John J. Quinn · Kyung-Soo Yi

Solid State Physics

Principles and Modern Applications

Springer

Solid State Physics

John J. Quinn · Kyung-Soo Yi

Solid State Physics

Principles and Modern Applications

Springer

Prof. John J. Quinn
University of Tennessee
Dept. Physics
Knoxville TN 37996
USA
jjquinn@utk.edu

Prof. Kyung-Soo Yi
Pusan National University
Dept. Physics
30 Jangjeon-Dong
Pusan 609-735
Korea, Republic of (South Korea)
ksyi@pusan.ac.kr

ISBN 978-3-540-92230-8 e-ISBN 978-3-540-92231-5
DOI 10.1007/978-3-540-92231-5
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009929177

Cover design: eStudio Calamar S.L.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

The book is dedicated to Betsy Quinn and Young-Sook Yi.

Preface

This textbook had its origin in several courses taught for two decades (1965–1985) at Brown University by one of the authors (JJQ). The original assigned text for the first semester course was the classic "Introduction to Solid State Physics" by C. Kittel. Many topics not covered in that text were included in subsequent semesters because of their importance in research during the 1960s through the 1980s. A number of the topics covered were first introduced in a course on "Many Body Theory of Metals" given by JJQ as a Visiting Lecturer at the University of Pennsylvania in 1961–1962, and later included in a course at Purdue University when he was a Visiting Professor (1964–1965). A sojourn into academic administration in 1984 removed JJQ from teaching for 8 years. On returning to a full time teaching–research professorship at the University of Tennessee, he again offered a 1 year graduate course in Solid State Physics. The course was structured so that the first semester (roughly the first half of the text) introduced all the essential concepts for students who wanted exposure to solid state physics. The first semester could cover topics from the first nine chapters. The second semester covered a selection of more advanced topics for students intending to do research in this field. One of the co-authors (KSY) took this course in Solid State Physics as a PhD student at Brown University. He added to and improved the lecture while teaching the subject at Pusan National University from 1984. The text is a true collaborative effort of the co-authors.

The advanced topics in the second semester are covered briefly, but thoroughly enough to convey the basic physics of each topic. References point the students who want more detail in the right direction. An entirely different set of advanced topics could have been chosen in the place of those in the text. The choice was made primarily because of the research interests of the authors.

In addition to Kittel's classic *Introduction to Solid State Physics*, 7th edn. (Wiley, New York, 1995), other books that influenced the evolution of this book are: *Methods of Quantum Field Theory in Statistical Physics* ed. by A.A. Abrikosov, L.P. Gorkov, I.E. Dzyaloshinsky (Prentice-Hall, Englewood,

NJ, 1963); *Solid State Physics* ed. by N.W. Ashcroft, N.D. Mermin (Saunder's College, New York, 1975); *Introduction to Solid State Theory* ed. by O. Madelung (Springer, Berlin, Heidelberg, New York, 1978); and *Fundamentals of Semiconductors* ed. by P.Y. Yu, M. Cardona (Springer, Berlin, Heidelberg, New York, 1995).

Many graduate students at Brown, Tennessee, and Pusan have helped to improve these lecture notes by pointing out sections that were difficult to understand, and by catching errors in the text. Dr. Alex Tselis presented the authors with his carefully written notes of the course at Brown when he changed his field of study to medical science. We are grateful to all the students and colleagues who have contributed to making the lecture notes better.

Both the co-authors want to acknowledge the encouragement and support of their families. The book is dedicated to them.

Knoxville and Pusan, *John J. Quinn*
August 2009 *Kyung-Soo Yi*

Contents

Part I

Basic Concepts in Solid-State Physics

1

Crystal Structures

1.1 Crystal Structure and Symmetry Groups

Although everyone has an intuitive idea of what a solid is, we will consider (in this book) only materials with a well-defined crystal structure. What we mean by a well-defined crystal structure is an arrangement of atoms in a *lattice* such that the atomic arrangement looks absolutely identical when viewed from two different points that are separated by a *lattice translation vector*. A few definitions are useful.

Lattice

A lattice is an infinite array of points obtained from three primitive translation vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$. Any point on the lattice is given by

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

Translation Vector

Any pair of lattice points can be connected by a vector of the form

$$\mathbf{T}_{n_1 n_2 n_3} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3. \tag{1.2}$$

The set of translation vectors form a group called the *translation group* of the lattice.

Group

A set of elements of any kind with a set of operations, by which any two elements may be combined into a third, satisfying the following requirements is called a *group*:

- The product (under group multiplication) of two elements of the group belongs to the group.

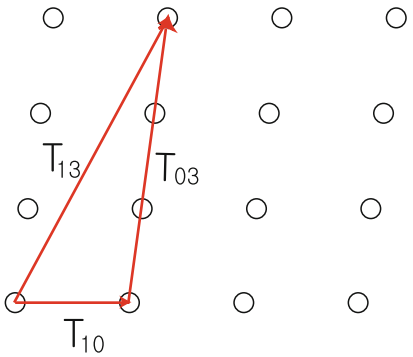


Fig. 1.1. Translation operations in a two-dimensional lattice

- The associative law holds for group multiplication.
- The identity element belongs to the group.
- Every element in the group has an inverse which belongs to the group.

Translation Group

The set of translations through any translation vector $\mathbf{T}_{n_1 n_2 n_3}$ forms a group. Group multiplication consists in simply performing the translation operations consecutively. For example, as is shown in Fig. 1.1, we have $\mathbf{T}_{13} = \mathbf{T}_{03} + \mathbf{T}_{10}$. For the simple translation group the operations commute, i.e., $\mathbf{T}_{ij}\mathbf{T}_{kl} = \mathbf{T}_{kl}\mathbf{T}_{ij}$ for, every pair of translation vectors. This property makes the group an *Abelian group*.

Point Group

There are other symmetry operations which leave the lattice unchanged. These are *rotations*, *reflections*, and the *inversion* operations. These operations form the *point group* of the lattice. As an example, consider the two-dimensional square lattice (Fig. 1.2). The following operations (performed about any lattice point) leave the lattice unchanged.

- E: identity
- R_1, R_3: rotations by $\pm 90°$
- R_2: rotation by $180°$
- m_x, m_y: reflections about x-axis and y-axis, respectively
- m_+, m_-: reflections about the lines $x = \pm y$

The multiplication table for this point group is given in Table 1.1. The operations in the first column are the first (right) operations, such as m_+ in $R_1 m_+ = m_y$, and the operations listed in the first row are the second (left) operations, such as R_1 in $R_1 m_+ = m_y$.

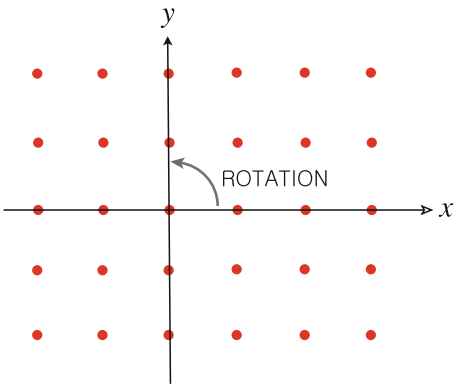


Fig. 1.2. The two-dimensional square lattice

Table 1.1. Multiplication table for the group 4 mm. The first (right) operations, such as m_+ in $R_1 m_+ = m_y$, are listed in the first column, and the second (left) operations, such as R_1 in $R_1 m_+ = m_y$, are listed in the first row

Operation	E	R_1	R_2	R_3	m_x	m_y	m_+	m_-
$E^{-1} = E$	E	R_1	R_2	R_3	m_x	m_y	m_+	m_-
$R_1^{-1} = R_3$	R_3	E	R_1	R_2	m_+	m_-	m_y	m_x
$R_2^{-1} = R_2$	R_2	R_3	E	R_1	m_y	m_x	m_-	m_+
$R_3^{-1} = R_1$	R_1	R_2	R_3	E	m_-	m_+	m_x	m_y
$m_x^{-1} = m_x$	m_x	m_+	m_y	m_-	E	R_2	R_1	R_3
$m_y^{-1} = m_y$	m_y	m_-	m_x	m_+	R_2	E	R_3	R_1
$m_+^{-1} = m_+$	m_+	m_y	m_-	m_x	R_3	R_1	E	R_2
$m_-^{-1} = m_-$	m_-	m_x	m_+	m_y	R_1	R_3	R_2	E

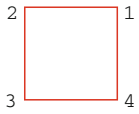


Fig. 1.3. Identity operation on a two-dimensional square

The multiplication table can be obtained as follows:

- Label the corners of the square (Fig. 1.3).
- Operating with a symmetry operation simply reorders the labeling. For example, see Fig. 1.4 for symmetry operations of $m_+, R_1,$ and m_x.

Therefore, $R_1 m_+ = m_y$. One can do exactly the same for all other products, for example, such as $m_y R_1 = m_+$. It is also very useful to note what happens to a point (x, y) under the operations of the point group (see Table 1.2). Note that under every group operation $x \to \pm x$ or $\pm y$ and $y \to \pm y$ or $\pm x$.

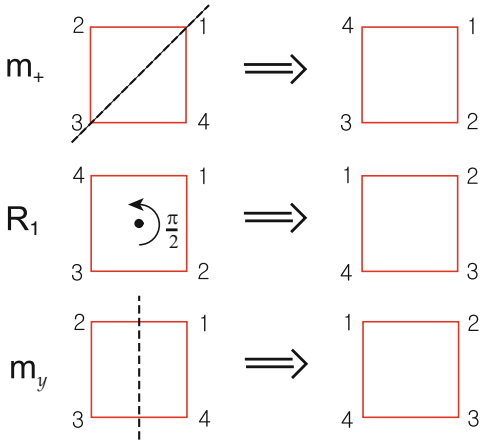


Fig. 1.4. Point symmetry operations on a two-dimensional square

Table 1.2. Point group operations on a point (x, y)

Operation	E	R_1	R_2	R_3	m_x	m_y	m_+	m_-
x	x	y	$-x$	$-y$	x	$-x$	y	$-y$
y	y	$-x$	$-y$	x	$-y$	y	x	$-x$

Fig. 1.5. The two-dimensional rectangular lattice

The point group of the two-dimensional square lattice is called 4 mm. The notation denotes the fact that it contains a fourfold axis of rotation and two mirror planes (m_x and m_y); the m_+ and m_- planes are required by the existence of the other operations. Another simple example is the symmerty group of a two-dimensional rectangular lattice (Fig. 1.5). The symmetry operations are E, R_2, m_x, m_y, and the multiplication table is easily obtained from that of 4 mm. This point group is called 2 mm, and it is a subgroup of 4 mm.

Allowed Rotations

Because of the requirement of translational invariance under operations of the translation group, the allowed rotations of the point group are restricted to certain angles. Consider a rotation through an angle ϕ about an axis through some lattice point (Fig. 1.6). If A and B are lattice points separated by a primitive translation $\mathbf{a}_1$, then A′ (and B′) must be a lattice point obtained by a rotation through angle ϕ about B (or $-\phi$ about A). Since A′ and B′

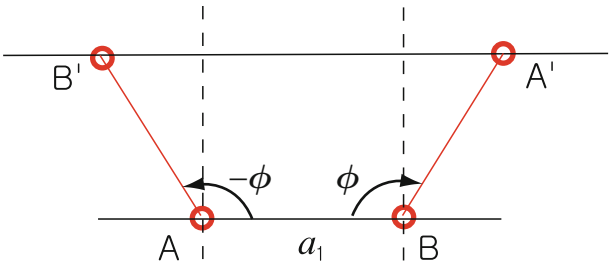


Fig. 1.6. Allowed rotations of angle ϕ about an axis passing through some lattice points A and B consistent with translational symmetry

Table 1.3. Allowed rotations of the point group

p	$\cos\phi$	ϕ	$n\,(=\lvert 2\pi/\phi\rvert)$
-1	1	0 or 2π	1
0	$\frac{1}{2}$	$\pm\frac{2\pi}{6}$	6
1	0	$\pm\frac{2\pi}{4}$	4
2	$-\frac{1}{2}$	$\pm\frac{2\pi}{3}$	3
3	-1	$\pm\frac{2\pi}{2}$	2

are lattice points, the vector B′A′ must be a translation vector. Therefore, we have

$$|\mathrm{B'A'}| = pa_1, \tag{1.3}$$

where p is an integer. But $|\mathrm{B'A'}| = a_1 + 2a_1 \sin\left(\phi - \frac{\pi}{2}\right) = a_1 - 2a_1 \cos\phi$. Solving for $\cos\phi$ gives

$$\cos\phi = \frac{1-p}{2}. \tag{1.4}$$

Because $-1 \leq \cos\phi \leq 1$, we see that p must have only the integral values -1, 0, 1, 2, 3. This gives for the possible values of ϕ listed in Table 1.3.

Although only rotations of $60°, 90°, 120°, 180°$, and $360°$ are consistent with translational symmetry, rotations through other angles are obtained in *quasicrystals* (e.g., fivefold rotations). The subject of quasicrystals, which do not have translational symmetry under the operations of the translation group, is an interesting modern topic in solid state physics which we will not discuss in this book.

Bravais Lattice

If there is only one atom associated with each lattice point, the lattice is called *Bravais lattice*. If there is more than one atom associated with each lattice point, the lattice is called *a lattice with a basis*. One atom can be considered to be located at the lattice point. For a lattice with a basis it is necessary to give the locations (or basis vectors) for the additional atoms associated with the lattice point.

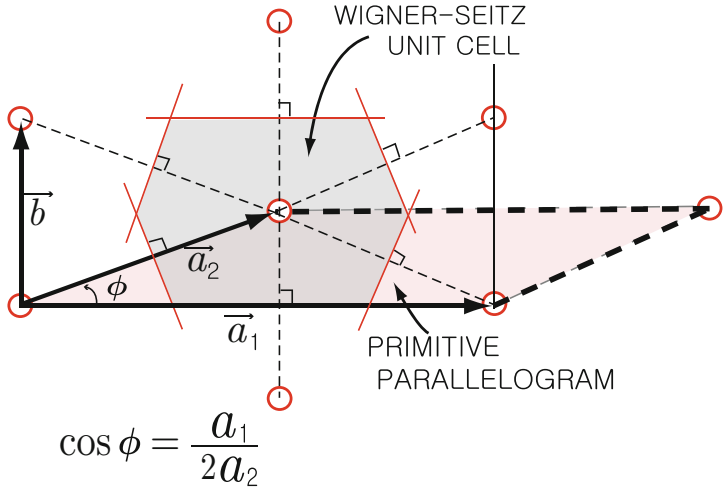


Fig. 1.7. Construction of the Wigner–Seitz cell of a two-dimensional centered rectangular lattice. Note that $\cos\phi = a_1/2a_2$

Primitive Unit Cell

From the three primitive translation vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, one can form a parallelepiped that can be used as a primitive unit cell. By stacking primitive unit cells together (like building blocks) one can fill all of space.

Wigner–Seitz Unit Cell

From the lattice point $(0, 0, 0)$ draw translation vectors to neighboring lattice points (to nearest, next nearest, etc., neighbors). Then, draw the planes which are perpendicular bisectors of these translation vectors (see, for example, Fig. 1.7). The interior of these intersecting planes (i.e., the space closer to $(0, 0, 0)$ than to any other lattice point) is called the *Wigner–Seitz unit cell*.

