, we know that v_g decreases monotonically with increasing ω . For small ω values, v_g is approximately a constant. Near the first Brillouin zone

boundary, v_g decreases rapidly and equals zero on the boundary. Therefore, $D(\omega)$ increases with ω, and approaches infinity at the zone boundary. The group velocity is in general different for different polarizations, and the density of states in Eq. (4.37) is for one polarization.

Now, we can expand our discussion to the 3-dimensional lattice. Taking the simple cubic lattice for an example because its symmetry is the easiest to visualize, we suppose that the lattice is periodical over a length of L along all three cubic axes. Let N be the number of atoms in the volume L^3, and a be the lattice constant, we have $L = N^{1/3} \cdot a$. The periodic condition requires

$$e^{i(k_x x + k_y y + k_z z)} = e^{i(k_x(x+L) + k_y(y+L) + k_z(z+L))}. \tag{4.38}$$

This leads to $k_x L = 2n_x\pi$, $k_y L = 2n_y\pi$, and $k_z L = 2n_z\pi$. So, we have

$$k_x = 0, \pm\frac{2\pi}{L} \pm \frac{4\pi}{L}, \cdots, \frac{N^{1/3}\pi}{L},$$

$$k_y = 0, \pm\frac{2\pi}{L} \pm \frac{4\pi}{L}, \cdots, \frac{N^{1/3}\pi}{L},$$

and

$$k_z = 0, \pm\frac{2\pi}{L} \pm \frac{4\pi}{L}, \cdots, \frac{N^{1/3}\pi}{L}. \tag{4.39}$$

Equation (4.39) defines a cubical network of N points uniformly distributed inside the first Brillouin zone of volume $\left(\frac{2\pi}{a}\right)^3$. The vector pointing from $\mathbf{k} = 0$ to any of the N points is a valid $\mathbf{k}$ state. For the basic cubical cell of $\left(\frac{2\pi}{L}\right)^3$, there is one $\mathbf{k}$ state on each of its 8 corners. Since each corner is shared among 8 such cubicles, there is one $\mathbf{k}$ state in the cubicle of $\left(\frac{2\pi}{L}\right)^3$. The density of states $\rho(\mathbf{k})$ is simply $1/\left(\frac{2\pi}{L}\right)^3$, i.e., $\left(\frac{L}{2\pi}\right)^3$ states per unit volume in $\mathbf{k}$ space. Counting the modes like we did in the one-dimension case, we end up again with N independent modes per polarization for N mobile atoms, and there are three polarizations. These results are true in general. They are easily recognizable in the simple cubic lattice, but actual symmetry and atomic basis needs to be considered in more detail for other lattice types.

The dispersion curve of the one-dimensional lattice shows that $\omega(k) =$ constant corresponds to two points in k space. In three dimensions, there will be a network of $\mathbf{k}$ points that satisfy $\omega(k_x, k_y, k_z) =$ constant. As the basic cubicle $\left(\frac{2\pi}{L}\right)^3$ shrinks to infinitesimal size in the limit of very large N,

the **k** points of constant ω becomes so numerous that they actually form a surface, i.e., the constant frequency surface. The number of k states located inside the shell between ω and $\omega + d\omega$ surfaces is simply the integration of $\rho(k) \cdot d^3k$ over this shell. It is also by definition equal to the density-of-states $D(\omega)$ multiplied by $d\omega$. So, we have

$$D(\omega)d\omega = \left(\frac{L}{2\pi}\right)^3 \int_{\text{shell}} d^3k = \left(\frac{L}{2\pi}\right)^3 \int_{\text{shell}} dS_\omega dk_\perp. \tag{4.40}$$

The differential dS_ω is an infinitesimal area on the constant frequency surface, and $dk_\perp$ is the differential **k** vector perpendicular to dS_ω. Noting that the derivative $\frac{d\omega}{dk_\perp}$ is the value of the gradient vector of ω, we may substitute $\frac{d\omega}{|\nabla_{\mathbf{k}}\omega|}$ for $dk_\perp$ and convert the shell integral to a surface integral, i.e.,

$$D(\omega)d\omega = \left(\frac{L}{2\pi}\right)^3 \int \frac{dS_\omega}{|\nabla_{\mathbf{k}}\omega|} \cdot d\omega = \left(\frac{L}{2\pi}\right)^3 \int \frac{dS_\omega}{v_g} \cdot d\omega. \tag{4.41}$$

Therefore, the density of states $D(\omega)$ per unit frequency in a volume $V(= L^3)$ is

$$D(\omega)d\omega = \frac{V}{(2\pi)^3} \int \frac{dS_\omega}{v_g}. \tag{4.42}$$

The integration is over a constant frequency surface in k space. The shape of the surface depends on lattice symmetry, and the group velocity is a function of **k** and polarization. It must be tedious to say the least, although possible in principle, to calculate $D(\omega)$ numerically for a given lattice. There are also methods that can extract $D(\omega)$ from experimental data. But, our goal here is to learn about the general common property of $D(\omega)$ in Debye approximation, which is independent from specific lattice structures, so that we can gain basic insights and intuitions about thermal phonons without tangling with tedious mathematics.

4.7 Debye Density of States, Debye Cutoff Frequency, and Debye Cutoff Wave Vector

Let us consider one polarization first. In Debye approximation, also known as long-wavelength or continuum approximation, the dispersion relation is reduced to a simple isotropic linear relation $\omega = v \cdot k$ because the wavelength is much longer than the lattice constant ($\lambda \gg a$). The constant

frequency surface becomes a spherical surface of radius k. The density of states reduces to

$$D(\omega) = \frac{V}{(2\pi)^3} \int \frac{dS_\omega}{v_g} = \frac{V}{(2\pi)^3} \cdot \frac{4\pi k^2 \cdot}{v} = \frac{V}{(2\pi)^3} \cdot \frac{4\pi \cdot \omega^2}{v^3}. \tag{4.43}$$

The total states inside the sphere of radius k is

$$\int D(\omega) d\omega = \frac{V}{(2\pi)^3} \int \frac{4\pi \cdot k^2}{v} \cdot v \cdot dk = \frac{V}{(2\pi)^3} \cdot \frac{4\pi}{3} \cdot k^3. \tag{4.44}$$

This is intuitively correct because the total states should equal to the density per unit volume multiplied by the volume of the sphere. Applying the constraint that the number of allowed modes must equal the number of mobile atoms N, we introduce the Debye cutoff wave-vector k_D and frequency ω_D by requiring

$$N = \frac{V}{6\pi^2} \cdot k_D^3 = \frac{V}{6\pi^2} \cdot \left(\frac{\omega_D}{v}\right)^3. \tag{4.45}$$

The Debye cutoff frequency and Debye cutoff wave-vector are respectively defined as

$$\omega_D = \left(\frac{6\pi^2 v^3 N}{V}\right)^{\frac{1}{3}} \qquad \text{and} \qquad k_D = \left(\frac{6\pi^2 N}{V}\right)^{\frac{1}{3}}. \tag{4.46}$$

In Debye model, we have $k \leq k_D$ and $\omega \leq \omega_D$. The Debye density-of-states is proportional to ω^2 and cuts off at the Debye frequency ω_D. Naturally, the above discussion should include all three polarizations. In Debye's long wavelength approximation, the velocity is considered not only independent of propagation directions, but also independent of polarizations. All we need to do is to multiply $D(\omega)$ by 3. Without much extra work, we can expand Debye's approach by including three isotropic velocities to count for the polarization differences. Denoting v_1 and v_2 as two transverse velocities and v_3 as the longitudinal one, we can simply write down the total Debye density-of-states as the sum of that of all three polarizations, i.e.,

$$D(\omega) = \frac{V}{2\pi^2} \cdot \left(\frac{\omega^2}{v_1^3} + \frac{\omega^2}{v_2^3} + \frac{\omega^2}{v_3^3}\right). \tag{4.47}$$

Each polarization has its own Debye cutoff frequency obtainable by substituting the corresponding velocity for v in $\omega_D = (\frac{6\pi^2 v^3 N}{V})^{1/3}$. For the hypothetical ratio of $v_1 : v_2 : v_3 = 1 : 1.3 : 1.6$, the composite $D(\omega)$ of Eq. (4.47) is plotted against ω in Fig. 1, showing three polarization peaks.

This exercise is simple but insightful because all three polarization peaks are indeed resolved in real crystals.

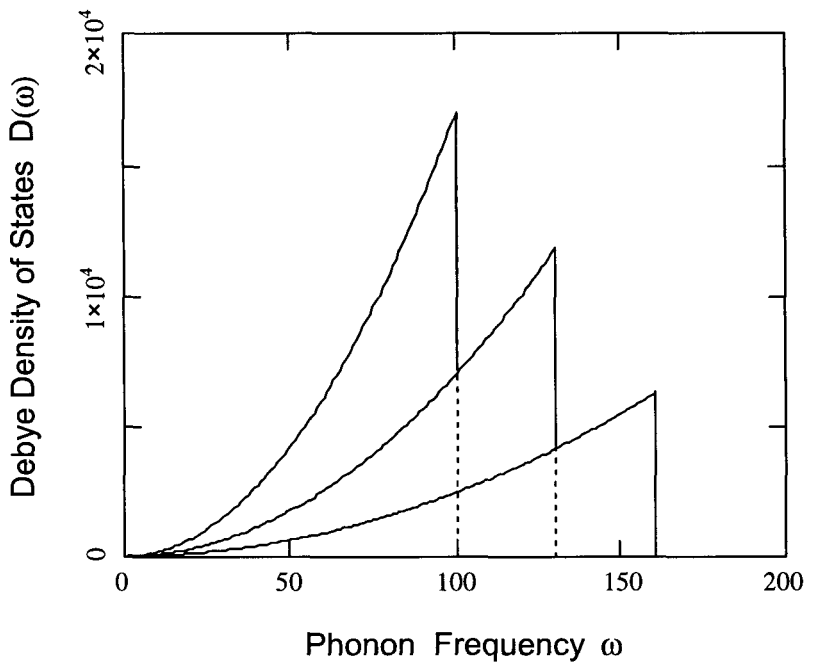


Fig. 4.1 Composite Debye density of states $D(\omega)$ of all three polarizations. The density of states for each polarization is proportional to ω^2/v^3, where the group velocity $v = v_1, v_2$, or v_3, respectively, as described in Eq. (4.47). Both $D(\omega)$ and ω are in arbitrary units, plotted according to the ratio $v_1 : v_2 : v_3 = 1 : 1.3 : 1.6$. The cutoff frequencies are set at ω_D, $\omega_D(v_2/v_1)$, and $\omega_D(v_3/v_1)$, respectively, and ω_D is arbitrarily set to equal to 100. The first peak is from all three polarizations. The second is from two, and the last is from only one polarization.

4.8 Van Hove Singularity and Phonon Density of States in Real Crystals

The Debye model works well in the long wavelength region where $\omega = vk$. For phonons of higher frequencies, the dispersion slows down the phonon group velocity and the validity of Debye model weakens. In the extreme, for

example, $D(\omega)$ goes to infinity as the constant frequency surface reaches the first Brillouin zone boundary where $v_g = 0$. Varying with crystal structures, there are many such singular points on the $D(\omega)$-versus-ω curve. They are the Van Hove singularities.

Various experimental methods can reveal the phonon density of states in crystals. The superconducting tunneling experiment is a very interesting one. Since the electron-phonon interaction is responsible for superconductivity in BCS superconductors, the electron-phonon coupling strength α and the phonon density-of-states $D(\omega)$ both play crucial roles in determining the superconducting transition temperature. It turns out that the second derivative of superconducting tunneling current directly reveals $\alpha^2 D(\omega)$. Fig. 4.2 shows the phonon spectrum $\alpha^2 D(\omega)$ obtained from superconducting tunnel junctions made of lanthanum (La). Two transverse peaks, one longitudinal peak, and Van Hove singularities are seen. This piece of data is taken as strong evidence that supports the electron-phonon interaction as the primary mechanism responsible for superconductivity in La.

4.9 Debye Model of the Lattice Heat Capacity and Debye Temperature

If we know the density-of-states $D(\omega)$, we can calculate the lattice thermal energy and heat capacity from $U = \int \hbar\omega D(\omega) n(\omega, T) d\omega$ because $n(\omega, T)$ is already known. Ideally, we should be able to calculate the exact $D(\omega)$ from Eq. (4.42), but obviously it is not an easy task. The next best is to make simplified models for $D(\omega)$. Einstein assumed $D(\omega) \propto \delta(\omega - \omega_0)$, and had some qualitative successes. Debye took the isotropic dispersion relation $\omega = v \cdot k$ from long wavelength phonons, and assumed its validity up to certain cutoff frequency ω_D in order to count for all phonon modes, i.e.,

$$D(\omega_D) = \frac{V\omega^2}{2\pi^2 v^3}, \qquad \text{and} \qquad \omega_D = \left(\frac{6\pi^2 v^3 N}{V}\right)^{\frac{1}{3}}. \tag{4.48}$$

Despite of its apparent over simplification for short wavelength phonons, Debye's model $D(\omega)$ is very successful in predicting lattice heat capacities, both qualitatively and quantitatively. It is especially valuable for making quick estimates when we know little about the real $D(\omega)$. Debye model also assumes the same group velocity for all three polarizations. Therefore,

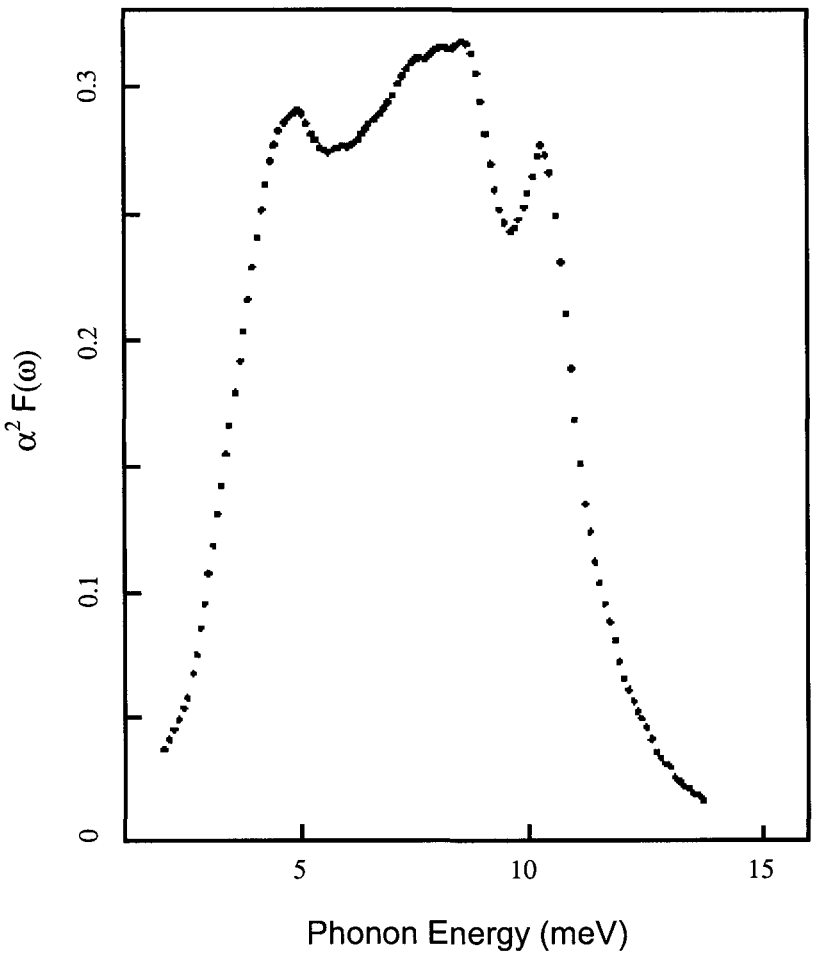


Fig. 4.2 The phonon spectrum of La, $\alpha^2 F(\omega)$, is plotted against the phonon energy $E(=\hbar\omega)$. The data is extracted from the second derivative of tunneling current flowing across superconducting tunnel junctions made of *La*. The *La* film is characterized by the energy gap $\Delta = 0.81meV$, the transition temperature $T_c = 4.9$ K, and the Debye temperature $k_B\theta = 12.2meV$. From L. F. Lou and W. J. Tomasch, Phys. Rev. Lett. 29, 858 (1972).

the total phonon energy of the acoustic phonon branch is

$$U = 3\int_0^{\omega_D} d\omega \left(\frac{V\omega^2}{2\pi^2 v^3}\right)\left(\frac{\hbar\omega}{e^{\frac{\hbar\omega}{k_B T}} - 1}\right)$$

$$= \frac{3V\hbar}{2\pi^2 v^3} \int_0^{\omega_D} d\omega \left(\frac{\omega^3}{e^{\frac{\hbar\omega}{k_B T}} - 1} \right). \tag{4.49}$$

The heat capacity can be calculated by taking the derivative dU/dT, and we have

$$C_v = \frac{dU}{dT} = \frac{3V\hbar^2}{2\pi^2 v^3 k_B T^2} \int_0^{\omega_D} d\omega \frac{\omega^4 \cdot e^{\frac{\hbar\omega}{k_B T}}}{\left(e^{\frac{\hbar\omega}{k_B T}} - 1\right)^2}. \tag{4.50}$$

Let

$$x = \frac{\hbar\omega}{k_B T}; \qquad x_D = \frac{\hbar\omega_D}{k_B T} = \frac{\theta}{T}, \tag{4.51}$$

where the parameter θ is the Debye temperature defined as

$$\theta = \frac{\hbar\omega_D}{k_B} = \frac{\hbar v}{k_B} \left(\frac{6\pi^2 N}{V} \right)^{\frac{1}{3}}. \tag{4.52}$$

In terms of the reduced parameters x, x_D, and θ, the thermal energy U and heat capacity $C_v (= dU/dT)$ become

$$U = \frac{3Vk_B^4 T^4}{2\pi^2 v^3 \hbar^3} \int_0^{x_D} dx \frac{x^3}{e^x - 1} = 9N(k_B T) \left(\frac{T}{\theta} \right)^3 \int_0^{x_D} dx \frac{x^3}{e^x - 1}, \tag{4.53}$$

and

$$C_v = 9Nk_B \left(\frac{T}{\theta} \right)^3 \int_0^{x_D} dx \frac{x^4 e^x}{(e^x - 1)^2}. \tag{4.54}$$

These integrals are given in NBS Handbook of mathematical functions AMS 55, and also in Jahnke-Emde-Losch tables. When $T \gg \theta$, the heat capacity approach $3Nk_B$, the Dulong-Petit law. At low temperatures where $T \ll \theta$, the integral upper limit goes to infinity, and we have

$$\int_0^{\infty} dx \frac{x^3}{e^x - 1} = \int_0^{\infty} dx \cdot x^3 \sum_{s=1}^{\infty} e^{-sx} = \sum_{s=1}^{\infty} \int_0^{\infty} dx \cdot x^3 e^{-sx}$$

$$= \sum_{s=1}^{\infty} \left[\frac{3!}{s^{3+1}} \right] = 6 \sum_{s=1}^{\infty} \frac{1}{s^4} = \frac{\pi^4}{15}, \tag{4.55}$$

The thermal energy of Eq. (4.53) becomes

$$U \approx 9N(k_B T) \left(\frac{T}{\theta} \right)^3 \left(\frac{\pi^4}{15} \right) = \frac{3}{5} \pi^4 N k_B \frac{T^4}{\theta^3}, \tag{4.56}$$

and the heat capacity can be calculated, i.e.,

$$C_v = \frac{dU}{dT} = \frac{12\pi^4}{5} N k_B \left(\frac{T}{\theta}\right)^3 = 234 N k_B \left(\frac{T}{\theta}\right)^3 . \tag{4.57}$$

This is the Debye T^3 law of heat capacity for crystalline insulators at low temperatures.

4.10 The Dominant Phonons

The total thermal energy of the lattice is calculated in Eq. (4.53), where the integrand $F(x) = x^3/(e^x - 1)$ contains all the energy dependence from the phonon energy, averaged phonon occupation number, and Debye phonon density-of-states. The pre-factor of the integration contains only constants and temperature. It is of great interest to plot $F(x)$ versus $x(= \frac{\hbar\omega}{k_B T})$ as seen in Fig. 4.3. The function $F(x)$ is dimensionless and can be viewed as the statistical weight distribution over x. The entire area A under the curve from $x = 0$ to x_D measures the contribution from all phonons to the total lattice energy. For example, we already know that $A = \frac{1}{3}(\frac{\theta}{T})^3$ when $T \gg \theta$ in the high temperature limit, and $A = \pi^4/15$ for low temperatures where $T \ll \theta$. The area ΔA under the curve over the segment $\Delta x(= \frac{\hbar\Delta\omega}{k_B T})$ is the contribution from phonons of energy $\hbar\omega$ at temperature T. Clearly, the largest ΔA is under the peak around $x \approx 3$. The phonons with energy $\hbar\omega \approx 3k_B T$ must contribute the most to the total thermal energy. They are the dominant phonons.

The dominant phonon concept is very useful in getting the order-of-magnitude estimates, and one can usually gain a good feel about the physics without detailed computations.

It is instructive to apply the dominant phonon concept to a quick estimate of lattice heat capacity. We will see how one can deduce the Debye T^3 law without much work at all. At low temperatures where $T \ll \theta$, very few phonons are getting excited to higher energies at all. In fact, the dominant phonons are around the energy of $\hbar\omega_0 \approx 3k_B T$, which is much less than $k_B\theta$. This condition put almost all the phonons in the long-wavelength realm where $\omega = v \cdot k$. So, we have

$$\hbar\omega_0 = \hbar v k_0 = 3k_B T,$$

and

$$\hbar\omega_D = \hbar v k_D = k_B\theta. \tag{4.58}$$

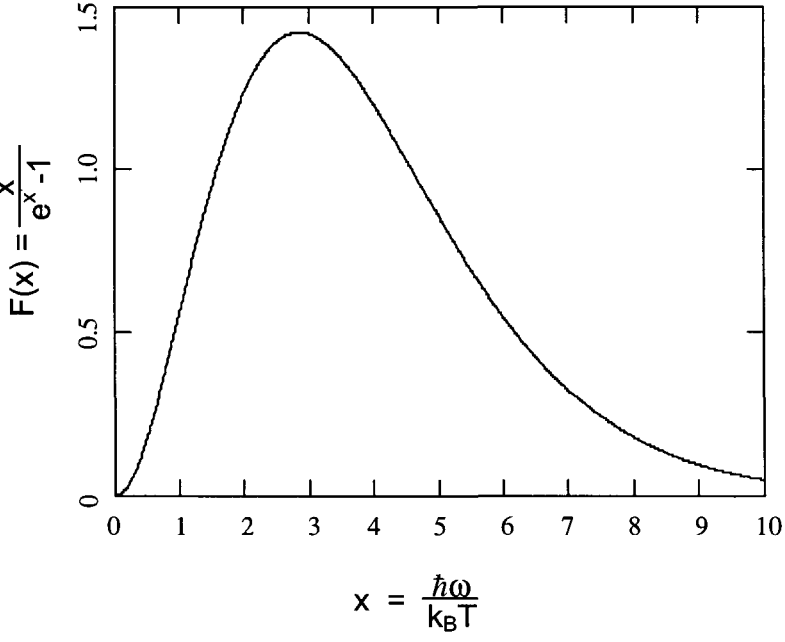


Fig. 4.3 The function $F(x) = x^3/(e^x - 1)$ is plotted versus x. The area under the curve represents the integration of $F(x)$ over x, which is the total of the statistically weighted contributions from all phonons at temperature T. The area under the peak around $x = 3$ is obviously the dominant fraction of the whole. Therefore, the dominant contribution to the total energy is from those phonons with energy $\hbar\omega \approx 3k_BT$. They are the dominant phonons. Note that the phonon population is actually the highest around $2k_BT$, where the function $x^2/(e^x - 1)$ peaks. But, the $3k_BT$ phonons dominate the energy contribution.

The phonon modes in the Debye sphere of radius k_D is only excited up to a smaller radius around k_0. The phonon population decreases rapidly for states above k_0. In the dominant phonon approximation, we will model the real phonon distribution with a bag of $3k_BT$ phonons, where the effective number of phonons Z multiplied by $3k_BT$ would be the correct total energy, i.e., $3k_BT \cdot Z = U$. Dividing Z by the number of states inside the sphere of radius k_0 including 3 polarizations, we will get the averaged number of phonons per k state per polarization, and let it be γ. If we spread all the Z phonons uniformly inside the Debye sphere, we will get $\gamma(\frac{k_0}{k_D})^3$ phonons per state, and note that $(\frac{k_0}{k_D})^3$ is equal to $27(\frac{T}{\theta})^3$ from Eq. (4.58). Now,

effectively speaking, our dominant phonon model has $3N$ states, $27\gamma(\frac{T}{\theta})^3$ phonons per state, and an energy of $3k_BT$ per phonon. Therefore, the total energy and Debye T^3 law are easily obtainable, and we have

$$U \cong 3k_BT \cdot 3N \cdot 27\gamma \left(\frac{T}{\theta}\right)^3 \Rightarrow C_v \cong 972\gamma N k_B \left(\frac{T}{\theta}\right)^3. \tag{4.59}$$

The parameter γ is determinable self consistently from its definition, i.e.,

$$\gamma = \frac{\frac{U}{3k_BT}}{3 \cdot \frac{V}{(2\pi)^3} \cdot \frac{4\pi}{3} \cdot k_0^3}. \tag{4.60}$$

Since we do not know U at the first place, we will take γ as an adjustable parameter. By setting $\gamma = 0.24$, we can actually obtain the exact same result as that derived in Eqs. (4.56) and (4.57). As a check for self-consistency, we get the same γ value by substituting U from Eq. (4.56) and k_0 from Eq. (4.58) into Eq. (4.60).

4.11 Crystal Momentum

When we heat up one end of a long iron rod, soon we will feel the heat at the other end. This is a common experience we all have, and can be generalized to almost all metals. Conduction electrons that conduct electricity in metals are therefore likely to conduct heat too. On the other hand, we have no conduction electrons in electric insulators, but many of them conduct heat anyway. In fact, some insulating single crystals of high purity can conduct heat as well as ordinary metals or even better. It shows that phonons must be involved in mediating the heat flow. It turns out that both conduction electrons and phonons are heat carriers, except that phonons may be more fundamental because they exist in all materials. Suppose the long iron rod is replaced with an electrically insulating single crystal. In essence, we effectively peel off the electron contribution and only look at phonons. The higher temperature at the hot end would cause the lattice to vibrate more violently with larger amplitudes, and create excess heat energy relative to the cold end. How does the heat energy flow toward the lower-temperature end? In other words, how does the larger amplitude decay at the hot end and how does the smaller amplitude grow at the cold end? The phonon picture comes handy for answering this question. Since the dominant phonons are around $3k_BT$, the hot end naturally has more phonons with higher energy than the cold end does. Hotter phonons will

flow toward the cold end in order to smooth out the heat distribution, so it can reach a uniform temperature. If a temperature gradient is maintained by external means, there will be a constant flow of phonons that carry the heat down stream.

The flow of heat, or flow of phonons, will eventually stop sometimes after the temperature gradient is removed. This indicates that phonons must collide with something including with each other among themselves. Otherwise, phonons will flow forever with no resistance, similar to electrons in electrical superconductors. In order to understand the collision processes, we need to know whether phonon has momentum. This is equivalent to ask whether a vibration mode carries momentum. If we know $\mathbf{k}$, $\omega_{\mathbf{k}}$, and T, both phonon property and that of the underlying lattice vibration mode are fully specified. The total atomic momentum in the vibration mode, if any, must be reflected as the phonon momentum multiplied by the number of phonons. To obtain the momentum in a vibration mode, we may calculate the momentum of each atom and sum them up. Let us take the one-dimensional lattice as an example, i.e.,

$$P = m\frac{d}{dt}\left(\sum_s q_s\right) = m\frac{d}{dt}\left(\sum_s q e^{iska}\right) = m\left(\frac{dq}{dt}\right)\cdot\frac{1-e^{iNka}}{1-e^{ika}}. \quad (4.61)$$

When $k = 0$, we have $\omega_k = vk = 0$. The atoms in this mode are not oscillating. The atoms of the entire crystal are either at rest or moving in phase with a constant speed $\frac{dq}{dt}$. Without external push at the first place, we can assume the atoms are at rest without loosing generality. So, we have $P = N \cdot m \cdot \frac{dq}{dt} = 0$ for $k = 0$. For all the other allowed states specified by $k = \pm\frac{2n\pi}{Na}$ with $n = 0, 1, \cdots, \frac{N}{2}$ inside the first Brillouin zone, it is rather easy to see $P = 0$. Since we have the net atomic momentum $P = 0$, the associated phonons can not have momentum either. Because all the $< n_k >$ phonons in one vibration mode are identical particles, if a single phonon had momentum, the total phonon momentum would have been $< n_k >$ times. The result would have contradicted $P = 0$. Therefore, phonons are massless energy excitations and they do not possess momentum.

During the collision of an external particle with the crystal, the momentum transferred may be large enough to create a phonon plus minute recoil of the crystal. The latter is equivalent to creating a uniform mode with a wave vector equal to $G = \frac{2n\pi}{a}$. From Eq. (4.61), we have $P = N \cdot m \cdot \frac{dq}{dt}$ for $k = G$. Now, $\frac{dq}{dt}$ is not zero because there is a real momentum $P(= \hbar G)$ transferred to the crystal. The crystal recoil energy is so small that rarely it

needs to be considered explicitly. But, the fact that the lattice can absorb or provide a $\mathbf{G}$ vector by such minute movement is very important in preserving the conservation of wave vectors during scattering. The quantity $\hbar\mathbf{G}$ is properly named as the crystal momentum.

In x-ray diffraction, photons are assumed to scatter from a rigid lattice, i.e., no phonons are emitted or absorbed. Since the scattering is elastic and the crystal recoil is negligible, the x-ray wave-vector gets bounced toward a different direction and its magnitude is not changed. So, the photon energy is conserved, i.e., $\hbar\omega(\mathbf{k}_o) = \hbar\omega(\mathbf{k}_i)$, provided that $\omega(\mathbf{k})$ does not have directional dependence. The diffraction obeys the von Laue law, i.e.,

$$\mathbf{k}_o = \mathbf{k}_i + \mathbf{G}. \tag{4.62}$$

The law selects the outgoing wave-vector $\mathbf{k}_o$ by adding a reciprocal lattice vector $\mathbf{G}$ to the incoming wave-vector $\mathbf{k}_i$.

The selection rule can be generalized to inelastic scattering of particles such as photons, electrons, and neutrons, et al, involving the absorption or emission of phonons. For instance, an incoming neutron of momentum $\hbar\mathbf{k}_i$ can be scattered into an outgoing neutron of momentum $\hbar\mathbf{k}_o$. If a phonon is emitted or absorbed in the scattering process, the wave vectors must satisfy the selection rule

$$\mathbf{k}_o = \mathbf{k}_i + \mathbf{k}'. \tag{4.63}$$

If the phonon wave-vector $\mathbf{k}'$ is outside the 1^{st} Brillouin zone, we should always replace $\mathbf{k}'$ with an equivalent $\mathbf{k}$ located inside the 1^{st} Brillouin zone. The mathematical procedure is simply to subtract a reciprocal lattice vector $\mathbf{G}$ from $\mathbf{k}'$. Physically, it means that the lattice prefers to transfer the extra energy and momentum to the whole crystal whenever it is allowed to. With $\mathbf{k}' = \mathbf{k} + \mathbf{G}$, the selection rule in Eq. (4.63) becomes

$$\mathbf{k}_o = \mathbf{k}_i + \mathbf{k} + \mathbf{G}. \tag{4.64}$$

Since the phonon with $\mathbf{k}'$ outside the first Brillouin zone does not exist, the scattering into the $\mathbf{k}'$ state must be understood as a composite process, namely, the recoil of the crystal as a whole in addition to the creation of one phonon. The former process involves the reciprocal lattice vector $\mathbf{G}$. The latter process creates a new phonon with $\mathbf{k}$ located inside the first Brillouin zone. Because the crystal recoil energy is negligible, the change of the neutron energy comes solely from the creation of the phonon , and

the energy of all particles is conserved, i.e.,

$$\begin{aligned}\hbar\omega(\mathbf{k}_o) &= \hbar\omega(\mathbf{k}_i) + \hbar\omega(\mathbf{k}') \\ &= \hbar\omega(\mathbf{k}_i) + \hbar\omega(\mathbf{k}+\mathbf{G}) \\ &= \hbar\omega(\mathbf{k}_i) + \hbar\omega(\mathbf{k}). \end{aligned} \tag{4.65}$$

The wave-vector selection rule in Eq. (4.64) multiplied by $\hbar$ becomes the familiar form of conservation of momentum. Therefore, the phonon behaves as if it possessed a momentum of $\hbar\mathbf{k}$. Since phonons do not have real physical momentum, the momentum-like quantity $\hbar\mathbf{k}$ is called the crystal momentum. The phonon crystal momentum, sometimes loosely referred to as phonon momentum, is merely a convenience for dealing with the scattering of phonons. What we really mean is the conservation of wave vectors.

Since the phonon frequency $\omega(\mathbf{k})$ is symmetrical when $\mathbf{k}$ is replaced by $-\mathbf{k}$, we have an equality of $< n_{\mathbf{k}} >=< n_{-\mathbf{k}} >$ for the phonon occupation number in these two modes. The total phonon momentum turns out to be zero in a region of thermal equilibrium.

$$\sum_k < n_{\mathbf{k}} > \hbar\mathbf{k} = \sum_{\mathbf{k}>0}(< n_{\mathbf{k}} > \hbar\mathbf{k} - < n_{-\mathbf{k}} > \hbar\mathbf{k}) = 0. \tag{4.66}$$

However, individual phonon gets scattered as if it had a crystal momentum $\hbar\mathbf{k}$. It should not be confused with the actual atomic momentum in crystals. Each individual atom does have real momentum, but the total momentum of all atoms always vanishes. Changing the phonon wave vector by adding or subtracting a reciprocal lattice vector does not change its vibration properties. We always have $\hbar\omega(\mathbf{k} \pm \mathbf{G}) = \hbar\omega(\mathbf{k})$.