Space Group

For a Bravais lattice, the space group is simply the product of the operations of the translation group and of the point group. For a lattice with a basis, there can be other symmetry operations. Examples are *glide planes* and *screw axes*; illustration of each is shown in Figs. 1.8 and 1.9, respectively.

Glide Plane

In Fig. 1.8, each unit cell contains six atoms and $T_{1/2}m_x$ is a symmertry operation even though neither $T_{1/2}$ nor m_x is an operation of the symmetry group.

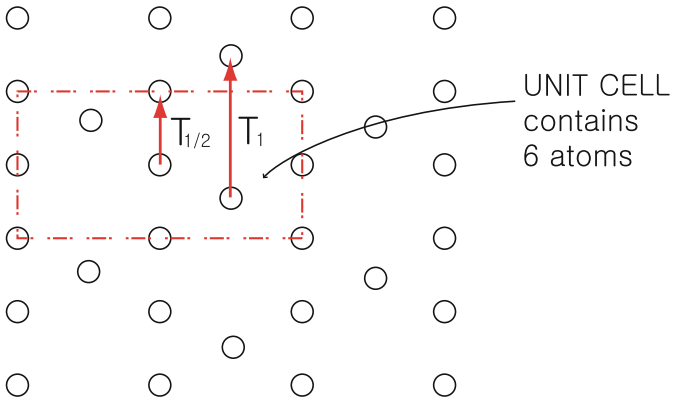


Fig. 1.8. Glide plane of a two-dimensional lattice. Each unit cell contains six atoms

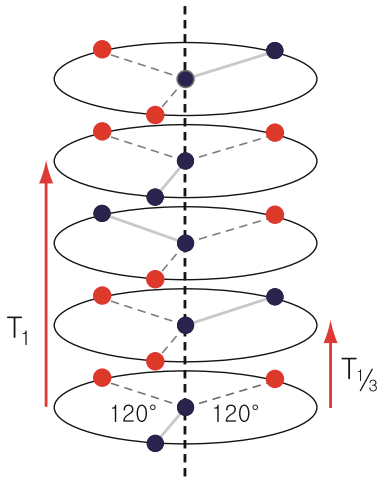


Fig. 1.9. Screw axis. Unit cell contains three layers and T_1 is the smallest translation. Occupied sites are shown by *solid dots*

Screw Axis

In Fig. 1.9, $T_{1/3}R_{120^\circ}$ is a symmetry operation even though $T_{1/3}$ and R_{120° themselves are not.

Two-Dimensional Bravais Lattices

There are only five different types of two-dimensional Bravais lattices.

1. Square lattice: primitive (P) one only
   It has $a = b$ and $\alpha = \beta = 90^\circ$.
2. Rectangular: primitive (P) and centered (C) ones
   They have $a \neq b$ but $\alpha = \beta = 90^\circ$.

3. Hexagonal: primitive (P) one only
   It has $a = b$ and $\alpha = 120°(\beta = 60°)$.
4. Oblique: primitive (P) one only
   It has $a \neq b$ and $\alpha \neq 90°$.

Three-Dimensional Bravais Lattices

There are 14 different types of three-dimensional Bravais lattices.

1. Cubic: primitive (P), body centered (I), and face centered (F) ones.
   For all of these $a = b = c$ and $\alpha = \beta = \gamma = 90°$.
2. Tetragonal: primitive and body centered (I) ones.
   For these $a = b \neq c$ and $\alpha = \beta = \gamma = 90°$. One can think of them as cubic lattices that have been stretched (or compressed) along one of the cube axes.
3. Orthorombic: primitive (P), body centered (I), face centered (F), and base centered (C) ones.
   For all of these $a \neq b \neq c$ but $\alpha = \beta = \gamma = 90°$. These can be thought of as cubic lattices that have been stretched (or compressed) by different amounts along two of the cube axes.
4. Monoclinic: primitive (P) and base centered (C) ones.
   For these $a \neq b \neq c$ and $\alpha = \beta = 90° \neq \gamma$. These can be thought of as orthorhombic lattices which have suffered shear distortion in one direction.
5. Triclinic: primitive (P) one.
   This has the lowest symmetry with $a \neq b \neq c$ and $\alpha \neq \beta \neq \gamma$.
6. Trigonal:
   It has $a = b = c$ and $\alpha = \beta = \gamma \neq 90° < 120°$. The primitive cell is a rhombohedron. The trigonal lattice can be thought of as a cubic lattice which has suffered shear distortion.
7. Hexagonal: primitive (P) one only.
   It has $a = b \neq c$ and $\alpha = \beta = 90°$, $\gamma = 120°$.

1.2 Common Crystal Structures

1. *Cubic*
   (a) Simple cubic (sc): Fig. 1.10
   For simple cubic crystal the lattice constant is a and the volume per atom is a^3. The nearest neighbor distance is also a, and each atom has six nearest neighbors. The primitive translation vectors are $\mathbf{a}_1 = a\hat{x}$, $\mathbf{a}_2 = a\hat{y}$, $\mathbf{a}_3 = a\hat{z}$.
   (b) Body centered cubic (bcc): Fig. 1.11
   If we take a unit cell as a cube of edge a, there are two atoms per cell (one at $(0, 0, 0)$ and one at $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$). The atomic volume is $\frac{1}{2}a^3$, and the nearest neighbor distance is $\frac{\sqrt{3}}{2}a$. Each atom has

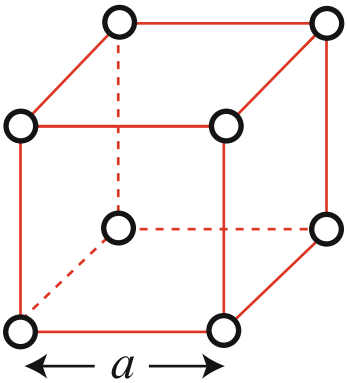


Fig. 1.10. Crystallographic unit cell of a simple cubic crystal of lattice constant a

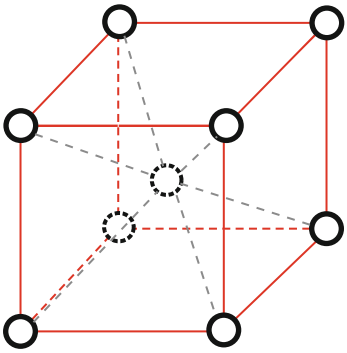


Fig. 1.11. Crystallographic unit cell of a body centered cubic crystal of lattice constant a

eight nearest neighbors. The primitive translations can be taken as $\mathbf{a}_1 = \frac{1}{2}a\,(\hat{x} + \hat{y} + \hat{z})$, $\mathbf{a}_2 = \frac{1}{2}a\,(-\hat{x} + \hat{y} + \hat{z})$, and $\mathbf{a}_3 = \frac{1}{2}a\,(-\hat{x} - \hat{y} + \hat{z})$. The parallelepiped formed by $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ is the primitive unit cell (containing a single atom), and there is only one atom per primitive unit cell.

(c) Face centered cubic (fcc): Fig. 1.12
If we take a unit cell as a cube of edge a, there are four atoms per cell; $\frac{1}{8}$ of one at each of the eight corners and $\frac{1}{2}$ of one on each of the six faces. The volume per atom is $\frac{a^3}{4}$; the nearest neighbor distance is $\frac{a}{\sqrt{2}}$, and each atom has 12 nearest neighbors. The primitive unit cell is the parallelepiped formed from the primitive translations $\mathbf{a}_1 = \frac{1}{2}a\,(\hat{x} + \hat{y})$, $\mathbf{a}_2 = \frac{1}{2}a\,(\hat{y} + \hat{z})$, and $\mathbf{a}_3 = \frac{1}{2}a\,(\hat{z} + \hat{x})$.

All three cubic lattices have the cubic group as their point group. Because the primitive translations are different, the simple cubic, bcc, and fcc lattices have different translation groups.

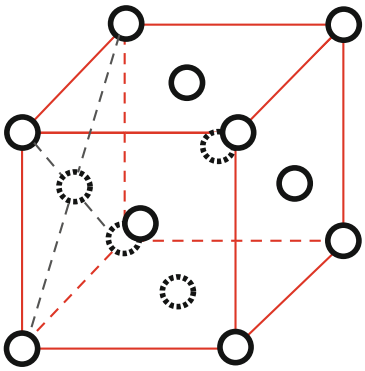


Fig. 1.12. Crystallographic unit cell of a face centered cubic crystal of lattice constant a

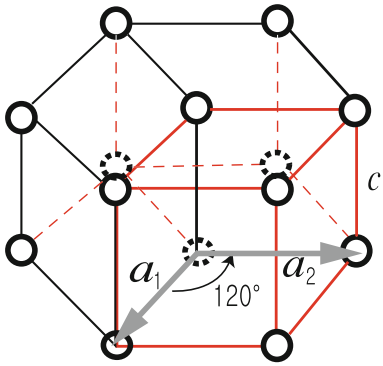


Fig. 1.13. Crystallographic unit cell of a simple hexagonal crystal of lattice constants a_1, a_2, and c

2. *Hexagonal*
   (a) Simple hexagonal: See Fig. 1.13.
   (b) Hexagonal close packed (hcp):
   This is a non-Bravais lattice. It contains two atoms per primitive unit cell of the simple hexagonal lattice, one at (0,0,0) and the second at $\left(\frac{1}{3}, \frac{2}{3}, \frac{1}{2}\right)$. The hexagonal close packed crystal can be formed by stacking the first layer (A) in a hexagonal array as is shown in Fig. 1.14. Then, the second layer (B) is formed by stacking atoms in the alternate triangular holes on top of the first layer. This gives another hexagonal layer displaced from the first layer by $\left(\frac{1}{3}, \frac{2}{3}, \frac{1}{2}\right)$. Then the third layer is placed above the first layer (i.e., at (0,0,1)). The stacking is then repeated ABABAB .... If one stacks ABCABC ..., where C is the hexagonal array obtained by stacking the third layer in the other set of triangular holes above the set B (instead of the set A), one gets an fcc lattice. The closest possible packing of the hcp atoms occurs

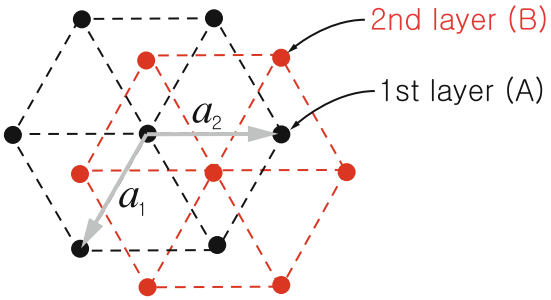


Fig. 1.14. Stacking of layers A and B in a hexagonal close packed crystal of lattice constants a_1, a_2, and c

when $\frac{c}{a} = \sqrt{8/3} \approx 1.633$. We leave this as an exercise for the reader. Zn crystalizes in a hcp lattice with $a = 2.66\,\text{Å}$ and $c = 4.96\,\text{Å}$ giving $\frac{c}{a} \approx 1.85$, larger than the ideal $\frac{c}{a}$ value.

3. *Zincblende Structure* This is a non-Bravais lattice. It is an FCC with two atoms per primitive unit cell located at (0,0,0) and $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$. The structure can be viewed as two interpenetrating fcc lattices displaced by one fourth of the body diagonal. Examples of the zincblende structure are ZnS (cubic phase), ZnO (cubic phase), CuF, CuCl, ZnSe, CdS, GaN (cubic phase), InAs, and InSb. The metallic ions are on one sublattice, the other ions on the second sublattice.
4. *Diamond Structure* This structure is identical to the zincblende structure, except that there are two identical atoms in the unit cell. This structure (unlike zincblende) has inversion symmetry about the point $\left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$. Diamond, Si, Ge, and gray tin are examples of the diamond structure.
5. *Wurtzite Structure* This structure is a simple hexagonal lattice with four atoms per unit cell, located at (0,0,0), $\left(\frac{1}{3}, \frac{2}{3}, \frac{1}{2}\right)$, $\left(0, 0, \frac{3}{8}\right)$, and $\left(\frac{1}{3}, \frac{2}{3}, \frac{7}{8}\right)$. It can be pictured as consisting of two interpenetrating hcp lattices separated by $\left(0, 0, \frac{3}{8}\right)$. In the wurtzite phase of ZnS, the Zn atoms sit on one hcp lattice and the S atoms on the other. ZnS, BeO, ZnO (hexagonal phase), CdS, GaN (hexagonal phase), and AlN are materials that can occur in the wurtzite structure.
6. *Sodium Chloride Structure* It consists of a face centered cubic lattice with a basis of two unlike atoms per primitive unit cell, located at (0,0,0) and $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$. In addition to NaCl, other alkali halide salts like LiH, KBr, RbI form crystals with this structure.
7. *Cesium Chloride Structure* It consists of a simple cubic lattice with two atoms per unit cell, located at (0,0,0) and $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$. Besides CsCl, CuZn ($\beta$-brass), AgMg, and LiHg occur with this structure.
8. *Calcium Fluoride Structure* It consists of a face centered cubic lattice with three atoms per primitive unit cell. The Ca ion is located at (0,0,0), the F atoms at $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ and $\left(\frac{3}{4}, \frac{3}{4}, \frac{3}{4}\right)$.

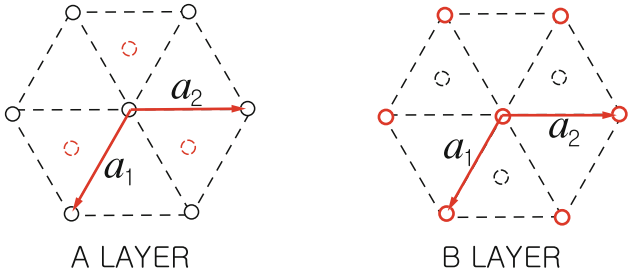


Fig. 1.15. Stacking of layers A and B in a graphite structure

9. *Graphite Structure* This structure consists of a simple hexagonal lattice with four atoms per primitive unit cell, located at (0,0,0), $\left(\frac{2}{3}, \frac{1}{3}, 0\right)$, $\left(0, 0, \frac{1}{2}\right)$, and $\left(\frac{1}{3}, \frac{2}{3}, \frac{1}{2}\right)$. It can be pictured as two interpenetrating HCP lattices separated by $\left(0, 0, \frac{1}{2}\right)$. It, therefore, consists of tightly bonded planes (as shown in Fig. 1.15) stacked in the sequence ABABAB . . .. The individual planes are very tightly bound, but the interplanar binding is rather weak. This gives graphite its well-known properties, like easily cleaving perpendicular to the c-axis.

Miller Indices

Miller indices are a set of three integers that specify the orientation of a crystal plane. The procedure for obtaining Miller indices of a plane is as follows:

1. Find the intercepts of the plane with the crystal axes.
2. Take the reciprocals of the three numbers.
3. Reduce (by multiplying by the same number) this set of numbers to the smallest possible set of integers.

As an example, consider the plane that intersects the cubic axes at A_1, A_2, A_3 as shown in Fig. 1.16. Then $x_i a_i = \overline{OA_i}$. The reciprocals of (x_1, x_2, x_3) are $\left(x_1^{-1}, x_2^{-1}, x_3^{-1}\right)$, and the Miller indices of the plane are $(h_1 h_2 h_3) = p\left(x_1^{-1}, x_2^{-1}, x_3^{-1}\right)$, where $(h_1 h_2 h_3)$ are the smallest possible set of integers $\left(\frac{p}{x_1}, \frac{p}{x_2}, \frac{p}{x_3}\right)$.

Indices of a Direction

A direction in the lattice can be specified by a vector $\mathbf{V} = u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3$, or by the set of integers $[u_1 u_2 u_3]$ chosen to have no common integral factor. For cubic lattices the plane $(h_1 h_2 h_3)$ is perpendicular to the direction $[h_1 h_2 h_3]$, but this is not true for general lattices.

Packing Fraction

The packing fraction of a crystal structure is defined as the ratio of the volume of atomic spheres in the unit cell to the volume of the unit cell.

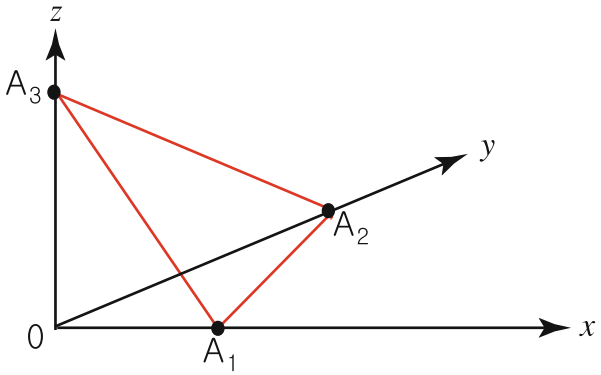


Fig. 1.16. Intercepts of a plane with the crystal axes

Examples

1. Simple cubic lattice:
   We take the atomic radius as $R = \frac{a}{2}$ (then neighboring atoms just touch). The packing fraction p will be given by

$$p = \frac{\frac{4}{3}\pi\left(\frac{a}{2}\right)^3}{a^3} = \frac{\pi}{6} \approx 0.52$$

2. Body centered cubic lattice:
   Here, we take $R = \frac{1}{2}\left(\frac{\sqrt{3}}{2}a\right)$, i.e., half the nearest neighbor distance. For the non-primitive cubic cell of edge a, we have two atoms per cell giving

$$p = \frac{2 \times \frac{4}{3}\pi\left(\frac{a\sqrt{3}}{4}\right)^3}{a^3} = \frac{\pi}{8}\sqrt{3} \approx 0.68$$

1.3 Reciprocal Lattice

If $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ are the primitive translations of some lattice, we can define the vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ by the condition

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}, \tag{1.5}$$

where $\delta_{ij} = 0$ if i is not equal to j and $\delta_{ii} = 1$. It is easy to see that

$$\mathbf{b}_i = 2\pi\frac{\mathbf{a}_j \times \mathbf{a}_k}{\mathbf{a}_i \cdot (\mathbf{a}_j \times \mathbf{a}_k)}, \tag{1.6}$$

where i, j, and k are different. The denominator $\mathbf{a}_i \cdot (\mathbf{a}_j \times \mathbf{a}_k)$ is simply the volume v_0 of the primitive unit cell. The lattice formed by the primitive translation vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ is called the *reciprocal lattice* (reciprocal to the lattice formed by $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$), and a reciprocal lattice vector is given by

$$\mathbf{G}_{h_1 h_2 h_3} = h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2 + h_3 \mathbf{b}_3. \tag{1.7}$$

Useful Properties of the Reciprocal Lattice

1. If $\mathbf{r} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3$ is a lattice vector, then we can write $\mathbf{r}$ as
$$\mathbf{r} = \sum_i (\mathbf{r} \cdot \mathbf{b}_i)\, \mathbf{a}_i. \tag{1.8}$$
2. The lattice reciprocal to $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ is $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$.
3. A vector $\mathbf{G_h}$ from the origin to a point (h_1, h_2, h_3) of the reciprocal lattice is perpendicular to the plane with Miller indices $(h_1 h_2 h_3)$.
4. The distance from the origin to the first lattice plane $(h_1 h_2 h_3)$ is $d\,(h_1 h_2 h_3) = 2\pi\, |\mathbf{G_h}|^{-1}$. This is also the distance between neighboring $\{h_1 h_2 h_3\}$ planes.

The proof of 3 is established by demonstrating that $\mathbf{G_h}$ is perpendicular to the plane $A_1 A_2 A_3$ shown in Fig. 1.16. This must be true if $\mathbf{G_h}$ is perpendicular to both $\overline{A_1 A_2}$ and to $\overline{A_2 A_3}$. But $\overline{A_1 A_2} = \overline{OA_2} - \overline{OA_1} = p\left(\frac{\mathbf{a}_2}{h_2} - \frac{\mathbf{a}_1}{h_1}\right)$. Therefore,

$$\mathbf{G_h} \cdot \overline{A_1 A_2} = (h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2 + h_3 \mathbf{b}_3) \cdot p \left(\frac{\mathbf{a}_2}{h_2} - \frac{\mathbf{a}_1}{h_1} \right), \tag{1.9}$$

which vanishes. The same can be done for $\overline{A_2 A_3}$. The proof of 4 is established by noting that

$$d(h_1 h_2 h_3) = \frac{\mathbf{a}_1}{h_1} \cdot \frac{\mathbf{G_h}}{|\mathbf{G_h}|}.$$

The first factor is just the vector $\overline{OA_1}$ for the situation where $p = 1$, and the second factor is a unit vector perpendicular to the plane $(h_1 h_2 h_3)$. Since $\mathbf{a}_1 \cdot \mathbf{G_h} = 2\pi h_1$, it is apparent that $d(h_1 h_2 h_3) = 2\pi\, |\mathbf{G_h}|^{-1}$.

1.4 Diffraction of X-Rays

Crystal structures are usually determined experimentally by studying how the crystal diffracts waves. Because the interatomic spacings in most crystals are of the order of a few Å's ($1\,\text{Å} = 10^{-8}$ cm), the maximum information can most readily be obtained by using waves whose wave lengths are of that

order of magnitude. Electromagnetic, electron, or neutron waves can be used to study diffraction by a crystal. For electromagnetic waves, $E = h\nu$, where E is the energy of the photon, $\nu = \frac{c}{\lambda}$ is its frequency and λ its wave length, and h is Planck's constant. For $\lambda = 10^{-8}$ cm, $c = 3 \times 10^{10}$ cm/s and $h = 6.6 \times 10^{-27}$ erg · s, the photon energy is equal to roughly 2×10^{-8} ergs or 1.24×10^{4} eV. Photons of energies of tens of kilovolts are in the X-ray range. For electron waves, $p = \frac{h}{\lambda} \simeq 6.6 \times 10^{-19}$ g · cm/s when $\lambda = 10^{-8}$ cm. This gives $E = \frac{p^2}{2m_e}$, where $m_e \simeq 0.9 \times 10^{-27}$g, of 2.4×10^{-10} ergs or roughly 150 eV. For neutron waves, we need simply replace m_e by $m_n = 1.67 \times 10^{-24}$ g to obtain $E = 1.3 \times 10^{-13}$ ergs $\simeq 0.08$ eV. Thus neutron energies are of the order of a tenth of an eV. Neutron scattering has the advantages that the low energy makes inelastic scattering studies more accurate and that the magnetic moment of the neutron allows the researcher to obtain information about the magnetic structure. It has the disadvantage that high intensity neutron sources are not as easily obtained as X-ray sources.

1.4.1 Bragg Reflection

We have already seen that we can discuss crystal planes in a lattice structure. Assume that an incident X-ray is specularly reflected by a set of crystal planes as shown in Fig. 1.17. Constructive interference occurs when the difference in path length is an integral number of wave length λ. It is clear that this occurs when

$$2d \sin\theta = n\lambda, \tag{1.10}$$

where d is the interplane spacing, θ is the angle between the incident beam and the crystal planes, as is shown on the figure, and n is an integer. Equation (1.10) is known as *Bragg's law*.

1.4.2 Laue Equations

A slightly more elegant discussion of diffraction from a crystal can be obtained as follows:

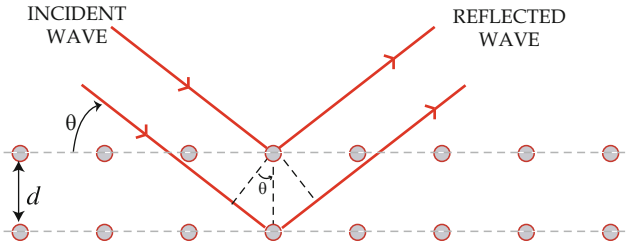


Fig. 1.17. Specular reflection of X-rays by a set of crystal planes separated by a distance d

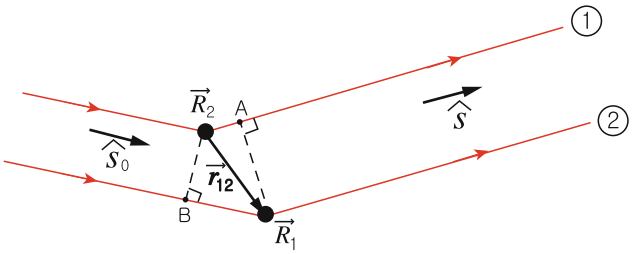


Fig. 1.18. Scattering of X-rays by a pair of atoms in a crystal

1. Let $\hat{s}_0$ be a unit vector in the direction of the incident wave, and $\hat{s}$ be a unit vector in the direction of the scattered wave.
2. Let $\mathbf{R}_1$ and $\mathbf{R}_2$ be the position vectors of a pair of atoms in a Bravais lattice, and let $\mathbf{r}_{12} = \mathbf{R}_1 - \mathbf{R}_2$.

Let us consider the waves scattered by $\mathbf{R}_1$ and by $\mathbf{R}_2$ and traveling different path lengths as shown in Fig. 1.18. The difference in path length is $|\,\overline{R_2A} - \overline{BR_1}\,|$. But this is clearly equal to $|\mathbf{r}_{12} \cdot \hat{s} - \mathbf{r}_{12} \cdot \hat{s}_0|$. We define $\mathbf{S}$ as $\mathbf{S} = \hat{s} - \hat{s}_0$; then the difference in path length for the two rays is given by

$$\Delta = |\mathbf{r}_{12} \cdot \mathbf{S}|\,. \tag{1.11}$$

For constructive interference, this must be equal to an integral number of wave length. Thus, we obtain

$$\mathbf{r}_{12} \cdot \mathbf{S} = m\lambda, \tag{1.12}$$

where m is an integer and λ is the wave length. To obtain constructive interference from every atom in the Bravais lattice, this must be true for every lattice vector $\mathbf{R}_n$. Constructive interference will occur only if

$$\mathbf{R_n} \cdot \mathbf{S} = \text{integer} \times \lambda \tag{1.13}$$

for every lattice vector $\mathbf{R_n}$ in the crystal. Of course there will be different integers for different $\mathbf{R_n}$ in general. Recall that

$$\mathbf{R_n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3. \tag{1.14}$$

The condition (1.13) is obviously satisfied if

$$\mathbf{a}_i \cdot \mathbf{S} = ph_i\lambda, \tag{1.15}$$

where h_i is the smallest set of integers and p is a common multiplier. We can obviously express $\mathbf{S}$ as

$$\mathbf{S} = (\mathbf{S} \cdot \mathbf{a}_1)\,\mathbf{b}_1 + (\mathbf{S} \cdot \mathbf{a}_2)\,\mathbf{b}_2 + (\mathbf{S} \cdot \mathbf{a}_3)\,\mathbf{b}_3. \tag{1.16}$$

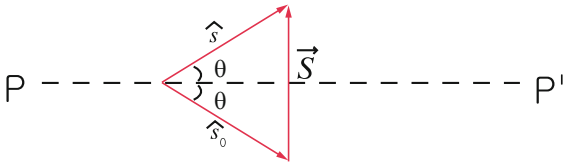


Fig. 1.19. Relation between the scattering vector $\mathbf{S} = \hat{s} - \hat{s}_0$ and the Bragg angle θ

Therefore, condition (1.13) is satisfied and constructive interference from every lattice site occurs if

$$\mathbf{S} = p\,(h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3)\,\lambda, \tag{1.17}$$

or

$$\frac{\mathbf{S}}{\lambda} = p\mathbf{G}_\mathbf{h}, \tag{1.18}$$

where $\mathbf{G_h}$ is a vector of the reciprocal lattice. Equation (1.18) is called the *Laue equation*.

Connection of Laue Equations and Bragg's Law

From (1.18) $\mathbf{S}$ must be perpendicular to the planes with Miller indices $(h_1h_2h_3)$. The distance between two planes of this set is

$$d(h_1h_2h_3) = \frac{2\pi}{|\mathbf{G_h}|} = p\frac{\lambda}{|\mathbf{S}|}. \tag{1.19}$$

We know that $\mathbf{S}$ is normal to the reflection plane PP′ with Miller indices $(h_1h_2h_3)$. From Fig. 1.19, it is apparent that $|\mathbf{S}| = 2\sin\theta$. Therefore, (1.19) can be written by

$$2d(h_1h_2h_3)\sin\theta = p\lambda,$$

where p is an integer. According to Laue's equation, associated with any reciprocal lattice vector $\mathbf{G_h} = h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3$, there is an X-ray reflection satisfying the equation $\lambda^{-1}S = pG_\mathbf{h}$, where p is an integer.

1.4.3 Ewald Construction

This is a geometric construction that illustrates how the Laue equation works. The construction goes as follows: See Fig. 1.20.

1. From the origin O of the reciprocal lattice draw the vector $\overline{\mathrm{AO}}$ of length λ^{-1} parallel to $\hat{s}_0$ and terminating on O.
2. Construct a sphere of radius λ^{-1} centered at A.

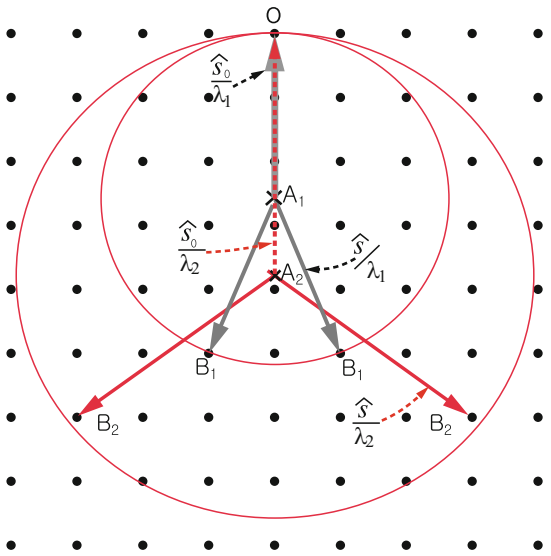


Fig. 1.20. Ewald construction for diffraction peaks

If this sphere intersects a point **B** of the reciprocal lattice, then $\overline{AB} = \frac{\hat{s}}{\lambda}$ is in a direction in which a diffraction maximum occurs. Since $\overline{A_1O} = \frac{\hat{s}_0}{\lambda_1}$ and $\overline{A_1B_1} = \frac{\hat{s}}{\lambda_1}$, $\frac{\mathbf{S}}{\lambda_1} = \frac{\hat{s}-\hat{s}_0}{\lambda_1} = \overline{OB_1}$ is a reciprocal lattice vector and satisfies the Laue equation. If a higher frequency X-ray is used, λ_2, A_2, and B_2 replace λ_1, A_1, and B_1. For a continuous spectrum with $\lambda_1 \geq \lambda \geq \lambda_2$, all reciprocal lattice points between the two sphere (of radii λ_1^{-1} and λ_2^{-1}) satisfy Laue equation for some frequency in the incident beam.