4.12 Umclapp and Normal Processes of Phonon Scattering

So far we have only considered independent harmonic oscillators governed by the quadratic potential. The phonon modes are independent from each other and there are no phonon-phonon interactions. The higher order terms in the Taylor-series expansion of the potential, if not negligible, will cause anharmonic behaviors involving phonon-phonon interactions. For example, we all have probably experienced with thermal expansion and thermal con-

ductivity, but most of us might not have wondered very much about them. In fact, the underlying mechanism involves phonon-phonon scattering risen from anharmonic forces in the lattice. One of the basic phonon-phonon scattering processes is the annihilation of two phonons to create a third one. If temperature is high enough that both $\mathbf{k}_1$ and $\mathbf{k}_2$ are close to the first Brillouin zone boundary, the wave vector of the third phonon that they produce can go beyond the first Brillouin zone as shown in Fig.4. Since phonons do not exist outside the first Brillouin zone, the extra momentum and energy must transfer to the host crystal, so that the newly created phonon will be inside the zone.

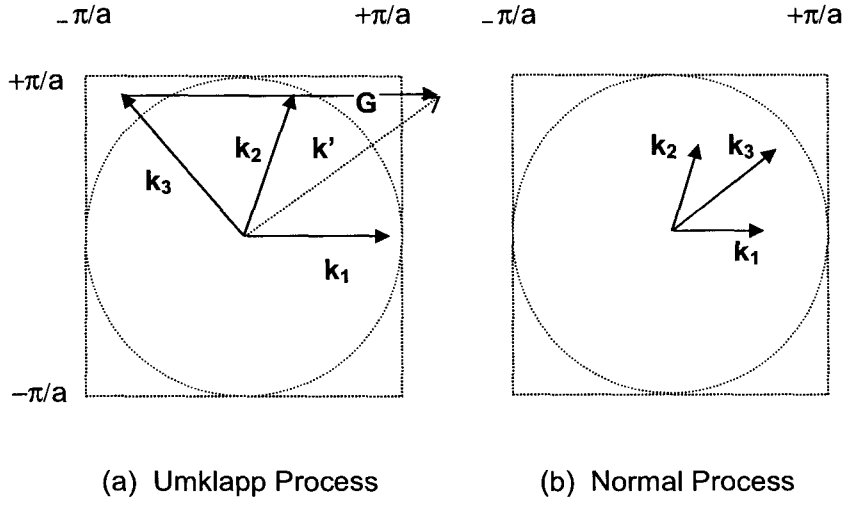


Fig. 4.4 Schematics of the first Brillouin zone of a square lattice in two dimensions. The radius of the dotted circle is π/a. When the phonon wave vector gets close to the vicinity of the circle, the Umklapp process occurs in phonon scattering. As show in (a), the extra $\hbar\mathbf{G}$ is transferred to the lattice as a whole in the form of the crystal momentum. If the wave vectors are far smaller than π/a as shown in (b), only normal process is involved.

In general, the conservation of wave vector in the three-phonon scattering will involve a $\mathbf{G}$ vector, i.e.,

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}' = \mathbf{k}_3 + \mathbf{G}. \tag{4.67}$$

When $\mathbf{G} \neq 0$, we have $\mathbf{k}_1 + \mathbf{k}_2 \neq \mathbf{k}_3$. In this case, the total phonon momentum before collision is not equal to the total phonon momentum

after the event. This process is called an Umklapp process. If $\mathbf{G} = 0$, we have $\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3$ where the total phonon momentum is conserved with no need of a reciprocal lattice vector. This is called a normal process.

Consider a long rod in thermal equilibrium. The net phonon momentum is zero. Now, if we give a push to all the phonons at one end by applying a heat pulse, every phonon gains an extra momentum of equal amount along the direction of the rod. The net momentum is no longer zero but biased with a constant value, which gives rise to a current density of phonons,

$$J = \sum_{\mathbf{k}} n_{\mathbf{k}} \hbar \delta \mathbf{k}. \tag{4.68}$$

If these phonons do not collide among themselves and with anything else, the thermal current will persist without decay. If there are collisions among phonons but all are normal processes, the net thermal current density will remain constant because the net momentum is not changed. The thermal current will also flow without loss in this case, and we do not need a temperature gradient to maintain the heat flow. We deliberately ignore the impurity, defect, electron, and boundary scattering in order to emphasize the fact that the momentum-conserved normal process will not cause thermal resistance. Everyday experience reminds us that a temperature gradient is always needed to conduct heat even in the purest material. So, there must be "momentum killing" processes. Common sense tells us that the obvious ones are the scattering of phonons off impurities, defects, boundary surfaces, and electrons. In fact, there is another important but less known scattering that is the Umklapp process of phonon-to-phonon scattering. All these scattering processes do not conserve phonon momentum. Therefore, they will slow down the net momentum of the phonon current. This is why an external driving force from a temperature gradient is needed if we want to maintain the thermal current.

4.13 Thermal Conductivity

Let us not balk down with the details of scattering processes now, and just characterize the scattering with a relaxation time τ and a mean free path ℓ. This is the relaxation-time scheme borrowed from the kinetic gas theory, where two consecutive collisions are separated by a relaxation time τ or by an averaged distance of mean-free-path ℓ. In a sphere of radius ℓ centered at x_0, phonons have an average velocity equal to v and the temperature in

the region is $T(x_0)$.

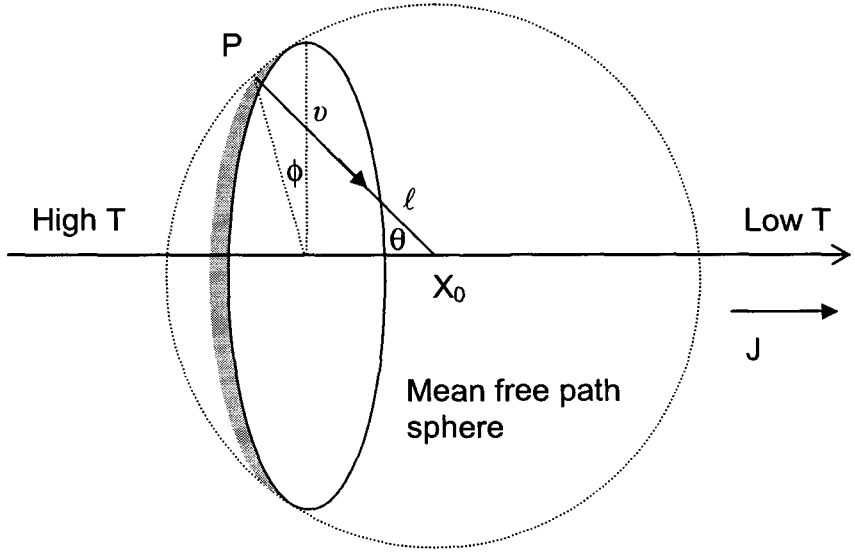


Fig. 4.5 Schematics of the phonon mean-free-path sphere, centered at x_0 in the phonon current. All the phonons that experienced scattering at x_0 are not likely to get scattered anywhere inside the sphere within the relaxation time τ. On the time scale much longer that τ, any small region inside the sphere is in local thermal equilibrium. The phonon energy density is uniform on the ring containing point P. But, on a parallel ring down stream at lower T, the energy density will be lower.

The situation is illustrated in Fig. 4.5. Let J be the averaged thermal current density flowing along the rod in the x-direction. The energy density on the spherical surface can be analyzed in terms of that on the ring with the same θ, i.e., $u(x_0 - \ell \cdot \cos\theta)$, and P is such a point on this circular ring. The x component of the thermal current density that flows toward x_0 from any point P on the ring of angle θ is $J(\theta)$. We have

$$J(\theta) = (v\cos\theta) \cdot u(x_0 - \ell\cos\theta). \tag{4.69}$$

Averaging contributions from all points on the surface of the mean-free-path sphere, we will have the averaged thermal current density, i.e.,

$$J = \frac{1}{4\pi \cdot \ell^2} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} J(\theta)(\ell d\theta)(\ell \sin\theta d\phi)$$

$$= \frac{1}{2}\int_0^{\pi} J(\theta)\cdot \sin\theta d\theta$$

$$= \frac{1}{2}\int_0^{\pi} (v\cos\theta)\cdot u(x_0 - \ell\cos\theta)\cdot \sin\theta d\theta. \tag{4.70}$$

Let $q = \cos\theta$. We have $dq = -\sin\theta d\theta$, and

$$J = \frac{1}{2}\int_{-1}^{1} v\cdot u(x_0 - \ell q)\cdot q dq. \tag{4.71}$$

Here, the variable $(x_0 - \ell q)$ varies from $(x_0 - \ell)$ to $(x_0 + \ell)$ on the x-axis around x_0. We can expand the function $u(x_0 - \ell q)$ with the Taylor series around $x_0 - \ell q = 0$,

$$u(x_0 - \ell q) = u(0) + (x_o - \ell q)\left(\frac{\partial u}{\partial(x_0 - \ell q)}\cdot\frac{\partial(x_0 - \ell q)}{\partial x}\right)_{x_0} + \cdots, \tag{4.72}$$

and the thermal current in Eq. (4.71) becomes

$$J = \frac{1}{2}\int_{-1}^{1} v\cdot\left(u(0) + x_0\left(\frac{\partial u}{\partial x}\right)_{x_0}\right) q dq - \frac{1}{2}\int_{-1}^{1} v\ell\left(\frac{\partial u}{\partial x}\right)_{x_0} q^2 dq. \tag{4.73}$$

The first integral is zero, so we have

$$J = -\frac{1}{2}v\ell\left(\frac{\partial u}{\partial x}\right)_{x_0}\cdot\frac{1}{3}q^3\Big|_{-1}^{1} = -\frac{1}{3}v\ell\left(\frac{\partial u}{\partial T}\right)_{x_0}\left(\frac{\partial T}{\partial x}\right)_{x_0}$$

$$= \frac{1}{3}C_v v\ell\left(-\frac{\partial T}{\partial x}\right)_{x_0} = \kappa\left(-\frac{\partial T}{\partial x}\right)_{x_0}. \tag{4.74}$$

Where κ is the thermal conductivity and $(\frac{\partial T}{\partial x})_{x_0}$ is the temperature gradient at x_0. The minus sign shows that the thermal current flows in the direction of decreasing temperature. Understanding the specific heat $C_v = (\frac{\partial u}{\partial T})_{x_0}$ is for the case of constant volume, the subscript v may be omitted for simplicity, and we have

$$\kappa = \frac{1}{3}\cdot Cv\ell$$

$$= \frac{1}{3}\cdot Cv^2\tau. \tag{4.75}$$

The thermal conductivity κ, specific heat C, and sound velocity v are widely measured macroscopic quantities in numerous materials. The mean free path ℓ and relaxation time τ are the phenomenological parameters that

gauge the effect of scattering in the microscopic world. The simple relationship $\kappa = \frac{1}{3}Cv\ell$ is usually applicable regardless of the detailed scattering mechanisms.

4.14 Phonon Mean Free Path

If we have several scattering mechanisms, the number of scattering per second $\frac{1}{\tau}$ is the sum of the scattering rate from each scattering mechanism, i.e.,

$$\frac{1}{\tau} = \frac{1}{\tau_{\text{phonon}}} + \frac{1}{\tau_{\text{imperfection}}} + \frac{1}{\tau_{\text{surface}}} + \frac{1}{\tau_{\text{electron}}}. \tag{4.76}$$

In the steady state, phonon drifts with a velocity v that is usually the sound velocity. The mean free path, $\ell = v\tau$, is therefore given by

$$\frac{1}{\ell} = \frac{1}{\ell_{\text{phonon}}} + \frac{1}{\ell_{\text{imperfection}}} + \frac{1}{\ell_{\text{surface}}} + \frac{1}{\ell_{\text{electron}}}. \tag{4.77}$$

Case 1. High temperature limit $T \gg \theta_D$

At temperatures higher than the Debye temperature $\theta_D (= \frac{\hbar\omega_D}{k_B})$, the factor $(\frac{\hbar\omega_{\mathbf{k}}}{k_B T})$ is much smaller than 1 for phonon energy $\hbar\omega_{\mathbf{k}} \leq \hbar\omega_D$. We have

$$< n_{\mathbf{k}} >= \frac{1}{e^{\frac{\hbar\omega_{\mathbf{k}}}{k_B T}} - 1} \approx \frac{k_B T}{\hbar\omega_{\mathbf{k}}}. \tag{4.78}$$

The average number of phonons available for scattering is proportional to T. More scattering means shorter relaxation time and shorter mean free path. Both τ and ℓ are inversely proportional to $n_{\mathbf{k}}$, and hence inversely proportional to T. At high temperatures, we have the Dulong-Petit law of specific heat, which says that the specific heat C approaches a constant. We have $\tau \propto \frac{1}{T}$ and $\ell \propto \frac{1}{T}$, leading to $\kappa \propto \frac{1}{T}$. Experimentally, the thermal conductivity at high temperatures is found to be $\kappa \propto \frac{1}{T^x}$ with the x value in somewhere between 1 and 2.

Case 2. Low temperature limit $T \ll \theta_D$

At temperatures much lower than the Debye temperature, only low-lying phonons are excited. The dominant phonon energy is much smaller than

the Debye energy and the corresponding wave vector is smaller than the Debye wave vector, i.e, $\omega \ll \omega_D$ and $k \ll k_D$.

Any phonon-phonon scattering among the low-lying phonons of small ω and $\mathbf{k}$ are unlikely to produce a third phonon of high energy. In fact, the emerging phonon must also have $\omega \ll \omega_D$ because the low temperature condition. This is possible only if the emerging $\mathbf{k}$ is smaller than $\mathbf{k}_D$. No reciprocal lattice vector is needed to conserve crystal momentum. Only normal processes of $\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3$ are involved if the Umklapp process is "frozen out".

We have learned that the normal processes with $\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3$ do not change the net phonon momentum. So, if a collection of phonons has some net velocity going down the rod, the normal processes will not change the net velocity. Therefore, thermal energy can transfer along with no need of a temperature gradient. In other words, the thermal conductivity becomes infinity. But the thermal conductivity in an ultra pure crystal remains to be finite even down to very low temperatures. There must be some momentum-destroying Umklapp process involved. The phonons participating in Umklapp processes must have their energy equal to $\hbar\omega_D$. The average number of phonons having this high energy at low temperatures is given by the Boltzmann tail.

$$< n_{\mathbf{k}} > \cong \frac{1}{e^{\frac{\theta_D}{T}} - 1} \approx e^{-\frac{\theta_D}{T}}. \tag{4.79}$$

Since the relaxation time and mean free path are both inversely proportional to the number of phonons available for scattering, we have

$$\tau \approx e^{\frac{\theta_D}{T}}, \qquad \text{and} \qquad \ell \approx e^{\frac{\theta_D}{T}}. \tag{4.80}$$

As the temperature decreases, the Umklapp process becomes unlikely to occur. Both the mean free path and relaxation time increase exponentially according to Eq. (4.80). Since we are dealing with insulator here, the electron scattering does not exist, $\ell_{electron} \to \infty$. As the Umklapp processes are frozen out at low temperatures, the thermal conductivity will become limited by the scattering of phonons off impurities and defects. In a very pure sample, the phonon mean-free-path can be so long that it will be limited by the sample dimension D, and $\kappa = \frac{1}{3}CvD$. This usually occurs in the low-temperature region where Debye T^3 law for specific heat is valid. So, we have thermal conductivity proportional to T^3. It is known as the Casimir limit.

Problems

1. When a quantum system is in statistical equilibrium with a heat bath, its wave function happens to be a wave packet described by $\Psi(t) = \sum_n c_n(t)\Phi_n$, where $\{\Phi_n\}$ is the complete set of orthonormal eigenfunctions for the steady state and $\{E_n\}$is the corresponding eigenvalues. Plot the probability $P(E)$ as a function of E and t. Define the probability $P(E_n)$.

2. Consider a system S_A at thermal equilibrium that is characterized by β_A and a set of energy $\{A_i\}$. The system S_B is also at thermal equilibrium that is characterized by β_B and a set of energy $\{B_j\}$.
   (a) Find the probability $P_A(A_i)$ for S_A to be at energy A_i, and $P_B(B_j)$ for S_B to be at energy B_j.
   (b) Bring S_A in loose contact with S_B and wait for them to reach equilibrium. The combined system S_C is characterized by β_C and $\{A_i + B_j\}$. Find the probability $P_C(A_i + B_j)$.
   (c) Show that both β_A and β_B approach β_C.
   (d) What is the thermodynamic parameter that β must be proportional to? What is the proportional constant?

3. Consider a system of phonons. The energy is described by $E_{k,n} = (n + 1/2)\hbar\omega_{k,n}$. Find the partition function Q for this system. Express Q as a function of Q_k.

4. Plot the averaged phonon occupation number $< n_k >$ versus the normalized phonon energy $\hbar\omega_k/k_BT$.

5. Plot the phonon population versus $x = \hbar\omega/k_BT$. At a given temperature T, Find the phonon energy for the most populated phonons.

6. Show that $\nabla_{\mathbf{k}}\omega(\mathbf{k})$ is perpendicular to the constant $\omega(\mathbf{k})$ surface.

7. In the x-ray diffraction experiment, the photon wave vector $\mathbf{k}$ gets scattered to $\mathbf{k} + \mathbf{G}$. Find the crystal's recoil velocity and the crystal's kinetic energy. Assume the crystal weighs 1 gm, and the photon has an energy equal to 10 KeV.

8. Plot the thermal conductivity of an insulator as a function of temper-

ature. What is the temperature dependence in different temperature ranges? Indicate the underlying scattering mechanisms.

Chapter 5

Free Electron Fermi Gas

Introduction and Summary

Chapter 5 has two parts. The first part investigates the free electron Fermi gas at $T = 0$ K. It is desirable to have a historic overview about the evolution of our understanding on electrons before delving into the main course of our study. We begin with a brief review of the historically important Drude model and the improved Sommerfeld model. The emphasis is to point out the underlying assumptions that are inevitably attached to the physics known at their respective time. Many concepts developed then are still very much alive today. In the Drude model, the electrons are considered as both free and independent. The electron will experience some sort of scattering that is parametrically described with the relaxation time. Both the classical mechanics and statistics govern here. In the Sommerfeld model, the electrons are considered as quantum particles that obey the Schrödinger equation and Fermi-Dirac statistics. Electrons are still treated as free electrons, but no longer truly independent because of Pauli exclusion principle. Sommerfeld model, a model of free electron Fermi gas, is the foundation of modern electron theories in solids. It is also by far the simplest tool to make quick estimates.

We move on to the study of the free electron Fermi gas at absolute zero temperature ($T = 0$ K). The objective is to define all the Fermi parameters, and then to gain an understanding of their respective order of magnitudes. The Schrödinger equation is set up and then solved under the assumption that the electrons are both free and independent. It is shown how the electron quantum states in the k space can be counted under the periodic boundary condition. The purpose is to derive the density of states. The free electron wave function is proved to be the eigenfunction to both the energy

and momentum operators. It turns out that the electron wave vector is related to both the physical momentum and energy. It is reviewed to show that the former relation can lead to the particle and wave duality. The latter is the energy dispersion relation or the energy band of free electrons. Stacking up the electrons in the quantum states in k space according to Pauli exclusion principle will lead to the definitions of all Fermi parameters. They are the Fermi wave-vector; Fermi energy; Fermi level; Fermi sphere; Fermi surface; Fermi velocity; and the Fermi temperature. An estimate to their order of magnitudes is given, and their physical meaning are explained.

The second part is devoted to the study of the free electron Fermi gas at finite temperatures. The objective is to extend our understanding to the more realistic case of finite temperatures, and then to learn from both the success and the failure of the model. Knowing the shortcomings of the model points to the need for improvements. The next step will take the periodic potential into consideration that is the subject of Chapter 6.

The Fermi-Dirac distribution is at the center of the free electron model. An understanding of where it comes from and what it really means deems to be necessary if one wants to be sure about what is going on with the calculation using this distribution. The statistical equilibrium between a small quantum system and a large heat bath discussed in Chapter 4 is very general. The general concepts of the Boltzmann factor, the Gibbs distribution, and the partition function are as valid for Fermions as they are for Bosons. But, the quantum differences between these two types of particles will take them to different distribution functions. Without the constraint of conservation of particles, the partition function of Bosons can be reduced to a product of factors; each belongs to a particular Boson energy. This useful form can not be obtained with Fermions unless the constraint of conservation of particles is also relaxed. At this juncture, the physical picture around the Fermi level within the range of the order of k_BT is discussed to show that a statistical distribution of electrons must exist. The system can be viewed as a subsystem of a single quantum level and the rest of the system. The subsystem can exchange both energy and particle with the rest of the system in order to reach statistical equilibrium. It is shown under this circumstance that the Fermi-Dirac distribution can be derived. It has the meaning of the probability to find one electron in that level when the combined total system is at thermal equilibrium. The chemical potential in the distribution function reflects the fact that this electron is coming from the given system. In essence, the statistical ensemble is enlarged to include subsystems of any number of electrons; each subsystem is in equilibrium

with the rest of the system. Now, the partition function can be obtained without the restriction on conserving electrons. The averaged number of electrons that can occupy one energy level is then calculated from the partition function. It turns out to be the same Fermi-Dirac distribution. Why this is the case is explained. The averaged total number of electrons peaks sharply for a given chemical potential. To force the averaged total number to equal the number of electrons in the system, the conservation of particles is observed in the statistical sense. This condition also reveals the chemical potential as a function of the temperature. The relationship between the chemical potential and the Fermi energy is explained.

After one understands where the Fermi-Dirac distribution comes from and what it means, it is time to apply. The total energy of electrons in the crystal can now be calculated as a function of temperature. The electronic heat capacity turns out to depend on the temperature linearly. This is what measured in experiments and the classical theory can not count for it. Under the influence of an external electric field, the entire Fermi sphere accelerates, but is slowed down by the scattering on the other hand. The end result is a drift velocity determined by the field and by the relaxation time of scattering. The current is easily calculable leading to the well-known Ohm's law. The Hall coefficient, determined to be negative in metals, shows why the electron must carry a negative charge. Also obtained is the electron concentration in the sample. These successes are very impressive for a model as simple as the free electron model. However, the same experiments show that the model has limitations. Things it does not explain include the effective mass, the temperature dependence of the relaxation time, and the positive Hall coefficient. The deficiencies call for a more sophisticated model that should include at least the periodic potential. That will be the subject to study in Chapter 6.

Free Electron Approximation at $T = 0$ K

5.1 Drude Model and Sommerfeld Model

Three years after J. J. Thomson discovered electron in 1897, Drude constructed his theory of electrical and thermal conduction. He applied the kinetic theory of gases to metals by considering electrons as particles similar to the gas molecules. The essential composition of Drude model includes:

1. The effect of ion-core potential is ignored, and conduction electrons are

assumed to move freely. This is the free electron approximation.

2. The electron-electron Coulomb interaction is neglected. This is the independent electron approximation.

3. Conduction electrons are treated as classical particles. Collision model is borrowed from the kinetic theory of gases. Electrons are accelerated in straight lines in between collisions by external electric field. The collision is instantaneous and can change the velocity of electrons abruptly. Electrons are assumed to bounce randomly among the rigid ion cores similar to balls in the pinball machine. There is no consideration of any electric interaction with the core potential, or with each other among themselves. Now, we know this collision model is incorrect, but the concept of relaxation time is still valid. As long as there are collisions in a system of particles, it is usually possible to characterize some of their statistical manifestation with a phenomenological relaxation time τ, and therefore avoid the need to know the detailed nature of the scattering mechanisms.

4. The electron will have traveled a time τ on average in between two consecutive collisions, independent of its position and velocity.

5. Electrons are assumed to achieve thermal equilibrium through collisions. The velocity emerging from a small volume in local equilibrium is isotropic, and only depends upon the local temperature. The hotter the region, the faster the electron will move.

Drude model is a scheme of relaxation time. All the unknown mechanisms of electron scattering are swept under the coverage of this phenomenological parameter. It makes intuitive sense if experiments are on the time scale much longer than the relaxation time. Drude model is successful in explaining many experimental results such as the electric and thermal conductivity. However, the Maxwell-Boltzmann statistics, the only statistics available at the time, is not actually appropriate for electrons. This is why Drude model fails to count for many other experimental results such as the low-temperature specific heat. Of course, Drude model also fails on explaining the effects where the periodic core potential is no longer negligible.

In the first quarter of twentieth century, leaping developments occurred in physics. Important milestones such as quantum mechanics, Pauli ex-

clusion principle, and Fermi-Dirac statistics all fell into place, and time was ripe for an improved new picture on metallic electrons. Sommerfeld took these new physics into consideration and improved upon the Drude model. The free electron approximation is still assumed, except electrons are treated with quantum mechanics. So, electrons are no longer independent in the sense that they expel each other via Pauli exclusion principle. The periodic core potential and the electron-electron Coulomb interactions are still ignored. The conduction electrons are modeled with a system of free electrons that obeys the Fermi statistics. It is properly named as the free-electron Fermi gas, or, in short the free electron model. Many concepts are conceived in the construction of free electron model, and they are of basic importance for the development of modern solid state theories.

5.2 Free Electron Quantum States and Density of States

The Schrödinger equation of a single free electron is

$$-\frac{\hbar}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\Psi(\mathbf{r})=\varepsilon_{\mathbf{k}}\Psi_{\mathbf{k}}(\mathbf{r}). \tag{5.1}$$

Without e-core and e-e interactions, the conduction electrons behave very much like independent free electrons. Each one is governed by the free-electron Schrödinger equation as described in Eq. (5.1), except that the quantum level $\varepsilon_{\mathbf{k}}$ will be filled according to the Pauli exclusion principle. The solution to this equation is the wave function of a single free electron and it is a plane wave, i.e.,

$$\Psi_{\mathbf{k}}(\mathbf{r})=e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{5.2}$$

Applying the periodic boundary condition over a translation vector $\mathbf{L}$,

$$\Psi_{\mathbf{k}}(\mathbf{r}+\mathbf{L})=\Psi_{\mathbf{k}}(\mathbf{r}), \tag{5.3}$$

we have

$$e^{ik_xL}=e^{ik_yL}=e^{ik_zL}=1, \tag{5.4}$$

which requires

$$k_x=\frac{2\pi n_x}{L},\qquad k_y=\frac{2\pi n_y}{L},\qquad \text{and}\qquad k_z=\frac{2\pi n_z}{L}. \tag{5.5}$$

The integers n_x, n_y, and n_z all vary from $-n$ to $+n$. The discrete values of k_x, k_y and k_z form a set of cubic space points. Each $\mathbf{k}$ vector pointing

to one such point represents a distinct electron state. There is one electron state per volume of $(2\pi/L)^3$ in the $\mathbf{k}$ space, and the allowed $\mathbf{k}$ states per unit volume in $\mathbf{k}$ space is

$$\frac{1}{\left(\frac{2\pi}{L}\right)^3} = \frac{V}{8\pi^3}. \tag{5.6}$$

This is the electron density-of-states in $\mathbf{k}$ space. To see the significance of the $\mathbf{k}$ vector, we note that the wave function $\Psi_{\mathbf{k}}(\mathbf{r})$ is not only an energy eigenfunction, but also an eigenfunction of the momentum operator $\mathbf{p}$, i.e.,

$$\mathbf{p} = \frac{\hbar}{i}\left(\frac{\partial}{\partial x}\cdot \mathbf{l}_x + \frac{\partial}{\partial y}\cdot \mathbf{l}_y + \frac{\partial}{\partial z}\cdot \mathbf{l}_z\right), \tag{5.7}$$

and

$$\mathbf{p}e^{i\mathbf{k}\cdot\mathbf{r}} = \hbar\mathbf{k}\cdot e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{5.8}$$

The momentum eigenvalue $\hbar\mathbf{k}$ is the momentum of the electron in the state $\Psi_{\mathbf{k}}(\mathbf{r})$. We have the electron momentum $\mathbf{p}$, velocity $\mathbf{v}$, and energy $\varepsilon_{\mathbf{k}}$ defined as

$$\mathbf{p} = \hbar\mathbf{k}, \quad \mathbf{v} = \mathbf{p}/m = \hbar\mathbf{k}/m, \tag{5.9}$$

and

$$\varepsilon_{\mathbf{k}} = \frac{\mathbf{p}^2}{2m} = \frac{1}{2}m\mathbf{v}^2 = \frac{\hbar^2\mathbf{k}^2}{2m}. \tag{5.10}$$

Along the direction of a given $\mathbf{k}$ vector, the wave function is obviously periodic on planes separating by the minimum distance of one wavelength λ. The periodicity requires

$$e^{ik(r+\lambda)} = e^{ikr} \quad \Rightarrow \quad k\lambda = 2\pi \quad \Rightarrow \quad \lambda = \frac{2\pi}{k} = \frac{2\pi\hbar}{p}. \tag{5.11}$$

This wavelength associated with the momentum is known as de Broglie wavelength. As a particle, the free conduction electron has a definite momentum $\hbar\mathbf{k}$. As de Broglie wave, the electron has a wavelength $\lambda = 2\pi/k$. Since the momentum or the wavelength is precisely defined, the position of the free electron is not confined to any finite spatial region. Substituting the wave function $\Psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}$ into the Schrödinger equation, we obtain the energy of the $\mathbf{k}$ state, i.e.,

$$\varepsilon_{\mathbf{k}} = \frac{\hbar^2}{2m}\mathbf{k}^2 = \frac{\hbar^2}{2m}(k_x^2 + k_y^2 + k_z^2). \tag{5.12}$$

5.3 Fermi Parameters – Fermi Wave Vector, Fermi Energy, Fermi Level, Fermi Sphere, Fermi Surface, Fermi Velocity, and Fermi Temperature

The allowed discrete values of k_x, k_y, and k_z are determined by the periodic boundary condition, and they form a set of cubic space points. Each **k** point is a quantum state available for electrons to occupy. In the ground state of N free electrons, we may place electrons into the **k** states, starting from the lowest energy level and continuing upward. According to Pauli exclusion principle, each **k** state can accommodate 2 electrons with opposite spins. First, we place two electrons in the $\mathbf{k} = 0$ state where the energy is the lowest at $\varepsilon_0 = 0$. Then, continue the placement to the next energy level and so forth. We will soon find that higher energy levels all contain many independent **k** states. So, the energy levels are degenerate. The degree of degeneracy increases as the energy increases. From the quadratic dependence of ε on **k**, it is obvious that all the degenerate **k** states lie on a spherical surface of constant energy. The filled states in **k** space will approach a sphere if the number of electrons is very large. After we accommodate all the electrons, the highest occupied energy level is called the Fermi level and its energy is the Fermi energy ε_F. The filled sphere in **k** space is the Fermi sphere and its radius is the Fermi wave vector k_F. The total number of allowed states within the Fermi sphere must be equal to one half of the electrons. Recalling that the electron density of states is $V/(8\pi^3)$ in **k** space and taking into account the spin degeneracy, we have

$$N = 2 \cdot \left(\frac{4\pi k_F^3}{3}\right) \cdot \left(\frac{V}{8\pi^3}\right) = \frac{k_F^3}{3\pi^2} \cdot V, \tag{5.13}$$

and

$$k_F = \left(3\pi^2 \cdot \frac{N}{V}\right)^{\frac{1}{3}}. \tag{5.14}$$

The Fermi wave vector k_F or the Fermi momentum $\hbar k_F$ is proportional to the cubic root of electron density $(N/V)^{1/3}$. The Fermi energy is

$$\varepsilon_F = \frac{\hbar^2}{2m}\left(3\pi^2 \cdot \frac{N}{V}\right)^{\frac{2}{3}}, \tag{5.15}$$

and ε_F is proportional to $(N/V)^{2/3}$. The constant-energy surface of radius k_F is the Fermi surface. Electrons on the Fermi surface all have Fermi

energy. The electron velocity on Fermi surface is the Fermi velocity, i.e.,

$$v_F = \frac{\hbar k_F}{m} = \frac{\hbar}{m}\left(3\pi^2 \cdot \frac{N}{V}\right)^{\frac{1}{3}}. \tag{5.16}$$

Fermi parameters of k_F, ε_F, and v_F are functions of the electron density N/V. In order to gain an order-of-magnitude insight on these parameters, it is convenient to define the electron volume in terms of a model of closely packed tiny electron spheres. Assuming r_s is the radius of the imagined electron sphere, we have

$$\frac{4\pi r_s^3}{3} = \frac{V}{N}, \quad \text{and} \quad r_s = \left(\frac{3}{4\pi} \cdot \frac{V}{N}\right)^{\frac{1}{3}}. \tag{5.17}$$

Knowing the electron density N/V, we can calculate r_s, and compare it with Bohr radius of hydrogen atom at ground state. The Bohr radius is

$$a_0 = \frac{\hbar^2}{me^2} = 0.529 \times 10^{-8} cm. \tag{5.18}$$

For most metals r_s/a_0 is between 2 and 3. We may rewrite k_F, v_F, and ε_F in terms of r_s/a_0 to gain some quantitative ideas. We have

$$k_F = \frac{3.36}{(\frac{r_s}{a_0})} \mathring{A}^{-1} \approx 1 \mathring{A}^{-1}. \tag{5.19}$$

From the Fermi wave vector k_F, we can deduce de Broglie wavelength from $\lambda = 2\pi/k_F$ for electrons on the Fermi surface. The wavelength turns out to be about a few angstroms or roughly the order of $1\mathring{A}$. The Fermi velocity is given by

$$v_F = \frac{\hbar k_F}{m} = \frac{4.20}{(\frac{r_s}{a_0})} \cdot 10^8 cm/sec \approx 10^8 cm/sec. \tag{5.20}$$

This is a substantial velocity about 1% of the speed of light. It is especially astonishing because we are dealing with the ground state at temperature $T = 0$ K. This high velocity is a consequence of Pauli exclusion principle. The classical gas molecules, which do not possess the mutual exclusion property, have the thermal velocity equal to about $10^7 cm/sec$ at room temperature. It is 10 times slower than the Fermi velocity of electrons at zero temperature. The Fermi energy is given by

$$\varepsilon_F = \frac{\hbar^2 k_F^2}{2m} = (\frac{e^2}{2a_0})(k_F a_0)^2 = \frac{50.1}{(\frac{r_s}{a_0})^2} eV. \tag{5.21}$$

The quantity $e^2/2a_0$ is known as Rydberg (Ry), which is the ground-state energy of hydrogen atom. We have $e^2/2a_0 = 1Ry = 13.6\ eV$. Since $k_F a_0$ is of the order of unity, most metallic elements will have the Fermi energy in the range of 1.5 to 15 eV. An alternative way to describe the Fermi energy is through the interesting concept of Fermi temperature, which is defined as

$$T_F = \frac{\varepsilon_F}{k_B} = \frac{58.2}{(\frac{r_s}{a_0})^2} \times 10^4 K^\circ. \tag{5.22}$$

The Fermi temperature T_F has nothing to do with the sample temperature under consideration, because the sample is assumed at 0 K . Pauli exclusion principle limits the electron occupation number to 2 for each **k** state. Extra electrons must occupy higher energy levels all the way up to ε_F. Fermi temperature is nothing but the temperature equivalent of Fermi energy. If there were no Pauli exclusion mechanism, as in the case of classical gas molecules, we would have had to heat the electrons up to $\sim 10^4$ K in order to gain an energy of the order of ε_F.