Wave Vector

It is often convenient to use the set of vectors $\mathbf{K_h} = 2\pi\mathbf{G_h}$. Then, the Ewald construction gives

$$\mathbf{q}_0 + \mathbf{K_h} = \mathbf{q}, \tag{1.20}$$

where $\mathbf{q}_0 = \frac{2\pi}{\lambda}\hat{s}_0$ and $\mathbf{q} = \frac{2\pi}{\lambda}\hat{s}$ are the wave vectors of the incident and scattered waves. Equation (1.20) says that wave vector is conserved up to 2π times a vector of the reciprocal lattice.

1.4.4 Atomic Scattering Factor

It is the electrons of an atom that scatter the X-rays since the nucleus is so heavy that it hardly moves in response to the rapidly varying electric field of the X-ray. So far, we have treated all of the electrons as if they were localized at the lattice point. In fact, the electrons are distributed about the nucleus of the atom (at position $\mathbf{r} = 0$, the lattice point) with a density $\rho(\mathbf{r})$. If you know

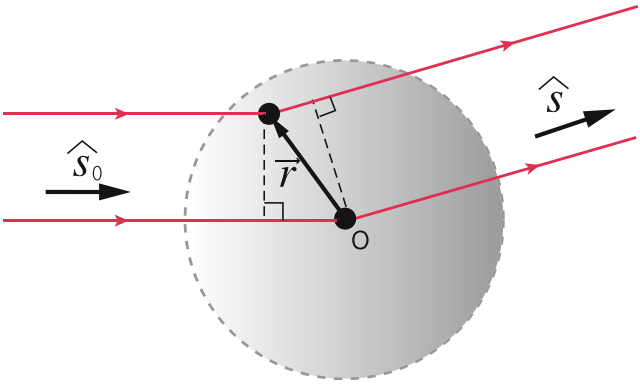


Fig. 1.21. Path difference between waves scattered at O and those at **r**

the wave function $\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_z)$ describing the z electrons of the atom, $\rho(\mathbf{r})$ is given by

$$\rho(\mathbf{r}) = \left\langle \sum_{i=1}^{z} \delta(\mathbf{r} - \mathbf{r}_i) \right\rangle = \left\langle \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_z) \left| \sum_{i=1}^{z} \delta(\mathbf{r} - \mathbf{r}_i) \right| \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_z) \right\rangle. \tag{1.21}$$

Now, consider the difference in path length Δ between waves scattered at O and those scattered at **r** (Fig. 1.21).

$$\Delta = \mathbf{r} \cdot (\hat{s} - \hat{s}_0) = \mathbf{r} \cdot \mathbf{S}. \tag{1.22}$$

The phase difference is simply $\frac{2\pi}{\lambda}$ times Δ, the difference in path length. Therefore, the scattering amplitude will be reduced from the value obtained by assuming all the electrons were localized at the origin O by a factor $z^{-1}f$, where f is given by

$$f = \int \mathrm{d}^3 r\, \rho(\mathbf{r})\, \mathrm{e}^{\frac{2\pi \mathrm{i}}{\lambda} \mathbf{r} \cdot \mathbf{S}}. \tag{1.23}$$

This factor is called the *atomic scattering factor*. If $\rho(\mathbf{r})$ is spherically symmetric we have

$$f = \int_0^\infty \int_{-1}^{1} 2\pi r^2 \mathrm{d}r\, \mathrm{d}(\cos\phi) \rho(r) \mathrm{e}^{\frac{2\pi \mathrm{i}}{\lambda} S r \cos\phi}. \tag{1.24}$$

Recall that $S = 2\sin\theta$, where θ is the angle between $\hat{s}_0$ and the reflecting plane PP′ of Fig. 1.19. Define μ as $\frac{4\pi}{\lambda}\sin\theta$; then f can be expressed as

$$f = \int_0^\infty \mathrm{d}r 4\pi r^2 \rho(r) \frac{\sin \mu r}{\mu r}. \tag{1.25}$$

If λ is much larger than the atomic radius, μr is much smaller than unity wherever $\rho(r)$ is finite. In that case $\frac{\sin \mu r}{\mu r} \simeq 1$ and $f \to z$, the number of electrons.

1.4.5 Geometric Structure Factor

So far we have considered only a Bravais lattice. For a non-Bravais lattice the scattered amplitude depends on the locations and atomic scattering factors of all the atoms in the unit cell. Suppose a crystal structure contains atoms at positions $\mathbf{r}_j$ with atomic scattering factors f_j. It is not difficult to see that this changes the scattered amplitude by a factor

$$F(h_1, h_2, h_3) = \sum_j f_j \mathrm{e}^{\frac{2\pi \mathrm{i}}{\lambda}\mathbf{r}_j \cdot \mathbf{S}(h_1 h_2 h_3)} \tag{1.26}$$

for the scattering from a plane with Miller indices $(h_1 h_2 h_3)$. In (1.26) the position vector $\mathbf{r}_j$ of the jth atom can be expressed in terms of the primitive translation vectors $\mathbf{a}_i$

$$\mathbf{r}_j = \sum_i \mu_{ji} \mathbf{a}_i. \tag{1.27}$$

For example, in a hcp lattice $\mathbf{r}_1 = (0, 0, 0)$ and $\mathbf{r}_2 = (\frac{1}{3}, \frac{2}{3}, \frac{1}{2})$ when expressed in terms of the primitive translation vectors. Of course, $\mathbf{S}(h_1 h_2 h_3)$ equal to $\lambda \sum_i h_i \mathbf{b}_i$, where $\mathbf{b}_i$ are primitive translation vectors in the reciprocal lattice. Therefore, $\frac{2\pi \mathrm{i}}{\lambda}\mathbf{r}_j \cdot \mathbf{S}(h_1 h_2 h_3)$ is equal to $2\pi \mathrm{i}\,(\mu_{j1} h_1 + \mu_{j2} h_2 + \mu_{j3} h_3)$, and the *structure amplitude* $F(h_1, h_2, h_3)$ can be expressed as

$$F(h_1, h_2, h_3) = \sum_j f_j \mathrm{e}^{2\pi \mathrm{i} \sum_i \mu_{ji} h_i}. \tag{1.28}$$

If all of the atoms in the unit cell are identical (as in diamond, Si, Ge, etc.) all of the atomic scattering factors f_j are equal, and we can write

$$F(h_1, h_2, h_3) = f \mathcal{S}(h_1 h_2 h_3). \tag{1.29}$$

The $\mathcal{S}(h_1 h_2 h_3)$ is called the *geometric structure amplitude*. It depends only on crystal structure, not on the atomic constituents, so it is the same for all hcp lattices or for all diamond lattices, etc.

Example

A useful demonstration of the geometric structure factor can be obtained by considering a bcc lattice as a simple cubic lattice with two atoms in the simple cubic unit cell located at (0,0,0) and $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Then

$$\mathcal{S}(h_1 h_2 h_3) = 1 + \mathrm{e}^{2\pi \mathrm{i}\left(\frac{1}{2}h_1 + \frac{1}{2}h_2 + \frac{1}{2}h_3\right)}. \tag{1.30}$$

If $h_1 + h_2 + h_3$ is odd, $\mathrm{e}^{\mathrm{i}\pi(h_1+h_2+h_3)} = -1$ and $\mathcal{S}(h_1 h_2 h_3)$ vanishes. If $h_1 + h_2 + h_3$ is even, $\mathcal{S}(h_1 h_2 h_3) = 2$. The reason for this effect is that the additional planes (associated with the body centered atoms) exactly cancel the scattering amplitude from the planes made up of corner atoms when $h_1 + h_2 + h_3$ is odd, but they add constructively when $h_1 + h_2 + h_3$ is even.

The scattering amplitude depends on other factors (e.g. thermal motion and zero point vibrations of the atoms), which we have neglected by assuming a perfect and stationary lattice.

1.4.6 Experimental Techniques

We know that constructive interference from a set of lattice planes separated by a distance d will occur when

$$2d \sin\theta = n\lambda, \tag{1.31}$$

where θ is the angle between the incident beam and the planes that are scattering, λ is the X-ray wave length, and n is an integer. For a given crystal the possible values of d are fixed by the atomic spacing, and to satisfy (1.31), one must vary either θ or λ over a range of values. Different experimental methods satisfy (1.31) in different ways. The common techniques are (1) the *Laue method*, (2) the *rotating crystal method*, and (3) the *powder method*.

Laue Method

In this method a single crystal is held stationary in a beam of continuous wave length X-ray radiation (Fig. 1.22). Various crystal planes select the appropriate wave length for constructive interference, and a geometric arrangement of bright spots is obtained on a film.

Rotating Crystal Method

In this method a monochromatic beam of X-ray is incident on a rotating single crystal sample. Diffraction maxima occur when the sample orientation relative to the incident beam satisfies Bragg's law (Fig. 1.23).

Powder Method

Here, a monochromatic beam is incident on a finely powdered specimen. The small crystallites are randomly oriented with respect to the incident beam, so that the reciprocal lattice structure used in the Ewald construction must be rotated about the origin of reciprocal space through all possible angles. This gives a series of spheres in reciprocal space of radii K_1, K_2, ... (we include the factor 2π in these reciprocal lattice vectors) equal to the smallest,

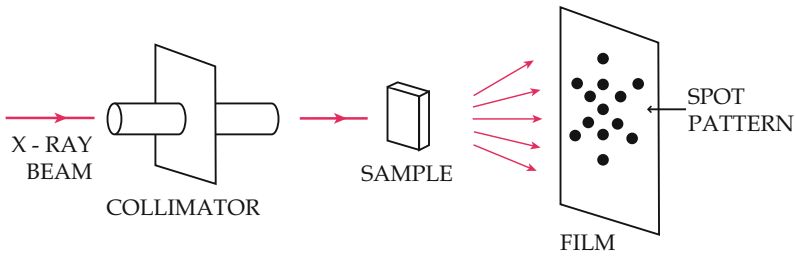


Fig. 1.22. Experimental arrangement of the Laue method

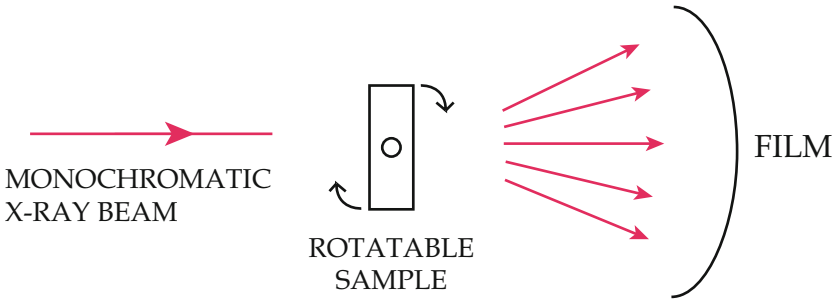


Fig. 1.23. Experimental arrangement of the rotating crystal method

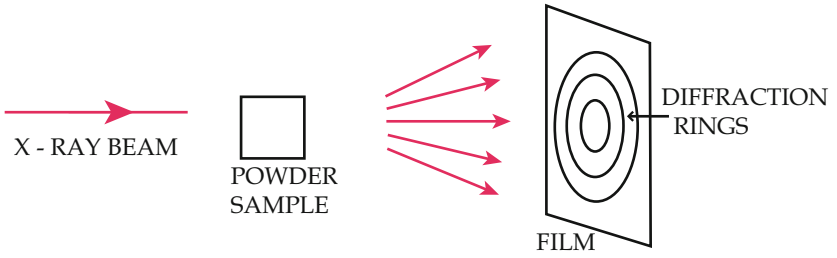


Fig. 1.24. Experimental arrangement of the powder method

next smallest, etc. reciprocal lattice vectors. The sequence of values $\frac{\sin(\phi_i/2)}{\sin(\phi_1/2)}$ give the ratios of $\frac{K_i}{K_1}$ for the crystal structure. This sequence is determined by the crystal structure. Knowledge of the X-ray wave length $\lambda = \frac{2\pi}{k}$ allows determination of the lattice spacing (Fig. 1.24).

1.5 Classification of Solids

1.5.1 Crystal Binding

Before considering even in a qualitative way how atoms bind together to form crystals, it is worthwhile to review briefly the periodic table and the ground state configurations of atoms. The single particle states of electrons moving in an effective central potential (which includes the attraction of the nucleus and some average repulsion associated with all other electrons) can be characterized by four quantum numbers: n, the principal quantum number takes on the values $1, 2, 3, \ldots$, l, the angular momentum quantum number takes on values $0, 1, \ldots, n-1$; m, the azimuthal quantum number (projection of l onto a given direction) is an integer satisfying $-l \leq m \leq l$; and σ, the spin quantum number takes on the values $\pm\frac{1}{2}$.