Free Electrons at Finite Temperatures

5.4 Fermi-Dirac Particles versus Bose-Einstein Particles

The concept of thermal equilibrium is the same regardless of whether the system consists of phonons or electrons. Many basic considerations we discussed for phonons in Chapter 4 can be carried over for electrons. For example, the key issue is still to find the probability distribution over all energy levels in the system when it reaches thermal equilibrium with a large heat bath. Temperature is defined the same way as the parameter that will reach a common value in two systems when they reach thermal equilibrium. The partition function plays the same normalization role that ensures unity when summing up the total probability. However, there are two important differences between electrons and phonons. The number of electrons is fixed in a given sample, while the number of phonons is unlimited. Pauli exclusion principle excludes the possibility for any two electrons to have exactly the same quantum numbers. Therefore, only two electrons of opposite spins can reside in the same quantum level, while unlimited phonons can live in the same state. These differences lead to different probability distributions. The averaged electron occupation number as a function of

energy and temperature is the Fermi-Dirac distribution, while the averaged phonon occupation number is given by the Planck distribution.

Consider a system of N identical particles, and assume that there is no interaction among them. Any two configurations that differ only by interchanging two or more particles are regarded as the same state. Therefore, the system is characterized if we know the number of particles n_a and the energy ε_a per particle in the quantum state specified by the subscript a. The issue here is to find the partition function Q under the condition of conservation of particles. For $\sum n_a = N$, there are further constraints on what number n_a can take, depending on what quantum particles the system is made of. We have

$n_a = 0, 1, 2, 3, 4, 5, \cdots$ for Bose-Einstein particles;
$n_a = 0, 1$ for Fermi-Dirac particles.

Pauli exclusion principle applies to Fermi particles, and there can only be zero or one particle in each state. Note that we consider the up and down spins as two distinct quantum states here. There is no such restriction on Bose particles. Let us first define the occupancy configuration as a specific particle distribution over the allowed energy quantum states. For example, $(n_1\varepsilon_1, n_2\varepsilon_2, \cdots)$ is one occupancy configuration, where the total energy is $\sum_a n_a\varepsilon_a$, and the probability factor is $\exp(-\beta\sum_a n_a\varepsilon_a)$. For both Bose and Fermi particles, the partition function begins from the same definition. It is the summation of all probability factors for every possible energy-level occupation configuration, i.e.,

$$Q = \sum_{n_1, n_2, \cdots} e^{-\beta \sum_a n_a \varepsilon_a}. \tag{5.23}$$

If we relax the constraint on particle conservation for a moment, the summation can run over all possible n_1, n_2,$\cdots$ independently. We can reduce Q to the more usable form rather easily. We have

$$\begin{aligned} Q &= \left(\sum_{n_1} e^{-\beta n_1 \varepsilon_1}\right)\left(\sum_{n_2} e^{-\beta n_2 \varepsilon_2}\right)\cdots \\ &= \prod_a \left(\sum_{n_a} e^{-\beta n_a \varepsilon_a}\right). \end{aligned} \tag{5.24}$$

For Bose particles, the number n_a can be anywhere from $0, 1, 2, \cdots$, to ∞.

The partition function reduces to

$$Q = \prod_a \left(\frac{1}{1 - e^{-\beta \varepsilon_a}} \right). \tag{5.25}$$

For Fermi particles, the number n_a can only be either 0 or 1. The partition function reduces to a different form of

$$Q = \prod_a \left(1 + e^{-\beta \varepsilon_a} \right). \tag{5.26}$$

The partition function in Eq. (5.25) is for Bose-Einstein statistics, and that in Eq. (5.26) is for the Fermi-Dirac statistics. Both are for the more relaxed case where the particle conservation condition is not imposed.

For the case of phonons, there is no restriction on the total number of phonons, and also no restriction on how many phonons can reside in one energy level. The partition function obtained above for the Bose-Einstein case is legitimately applicable to phonons, and we have seen the same result in Chapter 4. For other Boson or Fermion systems of fixed number of particles, The summation over n_1, $n_2, \cdots$, and n_a can not be treated independently because they must satisfy $\sum_a n_a = N$.

Let us go back to the case of conduction electrons, the number of electrons in a given sample is fixed. The ground state at $T = 0$ is obtained by placing electrons into the quantum states from the lowest energy up to the Fermi energy. Beyond the Fermi level, there are no more electrons available for placement. Obviously, there is only one configuration to match the electrons with the energy levels. All energy levels below the Fermi level are filled, and all those above are empty. The probability for having this configuration is 100%, and any other configuration can only have zero probability. This is an extreme case where n_1, $n_2, \cdots$, and n_a can not take all possible combinations of 0 and 1. In fact, they can not take any combination at all but one fixed configuration. The fixed number of electrons at $T = 0$ is apparently a very stringent requirement.

5.5 Electron Statistics — Equilibrium, Probability, Gibbs Distribution, Partition Function, Conservation of Particles, Chemical Potential, and Fermi-Dirac Distribution

When we solve the Schrödinger equation for free electrons, we know that there are numerous energy levels above the Fermi level. These levels are

available but empty at $T = 0$. Imaging the temperature is turned on slightly, no matter how low, electrons near the Fermi surface can get excited into these empty states above. To see the magnitude of this effect, let us do a quick order-of-magnitude comparison. The room temperature is about 300 K and the Fermi temperature is roughly 60000 K. In terms of energy, it is 26 meV versus 5 eV. So, the thermal energy of ordinary temperature is about 0.5% of the Fermi energy. Unless the sample is intentionally heated up to extremely high temperatures, we may safely assume $k_BT \leq 0.01\varepsilon_F$. The electrons with energy in the range of $(\varepsilon_F - k_BT) < \varepsilon_a < \varepsilon_F$ will have a good chance to get excited into states above the Fermi surface, leaving some empty states below. Naturally, the excited electrons will have a tendency to relax back to these empty states, but the thermal energy will keep sending up new ones. The result is a dynamic balance that gives rise to a characteristic electron distribution in the outmost thin shell of the Fermi sphere. The electrons deep inside the Fermi sphere will not be affected much at all, because there are no empty states available nearby within the energy of k_BT. In fact, the chance to find empty states for $\varepsilon_a < (\varepsilon_F - k_BT)$ diminishes quickly with decreasing ε_a. On the outside of Fermi sphere, the chance to find electrons with $\varepsilon_a > (\varepsilon_F + k_BT)$ also diminishes quickly with increasing ε_a. It appears that all the significant thermal actions are pretty much confined in the thin shell of $(\varepsilon_F - k_BT) < \varepsilon_a < (\varepsilon_F + k_BT)$ that encloses the Fermi surface. This thin shell can be viewed as a thermally smeared Fermi surface with a thickness of $\sim 2k_BT$. In dealing with thermally excited electrons, it makes more sense to measure their energy from $\varepsilon = \varepsilon_F$ in stead of $\varepsilon = 0$. After all, the electron energy to be compared to k_BT is $(\varepsilon_a - \varepsilon_F)$, not ε_a. When energy is measured from ε_F, the range of smeared Fermi surface becomes $-k_BT < (\varepsilon_a - \varepsilon_F) < +k_BT$.

The above discussion gives a qualitative description about what is happening. The derivation of the actual probability distribution still have to follow what we have discussed in Chapter 4 for systems in thermal equilibrium, except additional constraint conditions need to be imposed. The Boltzmann factor and Gibbs distribution derived in Chapter 4 are still applicable to the N-electron system. The Gibbs distribution for a specific occupancy configuration of $(n_1\varepsilon_1, n_2\varepsilon_2, \cdots)$ is

$$P(E,N) = \frac{e^{-\beta \sum_a n_a \varepsilon_a}}{Q} = \frac{e^{-\beta \sum_a n_a \varepsilon_a}}{\sum_{n_1,n_2,\cdots} e^{-\beta \sum_a n_a \varepsilon_a}}. \tag{5.27}$$

Subject to constraints of $\sum_a n_a\varepsilon_a = E$ and $\sum_a n_a = N$, the summation

goes over all quantum states with $n_a = 1$ or 0. Since there are numerous quantum states available at finite temperatures, the counting of all possible sets of N states that will yield the same total energy E must be a stunning task. As we have already discussed earlier, the evaluation of the partition function under these conditions is not so easy at all. Furthermore, the probability given by Eq. (5.27) deals with all N electrons simultaneously and it requires the specification of exactly where every electron is among the vast spectrum of quantum states. It is obviously of little value because the enormous effort needed just for book keeping. Remember that we are dealing with 10^{23} electrons. Therefore, we would rather like to know the occupancy probability for a single electronic state specified by ε_i with the rest of electrons in the background.

Let us divide the N-electron system into two subsystems A and B, where A contains n_i electrons occupying a single energy quantum state ε_i, and B contains $N - n_i$ electrons with a subtotal energy of $E - n_i\varepsilon_i$. The total energy E and total electron number N are conserved, but the two subsystems can exchange both energy and electrons in order to achieve thermal equilibrium. The likelihood of finding n_i electrons in the state ε_i is proportional to the number of ways that the subsystem B can have $N - n_i$ electrons with a subtotal energy of $E - n_i\varepsilon_i$. So, we have the probability given by

$$f(\varepsilon_i, n_i) \propto \Delta\Gamma(E - n_i\varepsilon_i, N - n_i) = e^{\ln \Delta\Gamma} = e^{\frac{S(E-n_i\varepsilon_i, N-n_i)}{k_B}}. \qquad (5.28)$$

The quantity $\Delta\Gamma$ is the number of states for the subsystem B to have $N - n_i$ electrons with a subtotal energy of $E - n_i\varepsilon_i$. The entropy of subsystem B is $S = k_B \cdot ln\Delta\Gamma$. We are not interested in the microscopic electron distribution in subsystem B, because knowing the entropy of subsystem B is sufficient to determine $f(\varepsilon_i, n_i)$ from Eq. (5.28). At a given temperature T, the volume of subsystem B is constant, but both E and N can vary due to the exchange with subsystem A. These changes are reflected in the corresponding entropy change, and we have $dS = (1/T)dE - (\mu/T)dN$ where μ is the chemical potential, or the thermodynamic potential per electron. Since both $n_i\varepsilon_i$ and n_i are infinitesimal quantities comparing respectively to E and N, the entropy S can be well approximated with the Taylor series expansion around E and N up to the linear terms, i.e.,

$$S(E - n_i\varepsilon_i, N - n_i) = S(E, N) - n_i\varepsilon_i \left(\frac{\partial S}{\partial E}\right)_N - n_i \left(\frac{\partial S}{\partial N}\right)_E$$

$$= S(E,N) - n_i\varepsilon_i\left(\frac{1}{T}\right) - n_i\left(-\frac{\mu}{T}\right). \qquad (5.29)$$

Substituting Eq. (5.29) into Eq. (5.28), the probability of finding n_i electrons in the state ε_i becomes

$$f(\varepsilon_i, n_i) = A \cdot e^{-\frac{n_i(\varepsilon_i-\mu)}{k_B T}}. \qquad (5.30)$$

Since n_i can only be 1 or 0, we either find one electron in the state ε_i or we do not. The total probability $f(\varepsilon_i,1) + f(\varepsilon_i,0)$ must be 1. We can easily solve the following coupled equations.

$$\begin{aligned} &f(\varepsilon_i,1) = A \cdot e^{\frac{-(\varepsilon_i-\mu)}{k_B T}}; \\ &f(\varepsilon_i,0) = A; \\ &f(\varepsilon_i,1) + f(\varepsilon_i,0) = 1. \end{aligned} \qquad (5.31)$$

The solution is

$$f(\varepsilon_i, n_i) = \frac{e^{\frac{-n_i(\varepsilon_i-\mu)}{k_B T}}}{1 + e^{-\frac{(\varepsilon_i-\mu)}{k_B T}}}, \qquad n_i = 1 \quad \text{or} \quad 0. \qquad (5.32)$$

When $n_i = 1$, the function $f(\varepsilon_i, 1)$ reduces to the Fermi-Dirac distribution $f(\varepsilon)$, i.e.,

$$f(\varepsilon) = \frac{1}{1 + e^{\frac{(\varepsilon-\mu)}{k_B T}}}. \qquad (5.33)$$

This is the probability of finding one electron in the state ε, while the other $N-1$ electrons exhaust all possible occupancy configurations. When $n_i = 0$, the function $f(\varepsilon_i, 0)$ reduces to $f(\varepsilon,0) = (1 - f(\varepsilon)) = 1/[1 + exp(-(\varepsilon - \mu)/k_B T)]$. It is the probability of finding no electron in the state ε. The same probability for one-electron state can be obtained from Eq. (5.27) by counting all N electrons among all quantum states, i.e.,

$$\begin{aligned} f(\varepsilon_i) = \frac{1}{1 + e^{\beta(\varepsilon_i-\mu)}} &= \sum_{\alpha} P(E,N) = \sum_{\alpha} \frac{e^{-\beta\sum_a n_a\varepsilon_a}}{Q} \\ &= \sum_{\alpha} \frac{e^{-\beta\sum_a n_a\varepsilon_a}}{\sum\limits_{n_1,n_2,\cdots} e^{-\beta\sum_a n_a\varepsilon_a}}. \end{aligned} \qquad (5.34)$$

The summation over α is over all N-electron occupancy configurations where the state ε_i is filled. In principle, it is possible to do the sum-

mation term by term, but we can hardly claim it as a very useful method, not to mention the desirable elegance and simplicity. On the other hand in the Fermi-Dirac distribution, the endless counting is avoided by making use of a single parameter μ. With the help of chemical potential μ, we have literally mapped the N-electron system into a one-electron state in equilibrium with the rest of the system that can be characterized with thermodynamic parameters. In this picture, any quantum state will have a finite probability of being occupied, or being not occupied, according to Fermi-Dirac distribution. The probability for an occupancy configuration of $(n_1\varepsilon_1, n_2\varepsilon_2, \cdots, n_a\varepsilon_a, \cdots)$ to occur is simply the product of $f(\varepsilon_1, n_1)\cdot f(\varepsilon_2, n_2)\cdots$, i.e., $\prod_a f(\varepsilon_a, n_a)$ calculable from Eq. (5.32). The factorial index, a, goes through all quantum states, some filled and some empty as specified by the occupancy configuration. The total number of electrons is not restricted to $\sum_a n_a = N$. In fact, the probability $\prod_a f(\varepsilon_a, n_a)$ can be calculated for $\sum_a n_a = 0, 1, 2, \cdots$, or any other number of electrons as long as these electrons are in thermal equilibrium characterized by T and μ. For the special case where $\sum_a n_a = N$ and $\sum_a n_a\varepsilon_a = E$, we have

$$P(E,N) = \prod_a \frac{e^{-\beta n_a(\varepsilon_a-\mu)}}{1+e^{-\beta(\varepsilon_a-\mu)}} = \frac{e^{-\beta\sum_a n_a(\varepsilon_a-\mu)}}{\prod_a\left(1+e^{-\beta(\varepsilon_a-\mu)}\right)}. \tag{5.35}$$

The same probability is given by Eq. (5.27). Combining the results, we have

$$\frac{e^{-\beta\sum_a n_a\varepsilon_a}}{\sum_{n_1,n_2,\cdots} e^{-\beta\sum_a n_a\varepsilon_a}} = \frac{e^{-\beta\sum_a n_a(\varepsilon_a-\mu)}}{\prod_a\left(1+e^{-\beta(\varepsilon_a-\mu)}\right)}. \tag{5.36}$$

The left hand side is the probability $P(E,N)$ when we treat all N electrons as N quantum particles. The right hand side is the same probability when we treat the same system as a one-electron state in equilibrium with the rest of the system that can be characterized with T and μ. When we set the equality in Eq. (5.36), both conditions of $\sum_a n_a = N$ and $\sum_a n_a\varepsilon_a = E$ are assumed for the given occupancy configuration $(n_1\varepsilon_1, n_2\varepsilon_2, \cdots, n_a\varepsilon_a, \cdots)$. The denominator of left-hand side is the partition function Q of the N-electron system, where the summation over n_1, n_2,$\cdots$, and so forth can not take random combination of 0 and 1 because $\sum_a n_a = N$. This limitation is lifted for the partition function $Q^{(\mu)}$ in the denominator of right-hand side, because $Q^{(\mu)}$ is enlarged to include ensembles of $\sum_a n_a = 0, 1, 2, \cdots, N, \cdots$ when energy is measured from μ. The summation over n_1, n_2,$\cdots$, and

so forth can take random combination of 0 and 1 in $Q^{(\mu)}$. Following the algebra steps given in Eqs. (5.24) and (5.26), we have

$$Q^{(\mu)} = \sum_{n_1,n_2,\cdots} e^{-\beta \sum_a n_a(\varepsilon_a - \mu)} = \prod_a (1 + e^{-\beta(\varepsilon_a - \mu)}). \tag{5.37}$$

According to Gibbs distribution, the probability of finding a system of electrons with the total energy $E = \sum_a n_a(\varepsilon_a - \mu)$ is

$$P(E) = \frac{e^{-\beta \sum_a n_a(\varepsilon_a - \mu)}}{Q^{(\mu)}} = \frac{e^{-\beta \sum_a n_a(\varepsilon_a - \mu)}}{\sum_{n_1,n_2,\cdots} e^{-\beta \sum_a n_a(\varepsilon_a - \mu)}} = \frac{e^{-\beta \sum_a n_a(\varepsilon_a - \mu)}}{\prod_a (1 + e^{-\beta(\varepsilon_a - \mu)})}. \tag{5.38}$$

The average $< \sum_a n_a >$ over ensembles $\sum_a n_a = 0, 1, 2, \cdots, N, \cdots$ will be sharply peaked for a given μ. We can select the appropriate μ value to ensure $< \sum_a n_a >= N$, the fixed number of electrons in our system. Here, the conservation of electrons is observed in the statistical sense. The averaged total number of electrons, $< N >=< \sum_a n_a >$, can be calculated from the partition function with μ as a parameter. To impose the condition of $< N >= N$ enables us to determine μ as a function of T.

Let us calculate the average of total number of electrons, $< N >$, over ensembles $\sum_a n_a = 0, 1, 2, \cdots, N, \cdots$ from $P(E)$ given in Eq. (5.38), i.e.,

$$< N >= \frac{1}{Q^{(\mu)}} \cdot \sum_{n_1,n_2,\cdots} (\sum_a n_a) \cdot e^{-\beta \sum_a n_a(\varepsilon_a - \mu)}. \tag{5.39}$$

This function is obviously related to the partial derivative of $Q^{(\mu)}$ with respect to μ,

$$\frac{\partial Q^{(\mu)}}{\partial \mu} = \sum_{n_1,n_2,\cdots} (\beta \sum_a n_a) \cdot e^{-\beta \sum_a n_a(\varepsilon_a - \mu)}. \tag{5.40}$$

Combining the above results, we have

$$\begin{aligned} < N > &= \frac{1}{Q^{(\mu)}} \cdot \frac{1}{\beta} \frac{\partial Q^{(\mu)}}{\partial \mu} \\ &= \frac{1}{\beta} \frac{\partial}{\partial \mu} \ln Q^{(\mu)}. \end{aligned} \tag{5.41}$$

The Helmholtz free energy, $g = U - TS = -\mu N$, can be expressed as a function of the partition function by the following definition,

$$e^{-\beta g} = Q^{(\mu)} \qquad \text{and} \qquad g = -(1/\beta) \ln Q^{(\mu)}. \tag{5.42}$$

The averaged total number of electrons, $< N >$ in Eq. (5.41), is simply reduced to

$$< N >= -\frac{\partial g}{\partial \mu}. \tag{5.43}$$

We can also calculate the averaged occupation number of one quantum state, $< n_a >$, over ensembles $\sum_a n_a = 0, 1, 2, \cdots, N, \cdots$ from $P(E)$ given in Eq. (5.38), i.e.,

$$< n_a >= \frac{1}{Q^{(\mu)}} \cdot \sum_{n_1, n_2, \cdots} n_a \cdot e^{-\beta \sum_a n_a(\varepsilon_a - \mu)}. \tag{5.44}$$

This function is obviously related to the partial derivative of $Q^{(\mu)}$ with respect to ε_a,

$$\frac{\partial Q^{(\mu)}}{\partial \varepsilon_a} = -\sum_{n_1, n_2, \cdots} \beta n_a \cdot e^{-\beta \sum_a n_a(\varepsilon_a - \mu)}. \tag{5.45}$$

Combining the results, we have

$$\begin{aligned} < n_a > &= \frac{1}{\beta Q^{(\mu)}} \left(-\frac{\partial Q^{(\mu)}}{\partial \varepsilon_a} \right) \\ &= -\frac{1}{\beta} \frac{\partial}{\partial \varepsilon_a} \ln Q^{(\mu)} \\ < n_a > &= \frac{\partial g}{\partial \varepsilon_a}. \end{aligned} \tag{5.46}$$

For Fermi gas, where $n_1 = 0$ or 1, $n_2 = 0$ or 1, $\cdots$, and so forth, we have

$$\begin{aligned} e^{-\beta g} &= Q^{(\mu)} \\ &= \sum_{n_1} e^{-\beta n_1(\varepsilon_1 - \mu)} \cdot \sum_{n_2} e^{-\beta n_2(\varepsilon_2 - \mu)} \cdot (\quad) \cdot (\quad) \cdots, \\ e^{-\beta g} &= \prod_i (1 + e^{-\beta(\varepsilon_i - \mu)}), \end{aligned} \tag{5.47}$$

and

$$g = -\frac{1}{\beta} \sum_i \ln(1 + e^{-\beta(\varepsilon_i - \mu)}). \tag{5.48}$$

The expectation value of electron occupation number can be obtained by differentiating the Helmholtz free energy with respect to ε_a, i.e.,

$$< n_a > = \frac{\partial g}{\partial \varepsilon_a}$$

$$= \frac{e^{-\beta(\varepsilon_a - \mu)}}{1 + e^{-\beta(\varepsilon_a - \mu)}} = \frac{1}{e^{\beta(\varepsilon_a - \mu)} + 1}. \tag{5.49}$$

This is the Fermi-Dirac distribution identical to what we have derived in Eq. (5.33), where the physical interpretation has been the probability for finding one electron in the state ε_a. Equation (5.49) gives a new meaning to the Fermi-Dirac distribution, namely, the expectation value of electron occupation number in the state ε_a. Stemming from the same mathematical function, both interpretations are better to be equivalent. Indeed they are. Since there is only one electron that may reside in the state ε_a and the corresponding probability is f, naturally the expectation value of electron occupation number is just $1 \cdot f$, and we have $< n_a >= f$.

5.6 Calculation of Chemical Potential as a Function of Temperature

In the case of $T = 0$, we have $< n_a >= 1$ for $\varepsilon_a < \mu$, and $< n_a >= 0$ for $\varepsilon_a > \mu$. So, the chemical potential μ must coincide with the Fermi energy level ε_F at $T = 0$. In order to find μ as a function of T, we may begin with the Helmholtz free energy, g, by computing $< N >$ from $-\partial g/\partial \mu$, and then impose the condition $< N >= N$ to comply with the conservation of electrons. The chemical potential $\mu(T)$ will result consequently. To facilitate the computation, it would be necessary to convert the summation in the g function of Eq. (5.48) into an integral first. Since the density of state in k space (number of states per unit volume in k space) is $V/8\pi^3$, and each state can accommodate 2 electrons with opposite spins, the number of electrons having the energy of $\frac{\hbar^2 k^2}{2m}$ will be $2(\frac{V}{8\pi^3})d^3k$. So, we can rewrite Eq. (5.48) in the integral form, i.e.,

$$g = -\frac{1}{\beta} \int \ln \left(1 + e^{-\beta\left(\frac{\hbar^2 k^2}{2m} - \mu\right)}\right) \left(\frac{V}{8\pi^3}\right) \cdot 2 \cdot d^3k. \tag{5.50}$$

The averaged number density ρ can be calculated from $< N >= -\partial g/\partial\mu$, and we have

$$\begin{aligned}\rho &= \frac{< N >}{V} = -\frac{1}{V}\frac{\partial g}{\partial \mu} \\ &= \frac{2}{8\pi^3}\int \frac{e^{-\beta\left(\frac{\hbar^2k^2}{2m}-\mu\right)}}{1+e^{-\beta\left(\frac{\hbar^2k^2}{2m}-\mu\right)}}\cdot d^3k \\ &= \frac{2}{8\pi^3}\int \frac{1}{1+e^{\beta\left(\frac{\hbar^2k^2}{2m}-\mu\right)}}\cdot d^3k. \end{aligned} \tag{5.51}$$

For $T = 0$, we have $\mu = \mu_0 \equiv \varepsilon_F$. The Fermi-Dirac distribution in the integrand of Eq. (5.51) is 1 for $\frac{\hbar^2k^2}{2m} < \mu_0$, and is 0 for $\frac{\hbar^2k^2}{2m} > \mu_0$. The number density can be integrated rather easily.

$$\begin{aligned}\rho &= \frac{2}{8\pi^3}\int_0^\infty \frac{1}{e^{\beta\left(\frac{\hbar^2k^2}{2m}-\mu_0\right)}+1} d^3k \\ &= \frac{2}{8\pi^3}\int_0^{k_F}\int_{\theta=0}^{2\pi}\int_{\phi=0}^{\pi} k^2\sin(\phi)d\phi d\theta dk \\ &= \left(\frac{2}{8\pi^3}\right)(4\pi)\int_0^{k_F} k^2dk \\ &= \left(\frac{2}{8\pi^3}\right)\left(\frac{4\pi}{3}\right)k_F^3 = \frac{k_F^3}{3\pi^2}. \end{aligned} \tag{5.52}$$

This is the same result obtained previously by counting the k states. Recall that the constant energy surface is a sphere in k-space, i.e.,

$$\varepsilon_F = \frac{\hbar^2k_F^2}{2m} = \frac{\hbar^2}{2m}(3\pi^2\rho)^{\frac{2}{3}}. \tag{5.53}$$

The density of states per unit energy per unit volume is given by

$$\begin{aligned}D(\varepsilon)d\varepsilon &= \frac{2}{8\pi^3}\cdot 4\pi k^2dk \\ &= \frac{2}{8\pi^3}\cdot(4\pi)\cdot\frac{2m}{\hbar^2}\cdot\varepsilon\cdot\sqrt{\frac{2m}{\hbar^2}}\cdot\frac{1}{2}\cdot\frac{1}{\sqrt{\varepsilon}}d\varepsilon \\ &= \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\sqrt{\varepsilon}d\varepsilon, \end{aligned} \tag{5.54}$$

and we have

$$D(\varepsilon) = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\sqrt{\varepsilon}. \tag{5.55}$$

For finite temperatures, we need to evaluate the number-density integral with respect to ε since the integrand is a function of ε. The integral can be easily transformed with the help of the density of states derived in Eq. (5.55). We have

$$\begin{aligned}
\rho &= \frac{2}{8\pi^3}\int_\phi\int_\theta\int_k \frac{1}{e^{\beta(\varepsilon-\mu)}+1}d^3k \\
&= \int_0^\infty \frac{1}{e^{\beta(\varepsilon-\mu)}+1}\cdot\frac{2}{8\pi^3}\cdot 4\pi k^2\cdot dk \\
&= \int_0^\infty \frac{1}{e^{\beta(\varepsilon-\mu)}+1}\left(\frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\right)\cdot\sqrt{\varepsilon}d\varepsilon \\
&= \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\int_0^\infty \frac{\sqrt{\varepsilon}d\varepsilon}{e^{\beta(\varepsilon-\mu)}+1}.
\end{aligned} \tag{5.56}$$

Let us label the integral in Eq. (5.56) as I, and we notice that I is of the following general form,

$$I = \int_0^\infty \frac{\sqrt{\varepsilon}d\varepsilon}{e^{\beta(\varepsilon-\mu)}+1} \quad \longrightarrow \quad I = \int_0^\infty \frac{q(\varepsilon)d\varepsilon}{e^{\beta(\varepsilon-\mu)}+1}. \tag{5.57}$$

Integral of this form can be integrated by parts analytically. We have

$$\begin{aligned}
I &= \int_0^\infty d\varepsilon = \int_0^\mu d\varepsilon + \int_\mu^\infty d\varepsilon \\
&= \int_0^\mu q(\varepsilon)\left[1 - \frac{1}{e^{-\beta(\varepsilon-\mu)}+1}\right]d\varepsilon + \int_\mu^\infty \frac{q(\varepsilon)}{e^{\beta(\varepsilon-\mu)}+1}d\varepsilon \\
&= \int_0^\mu q(\varepsilon)d\varepsilon + \int_\mu^\infty \frac{q(\varepsilon)d\varepsilon}{e^{\beta(\varepsilon-\mu)}+1} - \int_0^\mu \frac{q(\varepsilon)d\varepsilon}{e^{-\beta(\varepsilon-\mu)}+1}.
\end{aligned} \tag{5.58}$$

Now, we will change variables for the second integral and the third integral in Eq (5.58), respectively.

For second integral	For third integral
$x = \beta(\varepsilon - \mu)$	$x = -\beta(\varepsilon - \mu)$
$\varepsilon = \mu + \frac{x}{\beta}$	$\varepsilon = \mu - \frac{x}{\beta}$
$d\varepsilon = \frac{dx}{\beta}$	$d\varepsilon = -\frac{dx}{\beta}$

The integral I is transformed into the following form,

$$I = \int_0^{\mu} q(\varepsilon)d\varepsilon + \frac{1}{\beta}\int_0^{\infty} \frac{q(\mu + \frac{x}{\beta})dx}{e^x + 1} - \frac{1}{\beta}\int_0^{\beta\mu} \frac{q(\mu - \frac{x}{\beta})dx}{e^x + 1}. \tag{5.59}$$

For temperatures $T \ll T_F$, which is true for most of the circumstances, the energy $(\varepsilon - \mu)$ is of the order of $k_B T$ that is much smaller than the chemical potential $\mu(\sim k_B T_F)$. So we have $\frac{x}{\beta} \ll \mu$. The function $q(\mu \pm x/\beta)$ can be expanded in terms of Taylor series, and we may drop higher order terms beyond the linear term, i.e.,

$$q(\mu \pm \frac{x}{\beta}) \approx q(\mu) \pm \frac{x}{\beta}\frac{dq(\mu)}{d\mu}. \tag{5.60}$$

The integral I is further reduced to

$$\begin{aligned} I &= \int_0^{\mu} q(\varepsilon)d\varepsilon + \frac{2}{\beta^2}q'(\mu)\int_0^{\infty} \frac{x}{e^x + 1}dx \\ &= \int_0^{\mu} q(\varepsilon)d\varepsilon + \frac{2}{\beta^2}q'(\mu)\cdot\frac{\pi^2}{12}. \end{aligned} \tag{5.61}$$

The integral in Eq. (5.61) is not only useful in calculating the number density ρ, but also handy for calculating the energy density u. Assigning $q(\varepsilon) = \varepsilon^{3/2}$, we can calculate the energy density and the heat capacity, i.e.,

$$\begin{aligned} u &= \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\int_0^{\infty} \frac{\varepsilon^{\frac{3}{2}}d\varepsilon}{e^{\beta(\varepsilon-\mu)} + 1} \\ &= \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\left[\int_0^{\mu} \varepsilon^{\frac{3}{2}}d\varepsilon + \frac{2}{\beta^2}\frac{3}{2}\sqrt{\mu}\cdot\frac{\pi^2}{12}\right]; \\ C_v &= \frac{\partial U}{\partial T} = \frac{\partial u}{\partial T}\cdot V. \end{aligned} \tag{5.62}$$

We will calculate the specific heat differently later in this Chapter. Now, let us go back to the number density. We will substitute $q(\varepsilon) = \sqrt{\varepsilon}$ and $q'(\mu) = \frac{1}{2}\frac{1}{\sqrt{\mu}}$ into Eq. (5.61) to calculate I, and then substitute I into

Eq. (5.56) to calculate ρ. We have

$$\rho = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\left[\int_0^{\mu}\sqrt{\varepsilon}d\varepsilon + \frac{\pi^2}{6}(k_BT)^2\cdot\frac{1}{2\sqrt{\mu}}\right]$$

$$= \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\left[\left(\frac{2}{3}\right)\mu^{\frac{3}{2}} + \frac{\pi^2}{6}(k_BT)^2\cdot\frac{1}{2\sqrt{\mu}}\right]. \tag{5.63}$$

Since the number N is conserved and the volume is kept constant with no expansion with temperature, the number density ρ remains constant as temperature changes from $T = 0$ to a finite T. We already know ρ previously for the $T = 0$ case. Same result can be obtained by letting $T = 0$ in Eq. (5.63), i.e.,

$$\rho = \frac{2}{3}\left(\frac{1}{2\pi^2}\right)\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\varepsilon_F^{\frac{3}{2}}. \tag{5.64}$$

Since the number density is not changing with temperature, we can substitute the above expression for ρ in Eq. (5.63), and solve for μ as a function of T. The result is given here,

$$\mu^{\frac{3}{2}} = \varepsilon_F^{\frac{3}{2}}\left(1 - \frac{\pi^2}{12}(k_BT)^2\cdot\frac{1}{\frac{2}{3}\varepsilon_F^{\frac{3}{2}}\sqrt{\mu}}\right). \tag{5.65}$$

Since μ is of the order of ε_F, the second term in the bracket is very small because it is proportional to $(k_BT/\varepsilon_F)^2$. In the binomial series expansion, it is sufficiently accurate to keep the first two terms and neglect all higher order terms of $(k_BT/\varepsilon_F)^2$, i.e.,

$$\mu \approx \varepsilon_F\left(1 - \frac{\pi^2}{12}(k_BT)^2\cdot\frac{1}{\varepsilon_F^2}\right). \tag{5.66}$$

The chemical potential derived in Eqs. (5.65) and (5.66) is the energy required for removing an electron from the N-electron system. It is also the adjustable parameter chosen to satisfy the conservation of electrons. When $T = 0$, $\mu = \varepsilon_F$. When $T > 0$, the chemical potential monotonically decreases with increasing temperature, see Fig. 5.1. But, the quantity $(k_BT/\varepsilon_F)^2$ is so small for ordinary temperatures that we may safely replace μ with ε_F. It is a very good approximation for almost all practical purposes. Strictly speaking, the Fermi energy ε_F on Fermi surface only have meaning at $T = 0$. For finite temperatures, the Chemical potential should be used instead.