The energy of the single particle orbital is very insensitive to m and σ (in the absence of an applied magnetic field), but it depends strongly on n and l. Of course, due to the Pauli principle only one electron can occupy an orbital with given n, l, m, and σ. The periodic table is constructed by making an array of slots, with l value increasing from $l = 0$ as one moves to the left, and the value of $n + l$ increasing as one moves down. (Table 1.4) Of course, the correct number of slots must be allowed to account for the spin and azimuthal degeneracy $2(2l + 1)$ of a given l value. One then begins filling the slots from the top left, moving to the right, and then down when all slots of a given $(n + l)$ value have been used up. See Table 1.4, which lists the atoms (H, He, ...) and their atomic numbers in the appropriate slots. As the reader can readily observe, H has one electron, and it will occupy the $n = 1, l = 0(1s)$ state. Boron has five electrons and they will fill the $(1s)$ and $2s$ states with the fifth electron in the $2p$ state. Everything is very regular until Cr and Cu. These two elements have ground states in which one $4s$ electron falls into the $3d$ shell, giving for Cr the atomic configuration $(1s)^2(2s)^2(2p)^6(3s)^2(3p)^6(4s)^1(3d)^5$, and for Cu the atomic configuration $(1s)^2(2s)^2(2p)^6(3s)^2(3p)^6(4s)^1(3d)^{10}$. Other exceptions occur in the second transition series (the filling of the $4d$ levels) and in the third transition series (filling the $5d$ levels), and in the rare earth series (filling the $4f$ and $5f$ levels). Knowing this table allows one to write down the ground state electronic configuration of any atom. Note that the inert gases He, Ne, Kr, Rn, complete the shells $n = 1, n = 2, n = 3$, and $n = 4$, respectively. Ar and Xe are inert also; they complete the $n = 3$ shell (except for $3d$ electrons), and $n = 4$ shell (except for $4f$ electrons), respectively. Na, K, Rb, Cs, and Fr have one weakly bound s electron outside these closed shell configurations; Fl, Cl, Br, I and At are missing one p electron from the closed shell configurations. The alkali metals easily give up their loosely bound s electrons, and the halogens readily attract one p electron to give a closed shell configuration. The resulting $Na^+ - Cl^-$ ions form an ionic bond which is quite strong. Atoms like C, Si, Ge, and Sn have an $(np)^2(n + 1\ s)^2$ configuration. These four valence electrons can be readily shared with other atoms in covalent bonds, which are also quite strong.[1] Compounds like GaAs,

[1] In Table 1.4, we note exceptions (i)–(viii):

i Dropping a $4s$ electron into the $3d$ shell while filling $3d$ shell (Cr, Cu)
ii Dropping a $5s$ electron into the $4d$ shell while filling $4d$ shell (Nb, Mo, Ru, Rh, Ag)
iii Dropping both $5s$ electrons into the $4d$ shell while filling $4d$ shell (Pd)
iv Dropping both $6s$ electrons into the $5d$ shell while filling $5d$ shell (Pt)
v Dropping one $6s$ electron into the $5d$ shell while filling $5d$ shell (Au)
vi Adding one $5d$ electron before filling the entire $4f$ shell (La, Gd)
vii Adding one $6d$ electron before filling the entire $5f$ shell (Ac, Pa, U, Cm, Cf)
viii Adding two $5d$ electrons before filling the entire $5f$ shell (Th, Bk)

Table 1.4. Ground state electron configurations in a periodic table: Note that $n = 1, 2, 3, \ldots$; $l < n$; $m = -l, -l+1, \ldots, l$; $\sigma = \pm\frac{1}{2}$

																														$\leftarrow l$		$n + l \downarrow$
																														$l = 0$		
																														1 H	2 He	1 $(1s)$
																								$l = 1$						3 Li	4 Be	2 $(2s)$
																								5 B	6 C	7 N	8 O	9 F	10 Ne	11 Na	12 Mg	3 $(2p, 3s)$
														$l = 2$										13 Al	14 Si	15 P	16 S	17 Cl	18 Ar	19 K	20 Ca	4 $(3p, 4s)$
														21 Sc	22 Ti	23 V	$24^{(i)}$ Cr	25 Mn	26 Fe	27 Co	28 Ni	$29^{(i)}$ Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr	37 Rb	38 Sr	5 $(4p, 5s)$
$l = 3$														39 Y	40 Zr	$41^{(ii)}$ Nb	$42^{(ii)}$ Mo	43 Tc	$44^{(ii)}$ Ru	$45^{(ii)}$ Rh	$46^{(iii)}$ Pd	$47^{(ii)}$ Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe	55 Cs	56 Ba	6 $(4d, 5p, 6s)$
$57^{(vi)}$ La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	$64^{(vi)}$ Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	$78^{(iv)}$ Pt	$79^{(v)}$ Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn	87 Fr	88 Ra	7 $(4f, 5d, 6p, 7s)$
$89^{(vii)}$ Ac	$90^{(viii)}$ Th	$91^{(vii)}$ Pa	$92^{(vii)}$ U	93 Np	94 Pu	95 Am	$96^{(vii)}$ Cm	$97^{(viii)}$ Bk	$98^{(vii)}$ Cf	99 Es	100 F	101 Md	102 No																			8 $(5f, 6d, 7p, 8s)$

We note exceptions (i)–(viii):

i Dropping a $4s$ electron into the $3d$ shell while filling $3d$ shell (Cr, Cu)
ii Dropping a $5s$ electron into the $4d$ shell while filling $4d$ shell (Nb, Mo, Ru, Rh, Ag)
iii Dropping both $5s$ electrons into the $4d$ shell while filling $4d$ shell (Pd)
iv Dropping both $6s$ electrons into the $5d$ shell while filling $5d$ shell (Pt)
v Dropping one $6s$ electron into the $5d$ shell while filling $5d$ shell (Au)
vi Adding one $5d$ electron before filling the entire $4f$ shell (La, Gd)
vii Adding one $6d$ electron before filling the entire $5f$ shell (Ac, Pa, U, Cm, Cf)
viii Adding two $5d$ electrons before filling the entire $5f$ shell (Th, Bk)

GaP, GaSb, or InP, InAs, InSb, etc., are formed from column III and column V constituents. With the partial transfer of an electron from As to Ga, one obtains the covalent bonding structure of Si or Ge as well as some degree of ionicity as in NaCl. Metallic elements like Na and K are relatively weakly bound. Their outermost s electrons become almost free in the solid and act as a glue holding the positively charged ions together. The weakest bonding in solids is associated with weak Van der Waals coupling between the constituent atoms or molecules. To give some idea of the binding energy of solids, we will consider the binding of ionic crystals like NaCl or CsCl.

1.6 Binding Energy of Ionic Crystals

The binding energy of ionic crystals results primarily from the electrostatic interaction between the constituent ions. A rough order of magnitude estimate of the binding energy per molecule can be obtained by simply evaluating

$$\langle V\rangle = \frac{e^2}{R_0} = \frac{\left(4.8\times 10^{-10}\text{esu}\right)^2}{2.8\times 10^{-8}\text{cm}} \simeq 8\times 10^{-12}\text{ergs} \sim 5\text{eV}.$$

Here, R_0 is the observed interatomic spacing (which we take as $2.8\,\mathring{A}$, the spacing in NaCl). The experimentally measured value of the binding energy of NaCl is almost 8 eV per molecule, so our rough estimate is not too bad.

Interatomic Potential

For an ionic crystal, the potential energy of a pair of atoms i, j can be taken to be

$$\phi_{ij} = \pm\frac{e^2}{r_{ij}} + \frac{\lambda}{r_{ij}^n}. \tag{1.32}$$

Here, r_{ij} is the distance between atoms i and j. The $\pm$ sign depends on whether the atoms are like $(+)$ or unlike $(-)$. The first term is simply the Coulomb potential for a pair of point charges separated by r_{ij}. The second term accounts for *core repulsion*. The atoms or ions are not point charges, and when a pair of them gets close enough together their core electrons can repel one another. This core repulsion is expected to decrease rapidly with increasing r_{ij}. The parameters λ and n are phenomenological; they are determined from experiment.

Total Energy

The total potential energy is given by

$$U = \frac{1}{2}\sum_{i\neq j}\phi_{ij}. \tag{1.33}$$

It is convenient to define ϕ_i, the potential energy of the i^{th} atom as

$$\phi_i = \sum_j{}' \phi_{ij}. \tag{1.34}$$

Here, the prime on the sum implies that the term $i = j$ is omitted. It is apparent from symmetry considerations that ϕ_i is independent of i for an infinite lattice, so we can drop the subscript i. The total energy is then

$$U = \frac{1}{2}\, 2N\phi = N\phi, \tag{1.35}$$

where $2N$ is the number of atoms and N is the number of molecules.

It is convenient in evaluating ϕ to introduce a dimensionless parameter p_{ij} defined by $p_{ij} = R^{-1} r_{ij}$, where R is the distance between nearest neighbors. In terms of p_{ij}, the expression for ϕ is given by

$$\phi = \frac{\lambda}{R^n} \sum_j{}' p_{ij}^{-n} - \frac{e^2}{R} \sum_j{}' \left(\mp p_{ij}\right)^{-1}. \tag{1.36}$$

Here, the primes on the summations denote omission of the term $i = j$. We define the quantities

$$A_n = \sum_j{}' p_{ij}^{-n}, \tag{1.37}$$

and

$$\alpha = \sum_j \left(\mp p_{ij}\right)^{-1}. \tag{1.38}$$

The α and A_n are properties of the crystal structure; α is called the *Madelung constant*. The internal energy of the crystal is given by $N\phi$, where N is the number of molecules. The internal energy is given by

$$U = N\left[\lambda \frac{A_n}{R^n} - \alpha \frac{e^2}{R}\right]. \tag{1.39}$$

At the equilibrium separation R_0, $\left(\frac{\partial U}{\partial R}\right)_{R_0}$ must vanish. This gives the result

$$\lambda \frac{A_n}{R_0^n} = \alpha \frac{e^2}{nR_0}. \tag{1.40}$$

Therefore, the equilibrium value of the internal energy is

$$U_0 = N\phi_0 = -N\alpha \frac{e^2}{R_0}\left(1 - \frac{1}{n}\right). \tag{1.41}$$

Compressibility

The best value of the parameter n can be determined from experimental data on the compressibility κ. κ is defined by the negative of the change in volume per unit change in pressure at constant temperature divided by the volume.

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T. \tag{1.42}$$

The subscript T means holding temperature T constant, so that (1.42) is the isothermal compressibility. We will show that at zero temperature

$$\kappa^{-1} = V\left(\frac{\partial^2 U}{\partial V^2}\right)_{T=0}. \tag{1.43}$$

Equation (1.43) comes from the thermodynamic relations

$$F = U - TS, \tag{1.44}$$

and

$$dU = TdS - PdV. \tag{1.45}$$

By taking the differential of (1.44) and making use of (1.45), one can see that

$$dF = -PdV - SdT. \tag{1.46}$$

From (1.46) we have

$$P = -\left(\frac{\partial F}{\partial V}\right)_T. \tag{1.47}$$

Equation (1.42) can be written as

$$\kappa^{-1} = -V\left(\frac{\partial P}{\partial V}\right)_T = V\left(\frac{\partial^2 F}{\partial V^2}\right)_T. \tag{1.48}$$

But at $T = 0$, $F = U$ so that

$$\kappa^{-1} = V\left(\frac{\partial^2 F}{\partial V^2}\right)_{T=0} \tag{1.49}$$

is the inverse of the isothermal compressibility at $T = 0$. We can write the volume V as $2NR^3$ and use $\frac{\partial}{\partial V} = \frac{\partial R}{\partial V}\frac{\partial}{\partial R} = \frac{1}{6NR^2}\frac{\partial}{\partial R}$ in (1.39) and (1.43). This gives

$$\kappa_{T=0}^{-1} = \frac{\alpha e^2}{18R_0^4}(n-1), \tag{1.50}$$

or

$$n = 1 + \frac{18R_0^4}{\alpha e^2 \kappa}. \tag{1.51}$$

From the experimental data on NaCl, the best value for n turns out to be ~ 9.4.

Evaluation of the Madelung Constant

For simplicity let us start with a linear chain. Each positive (+) atom has two neighbors, which are negative (−) atoms, at $p_{01} = 1$. Therefore,

$$\alpha = \sum_j{}' \mp p_{ij}^{-1} = 2\left[1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots\right]. \tag{1.52}$$

If you remember that the power series expansion for $\ln(1 + x)$ is given by $-\sum_{n=1}^{\infty} \frac{(-x)^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$ and is convergent for $x \leq 1$, it is apparent that

$$\alpha = 2\ln 2. \tag{1.53}$$

If we attempt the same approach for NaCl, we obtain

$$\alpha = \frac{6}{1} - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \frac{6}{2} + \cdots. \tag{1.54}$$

This is taking six opposite charge nearest neighbors at a separation of one nearest neighbor distance, 12 same charge next nearest neighbors at $\sqrt{2}$ times that distance, etc. It is clear that the series in (1.54) converges very poorly. The convergence can be greatly improved by using a different counting procedure in which one works with groups of ions which form a more or less neutral array. The motivation is that the potential of a neutral assembly of charges falls off much more quickly with distance than that of a charged assembly.

Evjen's Method

We will illustrate *Evjen's method*[2] by considering a simple square lattice in two dimensions with two atoms per unit cell, one at $(0, 0)$ and one at $(\frac{1}{2}, \frac{1}{2})$. The crystal structure is illustrated in Fig. 1.25. The calculation is carried out as follows:

1. One considers the charges associated with different shells where the first shell is everything inside the first square, the second is everything outside the first but inside the second square, etc.
2. An ion on a face is considered to be half inside and half outside the square defined by that face; a corner atom is one quarter inside and three quarters outside.
3. The total Madelung constant is given by $\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \cdots$, where α_j is the contribution from the ith shell.

As an example, let us evaluate the total charge on the first few shells. The first shell has four atoms on faces, all with the opposite charge to the atom

[2] H.M. Evjen, Phys. Rev. **39**, 675 (1932).

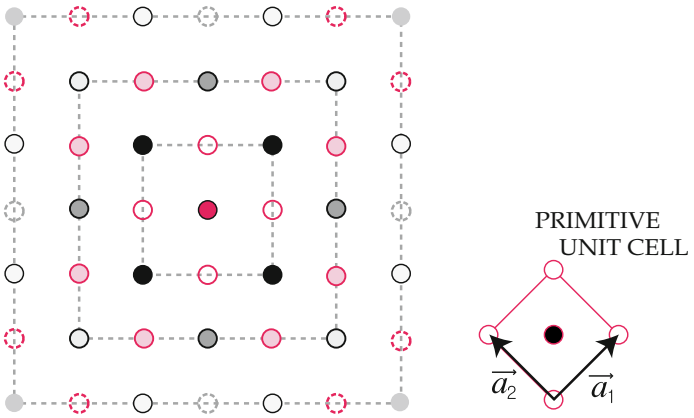


Fig. 1.25. Evjen method for a simple square lattice in two dimensions

at the origin and four corner atoms all with the same charge as the atom at the origin. Therefore, the charge of shell number one is

$$Q_1 = 4\left(\frac{1}{2}\right) - 4\left(\frac{1}{4}\right) = 1. \tag{1.55}$$

Doing the same for the second shell gives

$$Q_2 = 4\left(\frac{1}{2}\right) - 4\left(\frac{3}{4}\right) - 4\left(\frac{1}{2}\right) + 8\left(\frac{1}{2}\right) - 4\left(\frac{1}{4}\right) = 0. \tag{1.56}$$

Here the first two terms come from the remainder of the atoms on the outside of the first square; the next three terms come from the atoms on the inside of the second square. To get α_1 and α_2 we simply divide the individual charges by their separations from the origin. This gives

$$\alpha_1 = \frac{4\left(\frac{1}{2}\right)}{1} - \frac{4\left(\frac{1}{4}\right)}{\sqrt{2}} \simeq 1.293, \tag{1.57}$$

$$\alpha_2 = \frac{4\left(\frac{1}{2}\right)}{1} - \frac{4\left(\frac{3}{4}\right)}{\sqrt{2}} - \frac{4\left(\frac{1}{2}\right)}{2} + \frac{8\left(\frac{1}{2}\right)}{\sqrt{5}} + \frac{4\left(\frac{1}{4}\right)}{2\sqrt{2}} \simeq 0.314. \tag{1.58}$$

This gives $\alpha \simeq \alpha_1 + \alpha_2 \sim 1.607$. The readers should be able to evaluate α_3 for themselves.

Madelung Constant for Three-Dimensional Lattices

For a three-dimensional crystal, Evjen's method is essentially the same with the exception that

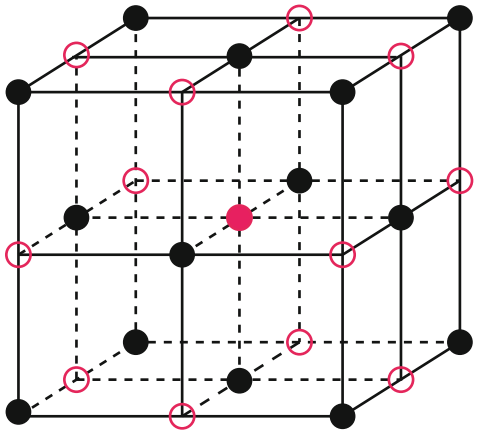


Fig. 1.26. Central atom and the first cube of the Evjen method for the NaCl structure

1. The squares are replaced by cubes.
2. Atoms on the face of a cube are considered to be half inside and half outside the cube; atoms on the edge are $\frac{1}{4}$ inside and $\frac{3}{4}$ outside, and corner atoms are $\frac{1}{8}$ inside and $\frac{7}{8}$ outside.