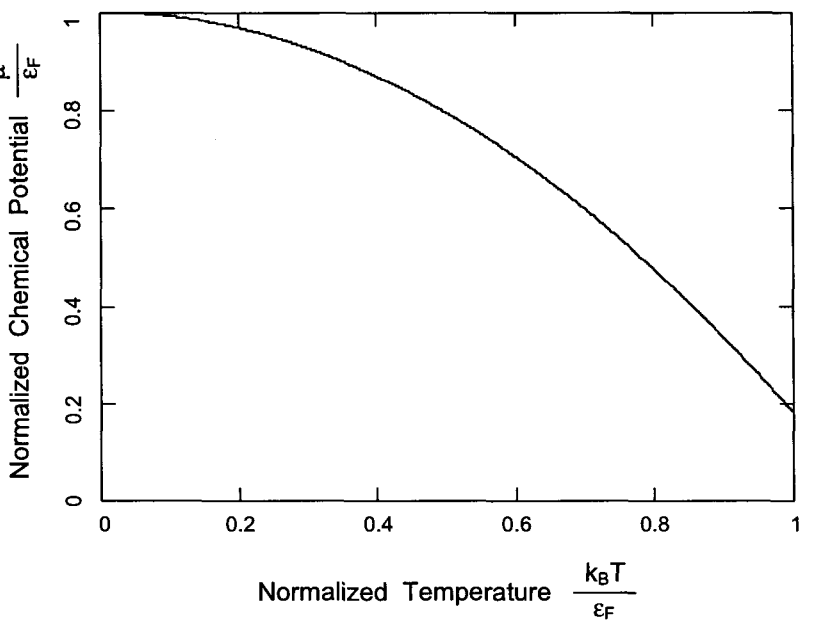


Fig. 5.1 Plot of the normalized chemical potential μ/ε_F versus the normalized temperature k_BT/ε_F from Eq. (5.66). The ratio μ/ε_F does not deviate much from 1 at all for $k_BT/\varepsilon_F < 0.1$. For extremely high temperature where $k_BT/\varepsilon_F \geq 1$, the approximations made to obtain Eq. (5.66) is no longer valid. Also, the constant volume assumption will most likely not hold either. In that case, we need to go back to Eq. (5.56) to look for solutions.

In Fig. 5.2, the Fermi-Dirac distribution function $f(\varepsilon)$ is plotted for a few selected temperatures as parameters. These plots are made for an arbitrarily chosen Fermi temperature, $T_F = 50000$ K. Note that T_F scales with the number of electrons and will vary from sample to sample. The smeared Fermi surface or more exactly the smeared chemical potential surface becomes more pronounced as temperature increases. The chemical potential can be read off the chart, see Fig. 5.2. For temperatures less than 5000 K, the chemical potential read off the chart is basically equal to ε_F. Unless the temperature is very high, μ is not much different from $\varepsilon_F (= k_BT_F)$.

We have derived the density of states $D(\varepsilon)$ in Eq. (5.55). Since all the states $D(\varepsilon)d\varepsilon$ have their occupation number equal to $f(\varepsilon)$, the number of electrons within the range from ε to $\varepsilon + d\varepsilon$ is $D(\varepsilon)f(\varepsilon)d\varepsilon$. Integrating from 0 to ε will give us the total number of electrons with energy smaller than

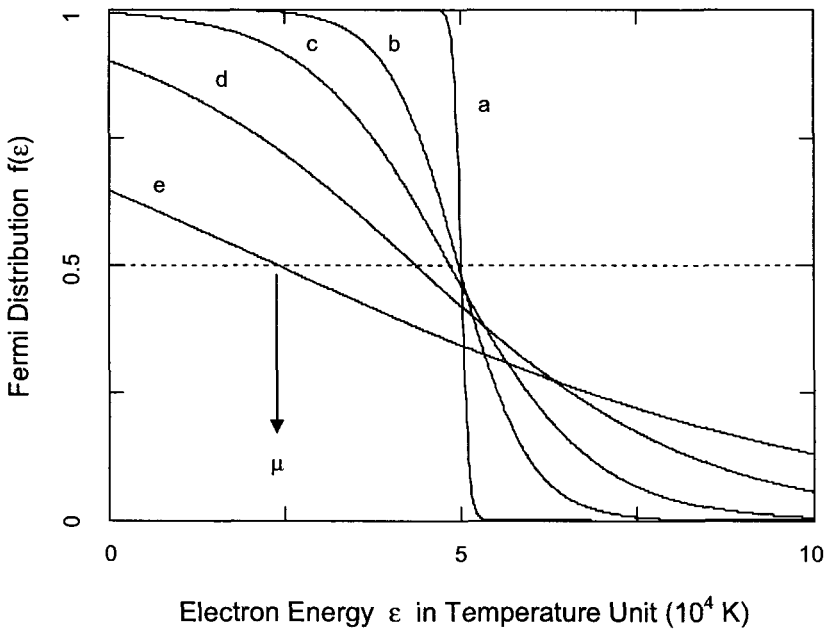


Fig. 5.2 The Fermi-Dirac distribution $f(\varepsilon)$, see Eqs. (5.33) and (5.49), is plotted versus the electron energy ε. The chemical potential μ is calculated first from Eq. (5.66) for a given temperature, and then $f(\varepsilon)$ is calculated for the same T and μ The Fermi temperature is assumed to be $T_F = 50000$ K. Curves a, b, c, d, and e are for $T = 500, 5000, 10000, 20000$, and 40000 K, respectively. The horizontal dash line intercepts $f(\varepsilon)$ at $\varepsilon = \mu$ and $f(\mu) = 1/2$, as indicated on curve e for an example. The higher temperature curve such as curve e should be used with caution, because (T/T_F) is no longer small. Approximations based on the assumption of small (T/T_F) are made in calculating both I and μ. For example, the biggest high-order term we dropped in Eq. (5.66) is proportional to $(T/T_F)^4$, roughly a 36% error for curve e.

or equal to ε . So, we have

$$N(\varepsilon) = \int_0^{\varepsilon} D(\varepsilon) f(\varepsilon) d\varepsilon, \tag{5.67}$$

and

$$D(\varepsilon) f(\varepsilon) = \frac{dN(\varepsilon)}{d\varepsilon}. \tag{5.68}$$

Letting the integration upper limit go to infinity in Eq. (5.67) will give us the total number of electrons N. At the ground state ($T = 0$), all N

electrons fill up the states below Fermi level, and all the states above are empty. At finite temperatures, for states deep inside the Fermi sphere, we have $\varepsilon \ll \mu$, $\exp((\varepsilon - \mu)/k_B T) \sim 0$, and $f(\varepsilon) \sim 1$. This verifies that states deep below the Fermi level are not affected by temperature and always stay filled. For states high above the Fermi level, we have $\varepsilon \gg \mu$, and $f(\varepsilon) \sim exp(-(\varepsilon - \mu)/k_B T)$. This is the high-energy tail of the distribution, see Fig. 5.2, and it turns into the Boltzmann distribution. This implies that electrons behave as if they were not Fermions, and Pauli exclusion principle is not very important for higher energy states. The underlying reason is the low electron density. Because there are only a small number of electrons in the huge number of states above the Fermi level, the density of electrons is so dilute that the probability for electrons to fall in the vicinity of the same quantum state is negligible. The electrons do not have to exclude each other, and the Fermi-Dirac distribution reduces to Boltzmann distribution.

5.7 Heat Capacity of Free Electrons

When temperature increases from 0 to T, the total energy gain in the free electron system is

$$U = \int_0^\infty \varepsilon \cdot D(\varepsilon) f(\varepsilon) d\varepsilon - \int_0^{\varepsilon_F} \varepsilon \cdot D(\varepsilon) d\varepsilon. \tag{5.69}$$

The first integral is the energy at temperature T. The second integral is the energy of the ground state at $T = 0$, where we have the Fermi-Dirac distribution $f(\varepsilon) = 1$ for $\varepsilon < \varepsilon_F$ and $f(\varepsilon) = 0$ for $\varepsilon > \varepsilon_F$. The heat capacity is

$$\begin{aligned} C &= \frac{dU}{dT} \\ &= \int_0^\infty \varepsilon D(\varepsilon) \frac{df}{dT} d\varepsilon - \frac{d}{dT} \int_0^{\varepsilon_F} \varepsilon D(\varepsilon) d\varepsilon \\ &= \int_0^\infty \varepsilon D(\varepsilon) \frac{df}{dT} d\varepsilon. \end{aligned} \tag{5.70}$$

One necessary condition that must be satisfied is the conservation of electrons, i.e.,

$$N = \int_0^\infty D(\varepsilon) f(\varepsilon) d\varepsilon = \text{constant}, \tag{5.71}$$

and

$$\varepsilon_F \frac{dN}{dT} = \int_0^\infty \varepsilon_F D(\varepsilon) \frac{df}{dT} d\varepsilon = 0. \tag{5.72}$$

Incorporating Eq. (5.72) into Eq. (5.70), the heat capacity can be rewritten as

$$C = \int_0^\infty (\varepsilon - \varepsilon_F) D(\varepsilon) \frac{df}{dT} d\varepsilon. \tag{5.73}$$

The integration can be carried out in temperature regions where approximations are justified. In the low temperature region, we have

$$\varepsilon_F \approx \mu, \qquad \text{and} \qquad \frac{k_B T}{\varepsilon_F} < 0.01.$$

Let us compute the derivative of Fermi-Dirac distribution first, i.e.,

$$\frac{df}{dT} = \frac{\varepsilon - \varepsilon_F}{k_B T^2} \frac{e^{\frac{\varepsilon - \varepsilon_F}{k_B T}}}{\left(e^{\frac{\varepsilon - \varepsilon_F}{k_B T}} + 1\right)^2}. \tag{5.74}$$

The derivative df/dT peaks around ε_F and falls off quickly when ε goes away from ε_F. So, we can replace $D(\varepsilon)$ with $D(\varepsilon_F)$. Letting $x = (\varepsilon - \varepsilon_F)/k_B T$, we can rewrite the heat capacity of Eq. (5.73) as

$$C \approx D(\varepsilon_F)({k_B}^2 T) \int_{\frac{-\varepsilon_F}{k_B T}}^\infty \frac{x^2 e^x}{(e^x + 1)^2} dx. \tag{5.75}$$

When we change variable from ε to x, the lower integration limit goes from $\varepsilon = 0$ to $x = -\varepsilon_F/k_B T$. In the low temperature region where $x = -\varepsilon_F/k_B T$ is a very large negative number, we have $e^x \approx e^{-\infty} = 0$, and we may approximate the lower limit by $-\infty$.

$$C = D(\varepsilon_F)({k_B}^2 T) \int_{-\infty}^\infty \frac{x^2 e^x}{(e^x + 1)^2} dx.$$

$$= D(\varepsilon_F)({k_B}^2 T) \frac{\pi^2}{3}. \tag{5.76}$$

For free electron Fermi gas at $T = 0$ K, we have

$$N = \frac{V}{3\pi^2} \left(\frac{2m\varepsilon}{\hbar^2}\right)^{\frac{3}{2}}, \qquad \ln N = \frac{3}{2} \ln \varepsilon + \text{constant}, \quad \text{and}$$

$$\frac{dN}{N} = \frac{3}{2}\frac{d\varepsilon}{\varepsilon} \qquad \Rightarrow \qquad D(\varepsilon) = \frac{dN}{d\varepsilon} = \frac{3}{2}\frac{N}{\varepsilon}. \tag{5.77}$$

Substituting $D(\varepsilon_F)$ of Eq. (5.77) into Eq. (5.76), we have

$$\begin{aligned} C &= \frac{1}{2}\pi^2 N k_B \cdot \frac{k_B T}{\varepsilon_F} \\ &= \frac{\pi^2}{2} N k_B \cdot \frac{T}{T_F}. \end{aligned} \tag{5.78}$$

This is the heat capacity of free electron Fermi gas in the temperature region where $k_B T \ll \varepsilon_F$. It is instructive to see why the temperature dependence turns out to be linear. Since only the electrons residing within an energy range of $\sim k_B T$ below the Fermi surface can be excited into the available levels above the Fermi level, the number of electrons that get excited is proportional to $\sim k_B T \cdot D(\varepsilon_F)$, and the total number of electron is roughly proportional to $\sim k_B T_F \cdot D(\varepsilon_F)$. The fraction of electrons that gets excited is proportional to $\sim (T/T_F) \cdot N$, and each such electron has gained the thermal energy of $k_B T$. So, we can estimate the total energy and heat capacity rather easily, i.e.,

$$U \approx N \left(\frac{T}{T_F}\right) k_B T, \tag{5.79}$$

and

$$C = \frac{dU}{dT} \approx N k_B \left(\frac{T}{T_F}\right). \tag{5.80}$$

The total heat capacity contains both contributions from the lattice and electrons. The electron part follows a linear T dependence, and phonons obey the Debye T^3 law, i.e.,

$$C = \gamma T + A T^3. \tag{5.81}$$

The coefficients γ and A contains material parameters and basic constants. Comparison between the experimental and theoretical heat capacity offers deep insights regarding our understanding of materials on the atomic scale.

5.8 Ohm's Law

Consider a system of free electrons in $\mathbf{k}$ space. The electrons fill up energy levels up to near the Fermi surface. In the thin shell of $\sim 2k_B T$ enclosing

the Fermi surface, thermal smearing occurs, but spherical symmetry is still maintained. At thermal equilibrium with no external electric field applied, the Fermi sphere is centered at $\mathbf{k} = \mathbf{0}$. From the spherical symmetry of the distribution of $\mathbf{k}$ states, it is rather easy to see why the net momentum $\sum \hbar\mathbf{k}$ associated with the whole sphere is zero. Therefore, the net velocity of all electrons is zero. This is consistent with the zero electric current since there is no external field. When an electrical field is applied, electrons will gain momentum in the direction of the field. So, the whole Fermi-sphere drifts in $\mathbf{k}$ space. Let us apply Newton's law to one electron, i.e.,

$$m\frac{d\mathbf{v}}{dt} = \hbar\frac{d\mathbf{k}}{dt} = -e\mathbf{E}. \tag{5.82}$$

Integrating over time, we have

$$\mathbf{k}(t) - \mathbf{k}(0) = -\frac{e\mathbf{E}t}{\hbar} \quad \longrightarrow \quad \mathbf{k}(t) = -\frac{e\mathbf{E}t}{\hbar};$$

$$\mathbf{v}(t) - \mathbf{v}(0) = -\frac{e\mathbf{E}t}{m} \quad \longrightarrow \quad \mathbf{v}(t) = -\frac{e\mathbf{E}t}{m}. \tag{5.83}$$

The velocity of individual electron, therefore, of the whole Fermi sphere, is accelerated steadily if there are no collisions of some sort to slow the electrons down. In reality, there are collisions with phonons, impurities, defects, and other electrons. These collisions stop electrons from having running-away velocities, and result in a steady drift velocity

$$\mathbf{v} = -\frac{e\mathbf{E}\tau}{m}. \tag{5.84}$$

The relaxation time τ is the averaged duration between two consecutive collisions. Electrons are only accelerated during the time τ before getting interrupted by the next collision. The collision slows down the velocity to a constant drift velocity. The current density is

$$\mathbf{J} = nq\mathbf{v} = \frac{ne^2\tau\mathbf{E}}{m} = \sigma\mathbf{E}. \tag{5.85}$$

This is the renounced Ohm's law, where the current density is proportional to the external electric field. The proportion constant σ is the D.C. conductivity and its inverse $\rho = 1/\sigma$ is the D.C. resistivity. We have

$$\sigma = \frac{ne^2\tau}{m}, \quad \text{and} \quad \rho = \frac{m}{ne^2\tau}. \tag{5.86}$$

5.9 Hall Effect

When an electrical field and a magnetic field are applied, the equation of motion of the free-electron Fermi sphere is

$$\hbar\left(\frac{d}{dt}+\frac{1}{\tau}\right)\delta\mathbf{k}=-e(\mathbf{E}+\frac{1}{c}\mathbf{v}\times\mathbf{B}). \tag{5.87}$$

As we have discussed in the Ohm's law, the whole Fermi sphere is accelerated by the electromagnetic force and is slowed down by collisions. The motion will settle in a steady state drifting. The drift momentum is $\hbar\delta\mathbf{k}$ and the drift velocity is $\mathbf{v}$. Knowing $m\mathbf{v}=\hbar\delta\mathbf{k}$, we have

$$m\left(\frac{d}{dt}+\frac{1}{\tau}\right)\mathbf{v}=-e(\mathbf{E}+\frac{1}{c}\mathbf{v}\times\mathbf{B}). \tag{5.88}$$

If $\mathbf{B}$ is the static magnetic field aligned along the z-axis, we can decompose the vector equation into three components, i.e.,

$$\begin{aligned}
m\left(\frac{d}{dt}+\frac{1}{\tau}\right)v_x&=-e(E_x+\frac{B}{c}v_y);\\
m\left(\frac{d}{dt}+\frac{1}{\tau}\right)v_y&=-e(E_y-\frac{B}{c}v_x);\\
m\left(\frac{d}{dt}+\frac{1}{\tau}\right)v_z&=-eE_z.
\end{aligned} \tag{5.89}$$

Take the case of static electrical field. The drift velocity components are given by

$$v_x=-\frac{e\tau}{m}(E_x+\frac{B}{c}v_y)=-\frac{e\tau}{m}E_x-\omega_c\tau v_y, \tag{5.90}$$

$$v_y=-\frac{e\tau}{m}(E_y-\frac{B}{c}v_x)=-\frac{e\tau}{m}E_y+\omega_c\tau v_x, \tag{5.91}$$

$$v_z=-\frac{e\tau}{m}E_z, \tag{5.92}$$

and the cyclotron frequency is

$$\omega_c=\frac{eB}{mc}. \tag{5.93}$$

In the conventional setup for Hall-effect measurement, if the electrical field is in the x direction to drive the current, the magnetic field is usually applied

in the z direction to bend the current toward y direction. Soon there will be a static electric field build up due to the accumulated charge. This is the Hall field, which cancels the Lorentz field. As a result, the current will end up to flow along the x direction in steady state. See Fig. (5.3) for the orientation of Hall effect.

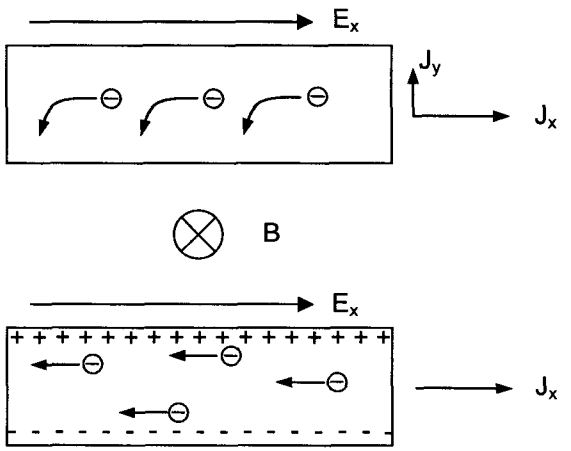


Fig. 5.3 Hall Effect. The magnetic field **B** is along the z direction going into the paper. The paper is in the x-y plane. The top figure shows the initial situation when the fields are just turned on. The bottom figure shows the steady state when the Hall field has already built up.

After the Hall field has built up and current flow reaches the steady state, we have $v_y = 0$ because the balance between Hall and Lorentz fields. The current in the z direction depends on E_z only, see Eq. (5.92). We have $v_z = 0$ because $E_z = 0$. The x-component current from Eq. (5.90) is

$$v_x = -\frac{e\tau}{m} E_x. \tag{5.94}$$

The same x-component current can also be determined from Eq. (5.91),

$$v_x = \frac{e\tau}{m\omega_c \tau} E_y = \frac{e}{m\omega_c} E_y. \tag{5.95}$$

Consequently, Eqs. (5.94) and (5.95) lead to the condition for Hall effect,

i.e.,

$$E_y = -\omega_c \tau E_x = -\frac{eB\tau}{mc} E_x. \tag{5.96}$$

The static field E_y is from the accumulated charges induced by the magnetic field. It counter balances the Lorentz field $\mathbf{v} \times \mathbf{B}$ that equals $v_x B$. We define the Hall coefficient R_H as

$$R_H = -\frac{E_y}{nev_x B} = \frac{-\frac{eB\tau}{mc} E_x}{\frac{ne^2\tau}{m} E_x B}.$$

$$R_H = -\frac{1}{nec}; \qquad \text{CGS unit} \tag{5.97}$$

$$R_H = -\frac{1}{ne}. \qquad \text{SI unit} \tag{5.98}$$

Experimentally, one can measure the x-direction current j_x and the y-direction voltage that gives E_y. The magnetic field B is calibrated ahead of time. So, the Hall coefficient R_H is easily determinable, and its inverse is the product of charge and charge carrier concentration, i.e.,

$$R_H = \frac{E_y}{j_x B} \qquad \text{and} \qquad nq = \frac{1}{R_H}. \tag{5.99}$$

We can calculate the electron concentration from the measured R_H. The sign of R_H shows clearly that the electron carries a negative charge. Hall Effect is an important tool that contributes to the development of semiconductors, primarily because it identifies another charge carrier called "hole" that carries a positive charge. Together with the resistivity measurement, see Eqs. (5.86) and (5.99), we can deduce the quantity $(\frac{e}{m})\tau$ easily. When both the carrier concentration $n(T)$ and relaxation time $\tau(T)$ are measured as functions of temperature, they provide quantitative insights regarding the quality and purity of the material. The Hall and resistivity measurements are the easiest but still the most fundament methods for characterizing semiconducting materials.

5.10 Some Limitations of the Free Electron model

The free electron model is very important. Its simplicity has made itself the most convenient model for quick estimates before going ahead with more involved investigations. Despite of its simplicity, the free electron model has explained quite a few properties of conduction electrons. However, the free

electron theory is only a simplified model. The periodic lattice potential, the electron-phonon interaction, and the electron-electron interaction are not considered.

The spirit of free electron model has further extended in the Landau Fermi Liquid theory. If those interactions the electron experiences can be considered as weak, the conduction electrons can still be mapped into a system of free quasi-electrons with an effective mass. Landau Fermi Liquid theory has been the starting point for many modern solid state theories including the BCS theory of superconductivity. Despite of its extraordinary successes, the power of Landau Fermi Liquid theory may have been finally getting challenges from the high T_c superconductivity. It is beyond the scope of this book to pursue further analyses on the Landau Fermi Liquid theory. Let us go back to comment on the problems with the free-electron Fermi gas model.

There are quite a few experimental results that the free electron theory does not explain satisfactorily. We will mention a few of them, which connect what we have learned to what we will learn in the next Chapter and beyond.

1. Electron Heat Capacity

   The free electron model predicts that the electronic heat capacity is linearly proportional to temperature T. The coefficient of this linear term is equal to

$$\gamma = \frac{1}{2}\pi^2 N {k_B}^2 \cdot \frac{1}{\varepsilon_F} = \left(\frac{1}{2}\pi^2 N {k_B}^2 \cdot \frac{2}{\hbar^2}\left(\frac{3\pi^2 N}{V}\right)^{-\frac{2}{3}}\right) \cdot m \quad \propto \quad m. \tag{5.100}$$

   From the specific heat experiment, one can calculate the electron mass from this expression and compare it to the free electron mass. The ratio m(calculated)/m(free) is rarely equal to one. For alkali and noble metals the ratio is close to one. For transition metals such as Fe and Mn, the ratio is smaller than one. For Bi and Sb, the ratio is larger than one. These results indicate that in most metals the electrons are not really as free as the true free electrons.

2. Ohm's Law

   In order to calculate the electrical conductivity and Ohm's law, we have

incorporated a phenomenological parameter of the relaxation time into the free electron model. All the unknown mechanisms of electron scattering are swept under the coverage of relaxation time, which gives rise to a resistive force to be included in the equation of motion. Although we do not know the origins of the resistive scattering, we do know intuitively that the resistive force must exist to counter balances the applied electrical force, so that the whole Fermi sphere can drift without a runaway velocity. The relaxation time, as an adjustable parameter in the theory, is to be determined by comparing the calculated electrical conductivity to the experimental value. The conductivity $\sigma = ne^2\tau/m$ calculated in this manner does not have explicit temperature dependence. In experiments, σ is very much temperature dependent. Adjusting the τ value in the theoretical conductivity to match experiments, we end up with a τ-versus-T curve. The revealed characteristic temperature dependence in τ points to the need for more detailed consideration of the scattering mechanisms. In other words, electrons might not be as free and as independent as assumed in the free electron model.

3. Hall Effect

   The free electron theory explains the Hall effect in metals very well. The Hall coefficient, $R_H = -\frac{1}{ne}$, does not have explicit temperature dependence. Experiments shows that the temperature dependence of R_H in semiconductors is very noticeable and characteristic. This implies that the carrier concentration n is a function of temperature. In some semiconductors, R_H can even be positive, indicating that the charge carrier carries a positive charge. The free electron theory can not count for these experimental facts at all. The current of positive charges sounds like a manifestation of electron vacancies over the positively charged background of ion cores. It indicates that we have to develop an improved view of the conduction electrons. They may not be completely free since there is the periodic ion-core potential in the background.

To improve upon the free electron theory, some incremental steps are logically obvious. They are listed here to set the stage for future considerations in the next Chapter and beyond.

1. Electrons are not really free. They move in the periodic potential of ion cores. Toward a better model, we must take this into account. In the next Chapter, we will treat the case of a weak periodic potential and develop the elementary theory of energy bands. Important concepts of the energy band theory, including the energy gap, band effective mass, Bragg scattering, and holes, will be introduced. We will see how much more understanding we can gain by simply turning on the weak periodic potential in the free electron Fermi gas.

2. The periodic ion cores are not static. The vibration amplitude, or the number of phonons, scales with temperature. For the temperature range where there are sufficient number of phonons, the electron-phonon scattering is responsible for the temperature dependence of the electrical conductivity observed in ultra pure metals. Even at extremely low temperatures, the electron-phonon interaction is still very important. For instance, it is the mechanism underneath the superconductivity of BCS superconductors. In addition, an electron can be dressed with phonons as a new excitation that can sometimes give rise to new experimental observations. Therefore, the electron-phonon interaction should be taken into account in the more advanced model for crystalline solids.

3. Finally, the electron-electron interaction may have to be considered. At the first glance, this interaction must be fairly strong, because the range of Coulomb force is rather long, and the nominal distance between electrons is only about 10^{-8} cm, see Eqs. (5.17) and (5.18). We may wonder why the free electron model works at all. Luckily, we have two major mechanisms working to reduce the occurrence of this interaction. The electron screening effectively reduces the range of the Coulomb force. In addition, the Pauli exclusion principle reduces the scattering probability. The end result is a rather long mean free path for electrons as if they were independent from each other. The typical mean free path in metals for e-e scattering is about ~ 1 μm at room temperature, and it is about ~ 10 cm at the liquid helium temperature. The effective strength of electron-electron scattering turns out to be weak after all. Landau's Fermi liquid theory is able to turn the interacting electron system into a free quasi-electron system. The later can be treated as the free electron system except that the quasi-electron has an effective mass. Usually, we can start with Landau's Fermi liq-

uid, so that we do not have to consider the electron-electron interaction very seriously. When Landau's Fermi liquid theory hits its limitation in some new materials, we may have to consider the electron-electron interaction from scratch.

Problems

1. Consider a two dimensional space. The free electron wave function is periodical over a distance of L in both the x and y directions.
   (a) Set up a computer program, so that you can generate a square lattice (k_x, k_y) in the two dimensional $\mathbf{k}$ space.
   (b) Calculate the number of quantum states in the belt bounded by $k = (\frac{2\pi}{L})(n+1)$ and $k = (\frac{2\pi}{L})(n-1)$ for 3 cases of $n = 10, 100$, and 1000.
   (c) Have your computer program count the exact number of states for the same 3 cases.
   (d) Plot the percentage error you made in (b) as a function of n. What can you say about the case when n is even larger? What is the energy inside the belt for the case of a very large n value?

2. The mass of our sun is $2 \times 10^{33} gm$. Estimate the number of electrons in the sun. If our sun turns into a white dwarf star in the future, this number of electrons will be confined in a sphere of radius $2 \times 10^{9} cm$. In the non-relativistic limit, find the Fermi energy in eV, Fermi temperature in degree K, and Fermi velocity in cm/sec for the white dwarf star.

3. Calculate the free electron energy density u and heat capacity C_v by using the integral given in Eq. (5.61). Compare your heat capacity with that given in Eq. (5.78).

4. Plot $\frac{df}{dT}$ in Eq. (5.74) as a function of $x = \frac{(\varepsilon - \varepsilon_F)}{k_B T}$.

5. . Plot $D(\varepsilon)$, $D(\varepsilon)f(\varepsilon)$ at $T = 0$, and $D(\varepsilon)f(\varepsilon)$ at $T = 300$ K.

This page is intentionally left blank

Chapter 6

Electron Energy Bands

Introduction and Summary

Chapter 6 has three parts. The main objective is to establish the concept of energy bands. In the first part, we will discuss the general influence of the periodic potential on the electron wave function. We begin with an overview of various many-body interactions in the crystal. The complexity of all forces points to the need for improved models over the free electron approximation. The lattice translation symmetry requires that the ion cores form a periodic potential. The valence electrons have a better chance to escape from their parent atoms because the flattened potential in between ion cores. These valence electrons will form a periodic charge cloud to provide further screening. When one electron in the crystal is singled out for investigation, all the interactions with other electrons and ion cores may be modeled with an effectively screened periodic potential. In this picture, the electrons are no longer free, but remain independent in the phenomenological sense. In addition, the ion cores are assumed to be static, i.e., the electron-phonon interaction is neglected for the moment. Whether this approach is valid or not can only be verified by experiments. There is no need to construct a specific effective potential at this point, because its general periodic nature is already sufficient to bring out the most important general properties; the Bloch wave function and the energy bands with gaps at the Brillouin zone boundaries.

The rest of the first part is devoted to Bloch theorem with two proofs and remarks. The Bloch theorem dictates what general form the electron wave function must take in a periodic potential. The first proof is from the general quantum mechanical arguments. The translation operator that commutes with the Hamiltonian can move the wave function to a new posi-

tion by a lattice vector. The new wave function can only be different from the old one by a phase factor because the underlying translation symmetry. Considering the relationship between the lattice and reciprocal lattice, the phase factor turns out to be a plane wave evaluated at this lattice vector. The Bloch wave function comes out naturally. This proof shows how basic and general the Bloch wave function really is. It dose not depend on specific potential and lattice as long as there is the translation symmetry.

The second proof comes as a necessary consequence when we try to solve the Schrödinger equation. The potential can be expressed as a Fourier summation of the plane waves over all reciprocal lattice vectors because the potential has the same periodicity of the lattice. The periodic boundary condition imposed on the wave function selects a complete set of orthogonal plane waves. The wave function can always be expanded in terms of a linear combination of these plane waves. In this scheme of plane wave expansion, the Schrödinger equation is decomposed into all the orthogonal $\mathbf{k}$ components. For each $\mathbf{k}$ wave, the wave function turns out to contain all plane waves with wave vectors related to $\mathbf{k}$ by the reciprocal lattice vectors $\mathbf{G}$. That is just the Bloch wave function. The algebraic effort also gets the Schrödinger equation sorted to the suitable form ready for us to solve it in the next section.

In the end of the first part, some general but important properties of the Bloch wave function are pointed out. The Bloch wave function is an amplitude-modulated plane wave, where the amplitude function has the periodicity of the lattice, but the wave function itself does not. The wave function differs in general by a phase factor when moved by a lattice vector. It ensures the electron density probability to have the same periodicity of the lattice. The free electron has the full translation symmetry of the free space, but Bloch electron only have the reduced translation symmetry of the lattice. As a consequence, the wave vector of Bloch electron is not equal to the free electron momentum, but to the crystal momentum of the electron. The energy eigenvalue of the Bloch wave function has multiple values for a given $\mathbf{k}$,which hints the existence of energy bands already.

The second part focuses on how to solve the Schrödinger equation for one electron in the weak periodic potential. The first objective is to demonstrate the formation of energy bands and gaps. The other objective is to see how the concepts of Brillouin zone, group velocity, crystal momentum, Bragg scattering, and electron effective mass all play out in the energy band.

At first, a brief explanation shows why a weak potential is not so bad an assumption to begin with. In the hypothetical experiment when the poten-

tial is turned off, the system must be reduced to the free electron case. All the other $\mathbf{k} - \mathbf{G}$ plane waves vanishes except the $\mathbf{k}$ wave in the Schrödinger equation. When the potential is turning on, one can intuitively follow the growth of other $\mathbf{k} - \mathbf{G}$ waves. This physical picture helps understand the perturbation treatment. The degree of degeneracy is defined in terms of how many free-electron energy levels are packed inside the energy range of the order of the potential. For the $\mathbf{k}$ vector far away from the Brillouin zone boundary, the energy level is shown to be non-degenerated. The electron energy will be shifted by an amount according to the second-order effect. For $\mathbf{k}$ vectors near or at the zone boundary, the energy level is degenerated and the effect is of the first order. For the general case of m-fold degeneracy, the Schrödinger equation reduces to a set of coupled linear equations of m unknown Fourier coefficients. The potential and the m energy values of the free electron enter the equation as known parameters. The energy eigenvalues can be obtained by solving the secular equation of m^{th} order. The case of two-fold degeneracy is solved to demonstrate this procedure. It reveals two energy bands with an energy gap opened up at the Brillouin zone boundary. This first-order effect diminishes when moving away from the boundary. At the zone boundary, the Fermi surface is torn apart to become perpendicular to the boundary plane, and also, the electron wave function splits into two types of standing waves; p-like in the upper band and s-like in the lower band. A calculation of the energy bands in the vicinity of the zone boundary is offered. It brings out the band effective mass that is lighter than the free electron mass, can be either positive or negative, and is proportional to the inverse of the curvature of the energy dispersion curve. Here, the electron could be accelerated in the opposite direction of the applied force. All these peculiar band effects are explained in great details in terms of Bragg scattering. The mixing of $\mathbf{k}$ and $\mathbf{k} - \mathbf{G}$ waves is shown graphically. The momentum calculated from the mixed wave function shows that electron can be scattered to go backward by exchanging momentum with the entire lattice. Following the variation of the ratio between the $\mathbf{k}$ and $\mathbf{k} - \mathbf{G}$ waves, the momentum change calculated from the force impulse is examined. The negative effective mass turns out to be the consequence of electron back scattering, so that Newton's law still holds for the external force, effective mass, and group acceleration. The fact that the external force equals the time derivative of the crystal momentum is also proved.

The third part discusses holes in semiconductors. The objective is to show how the concepts of energy band can be applied. Since holes are

excitations in the valence band with missing electrons, the entire valence band needs to be considered in order to characterize the properties of holes. A total of six properties are introduced; including the wave vector, energy, velocity, charge, current, and the effective mass.

Bloch Wave Function of the Conduction Electron

6.1 Equivalent Potential

In the model of free electron Fermi gas, the electrons are treated as independent free Fermi particles. The electron-electron and electron-nucleus forces are neglected. In real solids, it is not too hard to see that every electron must feel the Coulomb repelling from other electrons in addition to the Coulomb attraction from nuclei. For a fuller description of the electronic properties in solids, we need to include proper potential terms in the Hamiltonian to count for these forces. In general, the electron-nuclei potential is $\sum_j Z_a e/|\mathbf{r}_i - \mathbf{r}_j|$, where Z_a is the atomic number and the summation is over all nuclei. This potential describes the attractive force exerted on i^{th} electron by all nuclei. The electron-electron potential is $\sum_{j(\neq i)} e/|\mathbf{r}_i - \mathbf{r}_j|$, which summarize the repelling forces between the i^{th} electron and every other electron. In addition, there are electron-phonon interactions because of the vibrating lattice. The problem becomes a huge one of 10^{23} coupled Schrödinger equations. If the solid under consideration is a crystalline solid, the periodicity may help arrange the equations in a more orderly fashion, but it is still a problem of 10^{23} electrons. There is usually no easy solution without making some approximations. A lot of ingenuity goes into the art of making approximations. This is where a successful model separates itself from the less effective ones. We will start with an intuitively simple approximation to reduce the Hamiltonian to a one-electron equation, where the static periodic core potential is highlighted.

When 10^{23} atoms come together to grow into a crystal, the massive nuclei form the lattice with a positive charge of $Z_a e$ on each site. The electron orbitals around a nucleus are most likely to be torn into deformation by the proximity of other nuclei. The outer orbitals must be affected more severely than the deeper ones. Despite of the fact that the exact electron configuration is hard to know, there is a general periodic nature we can intuitively conceive about the electron density distribution. Attracted by the nuclei, the electron density must be high around all lattice sites

where nuclei reside. The electron density must be the lowest in between lattice sites where the Coulomb attractive force is the weakest. The electron density variation through out the crystal is therefore likely to have the same periodicity of the lattice. It is not a bad assumption to require $n(\mathbf{r}) = n(\mathbf{r} + \mathbf{R})$ where $\mathbf{R}$ is any lattice vector. In fact, this assumption is very successful in dealing with x-ray diffraction, see Chapter 2. The periodic lattice of positive charges is covered with the electron cloud of $n(\mathbf{r})$. Therefore, the Coulomb potential of the nuclei must be reduced because of electronic screening. To a test electron wondering through the crystal, we may calculate the forces from all the other electrons and 10^{23} nuclei one by one, and then add them up as the full many-body potential describes. Or, we may calculate the force from the screened periodic potential alone as if there were no other electrons. If the results are very close to each other, we can claim that the screened potential is an equivalent potential. In the sense it is equivalent to the full many-body potential. Obviously, this comparison is only academic because we do not know how to calculate the force from the full potential of the crystal. If we knew how, we would not need the approximation of an equivalent potential at the first place. The verification must therefore come from comparison with experiments. If the comparison is favorable, we will have confidence in the equivalent potential approximation, which reduces the many-body coupled Schrödinger equations to a single one-electron Schrödinger equation. This approximation is an approximation of independent electrons and static lattice. The dependence on other electrons is included in the screening that gives rise to an effective one-election potential $U(\mathbf{r})$. The lattice vibration is considered to be frozen in time because the much faster response time of the screening electrons. Therefore, the problem of 10^{23} electrons is reduced to a one-election problem. The question of how does one choose the best effective potential is a very complicated one. We will return to this question later if time allows. Or, you will have chances to learn about the related techniques in more advanced studies. Instead of going into the detailed features of $U(\mathbf{r})$, we will only discuss the general periodicity of $U(\mathbf{r})$ and its important implications. The electron-phonon scattering is also put off for future discussions.

Since the crystal is periodic, the potential must satisfy $U(\mathbf{r}+\mathbf{R}) = U(\mathbf{r})$, where $\mathbf{R}$ is a lattices vector. Qualitatively, we may expect the periodic crystal potential to be made of atomic potentials. The composite potential resembles the free atomic potential around each nucleus on the lattice site and the potential flattens in the region between ion cores because the

proximity of other atoms. As shown in Fig. 6.1, the flattened core po-

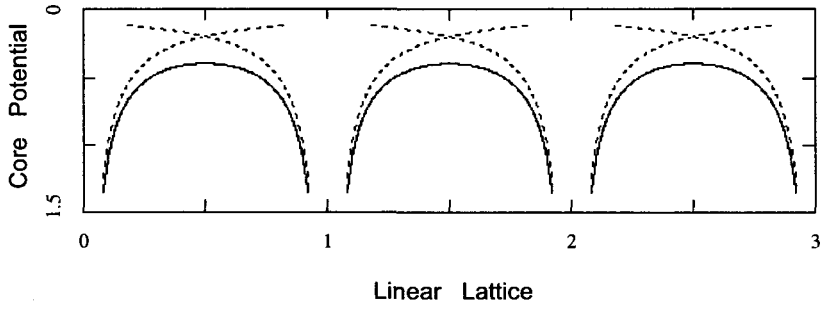


Fig. 6.1 Schematics of core potential in linear lattice: The atoms are located on sites labeled as 0, 1, 2, 3, and so forth. The dotted curves represent the atomic core potential in free space, proportional to $1/r$. The solid curves are the superimposed core potential when the atoms are brought together to form a crystal. No electron screening is considered in this figure. The net core charge is positive, and the potential for a positive test charge is higher when it approaches the core, as indicated by the arbitrary scale of the y-axis. For electrons, the potential energy near the core is lower because the negative electronic charge.

tential hints that outer orbital electrons will have a much greater chance to escape from the core potential in the crystal than from free standing atoms. Recalling the success of the free electron model, we may intuitively suspect that the outmost orbital electrons might not be bounded to a single core at all. It may move freely in the lattice with some characteristics of periodicity. This rather natural idea gives us a starting point to look for the electron wave function Ψ in the effective potential $U(\mathbf{r})$. Our improved understanding of crystals relies upon solving the one-electron Schrödinger equation,

$$\left(-\frac{\hbar}{2m}\nabla + U(\mathbf{r})\right)\Psi = \varepsilon\Psi. \tag{6.1}$$

The operator in the bracket is the Hamiltonian H, and we have $H\Psi = \varepsilon\Psi$.

6.2 Bloch Theorem — A Proof from the Vantage Point of Translation Symmetry in the Lattice

Bloch Theorem states that the electron wave function in a periodic potential is the product of a plane wave and a periodic function of the lattice, i.e.,

$$\Psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \cdot u_{\mathbf{k}}(\mathbf{r}), \tag{6.2}$$

and

$$u_{\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{\mathbf{k}}(\mathbf{r}). \tag{6.3}$$

The function $u_{\mathbf{k}}$ is periodic over any lattice vector $\mathbf{R}$. This implies that Bloch theorem can be written in another form

$$\Psi_{\mathbf{k}}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} \cdot \Psi_{\mathbf{k}}(\mathbf{r}). \tag{6.4}$$

We can prove the Bloch theorem from basic considerations of translation symmetry within the realm of quantum mechanics, without actually solving the Schrödinger equation in Eq. (6.1). Let us define the translation operator $T_{\mathbf{R}}$, which can shift the position variable of any function by $\mathbf{R}$ when operating on that function, i.e.,

$$T_{\mathbf{R}} f(\mathbf{r}) = f(\mathbf{r}+\mathbf{R}). \tag{6.5}$$

Since the crystal Hamiltonian is invariant under the translation of any lattice vector $\mathbf{R}$, we have

$$T_{\mathbf{R}} H\Psi = H(\mathbf{r}+\mathbf{R})\Psi(\mathbf{r}+\mathbf{R}) = H(\mathbf{r})\Psi(\mathbf{r}+\mathbf{R}) = H(\mathbf{r})T_{\mathbf{R}}\Psi(\mathbf{r}), \tag{6.6}$$

and

$$T_{\mathbf{R}} H = H T_{\mathbf{R}}. \tag{6.7}$$

Equation (6.7) shows that the translation operator $T_{\mathbf{R}}$ commutes with the Hamiltonian. The translation operator also has the following property,

$$T_{\mathbf{R}} T_{\mathbf{R}'} = T_{\mathbf{R}'} T_{\mathbf{R}} = T_{\mathbf{R}+\mathbf{R}'}, \tag{6.8}$$

because

$$T_{\mathbf{R}} T_{\mathbf{R}'}\Psi(\mathbf{r}) = T_{\mathbf{R}'} T_{\mathbf{R}}\Psi(\mathbf{r}) = \Psi(\mathbf{r}+\mathbf{R}+\mathbf{R}') = T_{\mathbf{R}+\mathbf{R}'}\Psi(\mathbf{r}). \tag{6.9}$$

Since H and $T_{\mathbf{R}}$ are commuting with each other, the eigenfunction of H is also an eigenfunction of $T_{\mathbf{R}}$. We have

$$T_{\mathbf{R}}\Psi = A(\mathbf{R})\Psi. \tag{6.10}$$

The eigenvalue of the translation operator is $A(\mathbf{R})$. By consecutive operations from Eqs. (6.9) and (6.10), we obtain

$$A(\mathbf{R}+\mathbf{R}') = A(\mathbf{R}) \cdot A(\mathbf{R}'). \tag{6.11}$$

From Eqs. (6.5) and (6.10), we notice that the wave function $\Psi(\mathbf{r}+\mathbf{R})$ is proportional to $\Psi(\mathbf{r})$ and the proportion constant is the eigenvalue of the translation operator $T_{\mathbf{R}}$, i.e.,

$$\Psi(\mathbf{r}+\mathbf{R}) = A(\mathbf{R}) \cdot \Psi(\mathbf{r}). \tag{6.12}$$

The translation symmetry dictates an identical probability for finding the electron at any two points related by a lattice vector $\mathbf{R}$, i.e., $\Psi^*\Psi(\mathbf{r}) = \Psi^*\Psi(\mathbf{r}+\mathbf{R})$. Therefore, the corresponding wave functions can only differ from each other by a phase factor. Applying this requirement to Eq. (6.12), we notice that $A(\mathbf{R})$ can only be a random phase factor, i.e., $A(\mathbf{R}) = e^{i\theta}$. Let $\mathbf{a}_i$ be the three primitive lattice vectors, and we may write

$$A(\mathbf{a}_i) = e^{i2\pi x_i}, \tag{6.13}$$

where x_i is an arbitrary constant to be determined. When $T_{\mathbf{a}_i}$ operates on $\Psi(\mathbf{r})$, the resulted wave function $\Psi(\mathbf{r}+\mathbf{a}_i)$ only shifts by a phase factor of $e^{i2\pi x_i}$ from $\Psi(\mathbf{r})$. Recall the lattice vector

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3,$$

then Eqs. (6.11) and (6.13) will lead to

$$\begin{aligned} A(\mathbf{R}) &= A(n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3) = A(n_1\mathbf{a}_1)A(n_2\mathbf{a}_2)A(n_3\mathbf{a}_3) \\ &= A(\mathbf{a}_1)^{n_1}A(\mathbf{a}_2)^{n_2}A(\mathbf{a}_3)^{n_3} = e^{i2\pi x_1 n_1} \cdot e^{i2\pi x_2 n_2} \cdot e^{i2\pi x_3 n_3} \\ &= e^{\mathbf{k}\cdot\mathbf{R}}. \end{aligned} \tag{6.14}$$

The last step of Eq. (6.14) is obtained from the orthogonal relationship between the reciprocal lattice vector $(\mathbf{b}_j)$ and the lattice vector $(\mathbf{a}_i)$, i.e.

$$\begin{aligned} \mathbf{a}_i \cdot \mathbf{b}_j &= 2\pi\delta_{ij}, \\ \mathbf{k} &= x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + x_3\mathbf{b}_3, \\ \mathbf{R} &= n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \end{aligned}$$

and

$$\mathbf{k} \cdot \mathbf{R} = (n_1x_1 + n_2x_2 + n_3x_3) \cdot 2\pi. \tag{6.15}$$

We have

$$T_{\mathbf{R}}\Psi(\mathbf{r}) = \Psi(\mathbf{r}+\mathbf{R}) = A(\mathbf{R})\Psi(\mathbf{r}) = e^{\mathbf{k}\cdot\mathbf{R}}\Psi(\mathbf{r}),$$

$$\Psi(\mathbf{r}+\mathbf{R}) = e^{\mathbf{k}\cdot\mathbf{R}}\Psi(\mathbf{r}), \tag{6.16}$$

or

$$\Psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{R}}u(\mathbf{r}). \tag{6.17}$$

The function $u(\mathbf{r})$ is a periodic function of the lattice, i.e., $u(\mathbf{r}+\mathbf{R}) = u(\mathbf{r})$. This is the proof of Bloch theorem from considerations of translation symmetry in the lattice, as well as from some basic properties of the Hamiltonian and wave function. For any eigenfunction $\Psi(\mathbf{r})$ of Eq. (6.1), there is a $\mathbf{k}$ vector that ensures $\Psi(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\Psi(\mathbf{r})$ for the Bravais lattice defined by $\mathbf{R}$. The eigenfunction can also be written as $\Psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{r})$ provided that $u(\mathbf{r})$ is a periodic function of the lattice. Note that the subscript $\mathbf{k}$ of $\Psi_{\mathbf{k}}$ and $u_{\mathbf{k}}$ can be added to specify the quantum state without affecting the proof. The Bloch wave function is very general. As long as the potential $U(\mathbf{r})$ is translation invariant, the electron wave function must be of the Bloch function form.

6.3 Bloch Theorem — A Proof from the Decomposition of Plane Waves in the Schrödinger Equation

The generalized periodic boundary condition for the electron wave function in a lattice is $\Psi(\mathbf{r}+N_i\mathbf{a}_i) = \Psi(\mathbf{r})$, where $i = 1$, 2, or 3 and $\mathbf{a}_i$ represents all three primitive vectors. The number of primitive cells N is equal to $N_1N_2N_3$. The wave function satisfying this boundary condition must also satisfy Bloch theorem where $\Psi(\mathbf{r}+N_i\mathbf{a}_i) = e^{iN_i\mathbf{k}\cdot\mathbf{a}_i}\Psi(\mathbf{r})$. Combining these two requirements, we end up with the necessary condition of $e^{iN_i\mathbf{k}\cdot\mathbf{a}_i} = 1$, which determines the allowed quantum states for the Bloch electrons. Recalling the relationship between $\mathbf{k}$ and $\mathbf{R}$ from Eq. (6.15), we have

$$e^{i\mathbf{k}\cdot\mathbf{R}} = e^{i2\pi N_i x_i} = 1 \longrightarrow x_i = \frac{m_i}{N_i},$$

and

$$\mathbf{k} = \frac{m_1}{N_1}\mathbf{b}_1 + \frac{m_2}{N_2}\mathbf{b}_2 + \frac{m_3}{N_3}\mathbf{b}_3. \tag{6.18}$$

The integers m_1, m_2, and m_3 in Eq. (6.18) define the allowed wave vectors for the Bloch wave function when the boundary condition of $\Psi(\mathbf{r}+N_i\mathbf{a}_i) = \Psi(\mathbf{r})$ is applied. When there is no additional boundary condition beside the translation invariance of $\Psi^*\Psi$, the Bloch wave function as expressed in Eq. (6.16) can have any $\mathbf{k}$ vector. The set of $\mathbf{k}$ vectors defined in Eq. (6.18) is just a subset.

Now, we can move on to solve the Schrödinger equation in Eq. (6.1). Along the way, we will see another proof of Bloch theorem. Since the potential in the one-electron Hamiltonian is a periodic function of the lattice, i.e. $U(\mathbf{r}+\mathbf{R}) = U(\mathbf{r})$ for any lattice vector $\mathbf{R}$, the plane-wave expansion of $U(\mathbf{r})$ will contain plane waves with the periodicity of the lattice. Consequently, the wave vectors must be limited to the reciprocal lattice vectors $\mathbf{G}$. We can write $U(\mathbf{r})$ in terms of

$$U(\mathbf{r}) = \sum_{\mathbf{G}} U_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}. \tag{6.19}$$

This is just the Fourier series that satisfies the condition $U(\mathbf{r}+\mathbf{R}) = U(\mathbf{r})$. Since a constant can always be added to the potential without changing the physics, we would choose the constant to make $U_0 = 0$ for convenience. The Fourier coefficients $U_{\mathbf{G}}$ can be obtained from the Fourier transform by integrating and averaging over the primitive cell of volume v, i.e.,

$$\begin{aligned} U_{\mathbf{G}} &= \frac{1}{v}\int e^{-i\mathbf{G}\cdot\mathbf{r}} U(\mathbf{r})\cdot d\mathbf{r} = \sum_{\mathbf{G}'} U_{\mathbf{G}'} \frac{1}{v}\int_0^{a_1,a_2,a_3} e^{-i(\mathbf{G}-\mathbf{G}')\cdot\mathbf{r}} d\mathbf{r} \\ &= \sum_{\mathbf{G}'} U_{\mathbf{G}'}\cdot\delta_{\mathbf{G}\mathbf{G}'}, \end{aligned} \tag{6.20}$$

and

$$U_0 = \frac{1}{v}\int U(\mathbf{r})\cdot d\mathbf{r} = 0. \tag{6.21}$$

Since the potential is real, that is $U^*(\mathbf{r}) = U(\mathbf{r})$. From the Fourier series expansion in Eq. (6.19) and the orthogonal property of Eq. (6.20) and Eq. (6.21), we will have $U_{-\mathbf{G}} = U_{\mathbf{G}}{}^*$. In the periodic lattice with translation symmetry, the inversion symmetry $U(\mathbf{r}) = U(-\mathbf{r})$ always holds. Similarly, we can deduce $U_{-\mathbf{G}} = U_{\mathbf{G}}$. These results are summarized here, i.e.,

$$\begin{aligned} U_{-\mathbf{G}} &= U_{\mathbf{G}}{}^*; \\ U_{\mathbf{G}} &= U_{-\mathbf{G}}\ ; \\ U_{\mathbf{G}} &= U_{\mathbf{G}}{}^*. \end{aligned} \tag{6.22}$$

It turns out that $U_{\mathbf{G}}$ is real and equal to $U_{-\mathbf{G}}$. We can also write the wave function in terms of plane wave expansion

$$\Psi(\mathbf{r}) = \sum_{\mathbf{q}} C_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}}. \tag{6.23}$$

Substituting Eqs. (6.19) and (6.23) into Eq. (6.1), the Schrödinger equation becomes

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + \sum_{\mathbf{G}} U_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}\right) \sum_{\mathbf{q}} C_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} = \varepsilon \sum_{\mathbf{q}} C_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}},$$

or

$$\sum_{\mathbf{q}} \frac{\hbar^2}{2m} q^2 C_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} + \sum_{\mathbf{G},\mathbf{q}} U_{\mathbf{G}} C_{\mathbf{q}} e^{i(\mathbf{G}+\mathbf{q})\cdot\mathbf{r}} = \varepsilon \sum_{\mathbf{q}} C_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}}. \tag{6.24}$$

Letting $\mathbf{q}' = \mathbf{G} + \mathbf{q}$ and change variables from $\mathbf{G} \to \mathbf{G}'$, and $\mathbf{q}' \to \mathbf{q}$, the second term of Eq. (6.24) transforms into the new notation.

$$\sum_{\mathbf{G},\mathbf{q}} U_{\mathbf{G}} C_{\mathbf{q}} e^{i(\mathbf{G}+\mathbf{q})\cdot\mathbf{r}} = \sum_{\mathbf{G},\mathbf{q}'} U_{\mathbf{G}} C_{\mathbf{q}'-\mathbf{G}} e^{i(\mathbf{q}')\cdot\mathbf{r}} = \sum_{\mathbf{G}',\mathbf{q}} U_{\mathbf{G}'} C_{\mathbf{q}-\mathbf{G}'} e^{i\mathbf{q}\cdot\mathbf{r}}. \tag{6.25}$$

Substituting the transformed expression for the second term in Eq. (6.24), the Schrödinger equation reduces to

$$\sum_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} \left(\left(\frac{\hbar^2}{2m} q^2 - \varepsilon\right) C_{\mathbf{q}} + \sum_{\mathbf{G}'} U_{\mathbf{G}'} C_{\mathbf{q}-\mathbf{G}'}\right) = 0. \tag{6.26}$$

At this point, we will impose the periodic boundary condition on the wave function by assuming that $e^{i\mathbf{q}\cdot\mathbf{r}}$ is periodic as moving from $\mathbf{r}$ to $\mathbf{r} + N_i\mathbf{a}_i$. The set of plane waves satisfying this periodic boundary condition is a complete set of orthogonal basis. Any function satisfying the same boundary condition can be expanded in terms of this orthogonal basis. This is what we did at the first place, expanding the wave function in terms of plane waves in Eq. (6.23). Let us change the variable $\mathbf{q}$ to $\mathbf{q}'$ in Eq. (6.23), multiply both sides of the equation with $e^{-i\mathbf{q}\cdot\mathbf{r}}$, and integrate over the volume $V = N_1\mathbf{a}_1 \cdot N_2\mathbf{a}_2 \times N_3\mathbf{a}_3$, i.e.,

$$C_{\mathbf{q}} = \frac{1}{V} \int e^{-i\mathbf{q}\cdot\mathbf{r}} \Psi(\mathbf{r}) \cdot d\mathbf{r}$$

$$= \sum_{\mathbf{q}'} C_{\mathbf{q}'} \frac{1}{V} \int_0^{N_1 a_1, N_2 a_2, N_3 a_3} e^{-i(\mathbf{q}-\mathbf{q}')\cdot\mathbf{r}} d\mathbf{r}$$

$$= \sum_{\mathbf{q}'} C_{\mathbf{q}'} \cdot \delta_{\mathbf{q}\mathbf{q}'}. \tag{6.27}$$

The boundary condition requires $(\mathbf{q} - \mathbf{q}') = \mathbf{q}'' = \sum_i (\frac{m_i}{N_i})\mathbf{b}_i$. Recalling that $(\frac{m_i}{N_i})\mathbf{b}_i \cdot N_i\mathbf{a}_i = m_i \cdot 2\pi$ and $\mathbf{b}_i \cdot \mathbf{a}_j = 0$, we can prove the orthogonal property in the last step of Eq. (6.27) by carrying out the integration. Let us take just one component as an illustration, i.e.,

$$\frac{1}{Na} \int_0^{Na} e^{-i(\mathbf{q}-\mathbf{q}')\cdot\mathbf{x}} d\mathbf{x} = \frac{1}{Na} \int_0^{Na} e^{-i\frac{m}{N}bx} d\mathbf{x} = \frac{1}{2m\pi} \int_0^{2m\pi} e^{-i\theta} d\theta$$

$$= 0 \qquad \text{when } \mathbf{q} \neq \mathbf{q}';$$

$$\frac{1}{Na} \int_0^{Na} e^{-i(\mathbf{q}-\mathbf{q}')\cdot\mathbf{x}} d\mathbf{x} = 1 \qquad \text{when } \mathbf{q} = \mathbf{q}'. \tag{6.28}$$

Similar integration applies to the proof of Eq. (6.20) for the potential $U(\mathbf{r})$, except that the reciprocal lattice vector $\mathbf{G}$ is involved there and the integration is over the primitive cell $v = \mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3$. The periodic condition for $U(\mathbf{r})$ is over one lattice spacing, and it is over $N_i\mathbf{a}_i$ for the wave function $\Psi(\mathbf{r})$.

Now, let us go back to the Schrödinger equation of Eq. (6.26). The entire factor in the bracket is a function of $\mathbf{q}$, and the equation is a special Fourier series that equals zero. Letting the function $f(\mathbf{q})$ represent the factor in the bracket of Eq. (6.26), we can write the Schrödinger equation in the following compact form, i.e.,

$$\sum_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} f(\mathbf{q}) = 0. \tag{6.29}$$

Changing the variable from $\mathbf{q}$ to $\mathbf{q}'$ in Eq. (6.29), multiplying both sides with $e^{-i\mathbf{q}\cdot\mathbf{r}}$, and integrating over the volume $V = N_1\mathbf{a}_1 \cdot N_2\mathbf{a}_2 \times N_3\mathbf{a}_3$, we have

$$\frac{1}{V} \sum_{\mathbf{q}'} f(\mathbf{q}') \int e^{-i(\mathbf{q}-\mathbf{q}')\cdot\mathbf{r}} d\mathbf{r} = \sum_{\mathbf{q}'} f(\mathbf{q}') \cdot \delta_{\mathbf{q}\mathbf{q}'} = 0,$$

and

$$f(\mathbf{q}) = 0. \tag{6.30}$$

Restoring $f(\mathbf{q})$ with the bracket quantity of Eq. (6.26), the Schrödinger equation is decomposed into terms specified by $\mathbf{q}$. So, we have

$$\left(\frac{\hbar^2}{2m}q^2 - \varepsilon\right) C_{\mathbf{q}} + \sum_{\mathbf{G}'} U_{\mathbf{G}'} C_{\mathbf{q}-\mathbf{G}'} = 0. \tag{6.31}$$

For convenience, we will change the variable from $\mathbf{q}$ to $\mathbf{k} - \mathbf{G}$, where $\mathbf{G}$ is the reciprocal lattice vector chosen to ensure that $\mathbf{k}$ lies in the first Brillouin zone.

$$\left(\frac{\hbar^2}{2m}(\mathbf{k}-\mathbf{G})^2 - \varepsilon\right) C_{\mathbf{k}-\mathbf{G}} + \sum_{\mathbf{G}'} U_{\mathbf{G}'} C_{\mathbf{k}-\mathbf{G}-\mathbf{G}'} = 0. \tag{6.32}$$

Since the summation over all $\mathbf{G}'$ or over all $\mathbf{G}' - \mathbf{G}$ make no difference, we will change $\mathbf{G}'$ to $\mathbf{G}' - \mathbf{G}$ in order to transform Eq. (6.32) to the following form,

$$\left(\frac{\hbar^2}{2m}(\mathbf{k}-\mathbf{G})^2 - \varepsilon\right) C_{\mathbf{k}-\mathbf{G}} + \sum_{\mathbf{G}'} U_{\mathbf{G}'-\mathbf{G}} C_{\mathbf{k}-\mathbf{G}'} = 0. \tag{6.33}$$

It should be emphasized that this equation is nothing but the Schrödinger equation in $\mathbf{k}$ space for the quantum state $\mathbf{k}$. The Schrödinger equation for different $\mathbf{k}$ states are separated as described by Eq. (6.33), and can be solved individually. This is a nice outcome from the boundary condition $\Psi(\mathbf{r}) = \Psi(\mathbf{r} + N_i\mathbf{a}_i)$. One important observation should be made here. Namely, only the coefficients $C_{\mathbf{k}}, C_{\mathbf{k}-\mathbf{G}'}, C_{\mathbf{k}-\mathbf{G}''}, C_{\mathbf{k}-\mathbf{G}'''}, \cdots$, and so forth are linked by the Schrödinger equation in Eq. (6.33). For each $\mathbf{k}$ in the first Brillouin zone, the wave function solution to the Schrödinger equation is a superposition of plane waves that can only contain wave vectors of $\mathbf{k}$ and $\mathbf{k}$ plus all $\mathbf{G}$ vectors. This wave function is a subset of the general Fourier series expansion in Eq. (6.23), Consequently, we have

$$\begin{aligned}\Psi_{\mathbf{r}}(\mathbf{r}) = \sum_{\mathbf{G}} C_{\mathbf{k}-\mathbf{G}} e^{i(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}} &= e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{G}} C_{\mathbf{k}-\mathbf{G}} e^{-i\mathbf{G}\cdot\mathbf{r}} \\ &= e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}),\end{aligned} \tag{6.34}$$

and

$$u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} C_{\mathbf{k}-\mathbf{G}} e^{-i\mathbf{G}\cdot\mathbf{r}}. \tag{6.35}$$

Equation (6.34) is in the form of Bloch wave function and $u_{\mathbf{k}}(\mathbf{r})$ of Eq. (6.35) satisfies the periodic condition of $u_{\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{\mathbf{k}}(\mathbf{r})$. We have reached the conclusion that the wave function solution of the one-electron Schrödinger equation is the Bloch wave function. This is not a surprising result, because the bloch wave function is assumed in selecting the allowed $\mathbf{k}$ states, see Eq. (6.18). Because the imposed boundary condition, the wave function of Eq. (6.34) is more restricted than the general Bloch wave function of Eq. (6.17). This is clearly stated by Eq. (6.35), where $u_{\mathbf{k}}(\mathbf{r})$ takes on a more specific functional form than the general $u_{\mathbf{k}}(\mathbf{r})$ in Eq. (6.17)

6.4 Some Remarks about Bloch Theorem

Bloch theorem gives the specific guideline on what wave-function form the electron must take in the periodical lattice potential. They are a special group of amplitude-modulated plane waves, i.e., $\Psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{r})$. The amplitude function $u(\mathbf{r})$ has the periodicity of the crystal lattice, and we have $u(\mathbf{r}) = u(\mathbf{r}+\mathbf{R})$ for any lattice vector $\mathbf{R}$. The wave function itself does not have this periodicity in general. Instead, we have $\Psi(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\Psi(\mathbf{r})$, i.e., the wave functions at two locations separated by a lattice vector $\mathbf{R}$ are different by a phase factor $e^{i\mathbf{k}\cdot\mathbf{R}}$. Consequently, we have

1. $\Psi(\mathbf{r}+\mathbf{R})^*\Psi(\mathbf{r}+\mathbf{R}) = \Psi(\mathbf{r})^*\Psi(\mathbf{r})$ for any lattice vector $\mathbf{R}$. The probability of finding the electron in the lattice has the same periodicity of the lattice.

2. The wave vector $\mathbf{k}$ is multi-valued. Adding or subtracting any reciprocal lattice vector $\mathbf{G}$ will not change the wave function, because $e^{i\mathbf{G}\cdot\mathbf{R}} = 1$.

3. For the special case when $\mathbf{k} = \mathbf{G}$, the wave function becomes periodic with the lattice, i.e. $\Psi(\mathbf{r}+\mathbf{R}) = \Psi(\mathbf{r})$ for any $\mathbf{R}$.