We illustrate the case of the NaCl structure as an example in the three dimensions (see Fig. 1.26).
For α_1 we obtain

$$\alpha_1 = \frac{6\ (\frac{1}{2})}{1} - \frac{12\ (\frac{1}{4})}{\sqrt{2}} + \frac{8\ (\frac{1}{8})}{\sqrt{3}} \simeq 1.456. \tag{1.59}$$

For α_2 we have the following contributions:

1. Remainder of the contributions from the atoms on the first cube
   $= \frac{6\ (\frac{1}{2})}{1} - \frac{12\ (\frac{3}{4})}{\sqrt{2}} + \frac{8\ (\frac{7}{8})}{\sqrt{3}}$
2. Atoms on the interior of faces of the second cube
   $= -\frac{6\ (\frac{1}{2})}{2} + \frac{6(4)\ (\frac{1}{2})}{\sqrt{5}} - \frac{6(4)\ (\frac{1}{2})}{\sqrt{6}}$
3. Atoms on the interior of edges of the second cube
   $= -\frac{12\ (\frac{1}{4})}{\sqrt{8}} + \frac{12(2)\ (\frac{1}{4})}{\sqrt{9}}$
4. Atoms on the interior of the coners of the second cube
   $= -\frac{8\ (\frac{1}{8})}{\sqrt{12}}$

Adding them together gives

$$\alpha_2 = \left(3 - \frac{9}{\sqrt{2}} + \frac{7}{\sqrt{3}}\right) + \left(-\frac{3}{2} + \frac{12}{\sqrt{5}} - \frac{12}{\sqrt{6}}\right) + \left(-\frac{3}{\sqrt{8}} + \frac{6}{3}\right) - \frac{1}{\sqrt{12}} \simeq 0.295. \tag{1.60}$$

Thus, to the approximation $\alpha \simeq \alpha_1 + \alpha_2$ we find that $\alpha \simeq 1.752$. The exact result for NaCl is $\alpha = 1.747558\ldots$, so Evjen's method gives a surprisingly accurate result after only two shells.

Results of rather detailed evaluations of α for several different crystal structures are $\alpha(\text{NaCl}) = 1.74756$, $\alpha(\text{CsCl}) = 1.76267$, $\alpha(\text{zincblende}) = 1.63806$, $\alpha(\text{wurtzite}) = 1.64132$. The NaCl structure occurs much more frequently than the CsCl structure. This may seem a bit surprising since $\alpha(\text{CsCl})$ is about 1% larger than $\alpha(\text{NaCl})$. However, core repulsion accounts for about 10% of the binding energy (see (1.41)). In the CsCl structure, each atom has eight nearest neighbors instead of the six in NaCl. This should increase the core repulsion by something of the order of 25% in CsCl. Thus, we expect about 2.5% larger contribution (from core repulsion) to the binding energy of CsCl. This negative contribution to the binding energy more than compensates the 1% difference in Madelung constants.

Problems

1.1. Demonstrate that

(a) The reciprocal lattice of a simple cubic lattice is simple cubic.
(b) The reciprocal lattice of a body centered cubic lattice is a face centered cubic lattice.
(c) The reciprocal lattice of a hexagonal lattice is hexagonal.

1.2. Determine the packing fraction of

(a) A simple cubic lattice
(b) A face centered lattice
(c) A body centered lattice
(d) The diamond structure
(e) A hexagonal close packed lattice with an ideal $\frac{c}{a}$ ratio
(f) The graphite structure with an ideal $\frac{c}{a}$.

1.3. The Bravais lattice of the diamond structure is fcc with two carbon atoms per primitive unit cell. If one of the two basis atoms is at (0,0,0), then the other is at $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

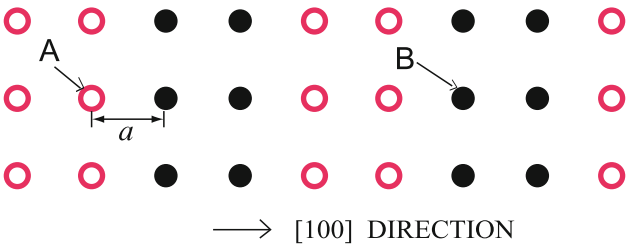
(a) Illustrate that a reflection through the (100) plane followed by a non-primitive translation through $[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$ is a glide-plane operation for the diamond structure.
(b) Illustrate that a fourfold rotation about an axis in diamond parallel to the x-axis passing through the point $(1, \frac{1}{4}, 0)$ (the screw axis) followed by the translation $[\frac{1}{4}, 0, 0]$ parallel to the screw axis is a screw operation for the diamond structure.

1.4. Determine the group multiplication table of the point group of an equilateral triangle.

1.5. CsCl can be thought of as a simple cubic lattice with two different atoms [at $(0, 0, 0)$ and $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$] in the cubic unit cell. Let f_+ and f_- be the atomic scattering factors of the two constituents.

(a) What is the structure amplitude $F(h_1, h_2, h_3)$ for this crystal?
(b) An X-ray source has a continuous spectrum with wave numbers $\boldsymbol{k}$ satisfying: $\boldsymbol{k}$ is parallel to the [110] direction and $2^{-1/2}\left(\frac{2\pi}{a}\right) \leq |\boldsymbol{k}| \leq 3 \times 2^{1/2}\left(\frac{2\pi}{a}\right)$, where a is the edge distance of the simple cube. Use the Ewald construction for a plane that contains the direction of incidence to show which reciprocal lattice points display diffraction maxima.
(c) If $f_+ = f_-$, which of these maxima disappear?

1.6. A simple cubic structure is constructed in which two planes of A atoms followed by two planes of B atoms alternate in the [100] direction.



1. What is the crystal structure (viewed as a non-Bravais lattice with four atoms per unit cell)?
2. What are the primitive translation vectors of the reciprocal lattice?
3. Determine the structure amplitude $F(h_1, h_2, h_3)$ for this non-Bravais lattice.

1.7. Powder patterns of three cubic crystals are found to have their first four diffraction rings at the values given below:

ϕ_{A}	30°	35°	50°	60°
ϕ_{B}	21°	29°	36°	42°
$\phi_{\circ}$	30°	50°	60°	74°

The crystals are monatomic, and the observer believes that one is body centered, one face centered, and one is a diamond structure.

1. What structures are the crystals A, B, and C?
2. The wave length λ of the incident X-ray is 0.95 Å. What is the length of the cube edge for the cubic unit cell in A, B, and C, respectively?

1.8. Determine the ground state atomic configurations of C(6), O(8), Al(13), Si(16), Zn(30), Ga(31), and Sb(51).

1.9. Consider $2N$ ions in a linear chain with alternating $\pm e$ charges and a repulsive potential AR^{-n} between nearest neighbors.

1. Show that, at the equilibrium separation R_0, the internal energy becomes

$$U(R_0) = -2\ln 2 \times \frac{Ne^2}{R_0}\left(1 - \frac{1}{n}\right).$$

2. Let the crystal be compressed so that $R_0 \to R_0(1-\delta)$. Show that the work done per unit length in compressing the crystal can be written $\frac{1}{2}C\delta^2$, and determine the expression for C.

1.10. For a BCC and for an FCC lattice, determine the separations between nearest neighbors, next nearest neighbors, ... down to the 5th nearest neighbors. Also determine the separations between nth nearest neighbors ($n = 1, 2, 3, 4, 5$) in units of the cube edge a of the simple cube.

Summary

In this chapter first we have introduced basic geometrical concepts useful in describing periodic arrays of objects and crystal structures both in real and reciprocal spaces assuming that the atoms sit at lattice sites.

A lattice is an infinite array of points obtained from three primitive translation vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$. Any point on the lattice is given by

$$\mathbf{n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3.$$

Any pair of lattice points can be connected by a vector of the form

$$\mathbf{T}_{n_1n_2n_3} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3.$$

Well defined crystal structure is an arrangement of atoms in a *lattice* such that the atomic arrangement looks absolutely identical when viewed from two different points that are separated by a *lattice translation vector*. Allowed types of Bravais lattices are discussed in terms of symmetry operations both in two and three dimensions. Because of the requirement of translational invariance under operations of the lattice translation, the rotations of $60^\circ, 90^\circ, 120^\circ, 180^\circ,$ and 360° are allowed.

If there is only one atom associated with each lattice point, the lattice of the crystal structure is called *Bravais lattice*. If more than one atom is associated with each lattice point, the lattice is called *a lattice with a basis*. If $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ are the primitive translations of some lattice, one can define a set of primitive translation vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ by the condition

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij},$$

where $\delta_{ij} = 0$ if i is not equal to j and $\delta_{ii} = 1$. It is easy to see that

$$\mathbf{b}_i = 2\pi \frac{\mathbf{a}_j \times \mathbf{a}_k}{\mathbf{a}_i \cdot (\mathbf{a}_j \times \mathbf{a}_k)},$$

where $i, j,$ and k are different. The lattice formed by the primitive translation vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ is called the *reciprocal lattice* (reciprocal to the lattice formed by $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$), and a reciprocal lattice vector is given by

$$\mathbf{G}_{h_1h_2h_3} = h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3.$$

Simple crystal structures and principles of commonly used experimental methods of wave diffraction are also reviewed briefly. Connection of Laue equations and Bragg's law is shown. Classification of crystalline solids are then discussed according to configuration of valence electrons of the elements forming the solid.

2

Lattice Vibrations

2.1 Monatomic Linear Chain

So far, in our discussion of the crystalline nature of solids we have assumed that the atoms sat at lattice sites. This is not actually the case; even at the lowest temperatures the atoms perform small vibrations about their equilibrium positions. In this chapter we shall investigate the vibrations of the atoms in solids. Many of the significant features of lattice vibrations can be understood on the basis of a simple one-dimensional model, a monatomic linear chain. For that reason we shall first study the linear chain in some detail.

We consider a linear chain composed of N identical atoms of mass M (see Fig. 2.1). Let the positions of the atoms be denoted by the parameters R_i, $i = 1, 2, \ldots, N$. Here, we assume an infinite crystal of vanishing surface to volume ratio, and apply *periodic boundary conditions*. That is, the chain contains N atoms and the Nth atom is connected to the first atom so that

$$R_{i+N} = R_i. \tag{2.1}$$

The atoms interact with one another (e.g., through electrostatic forces, core repulsion, etc.). The potential energy of the array of atoms will obviously be a function of the parameters R_i, i.e.,

$$U = U(R_1, R_2, \ldots, R_N). \tag{2.2}$$

We shall assume that U has a minimum $U\left(R_1^0, R_2^0, \ldots, R_N^0\right)$ for some particular set of values $\left(R_1^0, R_2^0, \ldots, R_N^0\right)$, corresponding to the equilibrium state of the linear chain. Define $u_i = R_i - R_i^0$ to be the deviation of the ith atom from its equilibrium position. Now expand U about its equilibrium value to obtain

$$\begin{aligned} U\left(R_1, R_2, \ldots, R_N\right) &= U\left(R_1^0, R_2^0, \ldots, R_N^0\right) + \textstyle\sum_i \left(\frac{\partial U}{\partial R_i}\right)_0 u_i \\ &+ \tfrac{1}{2!} \textstyle\sum_{i,j} \left(\frac{\partial^2 U}{\partial R_i \partial R_j}\right)_0 u_i u_j + \tfrac{1}{3!} \textstyle\sum_{i,j,k} \left(\frac{\partial^3 U}{\partial R_i \partial R_j \partial R_k}\right)_0 u_i u_j u_k + \cdots . \end{aligned} \tag{2.3}$$

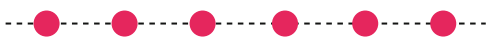


Fig. 2.1. Linear chain of N identical atoms of mass M

The first term is a constant which can simply be absorbed in setting the zero of energy. By the definition of equilibrium, the second term must vanish (the subscript zero on the derivative means that the derivative is evaluated at $u_1, u_2, \ldots, u_n = 0$). Therefore, we can write

$$U(R_1, R_2, \ldots, R_N) = \frac{1}{2!}\sum_{i,j} c_{ij} u_i u_j + \frac{1}{3!}\sum_{i,j,k} d_{ijk} u_i u_j u_k + \cdots , \tag{2.4}$$

where

$$\begin{aligned} c_{ij} &= \left(\frac{\partial^2 U}{\partial R_i \partial R_j}\right)_0 , \\ d_{ijk} &= \left(\frac{\partial^3 U}{\partial R_i \partial R_j \partial R_k}\right)_0 . \end{aligned} \tag{2.5}$$

For the present, we will consider only the first term in (2.4); this is called *the harmonic approximation*. The Hamiltonian in the harmonic approximation is

$$H = \sum_i \frac{P_i^2}{2M} + \frac{1}{2}\sum_{i,j} c_{ij} u_i u_j . \tag{2.6}$$

Here, P_i is the momentum and u_i the displacement from the equilibrium position of the ith atom.

Equation of Motion

Hamilton's equations

$$\begin{aligned} \dot{P}_i &= -\frac{\partial H}{\partial u_i} = -\sum_j c_{ij} u_j , \\ \dot{u}_i &= \frac{\partial H}{\partial P_i} = \frac{P_i}{M} \end{aligned} \tag{2.7}$$

can be combined to yield the equation of motion

$$M\ddot{u}_i = -\sum_j c_{ij} u_j . \tag{2.8}$$

In writing down the equation for $\dot{P}_i$, we made use of the fact that c_{ij} actually depends only on the relative positions of atoms i and j, i.e., on $|i - j|$. Notice that $-c_{ij}u_j$ is simply the force on the ith atom due to the displacement u_j of the jth atom from its equilibrium position. Now let $R_n^0 = na$, so that $R_n^0 - R_m^0 = (n-m)a$. We assume a solution of the coupled differential equations

of motion, (2.8), of the form

$$u_n(t) = \xi_q \mathrm{e}^{\mathrm{i}(qna - \omega_q t)}. \tag{2.9}$$

By substituting (2.9) into (2.8) we find

$$M\omega_q^2 = \sum_m c_{nm} \mathrm{e}^{\mathrm{i}q(m-n)a}. \tag{2.10}$$

Because c_{nm} depends only on $l = m - n$, we can rewrite (2.10) as

$$M\omega_q^2 = \sum_{l=1}^{N} c(l) \mathrm{e}^{\mathrm{i}qla}. \tag{2.11}$$

Boundary Conditions

We apply periodic boundary conditions to our chain; this means that the chain contains N atoms and that the Nth atom is connected to the first atom (Fig. 2.2). This means that the $(n + N)$th atom is the same atoms as the nth atom, so that $u_n = u_{n+N}$. Since $u_n \propto \mathrm{e}^{\mathrm{i}qna}$, the condition means that

$$\mathrm{e}^{\mathrm{i}qNa} = 1, \tag{2.12}$$

or that $q = \frac{2\pi}{Na} \times p$ where $p = 0, \pm 1, \pm 2, \ldots$. However, not all of these values of q are independent. In fact, there are only N independent values of q since there are only N degrees of freedom. If two different values of q, say q and q' give identical displacements for every atom, they are equivalent. It is easy to see that

$$\mathrm{e}^{\mathrm{i}qna} = \mathrm{e}^{\mathrm{i}q'na} \tag{2.13}$$

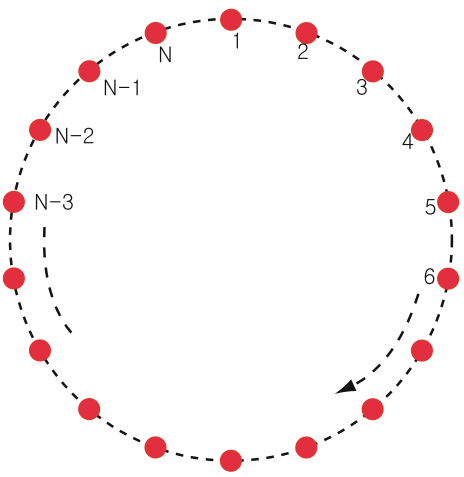


Fig. 2.2. Periodic boundary conditions on a linear chain of N identical atoms

for all values of n if $q' - q = \frac{2\pi}{a}l$, where $l = 0, \pm 1, \pm 2, \ldots$. The set of independent values of q are usually taken to be the N values satisfying $q = \frac{2\pi}{L}p$, where $-\frac{N}{2} \leq p \leq \frac{N}{2}$. We will see later that in three dimensions the independent values of $\mathbf{q}$ are values whose components (q_1, q_2, q_3) satisfy $q_i = \frac{2\pi}{L_i}p$, and which lie in the first Brillouin zone, the Wigner–Seitz unit cell of the reciprocal lattice.

Long Wave Length Limit

Let us look at the long wave length limit, where the wave number q tends to zero. Then $u_n(t) = \xi_0 \mathrm{e}^{-\mathrm{i}\omega_{q=0}t}$ for all values of n. Thus, the entire crystal is uniformly displaced (or the entire crystal is translated). This takes no energy if it is done very very slowly, so it requires $M\omega^2(0) = \sum_{l=1}^{N} c(l) = 0$, or $\omega(q=0) = 0$. In addition, it is not difficult to see that since $c(l)$ depends only on the magnitude of l that

$$M\omega^2(-q) = \sum_l c(l)\mathrm{e}^{-\mathrm{i}qla} = \sum_{l'} c(l')\mathrm{e}^{\mathrm{i}ql'a} = M\omega^2(q). \tag{2.14}$$

In the last step in this equation, we replaced the dummy variable l by l' and used the fact that $c(-l') = c(l')$. Equation (2.14) tells us that $\omega^2(q)$ is an even function of q which vanishes at $q = 0$. If we make a power series expansion for small q, then $\omega^2(q)$ must be of the form

$$\omega^2(q) = s^2q^2 + \cdots \tag{2.15}$$

The constant s is called the *velocity of sound*.

Nearest Neighbor Forces: An Example

So far, we have not specified the interaction law among the atoms; (2.15) is valid in general. To obtain $\omega(q)$ for all values of q, we must know the interaction between atoms. A simple but useful example is that of nearest neighbor forces. In that case, the equation of motion is

$$M\omega^2(q) = \sum_{l=1}^{1} c_l \mathrm{e}^{\mathrm{i}qla} = c_{-1}\mathrm{e}^{-\mathrm{i}qa} + c_0 + c_1\mathrm{e}^{\mathrm{i}qa}. \tag{2.16}$$

Knowing that $\omega(0) = 0$ and that $c_{-l} = c_l$ gives the relation $c_1 = c_{-1} = -\frac{1}{2}c_0$. Therefore, (2.16) is simplified to

$$M\omega^2(q) = c_0\left[1 - \left(\frac{\mathrm{e}^{\mathrm{i}qa} + \mathrm{e}^{-\mathrm{i}qa}}{2}\right)\right]. \tag{2.17}$$

Since $1 - \cos x = 2\sin^2\frac{x}{2}$, (2.17) can be expressed as

$$\omega^2(q) = \frac{2c_0}{M}\sin^2\frac{qa}{2}, \tag{2.18}$$

which is displayed in Fig. 2.3. By looking at the long wave length limit, the coupling constant is determined by $c_0 = \frac{2Ms^2}{a^2}$, where s is the velocity of sound.

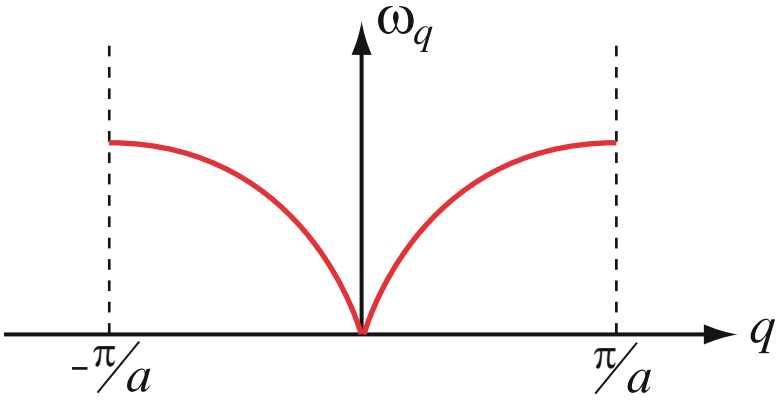


Fig. 2.3. Dispersion relation of the lattice vibration in a monatomic linear chain

2.2 Normal Modes

The general solution for the motion of the nth atom will be a linear combination of solutions of the form of (2.9). We can write the general solution as

$$u_n(t) = \sum_q \left[\xi_q \mathrm{e}^{\mathrm{i}qna - \mathrm{i}\omega t} + \mathrm{cc}\right], \tag{2.19}$$

where cc means the complex conjugate of the previous term. The form of (2.19) assures the reality of $u_n(t)$, and the $2N$ parameters (real and imaginary parts of ξ_q) are determined from the initial values of the position and velocity of the N atoms which specify the initial state of the system.

In most problems involving small vibrations in classical mechanics, we seek new coordinates p_k and q_k in terms of which the Hamiltonian can be written as

$$H = \sum_k H_k = \sum_k \left[\frac{1}{2M} p_k p_k^* + \frac{1}{2} M \omega_k^2 q_k q_k^*\right]. \tag{2.20}$$

In terms of these normal coordinates p_k and q_k, the Hamiltonian is a sum of N independent simple harmonic oscillator Hamiltonians. Because we use running waves of the form $\mathrm{e}^{\mathrm{i}qna - \mathrm{i}\omega_q t}$ the new coordinates q_k and p_k can be complex, but the Hamiltonian must be real. This dictates the form of (2.20).

The normal coordinates turn out to be

$$q_k = N^{-1/2} \sum_n u_n \mathrm{e}^{-\mathrm{i}kna}, \tag{2.21}$$

and

$$p_k = N^{-1/2} \sum_n P_n \mathrm{e}^{+\mathrm{i}kna}. \tag{2.22}$$

We will demonstrate this for q_k and leave it for the student to do the same for p_k. We can write (2.19) as

$$u_n(t) = \alpha \sum_k \xi_k(t) \mathrm{e}^{\mathrm{i}kna}, \tag{2.23}$$

where ξ_k is complex and satisfies $\xi^*_{-k} = \xi_k$. With this condition $u_n(t)$, given by (2.23), is real and α is simply a constant to be determined. We can write the potential energy $U = \frac{1}{2}\sum_{mn} c_{mn} u_m u_n$ in terms of the new coordinates ξ_k as follows.

$$U = \frac{1}{2}|\alpha|^2 \sum_{mn} c_{mn} \sum_k \xi_k \mathrm{e}^{\mathrm{i}kma} \sum_{k'} \xi_{k'} \mathrm{e}^{\mathrm{i}k'na}. \tag{2.24}$$

Now, let us use $k' = q - k$ to rewrite (2.24) as

$$U = \frac{1}{2}|\alpha|^2 \sum_{nkq} \left[\sum_m c_{mn} \mathrm{e}^{\mathrm{i}k(m-n)a} \right] \xi_k \xi_{q-k} \mathrm{e}^{\mathrm{i}qna}. \tag{2.25}$$

From (2.10) one can see that the quantity in the square bracket in (2.25) is equal to $M\omega_k^2$. Thus, U becomes

$$U = \frac{1}{2}|\alpha|^2 \sum_{nkq} M\omega_k^2 \xi_k \xi^*_{k-q} \mathrm{e}^{\mathrm{i}qna}. \tag{2.26}$$

The only factor in (2.26) that depends on n is $\mathrm{e}^{\mathrm{i}qna}$. It is not difficult to prove that $\sum_n \mathrm{e}^{\mathrm{i}qna} = N\delta_{q,0}$. We do this as follows: Define $S_N = 1 + x + x^2 + \cdots + x^{N-1}$; then $xS_N = x + x^2 + \cdots + x^N$ is equal to $S_N - 1 + x^N$.

$$xS_N = S_N - 1 + x^N. \tag{2.27}$$

Solving for S_N gives

$$S_N = \frac{1 - x^N}{1 - x}. \tag{2.28}$$

Now, let $x = \mathrm{e}^{\mathrm{i}qa}$. Then, (2.28) says

$$\sum_{n=0}^{N-1} \left(\mathrm{e}^{\mathrm{i}qa}\right)^n = \frac{1 - \mathrm{e}^{\mathrm{i}qaN}}{1 - \mathrm{e}^{\mathrm{i}qa}}. \tag{2.29}$$

Remember that the allowed values of q were given by $q = \frac{2\pi}{Na} \times$ integer. Therefore, $\mathrm{i}qaN = \mathrm{i}\frac{2\pi}{Na} aN \times$ integer, and $\mathrm{e}^{\mathrm{i}qaN} = \mathrm{e}^{2\pi\mathrm{i}\times\text{integer}} = 1$. Therefore, the numerator vanishes. The denominator does not vanish unless $q = 0$. When $q = 0$, $\mathrm{e}^{\mathrm{i}qa} = 1$ and the sum gives N. This proves that $\sum_n \mathrm{e}^{\mathrm{i}qna} = N\delta(q, 0)$ when $q = \frac{2\pi}{Na} \times$ integer. Using this result in (2.26) gives

$$U = \frac{1}{2}|\alpha|^2 \sum_k M\omega_k^2 \xi_k \xi_k^* N. \tag{2.30}$$

Choosing $\alpha = N^{-1/2}$ puts U into the form of the potential energy for N simple harmonic oscillators labeled by the k value. By assuming that P_n is proportional to $\sum_k p_k \mathrm{e}^{-\mathrm{i}kna}$ with $p_{-k}^* = p_k$, it is not difficult to show that (2.22) gives the kinetic energy in the desired form $\sum_k \frac{p_k p_k^*}{2M}$. The inverse of (2.21) and (2.22) are easily determined to be

$$u_n = N^{-1/2} \sum_k q_k \mathrm{e}^{\mathrm{i}kna}, \tag{2.31}$$

and

$$P_n = N^{-1/2} \sum_k p_k \mathrm{e}^{-\mathrm{i}kna}. \tag{2.32}$$

Quantization

Up to this point our treatment has been completely classical. We quantize the system in the standard way. The dynamical variables q_k and p_k are replaced by quantum mechanical operators $\hat{q}_k$ and $\hat{p}_k$ which satisfy the commutation relation

$$[p_k,\ q_{k'}] = -\mathrm{i}\hbar\delta_{k,k'}. \tag{2.33}$$

The quantum mechanical Hamiltonian is given by $H = \sum_k H_k$, where

$$H_k = \frac{\hat{p}_k \hat{p}_k^\dagger}{2M} + \frac{1}{2} M\omega_k^2 \hat{q}_k \hat{q}_k^\dagger. \tag{2.34}$$

$\hat{p}_k^\dagger$ and $\hat{q}_k^\dagger$ are the Hermitian conjugates of $\hat{p}_k$ and $\hat{q}_k$, respectively. They are necessary in (2.34) to assure that the Hamiltonian is a Hermitian operator. The Hamiltonian of the one-dimensional chain is simply the sum of N independent simple Harmonic oscillator Hamiltonians. As with the simple Harmonic oscillator, it is convenient to introduce the operators a_k and its Hermitian conjugate $a_k^\dagger$, which are defined by

$$q_k = \left(\frac{\hbar}{2M\omega_k}\right)^{1/2} \left(a_k + a_{-k}^\dagger\right), \tag{2.35}$$

$$p_k = \mathrm{i}\left(\frac{\hbar M\omega_k}{2}\right)^{1/2} \left(a_k^\dagger - a_{-k}\right). \tag{2.36}$$

The commutation relations satisfied by the a_k's and $a_k^\dagger$'s are

$$\left[a_k,\ a_{k'}^\dagger\right]_- = \delta_{k,k'} \text{ and } [a_k,\ a_{k'}]_- = \left[a_k^\dagger,\ a_{k'}^\dagger\right]_- = 0. \tag{2.37}$$

The displacement of the nth atom and its momentum can be written

$$u_n = \sum_k \left(\frac{\hbar}{2MN\omega_k}\right)^{1/2} \mathrm{e}^{\mathrm{i}kna}\left(a_k + a_{-k}^\dagger\right), \tag{2.38}$$

$$P_n = \sum_k \mathrm{i}\left(\frac{\hbar\omega_k M}{2N}\right)^{1/2} \mathrm{e}^{-\mathrm{i}kna}\left(a_k^\dagger - a_{-k}\right). \tag{2.39}$$

The Hamiltonian of the linear chain of atoms can be written

$$H = \sum_k \hbar\omega_k \left(a_k^\dagger a_k + \frac{1}{2}\right), \tag{2.40}$$

and its eigenfunctions and eigenvalues are

$$|n_1, n_2, \ldots, n_N\rangle = \frac{\left(a_{k_1}^\dagger\right)^{n_1}}{\sqrt{n_1!}} \cdots \frac{\left(a_{k_N}^\dagger\right)^{n_N}}{\sqrt{n_N!}} |0\rangle, \tag{2.41}$$

and

$$E_{n_1,n_2,\ldots,n_N} = \sum_i \hbar\omega_{k_i}\left(n_i + \frac{1}{2}\right). \tag{2.42}$$

In (2.41), $|0>=|0_1>|0_2>\cdots|0_N>$ is the ground state of the entire system; it is a product of ground state wave functions for each harmonic oscillator. It is convenient to think of the energy $\hbar\omega_k$ as being carried by an *elementary excitations* or *quasiparticle*. In lattice dynamics, these elementary excitations are called *phonons*. In the ground state, no phonons are present in any of the harmonic oscillators. In an arbitrary state, such as (2.41), n_1 phonons are in oscillator k_1, n_2 in k_2, …, n_N in k_N. We can rewrite (2.41) as $|n_1, n_2, \ldots, n_N\rangle = |n_1>|n_2>\cdots|n_N>$, a product of kets for each oscillator.