Applying the generalized periodic boundary condition requiring $\Psi(\mathbf{r}) = \Psi(\mathbf{r}+N_i\mathbf{a}_i)$, the wave vector $\mathbf{k}$ of Bloch electrons will be confined to the discrete values within the first Brillouin zone defined by

$$\mathbf{k} = \sum_{i=1}^{3} \frac{m_i}{N_i}\frac{2\pi}{a_i}\mathbf{l}_i, \quad m_i = 0, \pm 1, \pm 2, \cdots, \pm\frac{N_i}{2}. \tag{6.36}$$

For any $\mathbf{k}$ vector outside the zone, there is always an identical one inside the zone, obtainable by adding or subtracting an appropriate reciprocal lattice vector $\mathbf{G}$. The generalized periodic boundary condition in the lattice is also referred to as the Born-von Karman boundary condition.

We have applied the same boundary condition to electrons in the free space previously by requiring $\Psi(\mathbf{r}) = \Psi(\mathbf{r} + \mathbf{L})$ and arrived with a set of discrete $\mathbf{k}$ values similar to that of Eq. (6.36), except that $N_i\mathbf{a}_i$ is replaced with $\mathbf{L}$ since there is no crystal lattice. Because the momentum operator $(\hbar/i)\nabla$ commutes with the Hamiltonian in the free electron case, the wave function Ψ is also an eigenstate of the momentum operator. So, we have

$$\frac{\hbar}{i}\nabla\Psi = \mathbf{p}\Psi. \tag{6.37}$$

Since the free electron wave function Ψ is a plane wave proportional to $e^{i\mathbf{k}\cdot\mathbf{r}}$, we obtain the momentum eigenvalue $\mathbf{p} = \hbar\mathbf{k}$, i.e., the wave-vector quantity $\hbar\mathbf{k}$ is the electron momentum. In crystals, the presence of periodic potential limits the translation symmetry to a lesser degree in comparing to the full symmetry of free space. The momentum operator no longer commutes with the Hamiltonian. As a result, the energy eigenstate is not the simultaneous eigenstate of the momentum operator anymore. This can be shown by operating $(\hbar/i)\nabla$ on the Bloch wave function, i.e.

$$\frac{\hbar}{i}\nabla\Psi = \frac{\hbar}{i}\nabla\left(e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{r})\right) = \left[\hbar\mathbf{k} + \frac{\hbar}{i}\cdot\frac{1}{u(\mathbf{r})}\cdot\nabla u(\mathbf{r})\right]\Psi \neq \hbar\mathbf{k}\Psi. \tag{6.38}$$

This result shows that the momentum-like quantity $\hbar\mathbf{k}$ of the Bloch electron is not the true physical momentum $\mathbf{p}$ of the electron. Just like the free-electron momentum $\mathbf{p}(= \hbar\mathbf{k})$ that reflects the full translation symmetry of free space, the crystal momentum $\hbar\mathbf{k}(\neq \mathbf{p})$ of the Bloch electron characterizes the lesser degree of translation symmetry in the periodic lattice potential.

Bloch theorem requires that wave functions with their wave vectors related by reciprocal lattice vectors are identical, i.e., $\Psi_{\mathbf{k}} = \Psi_{\mathbf{k}+\mathbf{G}}$. This important property reflects in the separation of $\mathbf{k}$ vectors when the plane-wave expansion of the wave function is substituted into the Schrödinger equation. All coefficients $C_{\mathbf{k}-\mathbf{G}}$ for a given $\mathbf{k}$ are related through the Schrödinger equation in $\mathbf{k}$ space, i.e., Eq. (6.33)

$$\left(\frac{\hbar^2}{2m}(\mathbf{k}-\mathbf{G})^2 - \varepsilon\right)C_{\mathbf{k}-\mathbf{G}} + \sum_{\mathbf{G}'} U_{\mathbf{G}'-\mathbf{G}}C_{\mathbf{k}-\mathbf{G}'} = 0.$$

This represents n simultaneous linear equations with n unknowns of $C_{\mathbf{k}-\mathbf{G}}$. If there is a set of nontrivial solution for $C_{\mathbf{k}-\mathbf{G}}$, the determinant of these equations must be equal to zero. This condition will generate an equation of the n^{th} order of ε. There will be n roots for ε. So, for a given wave vector $\mathbf{k}$, we will have many energy values, as many as the number of non-vanishing $C_{\mathbf{k}-\mathbf{G}}$. The energy eigenvalue of Bloch electron is usually labeled with two quantum numbers as $\varepsilon_{n,\mathbf{k}}$, where n characterizes the multiple solutions for a given $\mathbf{k}$. For a fixed quantum number n, $\varepsilon_{n,\mathbf{k}}$ changes as a function of $\mathbf{k}$. In macroscopic samples, the spacing between adjacent $\mathbf{k}$ levels defined in Eq. (6.36) is actually infinitesimal. The energy spectrum $\varepsilon_n(\mathbf{k})$ approaches a continuous function of $\mathbf{k}$, and $\varepsilon_n(\mathbf{k})$ is called an energy band.

Electrons in the Weak Periodic Potential

6.5 One-electron Schrödinger Equation

Most of atomic orbital electrons stay closely attached to the nucleus to form the ion core. The ion cores form the basis over the underlying crystal lattice. The remaining electrons are valence electrons that spread out loosely in the crystal. The core electrons screen the nucleus of $Z_a e$ positive charges to result in the effective ion-core potential, which is further screened by valence electrons. We end up with an effective Coulomb potential $U(\mathbf{r})$ that has reduced range from the bare nucleus. Unless the test electron gets too close to the ion core, the periodic potential it sees is rather weak. Therefore, to model the crystal with independent valence electrons in a conceptual weak periodic potential may not be a bad starting point. Despite of the lack of interaction details, this simple one-electron theory will reveal the single most important consequence of the periodic potential, i.e., the concept of electron energy bands and energy gaps.

In the one-electron theory with the conceptual equivalent potential, the electron is treated independently from each other. The many-electron effects are mostly ignored except for the screening of the core potential. The Schrödinger equation reduces to a one-electron equation, and the electron wave function is of the form of Bloch function. For a given wave vector $\mathbf{k}$, the plane-wave expansion of the wave function only contains terms of $e^{i(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}}$, with $\mathbf{G}$ running through all reciprocal lattice vectors. The weak interaction of the electron with the periodic potential $U(\mathbf{r})$ can be considered as a perturbation that will modify the free-electron energy. Let us

denote the free-electron energy as $\varepsilon^0_{\mathbf{k}-\mathbf{G}} = \hbar^2(\mathbf{k}-\mathbf{G})^2/2m$, and write down the wave function and Schrödinger equation. We have

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} C_{\mathbf{k}-\mathbf{G}} e^{i(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}},$$

and

$$(\varepsilon^0_{\mathbf{k}-\mathbf{G}} - \varepsilon)C_{\mathbf{k}-\mathbf{G}} + \sum_{\mathbf{G}'} U_{\mathbf{G}'-\mathbf{G}} C_{\mathbf{k}-\mathbf{G}'} = 0. \tag{6.39}$$

For a given $\mathbf{G}$ vector in the Schrödinger equation, the summation over $\mathbf{G}'$ runs through all possible reciprocal lattice vectors. When $\mathbf{G}$ runs through all possible values, the Schrödinger equation generates a set of coupled equations. The number of equations equals the number of $\mathbf{G}$ vectors. The issue is to solve for all coefficients $C_{\mathbf{k}-\mathbf{G}}$, eigenvalue ε, and the wave function $\Psi_{\mathbf{k}}$. Let us examine one of the many Schrödinger equations by picking a specific $\mathbf{G}$, say $\mathbf{G} = \mathbf{G}_1$, and we have

$$(\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}_1})C_{\mathbf{k}-\mathbf{G}_1} = \sum_{\mathbf{G}} U_{\mathbf{G}-\mathbf{G}_1} C_{\mathbf{k}-\mathbf{G}}. \tag{6.40}$$

Here, $\mathbf{G}_1$ is a specific reciprocal lattice vector, not a variable that can run through all $\mathbf{G}$ values any more. We have also replaced the summation index $\mathbf{G}'$ with $\mathbf{G}$ on the right hand side for computation convenience. Let us imagine that we can turn the potential on and off. At this moment, let us turn the potential off to see what happens. The system must turn into free electrons with $U(\mathbf{r}) = 0$. The electron wave function reduces back to a single plane wave with the wave vector equal to $\mathbf{k} - \mathbf{G}_1$, and the energy returns to the free-electron energy. So, we have

$$C_{\mathbf{k}-\mathbf{G}} = 0 \qquad \text{when} \qquad \mathbf{G} \neq \mathbf{G}_1,$$

and

$$\varepsilon = \varepsilon^0_{\mathbf{k}-\mathbf{G}_1};$$

$$\Psi_{\mathbf{k}}(\mathbf{r}) = C_{\mathbf{k}-\mathbf{G}_1} \cdot e^{i(\mathbf{k}-\mathbf{G}_1)\cdot\mathbf{r}}. \tag{6.41}$$

Now, we will turn the potential on, slowly from zero to a weak $U(\mathbf{r})$, and watch what happens. It makes physical sense to anticipate the energy to deviate gradually from the free-electron value, and plane-wave components of other $\mathbf{k} - \mathbf{G}$ vectors to grow in the wave function. These anticipated asymptotic behaviors set some basic guidelines for us to work on the

problem. Namely, in the limit of vanishing $U(\mathbf{r})$, $C_{\mathbf{k}-\mathbf{G}}$ vanishes except for $\mathbf{G} = \mathbf{G}_1$, and ε approaches the corresponding free-electron value at $\varepsilon^0_{\mathbf{k}-\mathbf{G}_1} = \hbar^2(\mathbf{k} - \mathbf{G}_1)^2/2m$. For weak $U(\mathbf{r})$, ε should not be too far away from $\varepsilon^0_{\mathbf{k}-\mathbf{G}_1}$, and the magnitude of non-vanishing $C_{\mathbf{k}-\mathbf{G}}$ terms will depend on the spacing between $\varepsilon^0_{\mathbf{k}-\mathbf{G}_1}$ and $\varepsilon^0_{\mathbf{k}-\mathbf{G}}$.

To illustrate these points, we plot the free electron energy $\varepsilon = \hbar^2\mathbf{k}^2/2m$ as a function of $\mathbf{k}$ along one crystal axis in Fig. 6.2. Before we turn on the periodic potential, let us assume that the electron is in the free-electron state $\mathbf{k}_1 = \mathbf{k} - \mathbf{G}_1$ located near the first Brillouin zone boundary. Turning on the periodic potential will raise the probability to scatter this electron into other states, but Bloch theorem requires that the new states must relate to $\mathbf{k}_1$ by a $\mathbf{G}$ vector. Figure 6.2 shows two examples, involving the smallest $\mathbf{G}$ vector. For the $\mathbf{k}_1$ state located just inside the zone boundary, it will be scattered into the state $\mathbf{k}_1 + \mathbf{G}$ located just outside the zone. The difference in free-electron energy, $\Delta\varepsilon_1 = (\hbar^2{\mathbf{k}_1}^2/2m - \hbar^2(\mathbf{k}_1 + \mathbf{G})^2/2m)$, is small in this case. If $\Delta\varepsilon_1$ is comparable to or smaller than $U_{\mathbf{G}}$, the probability for this scattering to occur would be high and $C_{\mathbf{k}_1+\mathbf{G}}$ would be respectively large. Also, the correction $U_{\mathbf{G}}$ makes on ε_1 would be comparable to the size of $\Delta\varepsilon_1$. For the $\mathbf{k}_2$ state located deeper inside the zone, it will be scattered into the state $\mathbf{k}_2 + \mathbf{G}$ located further outside the zone. The difference in free-electron energy, $\Delta\varepsilon_2 = (\hbar^2{\mathbf{k}_2}^2/2m - \hbar^2(\mathbf{k}_2 + \mathbf{G})^2/2m)$, is large in this case, and we have $\Delta\varepsilon_2 \gg U_{\mathbf{G}}$. Energetically, it would not be favorable for this scattering to take place, and its probability of occurrence must be low. Consequently, $C_{\mathbf{k}_2+\mathbf{G}}$ would be small. The correction $U_{\mathbf{G}}$ makes on ε_2 in this case would be negligibly small in comparison with $\Delta\varepsilon_2$. If there is only one state in the energy range U, we will define it as non-degenerate. If there are m states bundled in the energy range U, even they do not have the same energy, we will define the situation as m-fold degeneracy. The perturbation effect of U on the electron energy and wave function depends on the degree of degeneracy.

6.6 Non-degenerate Case — A Second Order Effect

In the plane-wave expansion of the wave function, the allowed terms for a given $\mathbf{k}$ vector are those related to $\mathbf{k}$ by reciprocal lattice vectors $\mathbf{G}$. This is the statement of Bloch theorem when the periodic boundary condition of Eq. (6.18) is imposed. Each plane wave is related to a free electron state that has a free electron energy of $\varepsilon^0_{\mathbf{k}-\mathbf{G}} = \hbar^2(\mathbf{k} - \mathbf{G})^2/2m$. If there are no

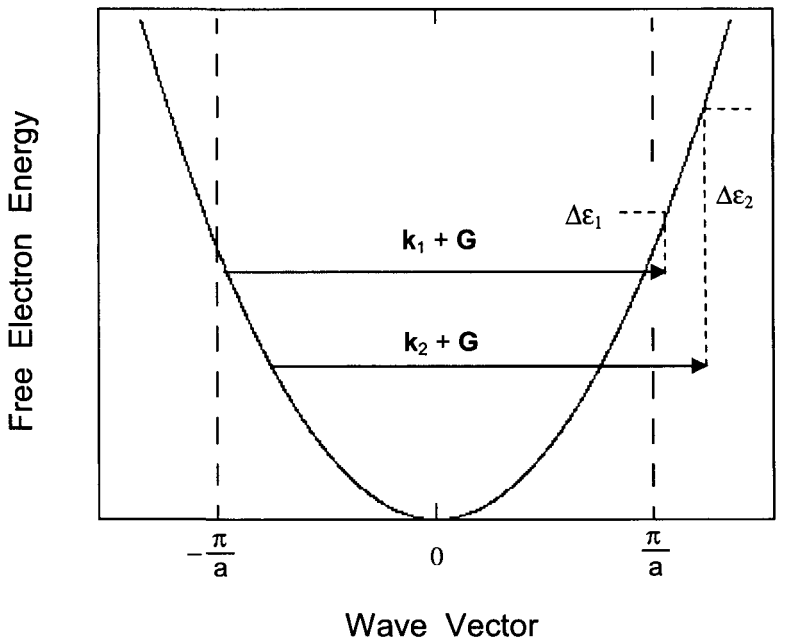


Fig. 6.2 The free-electron energy $\varepsilon \propto k^2$ curve. The state $\mathbf{k}_1$ near the first Brillouin zone boundary is scattered into $\mathbf{k}_1 + \mathbf{G}$ near the opposite zone boundary. $\Delta\varepsilon_1$ is the difference of free-electron energy between these two states. For $\Delta\varepsilon_1 < U_G$, the occurrence of this scattering is highly likely. On the other hand, for the state $\mathbf{k}_2$ somewhat away from the zone boundary, the probability to get scattered into $\mathbf{k}_2 + \mathbf{G}$ is relatively small, because $\Delta\varepsilon_2 \gg U_G$.

other free electron energy levels in the vicinity of $\varepsilon^0_{\mathbf{k}-\mathbf{G}_1} = \hbar^2(\mathbf{k} - \mathbf{G}_1)^2/2m$ within the range of U, we consider the energy level $\varepsilon^0_{\mathbf{k}-\mathbf{G}_1}$ to be non-degenerate. Mathematically the non-degenerate condition is written as:

$$|\varepsilon^0_{\mathbf{k}-\mathbf{G}_1} - \varepsilon^0_{\mathbf{k}-\mathbf{G}}| \gg U \qquad \text{for all } \mathbf{G} \neq \mathbf{G}_1. \tag{6.42}$$

In the limit of weak U, we know that $C_{\mathbf{k}-\mathbf{G}_1}$ is large in comparison with other non-vanishing $C_{\mathbf{k}-\mathbf{G}}$. The Schrödinger equation threads these Fourier coefficients together. We can solve for $C_{\mathbf{k}-\mathbf{G}}$ in terms of other $C_{\mathbf{k}-\mathbf{G}'}$ and $C_{\mathbf{k}-\mathbf{G}_1}$. Singling out the dominant term $C_{\mathbf{k}-\mathbf{G}_1}$ from the others, we have

$$\left(\frac{\hbar^2}{2m}(\mathbf{k} - \mathbf{G})^2 - \varepsilon\right) C_{\mathbf{k}-\mathbf{G}} + \sum_{\mathbf{G}'} U_{\mathbf{G}'-\mathbf{G}} C_{\mathbf{k}-\mathbf{G}'} = 0,$$

and

$$C_{\mathbf{k}-\mathbf{G}} = \frac{U_{\mathbf{G}_1-\mathbf{G}}}{\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}}} \cdot C_{\mathbf{k}-\mathbf{G}_1} + \sum_{\mathbf{G}' \neq \mathbf{G}_1} \frac{U_{\mathbf{G}'-\mathbf{G}}}{\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}}} \cdot C_{\mathbf{k}-\mathbf{G}'}. \tag{6.43}$$

The energy difference $(\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}})$ approaches $(\varepsilon^0_{\mathbf{k}-\mathbf{G}_1} - \varepsilon^0_{\mathbf{k}-\mathbf{G}})$ in the limit of $U \to 0$. The non-degenerate condition says that $|(\varepsilon^0_{\mathbf{k}-\mathbf{G}_1} - \varepsilon^0_{\mathbf{k}-\mathbf{G}})| \gg U$ for all $\mathbf{G} \neq \mathbf{G}_1$. Therefore, we expect that the ratio $U_{\mathbf{G}_1-\mathbf{G}}/(\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}})$ is much smaller than 1 for small U, and goes to zero as U goes to zero. By iterative substitution, we can expand $C_{\mathbf{k}-\mathbf{G}}$ in terms of a power series of $U/(\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}})$ multiplied by the dominant component $C_{\mathbf{k}-\mathbf{G}_1}$, i.e.,

$$C_{\mathbf{k}-\mathbf{G}} = \frac{U_{\mathbf{G}_1-\mathbf{G}}}{\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}}} \cdot C_{\mathbf{k}-\mathbf{G}_1} + \sum_{\mathbf{G}' \neq \mathbf{G}_1} \left(\frac{U_{\mathbf{G}'-\mathbf{G}}}{\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}}}\right)\left(\frac{U_{\mathbf{G}_1-\mathbf{G}'}}{\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}'}}\right) \cdot C_{\mathbf{k}-\mathbf{G}_1} + \cdots,$$

and

$$C_{\mathbf{k}-\mathbf{G}} \approx \frac{U_{\mathbf{G}_1-\mathbf{G}}}{\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}}} \cdot C_{\mathbf{k}-\mathbf{G}_1} + \Omega\left(\frac{U^2}{\Delta\varepsilon^2}\right). \tag{6.44}$$

The notation $\Omega\left(\frac{U^2}{\Delta\varepsilon^2}\right)$ represents the second order term of $U/(\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}})$. With the minor components $C_{\mathbf{k}-\mathbf{G}}$ calculated in terms of the dominant term $C_{\mathbf{k}-\mathbf{G}_1}$, we can substitute them back into the Schrödinger equation to eliminate all minor variables $C_{\mathbf{k}-\mathbf{G}}$. So, we have

$$\begin{aligned}(\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}_1})C_{\mathbf{k}-\mathbf{G}_1} &= \sum_{\mathbf{G}} U_{\mathbf{G}-\mathbf{G}_1} \cdot C_{\mathbf{k}-\mathbf{G}} \\ &\approx \sum_{\mathbf{G}} \frac{U_{\mathbf{G}_1-\mathbf{G}} \cdot U_{\mathbf{G}-\mathbf{G}_1}}{\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}}} \cdot C_{\mathbf{k}-\mathbf{G}_1} + \Omega\left(\frac{U^3}{\Delta\varepsilon^2}\right).\end{aligned} \tag{6.45}$$

Since the third order term of $\left(\frac{U^3}{\Delta\varepsilon^2}\right)$ is negligibly small, we can drop this term, eliminate $C_{\mathbf{k}-\mathbf{G}_1}$, and solve for ε, i.e.,

$$\begin{aligned}\varepsilon &\approx \varepsilon^0_{\mathbf{k}-\mathbf{G}_1} + \sum_{\mathbf{G}} \frac{|U_{\mathbf{G}_1-\mathbf{G}}|^2}{\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}}} \\ &\approx \varepsilon^0_{\mathbf{k}-\mathbf{G}_1} + \sum_{\mathbf{G}} \frac{|U_{\mathbf{G}_1-\mathbf{G}}|^2}{\varepsilon^0_{\mathbf{k}-\mathbf{G}_1} - \varepsilon^0_{\mathbf{k}-\mathbf{G}}}.\end{aligned} \tag{6.46}$$

Here we have taken the approximation $\varepsilon \approx \varepsilon^0_{\mathbf{k}-\mathbf{G}_1}$ in the second order term and dropped the 3^{rd} order term. Also note that $U_{\mathbf{G}-\mathbf{G}_1} = U_{\mathbf{G}_1-\mathbf{G}}$ from

Eq. (6.22). The effect of the weak potential U is to shift the free electron energy $\varepsilon^0_{\mathbf{k}-\mathbf{G}_1}$ by an amount proportional to U^2, if $\varepsilon^0_{\mathbf{k}-\mathbf{G}_1}$ is not degenerate. Whether $\varepsilon^0_{\mathbf{k}-\mathbf{G}_1}$ gets raised or lower depends on whether $\varepsilon^0_{\mathbf{k}-\mathbf{G}_1}$ is higher or lower than each contributing $\varepsilon^0_{\mathbf{k}-\mathbf{G}}$.

6.7 Degenerate Case — The First Order Effect

In the degenerate case, there are m reciprocal lattice vectors $(\mathbf{G}_1, \cdots, \mathbf{G}_m)$ with corresponding free-electron energy $(\varepsilon^0_{\mathbf{k}-\mathbf{G}_1}, \cdots, \varepsilon^0_{\mathbf{k}-\mathbf{G}_m})$, all within an energy range smaller than U but far apart from the other $\varepsilon^0_{\mathbf{k}-\mathbf{G}}$ on the scale of U. We have the algebra worked out in the following. The degenerate condition is given by

$$|\varepsilon^0_{\mathbf{k}-\mathbf{G}} - \varepsilon^0_{\mathbf{k}-\mathbf{G}_i}| \gg U, \quad \text{for } i = 1, \cdots, m, \text{ and } \mathbf{G} \neq \mathbf{G}_1, \cdots, \mathbf{G}_m. \tag{6.47}$$

For the specific $\mathbf{G}_i$ among the m degenerate levels, we group all $C_{\mathbf{k}-\mathbf{G}_j}$ of the m degenerate levels together in the right hand side of Schrödinger equation, i.e.,

$$(\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}_i}) C_{\mathbf{k}-\mathbf{G}_i} = \sum_{j=1}^{m} U_{\mathbf{G}_j-\mathbf{G}_i} C_{\mathbf{k}-\mathbf{G}_j} + \sum_{\mathbf{G}\neq\mathbf{G}_1,\cdots,\mathbf{G}_m} U_{\mathbf{G}-\mathbf{G}_i} C_{\mathbf{k}-\mathbf{G}}, \quad \text{where } i = 1, \cdots, m. \tag{6.48}$$

All m $C_{\mathbf{k}-\mathbf{G}_j}$ in the first summation are dominant ones, and all $C_{\mathbf{k}-\mathbf{G}}$ in the second summation are minor ones. Now, we want to express the minor $C_{\mathbf{k}-\mathbf{G}}$ in terms of the m dominant $C_{\mathbf{k}-\mathbf{G}_j}$. From Eq. (6.48), we have

$$C_{\mathbf{k}-\mathbf{G}} = \frac{1}{\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}}} \left(\sum_{j=1}^{m} U_{\mathbf{G}_j-\mathbf{G}} C_{\mathbf{k}-\mathbf{G}_j} + \sum_{\mathbf{G}'\neq\mathbf{G}_1,\cdots,\mathbf{G}_m} U_{\mathbf{G}'-\mathbf{G}} C_{\mathbf{k}-\mathbf{G}'} \right), \quad \text{where } \mathbf{G} \neq \mathbf{G}_1, \cdots, \mathbf{G}_m. \tag{6.49}$$

Substituting $C_{\mathbf{k}-\mathbf{G}}$ of Eq. (6.49) into Eq. (6.48), we have

$$(\varepsilon - \varepsilon^0_{\mathbf{k}-\mathbf{G}_i}) C_{\mathbf{k}-\mathbf{G}_i} = \sum_{j=1}^{m} U_{\mathbf{G}_j-\mathbf{G}_i} C_{\mathbf{k}-\mathbf{G}_j}$$

$$+\sum_{j=1}^{m}\left(\sum_{\mathbf{G}\neq\mathbf{G}_1,\cdots,\mathbf{G}_m}\frac{U_{\mathbf{G}-\mathbf{G}_i}U_{\mathbf{G}_j-\mathbf{G}}}{\varepsilon-\varepsilon^0_{\mathbf{k}-\mathbf{G}}}\right)\cdot C_{\mathbf{k}-\mathbf{G}_j}+\Omega\left(\frac{U^3}{\Delta\varepsilon^2}\right),$$

$$(\varepsilon-\varepsilon^0_{\mathbf{k}-\mathbf{G}_i})C_{\mathbf{k}-\mathbf{G}_i}=\sum_{j=1}^{m}U_{\mathbf{G}_j-\mathbf{G}_i}C_{\mathbf{k}-\mathbf{G}_j}+\Omega\left(\frac{U^2}{\Delta\varepsilon}\right)+\Omega\left(\frac{U^3}{\Delta\varepsilon^2}\right)+\cdots$$

$$\text{for } i=1,\cdots,m. \tag{6.50}$$

The right hand side is the correction terms to the free electron energy $\varepsilon^0_{\mathbf{k}-\mathbf{G}_i}$ in descending magnitudes. The leading term is proportional to U. The $\Omega\left(\frac{U^2}{\Delta\varepsilon}\right)$ and $\Omega\left(\frac{U^3}{\Delta\varepsilon^2}\right)$ terms are relatively small in comparison, so they can be dropped. We have the approximate result,

$$(\varepsilon-\varepsilon^0_{\mathbf{k}-\mathbf{G}_i})C_{\mathbf{k}-\mathbf{G}_i}=\sum_{j=1}^{m}U_{\mathbf{G}_j-\mathbf{G}_i}C_{\mathbf{k}-\mathbf{G}_j}\qquad\text{for } i=1,\cdots,m. \tag{6.51}$$

There are m coupled linear equations with m unknowns. In order to have non-trivial solutions for $C_{\mathbf{k}-\mathbf{G}_1}, C_{\mathbf{k}-\mathbf{G}_2},\cdots$, and $C_{\mathbf{k}-\mathbf{G}_m}$, the determinant of the coefficients must be zero. This condition leads to an equation of the m^{th} order for ε. The roots are the m characteristic energy eigenvalues, where the degeneracy is removed by the perturbation of U.

6.8 Two-fold Degeneracy and Energy Bands

The most illustrative special case of the general m-fold degeneracy is the two-fold degeneracy, where two free-electron levels are within a range smaller than U. The m coupled Schrödinger equations reduce to two coupled equations,

$$(\varepsilon-\varepsilon^0_{\mathbf{k}-\mathbf{G}_1})C_{\mathbf{k}-\mathbf{G}_1}=U_{\mathbf{G}_2-\mathbf{G}_1}C_{\mathbf{k}-\mathbf{G}_2},$$

and

$$(\varepsilon-\varepsilon^0_{\mathbf{k}-\mathbf{G}_2})C_{\mathbf{k}-\mathbf{G}_2}=U_{\mathbf{G}_1-\mathbf{G}_2}C_{\mathbf{k}-\mathbf{G}_1}. \tag{6.52}$$

Here, we have set $U_0=0$ by properly choosing an additive constant in the potential $U(\mathbf{r})$. The free-electron energies $\varepsilon^0_{\mathbf{k}-\mathbf{G}_1}$ and $\varepsilon^0_{\mathbf{k}-\mathbf{G}_2}$ satisfy the condition $|\varepsilon^0_{\mathbf{k}-\mathbf{G}_1}-\varepsilon^0_{\mathbf{k}-\mathbf{G}_2}|\ll U$ because of the underlying two-fold degeneracy. Letting $\mathbf{q}=\mathbf{k}-\mathbf{G}_1$ and $\mathbf{G}=\mathbf{G}_2-\mathbf{G}_1$, we have

$$(\varepsilon-\varepsilon^0_{\mathbf{q}})C_{\mathbf{q}}=U_{\mathbf{G}}C_{\mathbf{q}-\mathbf{G}},$$

and

$$(\varepsilon - \varepsilon^0_{\mathbf{q}-\mathbf{G}})C_{\mathbf{q}-\mathbf{G}} = U_{-\mathbf{G}}C_{\mathbf{q}} = U^*_{\mathbf{G}}C_{\mathbf{q}}. \tag{6.53}$$

Two-fold degeneracy requires $\varepsilon^0_{\mathbf{q}} \approx \varepsilon^0_{\mathbf{q}-\mathbf{G}}$, and $|\varepsilon^0_{\mathbf{q}} - \varepsilon^0_{\mathbf{q}-\mathbf{G}'}| \gg U$ for all other $\mathbf{G}' \neq \mathbf{G}$. The condition $\varepsilon^0_{\mathbf{q}} = \varepsilon^0_{\mathbf{q}-\mathbf{G}}$ implies $|\mathbf{q}| = |\mathbf{q} - \mathbf{G}|$. Similar to von Laue's diffraction rule for x-ray diffraction, the electron wave vectors $\mathbf{q}$ and $\mathbf{q} - \mathbf{G}$ must point on the plane bisecting the reciprocal lattice vector $\mathbf{G}$, where the electron gets scattered elastically by the periodical potential between the $\mathbf{q}$ and $\mathbf{q} - \mathbf{G}$ states. The bisecting plane by definition is a Brillouin zone boundary in $\mathbf{k}$ space. It is also known as the Bragg plane because reflections from the plane satisfy the Bragg law. The following figure illustrates the geometry of all wave vectors involved in the two-fold degeneracy case. Now let us solve the coupled Schrödinger equations in the

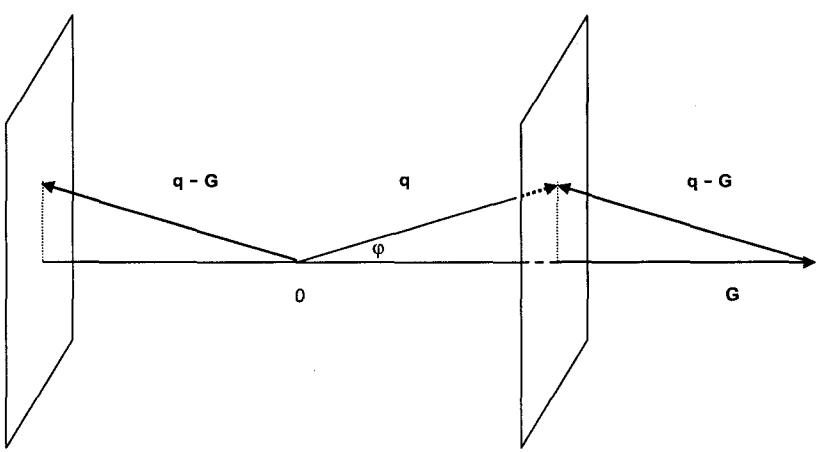


Fig. 6.3 Schematics of portions of the Brillouin zone boundary. The zone boundary is made of bisecting planes of $\mathbf{G}$ vectors. If the $\mathbf{G}$ vectors are the shortest reciprocal lattice vectors, the boundary surface belongs to the first Brillouin zone. For any wave vector $\mathbf{q}$ pointing on the boundary, there is the wave vector $\mathbf{q} - \mathbf{G}$ pointing on the opposite boundary. The $\mathbf{q} - \mathbf{G}$ vector is sometimes shown as being shifted in parallel to the end of the $\mathbf{G}$ vector, as shown in the right side of the figure. From the geometry, we have $|\mathbf{q}| = |\mathbf{q} - \mathbf{G}|$, and $|\mathbf{q} \cdot \mathbf{G}|/G = G/2$. The dash line on the boundary plane is equal to $|\mathbf{q} - \mathbf{G}/2|$. With the help of Fig 6.2, we can verify $\varepsilon^0_q = \varepsilon^0_{\mathbf{q}-\mathbf{G}}$.

case of two-fold degeneracy. If both $C_{\mathbf{q}}$ and $C_{\mathbf{q}-\mathbf{G}}$ have non-zero solutions,

the determinant of coefficients must equal zero, i.e.,

$$\begin{pmatrix} \varepsilon - \varepsilon^0_{\mathbf{q}} & -U_{\mathbf{G}} \\ -U^*_{\mathbf{G}} & \varepsilon - \varepsilon^0_{\mathbf{q}-\mathbf{G}} \end{pmatrix} = 0,$$

and

$$(\varepsilon - \varepsilon^0_{\mathbf{q}})(\varepsilon - \varepsilon^0_{\mathbf{q}-\mathbf{G}}) = |U_{\mathbf{G}}|^2. \tag{6.54}$$

The roots of this quadratic equation of ε are

$$\varepsilon = \frac{1}{2}(\varepsilon^0_{\mathbf{q}} + \varepsilon^0_{\mathbf{q}-\mathbf{G}}) \pm \sqrt{\left(\frac{\varepsilon^0_{\mathbf{q}} - \varepsilon^0_{\mathbf{q}-\mathbf{G}}}{2}\right)^2 + |U_{\mathbf{G}}|^2}. \tag{6.55}$$

The electron energy $\varepsilon(\mathbf{q})$, when plotted as a function of $\mathbf{q}$, reveals the energy bands. In general, $\mathbf{q}$ is not necessarily in the direction of $\mathbf{G}$, see Fig 6.3. When it does, Equation (6.55) becomes a simpler special case that can be easily plotted as shown in Fig. 6.4. In this figure, we see two energy bands separated by an energy gap located at $\mathbf{q} = \frac{\mathbf{G}}{2}$. The upper band is $\varepsilon(\mathbf{q})$ with the + square root in Eq. (6.55), and the lower band is for the − counter part. Regions of interests are at the Brillouin zone boundary, near the boundary in the gray area, and far away from the boundary.

6.8.1 *Energy gap*

For the case of two-fold degeneracy where $\varepsilon^0_{\mathbf{q}} = \varepsilon^0_{\mathbf{q}-\mathbf{G}}$, and $\mathbf{q}$ pointing on the Brillouin zone boundary, we have

$$\begin{aligned} \varepsilon &= \frac{1}{2}(\varepsilon^0_{\mathbf{q}} + \varepsilon^0_{\mathbf{q}-\mathbf{G}}) \pm |U_{\mathbf{G}}| \\ &= \varepsilon^0_{\mathbf{q}} \pm |U_{\mathbf{G}}|. \end{aligned} \tag{6.56}$$

The weak periodic potential removes the degeneracy by opening an energy gap. Namely, one energy level is lowered by $|U_G|$ from the corresponding free-electron level and the other is raised by the same amount. This perturbation effect from the periodic ion cores is of the first order of U.

6.8.2 *Nearly free electron states*

For other reciprocal lattice vector $\mathbf{G}'$, either both $\mathbf{q}$ and $\mathbf{q} - \mathbf{G}'$ are not pointing on the same Brillouin zone boundary, or both are far away from any Brillouin zone boundary.

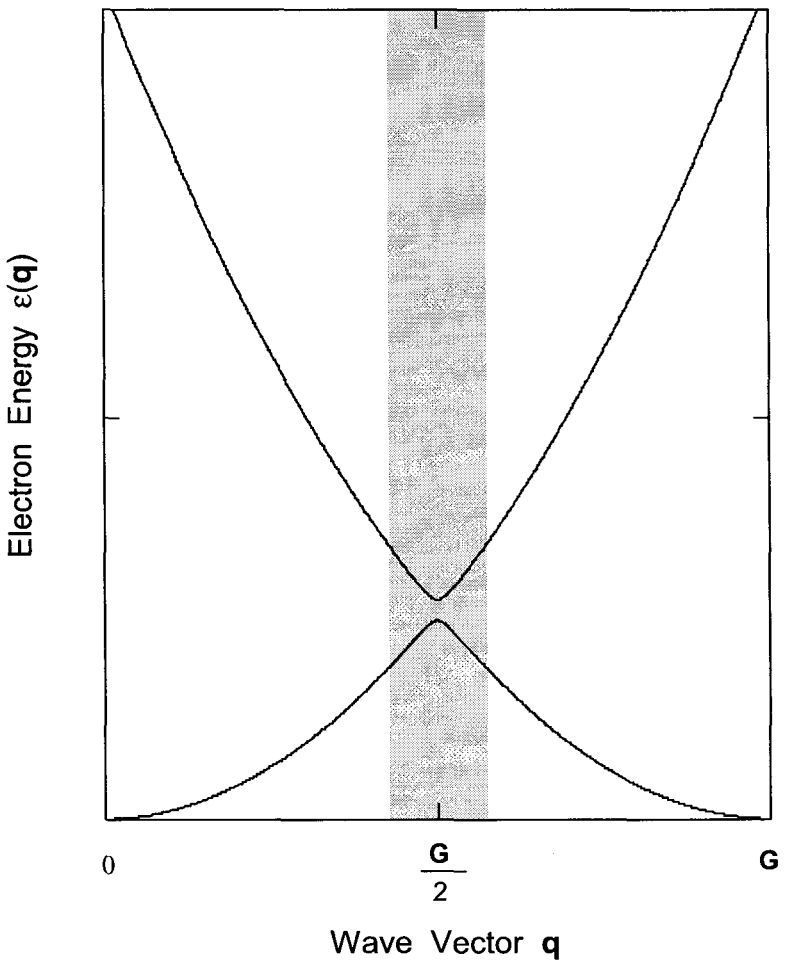


Fig. 6.4 The electron energy band, $\varepsilon(\mathbf{q})$, is in general along certain $\mathbf{q}$ vector that makes an angle of ϕ with $\mathbf{G}$. For the case of $\phi = 0$, $\mathbf{q}$ will be in the direction of $\mathbf{G}$ as shown here. The Brillouin zone boundary is perpendicular to the paper, located at $|\mathbf{q}| \cos\phi = |\mathbf{G}/2|$. Only in the gray area the energy is affected by the periodical potential. The electron energy spectrum is basically that of free electrons outside the gray area.

From analysis using similar diagrams of Figs. 6.2 and 6.3, we can see that $|\mathbf{q}| \neq |\mathbf{q} - \mathbf{G}'|$, and $\Delta\varepsilon = |\varepsilon^0_{\mathbf{q}} - \varepsilon^0_{\mathbf{q}-\mathbf{G}'}| \gg U_{\mathbf{G}'}$. Taking this condition with

Eq. (6.55), we have

$$\varepsilon = \frac{1}{2}(\varepsilon^0_{\mathbf{q}} + \varepsilon^0_{\mathbf{q}-\mathbf{G}'}) \pm \frac{1}{2}(\varepsilon^0_{\mathbf{q}} - \varepsilon^0_{\mathbf{q}-\mathbf{G}'}) = \varepsilon^0_{\mathbf{q}} \quad \text{or} \quad \varepsilon^0_{\mathbf{q}-\mathbf{G}'}. \tag{6.57}$$

The first-order perturbation effect vanishes for this case. The energy eigenvalue roots reduce to their respective free electron values within the first-order theory. But, if we expand our scope to include higher order effects, there is actually a small energy shift caused by the perturbation of U on the non-degenerate free-electron levels. It is a second-order effect that is too small to show up in the first-order approximation of Eq. (6.57). We have already calculated that effect previously in Eq. (6.46). The weak periodic potential has the strongest effect on free-electron levels whose wave vectors are pointing on or near the Brillouin zone boundary, where the energy shift due to the perturbation is of the first-order $\propto U$. For electron states with wave vectors located far away from the zone boundary, it is a second-order effect $\propto U^2/\Delta\varepsilon$.

6.8.3 *Fermi surface*

For the exact degenerate case where $\varepsilon^0_{\mathbf{q}} = \varepsilon^0_{\mathbf{q}-\mathbf{G}}$, it requires $|\mathbf{q}| = |\mathbf{q} - \mathbf{G}|$. So, $\mathbf{q} - \mathbf{G}/2$ is on the Brillouin zone boundary plane that bisects the $\mathbf{G}$ vector, see Fig. 6.3 for the geometry. Let us write the energy ε of Eq. (6.56) as an explicit function of $\mathbf{q}$ and $\mathbf{G}$, and then differentiate ε with respect to $\mathbf{q}$. We have

$$\varepsilon = \frac{1}{2}\left[\frac{\hbar^2}{2m}\mathbf{q}^2 + \frac{\hbar^2}{2m}(\mathbf{q} - \mathbf{G})^2\right] \pm |U_{\mathbf{G}}|,$$

and

$$\frac{\partial\varepsilon}{\partial q} = \frac{\hbar^2}{2m}\frac{1}{2}[2q + 2(q - G)] = \frac{\hbar^2}{m}(q - \frac{1}{2}G). \tag{6.58}$$

The partial derivative is taken after the inner product so that q and $q-G$ are scalars. In three dimensions, we have the gradient $\nabla_{\mathbf{q}}\varepsilon = (\hbar^2/m)(\mathbf{q}-\mathbf{G}/2)$, which lies on the Brillouin zone boundary plane, as shown in Fig. 6.3. Since $\mathbf{q} - \mathbf{G}/2$ is the gradient vector of $\varepsilon(\mathbf{q})$, it must be also perpendicular to the surface described by $\varepsilon(\mathbf{q}) \approx \varepsilon^0_{\mathbf{q}} =$ constant. Therefore, the constant energy surface $\varepsilon^0_{\mathbf{q}}$, which is a spherical surface for free electrons, must be distorted to become perpendicular to Brillouin zone boundary planes at their intersections. The most important surface of constant energy is the Fermi surface. The Fermi surface is torn by the weak periodic potential

near the Brillouin zone boundary, where it is no longer a spheric surface any more.

6.8.4 *Standing waves of electron states*

The boundary plane of a Brillouin zone is the bisecting plane of a reciprocal lattice vector $\mathbf{G}$. The electron wave vector $\mathbf{q}$ pointing on this plane is not in the direction of $\mathbf{G}$ in general. When it is, we have $\mathbf{q} = \mathbf{G}/2$ and $\mathbf{q}-\mathbf{G}/2 = 0$. This is the special case where the group velocity $\partial\varepsilon/\partial q = 0$, see Eq. (6.58), and we have standing waves. To calculate the standing waves, let us recall the enegy ε from Eq. (6.56) and the coupled Schrödinger equations from Eq. (6.53), i.e.,

$$\varepsilon = \varepsilon^0_{\mathbf{q}} \pm |U_{\mathbf{G}}| = \varepsilon^0_{\mathbf{q}-\mathbf{G}} \pm |U_{\mathbf{G}}| = \varepsilon^0_{\mathbf{G}/2} \pm |U_{\mathbf{G}}|,$$

and

$$\begin{cases} (\varepsilon - \varepsilon^0_{\mathbf{q}})C_{\mathbf{q}} = U_{\mathbf{G}}C_{\mathbf{q}-\mathbf{G}}; \\ (\varepsilon - \varepsilon^0_{\mathbf{q}-\mathbf{G}})C_{\mathbf{q}-\mathbf{G}} = U^*_{\mathbf{G}}C_{\mathbf{q}}. \end{cases}$$

Substituting $\varepsilon = \varepsilon^0_{\mathbf{q}} \pm |U_{\mathbf{G}}|$ into the first of the coupled equations, and $\varepsilon = \varepsilon^0_{\mathbf{q}-\mathbf{G}} \pm |U_{\mathbf{G}}|$ into the second, we obtain

$$C_{\mathbf{q}} = \pm\frac{U_{\mathbf{G}}}{|U_{\mathbf{G}}|}C_{\mathbf{q}-\mathbf{G}} \quad \text{and} \quad C_{\mathbf{q}-\mathbf{G}} = \pm\frac{U^*_{\mathbf{G}}}{|U_{\mathbf{G}}|}C_{\mathbf{q}}. \tag{6.59}$$

The $+$ and $-$ signs are for the upper and lower bands, respectively. From Eq. (6.22), we know that U_G is real. Since the core potential $U(\mathbf{r})$ is negative to electrons, the ratio $U_G/|U_G|$ is equal to -1. So, we have

$$C_{\mathbf{q}-\mathbf{G}} = \pm C_{\mathbf{q}} \quad \begin{cases} + & \text{for lower band:} \quad \varepsilon = \varepsilon^0_{\mathbf{q}} - |U_{\mathbf{G}}|; \\ - & \text{for upper band:} \quad \varepsilon = \varepsilon^0_{\mathbf{q}} + |U_{\mathbf{G}}|. \end{cases} \tag{6.60}$$

The $\mathbf{q}$-state wave function $\Psi_{\mathbf{q}}(\mathbf{r})$ is

$$\begin{aligned} \Psi_{\mathbf{q}}(\mathbf{r}) &= \sum_{\mathbf{G}'} C_{\mathbf{q}-\mathbf{G}'}e^{i(\mathbf{q}-\mathbf{G}')\cdot\mathbf{r}} = C_{\mathbf{q}}e^{i\mathbf{q}\cdot\mathbf{r}} \pm C_{\mathbf{q}}e^{i(\mathbf{q}-\mathbf{G})\cdot\mathbf{r}} \\ &= C_{\mathbf{q}}\left(e^{i\frac{\mathbf{G}}{2}\cdot\mathbf{r}} \pm e^{-i\frac{\mathbf{G}}{2}\cdot\mathbf{r}}\right) \end{aligned}$$

$$= \begin{cases} 2C_{\mathbf{q}} \cos(\frac{\mathbf{G}}{2} \cdot \mathbf{r}) & \text{for} \quad \varepsilon = \varepsilon^0_{\mathbf{q}} - |U_{\mathbf{G}}|; \\ 2iC_{\mathbf{q}} \sin(\frac{\mathbf{G}}{2} \cdot \mathbf{r}) & \text{for} \quad \varepsilon = \varepsilon^0_{\mathbf{q}} + |U_{\mathbf{G}}|. \end{cases} \tag{6.61}$$

The electron density of the $\mathbf{q}$ state is proportional to $\Psi^*_{\mathbf{q}}\Psi_{\mathbf{q}}$. As $\mathbf{r}$ approaches zero on any lattice sites, we are getting into the ion core, where the cosine factor approaches unity because $\frac{1}{2}\mathbf{G}\cdot\mathbf{r} = n\pi$, and the sine factor vanishes. This situation is very much like atomic orbital levels. The s-level wave function is a cosine wave that does not vanish near the center of the atom, and the sine wave of the p level does. We have the s-like charge density that peaks at the ion core, and p-like charge density vanishes there. While in between ion cores, the s-like charge density vanishes and the p-like charge density peaks. From the charge density distribution, we can readily see that the s-like wave function is associated with the lower energy band and the p-like wave is with the higher energy band. We have

$$\Psi^*_{\mathbf{q}}(\mathbf{r})\Psi_{\mathbf{q}}(\mathbf{r}) \propto \left(\cos\frac{1}{2}\mathbf{G}\cdot\mathbf{r}\right)^2 \qquad \text{s-like for} \qquad \varepsilon = \varepsilon^0_{\mathbf{q}} - |U_{\mathbf{G}}|,$$

and

$$\Psi^*_{\mathbf{q}}(\mathbf{r})\Psi_{\mathbf{q}}(\mathbf{r}) \propto \left(\sin\frac{1}{2}\mathbf{G}\cdot\mathbf{r}\right)^2 \qquad \text{p-like for} \qquad \varepsilon = \varepsilon^0_{\mathbf{q}} + |U_{\mathbf{G}}|. \tag{6.62}$$

6.9 Energy Bands Near Brillouin Zone Boundaries

The exact two-fold degeneracy at the Brillouin zone boundary offers a good idea about where the energy gap comes from, and about its magnitude. The electron states $\varepsilon(\mathbf{q})$ away from Brillouin boundaries are not affected by the periodic potential to the first order and can be treated as free electron states $\varepsilon^0_{\mathbf{q}}$ for practical purposes. In order to connect these two regions, we need to know what happens in between, i.e., in the vicinity of Brillouin zone boundaries inside the gray region of Fig. 6.4. The energy eigenvalue roots given in Eq. (6.55) are actually solved for the nearly two-fold degeneracy under the condition of $|\varepsilon^0_{\mathbf{q}} - \varepsilon^0_{\mathbf{q}-\mathbf{G}}| \ll U$. It implies that $\mathbf{q}$ and $\mathbf{q}-\mathbf{G}$ are very close to the Brillouin zone boundary, but not necessarily pointing right on the boundary, see Fig. 6.2 for the picture. We want to examine the band structure in the vicinity around the Brillouin zone boundary.

Write the free-electron energy explicitly in terms of $\varepsilon^0_{\mathbf{q}} = (\hbar\mathbf{q})^2/2m$ and $\varepsilon^0_{\mathbf{q}-\mathbf{G}} = \hbar^2(\mathbf{q}-\mathbf{G})^2/2m$, substitute them into Eq. (6.55), re-arrange terms

with some algebraic manipulations, and we have

$$\varepsilon = \frac{\hbar^2}{2m}(\mathbf{q}^2 - \mathbf{q}\cdot\mathbf{G} + \frac{\mathbf{G}^2}{2}) \pm \sqrt{\left(\frac{\frac{\hbar^2}{2m}\cdot 2\mathbf{G}\cdot(\mathbf{q}-\frac{\mathbf{G}}{2})}{2}\right)^2 + |U_{\mathbf{G}}|^2}$$

$$= \frac{\hbar^2}{2m}\left((\mathbf{q}-\frac{\mathbf{G}}{2})^2 + \frac{\mathbf{G}^2}{4}\right) \pm \sqrt{4\left(\frac{\hbar^2}{2m}\cdot\frac{\mathbf{G}}{2}\cdot(\mathbf{q}-\frac{\mathbf{G}}{2})\right)^2 + |U_{\mathbf{G}}|^2}. \tag{6.63}$$

Define $\mathbf{q}$ tilde as

$$\tilde{\mathbf{q}} = \mathbf{q} - \frac{\mathbf{G}}{2}, \qquad \text{and let} \qquad \lambda = \frac{\hbar^2(\mathbf{G}/2)^2}{2m}.$$

We can rewrite the energy ε as

$$\varepsilon = \frac{\hbar^2}{2m}(\tilde{\mathbf{q}}^2 + \frac{\mathbf{G}^2}{4}) \pm \sqrt{4\lambda\frac{\hbar^2\tilde{\mathbf{q}}^2}{2m}\cos^2\theta + |U_{\mathbf{G}}|^2}. \tag{6.64}$$

For the case where

$$4\lambda\frac{\hbar^2\tilde{\mathbf{q}}^2}{2m}\cos^2\theta \ll |U_{\mathbf{G}}|^2,$$

the energy is reduced to

$$\varepsilon \approx \lambda + \frac{\hbar^2\tilde{\mathbf{q}}^2}{2m} \pm |U_{\mathbf{G}}|\left(1 + 2\frac{\lambda}{|U_{\mathbf{G}}|^2}\frac{\hbar^2\tilde{\mathbf{q}}^2}{2m}\cos^2\theta\right)$$

$$\approx (\lambda \pm |U_{\mathbf{G}}|) + \frac{\hbar^2\tilde{\mathbf{q}}^2}{2m}\left(1 \pm \frac{2\lambda}{|U_{\mathbf{G}}|}\cos^2\theta\right). \tag{6.65}$$

The last equation is the eigenvalue ε solved as a function of $(\mathbf{q} - \mathbf{G}/2)$, or $\tilde{\mathbf{q}}$, in the vicinity near the Brillouin zone boundary at $\mathbf{G}/2$. The correction term depends on both $\tilde{\mathbf{q}}$ and the angle θ between $\tilde{\mathbf{q}}$ and $\mathbf{G}$. If $\tilde{\mathbf{q}}$ is not zero but lies on the zone boundary plane, we have $\theta = 90°$ and Eq. (6.65) reduces to Eq. (6.56). Same is true for the case when $\tilde{\mathbf{q}} = 0$. Consistently, we are back to the exact two-fold degenerate case. The solution in Eq. (6.65) can be further simplified if we let $\varepsilon_+ = \lambda + |U_G|$ for the upper band and $\varepsilon_- = \lambda - |U_G|$ for the lower band. We have

$$\varepsilon \approx \varepsilon_+ + \frac{\hbar^2\tilde{\mathbf{q}}^2}{2m_+^*} \approx \varepsilon_+ + \frac{\hbar^2\tilde{\mathbf{q}}^2}{2m^*} \qquad \text{for the upper band,} \tag{6.66}$$

and

$$\varepsilon \approx \varepsilon_- + \frac{\hbar^2\tilde{\mathbf{q}}^2}{2m_-^*} \approx \varepsilon_- - \frac{\hbar^2\tilde{\mathbf{q}}^2}{2|m^*|} \qquad \text{for the lower band,} \quad (6.67)$$

where the effective mass m_+^*, m_-^*, and m^* are defined by

$$m_\pm^* = \frac{m}{\left(1 \pm \frac{2\lambda}{|U_\mathbf{G}|}\cos^2\theta\right)} \approx \begin{cases} \pm\frac{m}{\frac{2\lambda}{|U_\mathbf{G}|}} = m^* & \text{near or on the zone boundary;} \\ m & \text{far away from the zone boundary.} \end{cases} \quad (6.68)$$

When we move away from the Brillouin zone boundary but still in the vicinity, the upper band energy ε_+ is corrected by adding a term $\hbar^2(\mathbf{q}-\mathbf{G}/2)^2/2m^*$ and the lower band energy ε_- is corrected by subtracting the same term. Here, m^* is the band effective mass, which is reduced from the free-electron mass m approximately by a factor of $2\lambda/|U_G|$. More interestingly to note is the sign of effective mass. It is positive in the upper band and turns into a negative mass in the lower band. The effective mass is not the real mass, but a manifestation of electron propagation in the periodic potential.

The electron energy bands described by Eqs. (6.66) and (6.67), and the band effective mass of Eq. (6.68) are valid in the vicinity of Brillouin zone boundaries, where the condition $|\varepsilon_\mathbf{q}^0 - \varepsilon_{\mathbf{q}-\mathbf{G}}^0| < U$ is satisfied. When we move further away from the boundary to regions where this condition is no longer true, both the electron energy and mass return to their free-electron values within the first order approximation. Equations (6.66) and (6.67) are plotted in Fig. 6.5 together with the free electron spectrum.

6.10 Band Effective Mass, Group Velocity, Crystal Momentum, and Bragg Scattering

It turns out that the curvature of the band has a lot to do with the band effective mass. The time dependent wave function describing the electron in the state $\mathbf{k}$ and energy ε has a factor proportional to $e^{i\cdot\varepsilon/\hbar\cdot t}$. When $-i\hbar\partial/\partial t$ operates on this part of the wave function, the energy ε comes out as the eigenvalue. Comparing to $e^{i\omega t}$, $\varepsilon/\hbar$ must be the frequency ω of the wave function. The group velocity $\mathbf{v}_g$ of the electron wave is $\nabla_\mathbf{k}\omega(\mathbf{k})$, so we

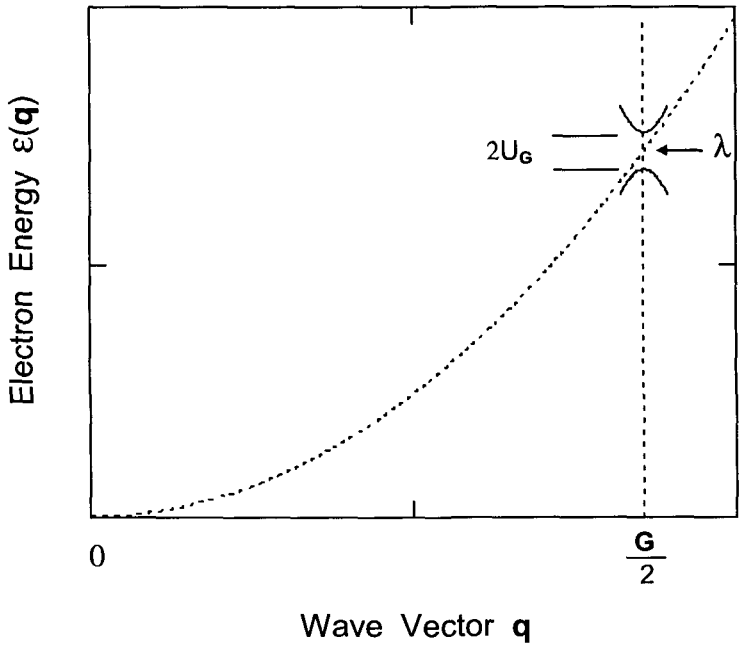


Fig. 6.5 Plot of the electron bands in the vicinity of a Brillouin zone boundary. In making the plot, we have assumed that $\hbar^2/2m = 1$, $U/\lambda = 0.05$, $m^*/m = 0.025$, q is in the direction of G, and $\tilde{q}$ is cut off at 4% of $G/2$. The dotted curve is the free electron spectrum. The dotted vertical line is the side view of the Brillouin zone boundary. The two parabolic curves in solid lines are the upper and lower bands. The energy gap is $2U_G$, and the free electron energy λ is at the center of the energy gap. Both bands approach the free electron spectrum asymptotically outside the region of 4% of $G/2$. The parabolic bands as shown here in the figure are part of the full band diagram in repeated zone scheme of Fig. fig6.4.

have $\mathbf{v}_g = 1/\hbar \cdot \nabla_{\mathbf{k}}\varepsilon(\mathbf{k})$. The dispersion of the energy band $\varepsilon(\mathbf{k})$ determines the group velocity of an electron wave packet. To simplify the mathematics, we will choose a certain symmetry direction, so that $\mathbf{v}_g = 1/\hbar \cdot \nabla_{\mathbf{k}}\varepsilon(\mathbf{k})$ will be reduced to $v_g = 1/\hbar \cdot d\varepsilon(k)/dk$. From

$$v_g = \frac{1}{\hbar}\frac{d\varepsilon}{dk} \qquad \text{and} \qquad \hbar\frac{dk}{dt} = F,$$

we have

$$\frac{dv_g}{dt} = \frac{1}{\hbar}\frac{d^2\varepsilon}{dk^2}\frac{dk}{dt} = \frac{1}{\hbar^2}\frac{d^2\varepsilon}{dk^2} \cdot F.$$

Here, we have discovered the proportional relationship between the external force and the time derivative of the electron group velocity, i.e.,

$$F = \frac{\hbar^2}{\frac{d^2\varepsilon}{dk^2}} \cdot \frac{dv_g}{dt}.$$

This is nothing but Newton's second law, which defines the effective mass as

$$m^* = \frac{\hbar^2}{\frac{d^2\varepsilon}{dk^2}},$$

or

$$\frac{1}{m^*} = \frac{1}{\hbar^2}\left(\frac{d^2\varepsilon}{dk^2}\right). \tag{6.69}$$

Equation (6.69) states that the effective mass is inversely proportional to the band curvature, i.e.,

$$\frac{1}{\text{Effective mass}} \propto \text{curvature of } \varepsilon(\mathbf{k}).$$

Near the top of the lower band around the Brillouin zone boundary, $\varepsilon(k)$ is concave downwards. The curvature $d^2\varepsilon/dk^2$ is negative and the band effective mass is negative. Near the bottom of the upper band around the Brillouin zone boundary, $\varepsilon(k)$ is concave upwards. The curvature $d^2\varepsilon/dk^2$ is positive and the band effective mass is positive. These results of having positive and negative effective mass are consistent with the definition of effective mass made in eigenvalue calculations earlier. The effective mass is reduced approximately by the factor of $2\lambda/|U_G|$. Effective mass can be obtained from the experimentally measured $\varepsilon(k)$. Larger curvature in $\varepsilon(k)$ will end up with smaller band effective mass and vise versa.

What is the physical meaning of the effective mass? Accepting the negative band effective mass, a seemly incomprehensible consequence results from the equation of motion $F = m^* dv_g/dt$. Namely, the acceleration dv_g/dt is in the opposite direction of the applied force F. What this means? Furthermore, with a nominal value of $\lambda \approx 20eV$ and $|U_G| \approx 0.2eV$, we have $m^*/m \approx 0.01$. The implication is that the electron in the energy band would be accelerated 100 times faster than the free electron by the same force. How can this be?

All these puzzling questions can be answered if we remember that Bragg reflection is involved with a large momentum transfer from the lattice to

the electron. First, let us go back to the Schrödinger equation connecting $C_{\mathbf{q}}$ and $C_{\mathbf{q}-\mathbf{G}}$. We recall

$$\frac{C_{\mathbf{q}-\mathbf{G}}}{C_{\mathbf{q}}} = \frac{\varepsilon - \varepsilon_{\mathbf{q}}^0}{U_{\mathbf{G}}}. \tag{6.70}$$

In order to see the change in the ratio $C_{\mathbf{q}-\mathbf{G}}/C_{\mathbf{q}}$ as a function of $\mathbf{q}$, we have blown up the energy gap region of Fig. 6.4 and shown it in Fig. 6.6. From Eq. (6.70) and Fig. 6.6, we can follow the change in the ratio of $C_{\mathbf{q}-\mathbf{G}}/C_{\mathbf{q}}$ qualitatively.

Since the periodic core potential U_G is negative for electrons, the Fourier coefficient $C_{\mathbf{q}}$ and $C_{\mathbf{q}-\mathbf{G}}$ are of the same sign for the lower band where $\varepsilon < \varepsilon_{\mathbf{q}}^0$. They have opposite signs for the upper band where $\varepsilon > \varepsilon_{\mathbf{q}}^0$. The ratio $|C_{\mathbf{q}-\mathbf{G}}/C_{\mathbf{q}}|$ reaches the maximum value of 1 when $\varepsilon - \varepsilon_{\mathbf{q}}^0 = \pm|U_G|$. This happens when $\mathbf{q}$ is pointing on the Brillouin zone boundary, which is the exact 2-fold degenerate case we discussed previously. The s-like wave function in the lower band comes from $C_{\mathbf{q}} = C_{\mathbf{q}-\mathbf{G}}$ and the p-like wave function in upper band comes from $C_{\mathbf{q}} = -C_{\mathbf{q}-\mathbf{G}}$, see Eqs. (6.61) and (6.62). When $\mathbf{q}$ moves far away from the zone boundary in either direction, ε approaches $\varepsilon_{\mathbf{q}}^0$ and the ratio $|C_{\mathbf{q}-\mathbf{G}}/C_{\mathbf{q}}|$ approaches zero. In the vicinity of Brillouin zone boundaries, the ratio $|C_{\mathbf{q}-\mathbf{G}}/C_{\mathbf{q}}|$ is between 1 and ~ 0, depending on how close $\mathbf{q}$ is to the boundary.

To be more quantitative, we can go ahead to calculate the electron momentum from the wave function. Following the behavior of the electron momentum, we will learn about the effect of Bragg scattering, electron group velocity, electron crystal momentum, and the physical meaning of the electron effective mass. Let us recall the electron wave function, i.e.,

$$\Psi_{\mathbf{q}}(\mathbf{r}) = C_{\mathbf{q}} \cdot e^{i\mathbf{q}\cdot\mathbf{r}} + C_{\mathbf{q}-\mathbf{G}} \cdot e^{i(\mathbf{q}-\mathbf{G})\cdot\mathbf{r}},$$

and the electron momentum is calculable from

$$\mathbf{p} = < \Psi_{\mathbf{q}} | \frac{\hbar}{i} \nabla | \Psi_{\mathbf{q}} > .$$

Making use of the identity

$$|C_{\mathbf{q}}|^2 + |C_{\mathbf{q}-\mathbf{G}}|^2 = 1,$$

and the periodical property

$$\frac{1}{V_{\text{cell}}} \int_{\text{cell}} e^{i\mathbf{G}\cdot\mathbf{r}} dxdydz = 0,$$

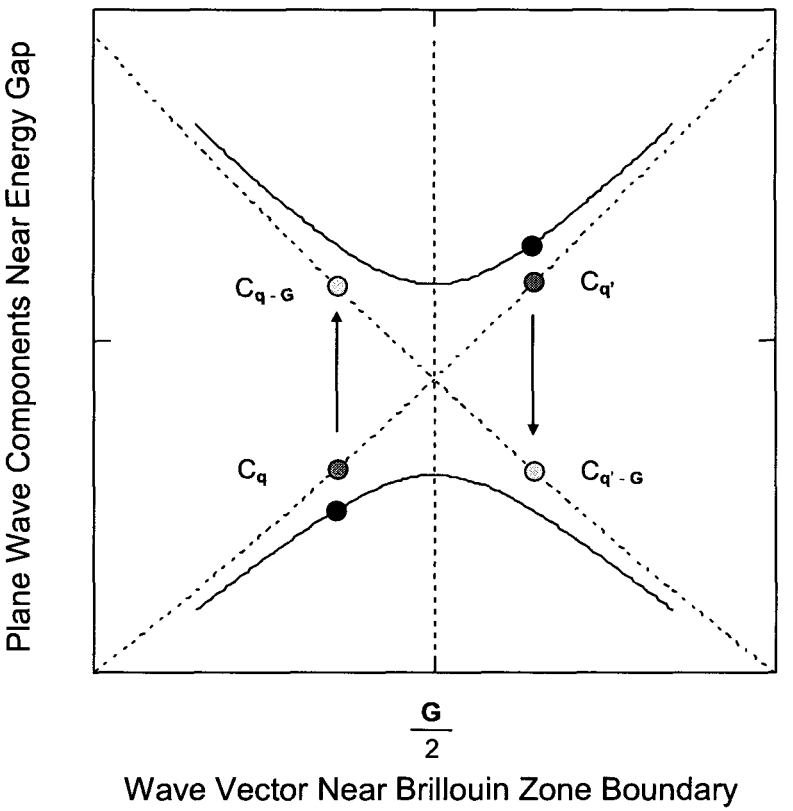


Fig. 6.6 Energy bands are shown with two contributing plane-wave components in the vicinity of Brillouin zone boundary. The solid dots are two electrons in the lower and upper bands, respectively. Near the zone boundary, the free electron states $C_{\mathbf{q}}$, and $C_{\mathbf{q}'}$ are scattered into $C_{\mathbf{q}-\mathbf{G}}$, and $C_{\mathbf{q}'-\mathbf{G}}$, respectively. The ratio $C_{\mathbf{q}-\mathbf{G}}/C_{\mathbf{q}}$ varies from 0 to 1 when the electron approaches the zone boundary in the lower band. The ratio $C_{\mathbf{q}'-\mathbf{G}}/C_{\mathbf{q}'}$ varies from -1 to 0 when the electron moves away from the zone boundary in the upper band.

we obtain

$$\mathbf{p} = \hbar\mathbf{q} - |C_{\mathbf{q}-\mathbf{G}}|^2 \cdot \hbar\mathbf{G}, \tag{6.71}$$

or, equivalently in a different form of

$$\mathbf{p} = |C_{\mathbf{q}}|^2\hbar\mathbf{q} + |C_{\mathbf{q}-\mathbf{G}}|^2 \cdot (\hbar\mathbf{q} - \hbar\mathbf{G}). \tag{6.72}$$

This is the electron momentum for the case where the electron wave function is the linear combination of two states of $\mathbf{q}$ and $\mathbf{q} - \mathbf{G}$. When we move far away from the Brillouin zone boundary, $|C_{\mathbf{q}-\mathbf{G}}|^2$ goes to zero and $|C_{\mathbf{q}}|^2$ equals to 1. Equation (6.72) tells us that the electron momentum reduces back to the free electron momentum $\hbar\mathbf{q}$. When we move into the vicinity around the zone boundary, the electron momentum is no longer the free-electron momentum, because the free-electron wave function is mixed with the reflected wave from Bragg scattering. At the Brillouin zone boundary, we have $|C_{\mathbf{q}-\mathbf{G}}|^2 = |C_{\mathbf{q}}|^2 = 1/2$, and the electron momentum becomes

$$\mathbf{p} = \frac{1}{2}(\hbar\mathbf{q} + (\hbar\mathbf{q} - \hbar\mathbf{G})). \tag{6.73}$$

In this case, the electron momentum $\mathbf{p}$ lies on the zone boundary and is perpendicular to the $\mathbf{G}$ vector. The momentum component parallel to $\mathbf{G}$ is completely stopped by the Bragg scattering and diminishes to zero. The graphic representation of this situation is depicted in Fig. 6.3.

With a little imagination when looking at this figure, we can see that $\mathbf{p}$ will make an angle larger than 90° with $\mathbf{G}$ if $\mathbf{q}$ is inside the zone. This angle increases if $\mathbf{q}$ is more parallel to $\mathbf{G}$. When $\mathbf{q}$ is parallel to $\mathbf{G}$, this angle will be 180°. This is to say that the electron momentum $\mathbf{p}$ is in the opposite direction of $\hbar\mathbf{q}$, which is the momentum of the incoming free electron. After the Bragg back scattering from the recoil lattice, the incoming free-electron wave is mixed with the reflected wave, and the momentum not only gets slowed down to $\mathbf{p}$ from $\hbar\mathbf{q}$, but also gets its direction reversed. See Eqs. (6.71), (6.72), and (6.73) as well as Figs. 6.6 and 6.7.

Still in Fig. 6.3, if $\mathbf{q}$ is just outside the zone, the angle between $\mathbf{p}$ and $\mathbf{G}$ would be smaller than 90°. It decreases as $\mathbf{q}$ goes more parallel to $\mathbf{G}$. In the case when $\mathbf{q}$ is parallel to $\mathbf{G}$, the angle becomes 0°. In this case of Bragg forward scattering, the momentum gets slowed down to $\mathbf{p}$, but its direction remains unchanged from $\hbar\mathbf{q}$.

To make further analysis, we may decompose $\mathbf{p}$ and $\mathbf{q}$ into parallel and perpendicular components with respect to $\mathbf{G}$. The perpendicular component is

$$\mathbf{p}_\perp = \hbar\mathbf{q}_\perp,$$

and the parallel component is

$$\mathbf{p}_\parallel = \hbar\mathbf{q}_\parallel - |C_{\mathbf{q}-\mathbf{G}}|^2 \cdot \hbar\mathbf{G}. \tag{6.74}$$

Only the parallel momentum $\mathbf{p}_\parallel$ is affected by the Bragg scattering. When

an external force impulse $\mathbf{F}\Delta t$ is applied, the electron will go from the state $\mathbf{q}$ to $\mathbf{q}+\Delta\mathbf{q}$. The wave function evolves to

$$\Psi_{\mathbf{q}+\Delta\mathbf{q}}(\mathbf{r}) = C_{\mathbf{q}+\Delta\mathbf{q}} \cdot e^{i(\mathbf{q}+\Delta\mathbf{q})\cdot\mathbf{r}} + C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}} \cdot e^{i(\mathbf{q}+\Delta\mathbf{q}-\mathbf{G})\cdot\mathbf{r}}. \tag{6.75}$$

The component $C_{\mathbf{q}+\Delta\mathbf{q}}$ decreases and $C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}$ will increase if $\Delta\mathbf{q}$ is an increment toward the zone boundary ($\mathbf{q}_{\|}+\Delta\mathbf{q}_{\|}$ approaches closer to $\frac{1}{2}\mathbf{G}$). The reverse is true when $\Delta\mathbf{q}$ is away from the zone boundary ($\mathbf{q}_{\|}+\Delta\mathbf{q}_{\|}$ deviates more from $\frac{1}{2}\mathbf{G}$). We may proceed to calculate the momentum change $\Delta\mathbf{p}$ from the electron wave function or just simply differentiate the electron momentum $\mathbf{p}$ to obtain $\Delta\mathbf{p}$. The electron momentum change $\Delta\mathbf{p}$ comes from the redistribution between its two components in the wave function when $\mathbf{q}$ goes to $\mathbf{q}+\Delta\mathbf{q}$. We have

$$\Delta\mathbf{p} = \hbar\Delta\mathbf{q} - \hbar\mathbf{G}\cdot(|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2). \tag{6.76}$$

The change of electron momentum parallel to $\mathbf{G}$ is

$$\Delta\mathbf{p}_{\|} = \hbar\Delta\mathbf{q}_{\|} - \hbar\mathbf{G}(|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2), \tag{6.77}$$

which can be in the direction of $\hbar\Delta\mathbf{q}_{\|}$ or in its opposite direction, depending on whether $(|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2)$ is negative or positive. The factor $(|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2)$ is the differential change of $|C_{\mathbf{q}-\mathbf{G}}|^2$. Its value near the Brillouin zone boundary can be a good fraction of 1, and can be either positive or negative. In order to see how $|C_{\mathbf{q}-\mathbf{G}}|^2$ changes when $\mathbf{q}$ goes to $\mathbf{q}+\Delta\mathbf{q}$, we need to solve for $C_{\mathbf{q}-\mathbf{G}}$ near the Brillouin zone boundary. After some algebraic manipulation, we obtain

$$\frac{C_{\mathbf{q}-\mathbf{G}}}{C_{\mathbf{q}}} \approx \frac{2\lambda}{|U_{\mathbf{G}}|}\frac{|\tilde{\mathbf{q}}|}{|\mathbf{G}/2|}\cos\theta \mp \sqrt{\frac{4\lambda^2}{|U_{\mathbf{G}}|^2}\left(\frac{|\tilde{\mathbf{q}}|}{|\mathbf{G}/2|}\right)^2\cos^2\theta + 1}. \tag{6.78}$$

The minus sign is for the upper band and the plus sign is for the lower band. The angle θ is between $\mathbf{q}-\mathbf{G}/2$ and $\mathbf{G}/2$. The absolute value of the ratio $|C_{\mathbf{q}-\mathbf{G}}/C_{\mathbf{q}}|$ approaches 1 when $\mathbf{q}$ moves toward the Brillouin zone boundary where $\theta \to 90°$. Its value diminishes when $\mathbf{q}$ moves far away from the zone boundary. In Fig. 6.7, we plot this ratio as a function of θ along a randomly selected $\mathbf{q}$ vector that makes an angle of ϕ with $\mathbf{G}$.

For electrons in the lower band just inside the Brillouin zone boundary where $\theta > 90°$, $|C_{\mathbf{q}-\mathbf{G}}|^2$ increases to $|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2$ when $\mathbf{q}_{\|}$ goes to $\mathbf{q}_{\|}+\Delta\mathbf{q}_{\|}$ toward the boundary. The factor $(|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2)$ is positive, and the term $-(|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2)\cdot\hbar\mathbf{G}$ is in the opposite direction of $\hbar\Delta\mathbf{q}_{\|}$, see Eq. (6.77). In case of $\mathbf{q}_{\|}$ moving away from the boundary, the

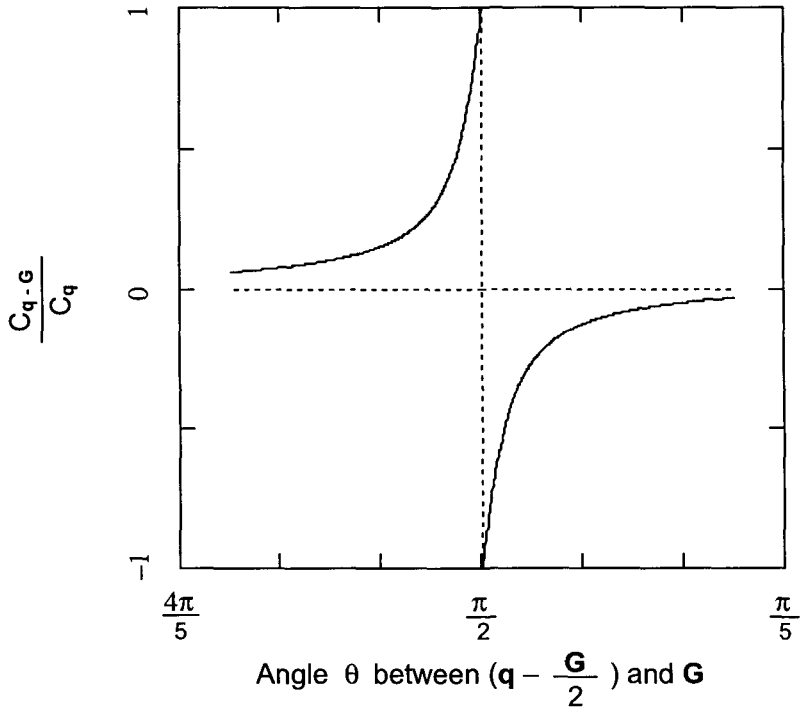


Fig. 6.7 Plane-wave component ratio $C_{\mathbf{q}-\mathbf{G}}/C_{\mathbf{q}}$ of Eq. (6.78) is plotted in the vicinity of Brillouin zone boundary. The ratio varies along a randomly selected $\mathbf{q}$ vector that makes an angle of $\phi = \pi/12$ with $\mathbf{G}$. The vector $\tilde{\mathbf{q}}$ is on the Brillouin zone boundary when the angle $\theta = \pi/2$. This is the case where the electron can be either in the lower band as the s-wave, or in the upper band as the p-wave. When $\theta > \pi/2$, the electron state is inside the Brillouin zone and the electron is primarily in the lower band. When $\theta < \pi/2$, the electron state is outside the Brillouin zone and the electron is primarily in the upper band.

factor $(|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2)$ is negative, and the term $-(|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2) \cdot \hbar\mathbf{G}$ is also in the opposite direction of $\hbar\Delta\mathbf{q}_{\|}$, see Eq.(6.77). In either case, the factor $-(|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2) \cdot \hbar\mathbf{G}$ is much larger than $\hbar\Delta\mathbf{q}_{\|}$ in general, and is in its opposite direction. Therefore, the electron momentum change $\Delta\mathbf{p}_{\|}$ near the zone boundary in the lower band is in the opposite direction of $\hbar\Delta\mathbf{q}_{\|}$ and much larger.

The group velocity $\mathbf{v}_g$ is the velocity of the wave packet that represents the electron. The electron is accelerated according to the total force exerted

on it, i.e., $\mathbf{F}_t = m d\mathbf{v}_g/dt$, where $\mathbf{F}_t$ is the total of the external force $\mathbf{F}$ and the force $\mathbf{F}_\ell$ from the recoiling lattice. Here, the mass m is the free electron mass, not the band effective mass. The total force $\mathbf{F}_t$ exerted on the electron must also equal $d\mathbf{p}/dt$. So, we have $\mathbf{v}_g = \mathbf{p}/m$. For electrons near the zone boundary in the lower band, $\Delta\mathbf{p}_\parallel$ is in the opposite direction of $\hbar\Delta\mathbf{q}_\parallel$, and much bigger than it normally. Therefore, the acceleration $d\mathbf{v}_g/dt$ $(= 1/m \cdot d\mathbf{p}/dt)$ will make a large angle with $\hbar\Delta\mathbf{q}_\parallel$, greater than 90° and in many cases close to 180°. Magnitude wise, we have $|d\mathbf{v}_g/dt| \gg 1/m \cdot |\hbar \cdot d\mathbf{q}/dt|$.

The total force $\mathbf{F}_t$ on the electron is from the externally applied force $\mathbf{F}$ and the force from the lattice $\mathbf{F}_\ell$. When the crystal is considered as a whole system, the recoil momentum of the lattice cancels the electron momentum caused by $\mathbf{F}_\ell$, that is the second term in Eq. (6.76). To an observer in the laboratory, the relationship $\mathbf{F} = \hbar \cdot d\mathbf{q}/dt$ always holds regardless whether the electron is free or in crystals. The quantity $\hbar \cdot d\mathbf{q}/dt$ is the change rate of the real electron momentum if the electron is free, but it is not the case for electrons near Brillouin zone boundaries.

When the Bragg reflection is involved, the electron momentum gets changed quite a lot from the recoiling lattice. From the discussion that leads to Eq. (6.69), we have $\mathbf{F} = \hbar \cdot d\mathbf{q}/dt = m^* d\mathbf{v}_g/dt$. Since $m \cdot d\mathbf{v}_g/dt$ is in the opposite direction of $\mathbf{F}$ $(= \hbar \cdot d\mathbf{q}/dt)$ and bigger than $\mathbf{F}$, the effective mass m^* must be negative and smaller than m. This is consistent with what we have calculated in Eq. (6.68). Namely, the electron effective mass m^* in the lower band is given by $m^* = -m/(2\lambda/|U_\mathbf{G}|)$. The effective mass m^*, a manifestation of the energy band, compensates the large change in $d\mathbf{v}_g/dt$ caused by Bragg scattering, so that the Newton's second law still holds for the externally applied force, i.e., $\mathbf{F} = m^* d\mathbf{v}_g/dt$.

For electrons in the upper band just outside the Brillouin zone boundary where $\theta < 90°$, the factor $(|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2)$ is negative for positive $\Delta\mathbf{q}_\parallel$, i.e., for electrons moving away from the zone boundary. The factor $(|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2)$ is positive for negative $\Delta\mathbf{q}_\parallel$ when the electron moves toward the boundary. Therefore, the lattice contribution to $\Delta\mathbf{p}_\parallel$ in this vicinity is in the direction of $\Delta\mathbf{q}_\parallel$. It is a small-angle forward scattering. To add the large vector $-\hbar\mathbf{G} \cdot (|C_{\mathbf{q}+\Delta\mathbf{q}-\mathbf{G}}|^2 - |C_{\mathbf{q}-\mathbf{G}}|^2)$ to $\hbar\Delta\mathbf{q}_\parallel$, we obtain $\Delta\mathbf{p}_\parallel$ that is in the direction of $\Delta\mathbf{q}_\parallel$ and much larger than $\hbar\Delta\mathbf{q}$. Knowing $|\Delta\mathbf{p}| \gg |\hbar\Delta\mathbf{q}|$ and $\Delta\mathbf{p}/m = \Delta\mathbf{v}_g$, we conclude that the acceleration $d\mathbf{v}_g/dt$ is larger than $1/m \cdot \hbar\Delta\mathbf{q}/dt$ and points to the forward direction of $\Delta\mathbf{q}_\parallel$. For getting $m^* d\mathbf{v}_g/dt = \mathbf{F}$, the effective mass m^* in the upper band must reduce from the free electron value, but remains positive. This is why we

have $m^* = +m/(2\lambda/|U_{\mathbf{G}}|)$.

An Example Application of the Energy Band Concept

6.11 Holes in Semiconductors

Holes are excitations of the valence band. We should always consider the entire valence band with a missing electron to determine the properties of the hole.

6.11.1 *Wave vector* $\mathbf{k}_h = -\mathbf{k}_e$

A completely filled valence band has a net crystal momentum equal to zero. As shown in Fig. 6.8, the missing electron in the valence band is not missing at all, but gets excited into the conduction band. If this electron has a crystal momentum equal to $\hbar\mathbf{k}_e$, the valence band now has one unpaired electron at $-\mathbf{k}_e$. The net crystal momentum of the valence band becomes $-\hbar\mathbf{k}_e$. Since a hole is considered as the valence band with one missing electron. The hole must have a crystal momentum equal to $-\hbar\mathbf{k}_e$, i.e.,

$$\mathbf{k}_h = -\mathbf{k}_e. \tag{6.79}$$

6.11.2 *Energy* $\varepsilon_h(\mathbf{k}_h) = -\varepsilon_e(\mathbf{k}_e)$

For a completely filled valence band, we can add the energy of all electrons to find its total energy. If the total energy is E, the energy will increase to $E-\varepsilon_e(\mathbf{k}_e)$ when the electron of energy $\varepsilon_e(\mathbf{k}_e)$ is excited into the conduction band. Note that $\varepsilon_e(\mathbf{k}_e)$ is negative for electrons. The completely filled valence band is the state of zero holes, and the valence band with one missing electron is the state of one hole. Obviously, the energy of the hole should be defined as the extra energy in the one-hole state over the state of zero holes. We have

$$\varepsilon_h(\mathbf{k}_h) = (E - \varepsilon_e(\mathbf{k}_e)) - E = -\varepsilon_e(\mathbf{k}_e). \tag{6.80}$$

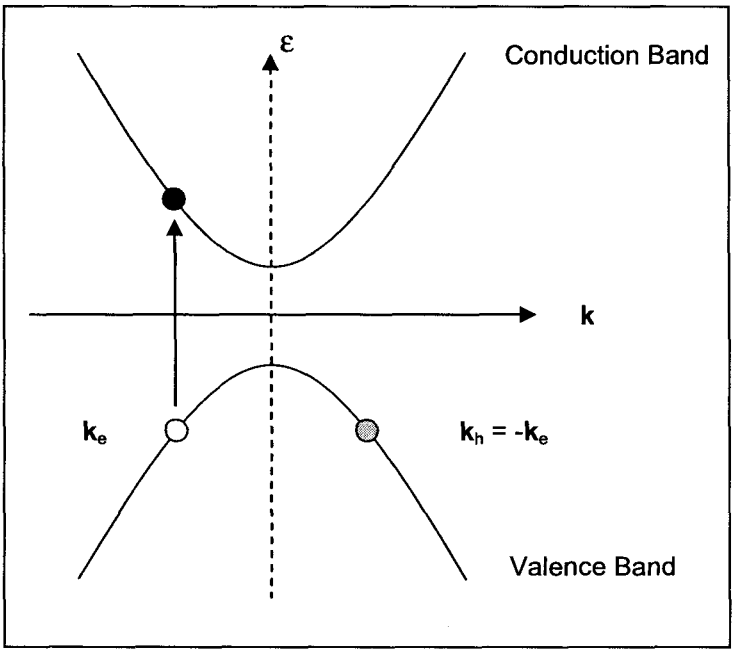


Fig. 6.8 Band diagram in the vicinity of Brillouin zone boundary. The wave vector $\mathbf{k}$ is measured from the boundary located at $\mathbf{G}/2$. In this diagram, $\mathbf{k}_e$ is negative. The empty circle represents the missing electron. The gray circle is the unpaired electron. The solid black circle is the electron excited into the conduction band.

6.11.3 *Velocity* $\mathbf{v}_h = \mathbf{v}_e$

Knowing the wave vector and energy of the hole, we can construct a corresponding band for holes from the underlying electron band. As shown in the Fig. 6.9, the negative branch is the electron band that is the valence band we are accustomed to, and the positive branch is the conceptual hole-band.

The electron to be excited away has a wave vector $\mathbf{k}_e$ and energy $\varepsilon_e(\mathbf{k}_e)$ in the valence band. Its velocity in the valence band is $(\nabla_{\mathbf{k}}\varepsilon_e)_{\mathbf{k}=\mathbf{k}_e}$. When this electron gets excited to the conduction band, the valence band behaves like a hole that has a wave vector $\mathbf{k}_h = -\mathbf{k}_e$ and energy $\varepsilon_h(\mathbf{k}_h) = -\varepsilon_e(\mathbf{k}_e)$. Therefore, we may consider holes as particles in an equivalent energy band. If we consider the k-axis as a horizontal mirror, the hole band can be obtained by taking the mirror image of the valence band and then rotating

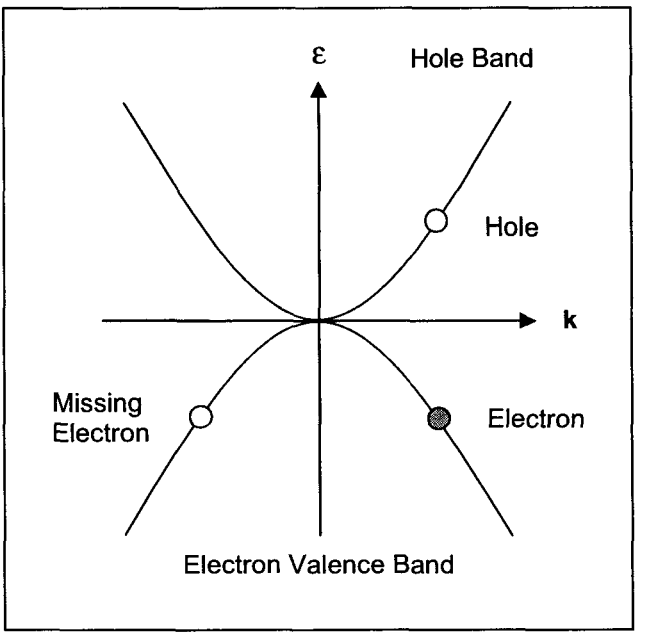


Fig. 6.9 Hole band constructed from the valence band with a missing electron. The empty circle in the valence band is at $(\mathbf{k}_e, \varepsilon_e(\mathbf{k}_e))$ where the electron is missing. The gray circle is the unpaired electron. The empty circle in the hole band is the hole with $\mathbf{k}_h = -\mathbf{k}_e$ and $\varepsilon_h(\mathbf{k}_h) = -\varepsilon_e(\mathbf{k}_e)$.

180° about the ε-axis. This symmetry operation satisfies $\mathbf{k}_h = -\mathbf{k}_e$ and $\varepsilon_h(\mathbf{k}_h) = -\varepsilon_e(\mathbf{k}_e)$. The resulted hole-band is shown as the positive branch in Fig. 6.9. The hole velocity in the hole band is $(\nabla_{\mathbf{k}}\varepsilon_h)_{\mathbf{k}=\mathbf{k}_h}$. Since the slope at $\mathbf{k}_e$ in the valence band is the same as the slope at $\mathbf{k}_h$ in the hole band, we have

$$(\nabla_{\mathbf{k}}\varepsilon_h)_{\mathbf{k}=\mathbf{k}_h} = (\nabla_{\mathbf{k}}\varepsilon_e)_{\mathbf{k}=\mathbf{k}_e},$$

i.e.,

$$\mathbf{v}_h = \mathbf{v}_e. \tag{6.81}$$

6.11.4 *Charge* $q_h = -q_e$

If we denote the electron charge q_e as $-e$, the hole charge q_h would be $+e$, i.e., $q_h = e = -q_e$. We can see this property of the hole by counting the

total electron charges in the valence band. The completely filled valence band is the state of zero holes, which has a total charge of $N \cdot q_e$. The state of one hole is the valence band with one electron missing, which has a total charge of $(N-1) \cdot q_e$. Therefore, the charge of a single hole is given by

$$q_h = (N-1) \cdot q_e - N \cdot q_e = -q_e = +e. \tag{6.82}$$

We can see this result from another point of view. Let us consider the equation of motion of an electron in the valence band under the influence of an electromagnetic field.

$$\hbar \frac{d\mathbf{k}_e}{dt} = q_e \cdot \left(\mathbf{E} + \frac{1}{c} \mathbf{v}_e \times \mathbf{B} \right), \tag{6.83}$$

and

$$\hbar \frac{d\mathbf{k}_e}{dt} = -e \cdot \left(\mathbf{E} + \frac{1}{c} \mathbf{v}_e \times \mathbf{B} \right). \tag{6.84}$$

We know that a hole with $\mathbf{k}_h = -\mathbf{k}_e$ and $\mathbf{v}_h = \mathbf{v}_e$ would be created if this electron gets taken away from the valence band. Mathematically, the above equation must hold true when we substitute $\mathbf{k}_e$ with an identical vector $-\mathbf{k}_h$, and $\mathbf{v}_e$ with an identical vector $\mathbf{v}_h$. So, we have

$$-\hbar \frac{d\mathbf{k}_h}{dt} = -e \cdot \left(\mathbf{E} + \frac{1}{c} \mathbf{v}_h \times \mathbf{B} \right), \tag{6.85}$$

and

$$\hbar \frac{d\mathbf{k}_h}{dt} = e \cdot \left(\mathbf{E} + \frac{1}{c} \mathbf{v}_h \times \mathbf{B} \right). \tag{6.86}$$

Let $q_h = e$, and we have

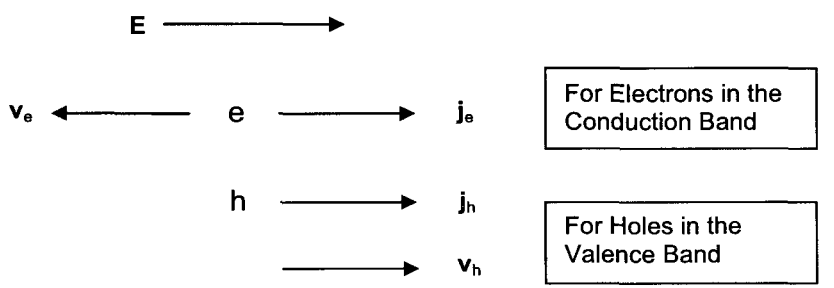
$$\hbar \frac{d\mathbf{k}_h}{dt} = q_h \cdot \left(\mathbf{E} + \frac{1}{c} \mathbf{v}_h \times \mathbf{B} \right). \tag{6.87}$$

Comparing Eqs (6.83) and (6.87), we found that they are of exactly the same mathematical form that describes the motion of a charge particle. Similar to Eq. (6.83) that describes the motion of the electron with a charge of $q_e = -e$, Eq. (6.87) is for the hole as if it were a particle having a charge of $q_h = +e$.

6.11.5 *Current* $\mathbf{j}_e = \mathbf{j}_h$

The net velocity of electrons in a completely filled valence band is zero because the symmetry $\varepsilon_e(\mathbf{k}_e) = \varepsilon_e(-\mathbf{k}_e)$. This is true for zero external fields. When the field is not zero, the net velocity is still zero because there are no available states for the electrons to move in a filled band. When the valence band is missing one electron, the whole band has a net velocity carried by the unpaired electron with or without applied electrical field. The current carried by the unpaired electron is

$$\begin{aligned} \mathbf{j}_e &= -e \cdot \mathbf{v}(\mathbf{k}_e) && \text{(unpaired electron)} \\ &= (-e)\,(-\mathbf{v}(-\mathbf{k}_e)) && \text{(missing electron)} \\ &= e \cdot \mathbf{v}(\mathbf{k}_h) && \text{(hole)} \\ &= \mathbf{j}_h. \end{aligned} \tag{6.88}$$



6.11.6 *Effective mass* $m_h^* = -m_e^*$

We recall that the band effective mass is inversely proportional to the band curvature. In the band diagram of Fig. 6.9, the curvature of the hole band is positive and the curvature of the valence band is negative. But, at the corresponding band locations of the hole and the missing electron, the absolute values of the curvature are the same. So, we have

$$m_h^* = \frac{\hbar^2}{\left(\frac{d^2\varepsilon}{dk^2}\right)_{k_h}} = -\frac{\hbar^2}{\left(\frac{d^2\varepsilon}{dk^2}\right)_{k_e}} = -m_e^*. \tag{6.89}$$

An alternative way to see $m_h^* = -m_e^*$ is to consider the valence band with one missing electron as the state of one hole, and the completely filled

valence band as the state of zero hole. Subtracting the total mass of the zero-hole state from that of the one-hole state, we would logically end up with the effective mass of one hole, i.e.,

$$m_h^* = \sum_{i=1}^{N-1} m_{e,i}^* - \sum_{i=1}^{N} m_{e,i}^* = -m_e^*. \tag{6.90}$$

Problems

1. When degeneracy reduces to none, i.e., m = 1 in Eq. (6.50), show the correction to the energy in Eq. (6.50) reduces to the same second order term described in Eq. (6.46) for the non-degenerate case.

2. For a linear lattice, draw a representative wave function for a free electron, and draw a corresponding Bloch wave function.

3. Modify Fig. 6.2 to show the effect of the weak periodic potential.

4. When we cut an onion, the cross section shows a series of concentric onion rings. Each ring represents a trace on the surface of a constant free-electron energy ε. Superimpose these rings with Fig. 6.3. Find the ring with radius equal to the $\mathbf{q}$ vector shown in Fig. 6.3, and show the effect of the weak periodical potential on this ring. Plot ε versus $q \cdot \cos\phi$, where ϕ is any fixed angle shown in Fig. 6.3.

5. Show $\nabla_{\mathbf{q}}\varepsilon = (\frac{\hbar^2}{m})(\mathbf{q} - \frac{\mathbf{G}}{2})$.

6. For a two-dimensional square lattice, if the Fermi surface intersects with the first Brillouin zone boundaries, draw
   (a) The First Brillouin zone and the free electron Fermi surface.
   (b) The distorted and torn Fermi surface.
   (c) Reduce the Fermi surface of (a) and (b) into the First Brillouin zone.
   (d) Show the direction of electron velocity on the Fermi surface, especially at the intersections between the Fermi surface and zone boundaries.

7. Show $\frac{(\varepsilon - \varepsilon_{\mathbf{q}}^0)}{U_{\mathbf{G}}} = \frac{U_{\mathbf{G}}^*}{(\varepsilon - \varepsilon_{\mathbf{q}-\mathbf{G}}^0)}$.

8. (a) Show

$$\frac{C_{\mathbf{q}-\mathbf{G}}}{C_{\mathbf{q}}} \approx \mp 1 + \frac{2\lambda}{|U_{\mathbf{G}}|}\frac{|\tilde{\mathbf{q}}|}{|\mathbf{G}/2|}\cos\theta \mp \frac{2\lambda^2}{|U_{\mathbf{G}}|^2}\left(\frac{|\tilde{\mathbf{q}}|}{|\mathbf{G}/2|}\right)^2 \cos^2\theta.$$

(b) Calculate both $C_{\mathbf{q}}$ and $C_{\mathbf{q}-\mathbf{G}}$, and plot them as functions of θ.

This page is intentionally left blank

Bibliography

The topics discussed in this book are the classical fundamentals, which have been well established and can be found in various required courses for physics students in the past half century or so, if it is not traced further back in time. But, it is still an art to pull them together in order to explain the concepts of phonons and electrons clearly. There are many textbooks that cover the same material. Of course, the philosophy of approach and style of presentation are quite different. This book takes a focused approach. Just a few topics are selected, which are believed to be of higher priority for the beginning students to learn first. On the selected subjects, the treatment is filled with much more details than usual. The objective is to prepare the students for a smoother sail when taking on other textbooks of solid state physics. The following books are used as resources and references, where broader coverage can also be found.

[1] Charles Kittel, *Introduction to Solid State Physics*, Seventh Edition, John Wiley and Sons, Inc., New York, 1996.

[2] Neil W. Ashcroft and N. David Mermin, *Solid State Physics*, Holt, Rinehart, and Winston, New York, 1976.

[3] Huang Kun and Han Lu Qi, *Solid State Physics*, Eighth Printing, the Publishing Company for High Education and Xin Hua Book Store, Beijing, 1997. (in Chinese).

[4] J. M. Ziman, *Electrons and Phonons*, Oxford University Press, London, 1958.

[5] Albert Messiah, *Quantum mechanics*, Dover Publication, Inc., Mineola, New York, 2000

[6] Richard P. Feynman, *Statistical Mechanics,A set of Lectures*, Seventh Printing, (Frontiers in Physics Lecture Note Series, David Pines, Editor), The

Benjamin/Cummings Publishing Company, Inc., Reading Massachusetts, 1992.

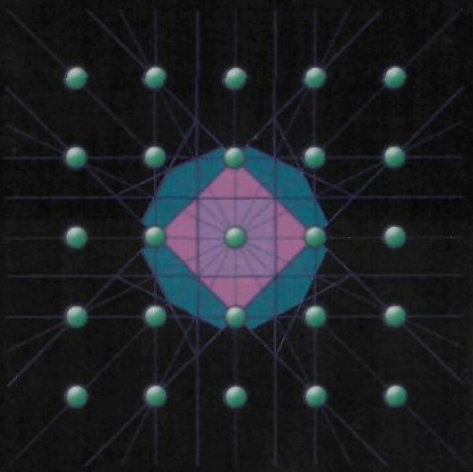
[7] L. D. Landau and E. M. lifshitz, *Statistical Physics (Course of Theoretical Physics, Volume 5)*, Butterworth-Heinemann, Oxford, 1999.

[8] Kerson Huang, *Statistical Mechanics*, Fourth Printing, John Wiley & Sons, Inc., New York, 1967.

[9] Hugh D. Young, *Statistical Treatment of Experimental Data*, McGraw-Hill Book Company, New York, 1962.

[10] Mary L. Boas, *Mathematical Methods in Physical Sciences*, John Wiley & Sons, Inc., New York, 1966.

Index



Introduction to Phonons and Electrons

This book focuses on phonons and electrons, which the student needs to learn first in solid state physics. The required quantum theory and statistical physics are derived from scratch. Systematic in structure and tutorial in style, the treatment is filled with detailed mathematical steps and physical interpretations. This approach ensures a self-sufficient content for easier teaching and learning. The objective is to introduce the concepts of phonons and electrons in a more rigorous and yet clearer way, so that the student does not need to relearn them in more advanced courses. Examples are the transition from lattice vibrations to phonons and from free electrons to energy bands.

The book can be used as the beginning module of a one-year introductory course on solid state physics, and the instructor will have a chance to choose additional topics. Alternatively, it can be taught as a stand-alone text for building the most-needed foundation in just one semester.

ISBN 981-238-439-1

9 789812 384393

www.worldscientific.com

5327 hc