2.3 Mössbauer Effect

With the simple one-dimensional harmonic approximation, we have the tools necessary to understand the physics of some interesting effects. One example is *the Mössbauer effect*.[1] This effect involves the decay of a nuclear excited state via γ-ray emission (see Fig. 2.4). First, let us discuss the decay of a nucleus in a free atom; to be explicit, let us consider the decay of the excited state of Fe^{57} via emission of a 14.4 keV γ-ray.

$$Fe^{57^*} \longrightarrow Fe^{57} + \gamma \tag{2.43}$$

The excited state of Fe^{57} has a lifetime of roughly 10^{-7} s. The uncertainty principle tells us that by virtue of the finite lifetime $\Delta t = \tau = 10^{-7}$ s, there

[1] R. L. Mössbauer, D.H. Sharp, Rev. Mod. Phys. **36**, 410 (1964).

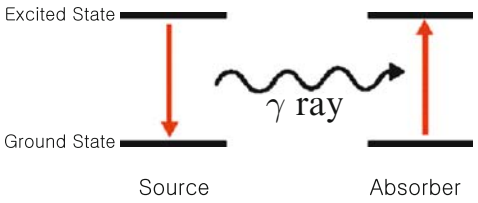


Fig. 2.4. The exact transition energy is required to be reabsorbed because of the very sharply defined nuclear energy states

is an uncertainty ΔE in the energy of the excited state (or a natural linewidth for the γ-ray) given by $\Delta E = \frac{\hbar}{\Delta t}$. Using $\Delta t = 10^{-7}\,\mathrm{s}$ gives $\Delta\omega = 10^{7}\,\mathrm{s}^{-1}$ or $\Delta(\hbar\omega) \simeq 6 \times 10^{-9}\,\mathrm{eV}$. Thus, the ratio of the linewidth $\Delta\omega$ to the frequency ω is $\frac{\Delta\omega}{\omega} \simeq 4 \times 10^{-13}$.

In a resonance experiment, the γ-ray source emits and the target resonantly absorbs the γ-rays. Unfortunately, when a γ-ray is emitted or absorbed by a nucleus, the nucleus must recoil to conserve momentum. The momentum of the γ-ray is $p_\gamma = \frac{\hbar\omega}{c}$, so that the nucleus must recoil with momentum $\hbar K = p_\gamma$ or energy $E(K) = \frac{\hbar^2 K^2}{2M}$ where M is the mass of the atom. The recoil energy is given by $E(K) = \frac{\hbar^2\omega^2}{2Mc^2} = \frac{(\hbar\omega)^2}{2(M/m)mc^2}$. But $mc^2 \simeq 0.5 \times 10^6\,\mathrm{eV}$ and the ratio of the mass of Fe^{57} to the electron mass m is $\sim 2.3 \times 10^5$, giving $E(K) \simeq 2 \times 10^{-3}\,\mathrm{eV}$. This recoil energy is much larger than the energy uncertainty of the γ-ray ($6 \times 10^{-9}\,\mathrm{eV}$). Because of the recoil on emission and absorption, the γ-ray is short by $4 \times 10^{-3}\,\mathrm{eV}$ of energy necessary for resonance absorption. Mössbauer had the idea that if the nucleus that underwent decay was bound in a crystal (containing $\sim 10^{23}$ atoms) the recoil of the entire crystal would carry negligible energy since the crystal mass would replace the atomic mass of a single Fe^{57} atom. However, the quantum mechanical state of the crystal might change in the emission process (via emission of phonons). A typical phonon has a frequency of the order of $10^{13}\,\mathrm{s}^{-1}$, much larger than $\Delta\omega = 10^{7}\,\mathrm{s}^{-1}$ the natural line width. Therefore, in order for resonance absorption to occur, the γ-ray must be emitted without simultaneous emission of phonons. This *no phonon γ-ray emission* occurs a certain fraction of the time and is referred to as *recoil free fraction*. We would like to estimate the recoil free fraction.

As far as the recoil-nucleus is concerned, the effect of the γ-ray emission can be represented by an operator H' defined by

$$H' = C\mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\hat{\mathbf{R}}_N}, \tag{2.44}$$

where C is some constant, $\hbar K$ is the recoil momentum, and $\hat{R}_N$ is the position operator of the decaying nucleus. This expression can be derived using the semiclassical theory of radiation, but we simply state it and demonstrate that it is plausible by considering a free nucleus.

Recoil of a Free Nucleus

The Hamiltonian describing the motion of the center of mass of a free atom is

$$H_0 = \frac{P^2}{2M}. \tag{2.45}$$

The eigenstates of H_0 are plane waves

$$|k\rangle = V^{-1/2} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}_N}$$

whose energy is

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

Operating on an initial state $|k>$ with H' gives a new eigenstate proportional to $|k+K>$. The change in energy (i.e., the recoil energy) is

$$\Delta E = E(k+K) - E(k) = \frac{\hbar^2}{2M}\left(2\mathbf{k}\cdot\mathbf{K} + K^2\right).$$

For a nucleus that is initially at rest, $\Delta E = \frac{\hbar^2 K^2}{2M}$, exactly what we had given previously.

Mössbauer Recoil Free Fraction

When the atom whose nucleus emits the γ-ray is bound in the crystal, the initial and final eigenstates must describe the entire crystal. Suppose the initial eigenstate is

$$|n_1, n_2, \ldots, n_N> = \prod_i n_i^{-1/2} \left(a_{k_i}^\dagger\right)^{n_i} |0> .$$

In evaluating H' operating on this state, we write $R_N = R_N^0 + u_N$ to describe the center of mass of the nucleus which emits the γ-ray. We can choose the origin of our coordinate system at the position R_N^0 and write

$$R_N = u_N = \sum_k \left(\frac{\hbar}{2MN\omega_k}\right)^{1/2} \left(a_k + a_{-k}^\dagger\right). \tag{2.46}$$

Because k is a dummy variable to be summed over, and because $\omega_k = \omega_{-k}$, we can replace $a_{-k}^\dagger$ by $a_k^\dagger$ in (2.46).

The probability of a transition from initial state $|n_i> = |n_1, n_2, \ldots, n_N>$ to final state $|m_f> = |m_1, m_2, \ldots, m_N>$ is proportional to the square of the matrix element $\langle m_f | H' | n_i \rangle$. This result can be established by using time dependent perturbation theory with H' as the perturbation. Let us write this probability as $P(m_f, n_i)$. Then $P(m_f, n_i)$ can be expressed as

$$P(m_f, n_i) = \alpha \left| \left\langle m_f \left| C\mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{R}_N} \right| n_i \right\rangle \right|^2. \tag{2.47}$$

In (2.47) α is simply a proportional constant, and we have set $H' = Ce^{i\mathbf{K}\cdot\mathbf{R}_N}$. Because $P(m_f, n_i)$ is the probability of going from $|n_i>$ to $|m_f>$, $\sum_{m_f} P(m_f, n_i) = 1$. This condition gives the relation

$$\alpha|C|^2 \sum_{m_f} \left\langle m_f \left| e^{i\mathbf{K}\cdot\mathbf{R}_N} \right| n_i \right\rangle^* \left\langle m_f \left| e^{i\mathbf{K}\cdot\mathbf{R}_N} \right| n_i \right\rangle = 1. \tag{2.48}$$

Because $e^{i\mathbf{K}\cdot\mathbf{R}_N}$ is Hermitian, $\left\langle m_f \left| e^{i\mathbf{K}\cdot\mathbf{R}_N} \right| n_i \right\rangle^*$ is equal to $\left\langle n_i \left| e^{-i\mathbf{K}\cdot\mathbf{R}_N} \right| m_f \right\rangle$. We use this result in (2.48) and make use of the fact that $|m_f\rangle$ is part of a complete orthonormal set so that $\sum_{m_f} |m_f\rangle\langle m_f|$ is the unit operator to obtain

$$\alpha|C|^2 \left[\left\langle n_i \left| e^{-i\mathbf{K}\cdot\mathbf{R}_N} \times e^{i\mathbf{K}\cdot\mathbf{R}_N} \right| n_i \right\rangle \right]^2 = 1.$$

This is satisfied only if $\alpha|C|^2 = 1$, establishing the result

$$P(m_f, n_i) = \left| \left\langle m_f \left| e^{i\mathbf{K}\cdot\mathbf{R}_N} \right| n_i \right\rangle \right|^2. \tag{2.49}$$

Evaluation of $P(n_i, n_i)$

The probability of γ-ray emission without any change in the state of the lattice is simply $P(n_i, n_i)$. We can write R_N in (2.49) as

$$R_N = \sum_k \beta_k \left(a_k + a_k^\dagger \right), \tag{2.50}$$

where we have introduced $\beta_k = \left(\frac{\hbar}{2MN\omega_k} \right)^{1/2}$. If we write $|n_i>= |n_1> |n_2> \cdots |n_N>$, then

$$\left\langle n_i \left| e^{i\mathbf{K}\cdot\mathbf{R}_N} \right| n_i \right\rangle = \langle n_1| < n_2| \cdots < n_N| [e^{iK\sum_k \beta_k (a_k + a_k^\dagger)}] |n_1> |n_2> \cdots |n_N\rangle. \tag{2.51}$$

The operator a_k and $a_k^\dagger$ operates only on the kth harmonic oscillator, so that (2.51) can be rewritten

$$\left\langle n_i \left| e^{i\mathbf{K}\cdot\mathbf{R}_N} \right| n_i \right\rangle = \prod_k \langle n_k | e^{iK\beta_k (a_k + a_k^\dagger)} | n_k \rangle. \tag{2.52}$$

Each factor in the product can be evaluated by expanding the exponential in power series. This gives

$$\begin{aligned} \langle n_k | e^{iK\beta_k (a_k + a_k^\dagger)} | n_k \rangle = 1 &+ \frac{(iK\beta_k)^2}{2!} \langle n_k | a_k a_k^\dagger + a_k^\dagger a_k | n_k \rangle \\ &+ \frac{(iK\beta_k)^4}{4!} \langle n_k | (a_k + a_k^\dagger)^4 | n_k \rangle + \cdots . \end{aligned} \tag{2.53}$$

The result for this matrix element is

$$\langle n_k | e^{iK\beta_k (a_k + a_k^\dagger)} | n_k \rangle = 1 - \frac{E(K)}{\hbar\omega_k} \frac{n_k + \frac{1}{2}}{N} + O(N^{-2}). \tag{2.54}$$

We shall neglect terms of order N^{-2}, $N^{-3}, \ldots$, etc. in this expansion. With this approximation we can write

$$\langle n_i | \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{R}_N} | n_i \rangle \simeq \prod_k \left[1 - \frac{E(K)}{\hbar\omega_k} \frac{n_k + \frac{1}{2}}{N} \right] . \tag{2.55}$$

To terms of order N^{-1}, the product appearing on the right-hand side of (2.55) is equivalent to $\mathrm{e}^{-\frac{E(K)}{N} \sum_k \frac{n_k+\frac{1}{2}}{\hbar\omega_k}}$ to the same order. Thus, for the recoil free fraction, we find

$$P(n_i, n_i) = \mathrm{e}^{-2\frac{E(K)}{N} \sum_k \frac{n_k+\frac{1}{2}}{\hbar\omega_k}} . \tag{2.56}$$

Although we have derived (2.56) for a simple one-dimensional model, the result is valid for a real crystal if sum over k is replaced by a three-dimensional sum over all $\mathbf{k}$ and over the three polarizations. We will return to the evaluation of the sum later, after we have considered models for the phonon spectrum in real crystals.

2.4 Optical Modes

So far, we have restricted our consideration to a monatomic linear chain. Later on, we shall consider three-dimensional solids (the added complication is not serious). For the present, let us stick with the one-dimensional chain, but let us generalize to the case of two atoms per unit cell (Fig. 2.5).

If atoms A and B are identical, the primitive translation vector of the lattice is a, and the smallest reciprocal vector is $K = \frac{2\pi}{a}$. If A and B are distinguishable (e.g. of slightly different mass) then the smallest translation vector is $2a$ and the smallest reciprocal lattice vector is $K = \frac{2\pi}{2a} = \frac{\pi}{a}$. In this case, the part of the ω vs. q curve lying outside the region $|q| \leq \frac{\pi}{2a}$ must be translated (or folded back) into the first Brillouin zone (region between $-\frac{\pi}{2a}$ and $\frac{\pi}{2a}$) by adding or subtracting the reciprocal lattice vector $\frac{\pi}{a}$. This results in the spectrum shown in Fig. 2.6. Thus, for a non-Bravais lattice, the phonon spectrum has more than one branch. If there are p atoms per primitive unit cell, there will be p branches of the spectrum in a one-dimensional crystal. One branch, which satisfies the condition that $\omega(q) \to 0$ as $q \to 0$ is called the *acoustic branch* or *acoustic mode*. The other $(p-1)$ branches are called *optical branches* or *optical modes*. Due to the difference between the

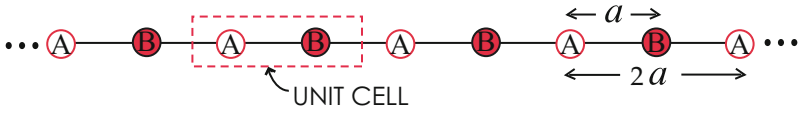


Fig. 2.5. Linear chain with two atoms per unit cell

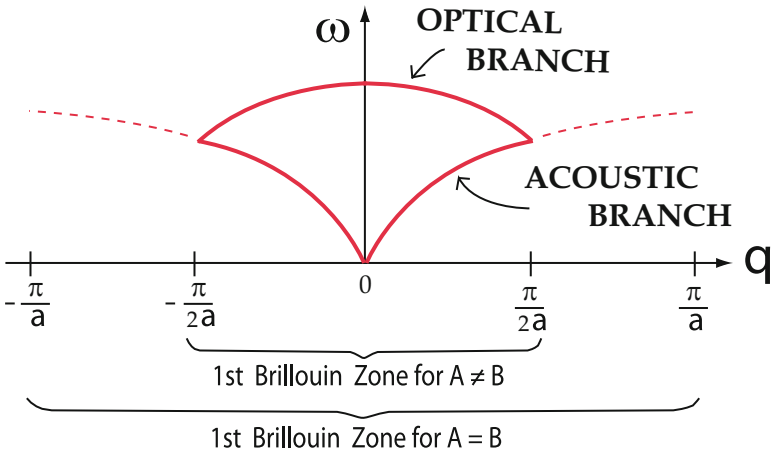


Fig. 2.6. Dispersion curves for the lattice vibration in a linear chain with two atoms per unit cell

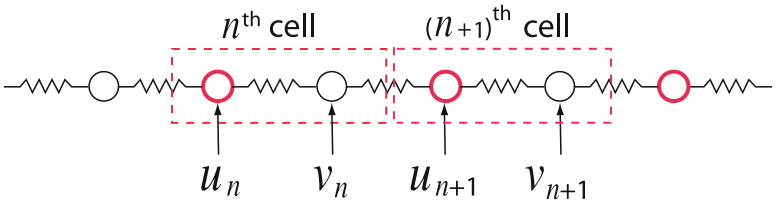


Fig. 2.7. Unit cells of a linear chain with two atoms per cell

pair of atoms in the unit cell when A $\neq$ B, the degeneracy of the acoustic and optical modes at $q = \pm\frac{q}{2a}$ is usually removed. Let us consider a simple example, the linear chain with nearest neighbor interactions but with atoms of mass M_1 and M_2 in each unit cell. Let u_n be the displacement from its equilibrium position of the atom of mass M_1 in the nth unit cell; let v_n be the corresponding quantity for the atom of mass M_2. Then, the equations of motion are

$$M_1\ddot{u}_n = K\left[(v_n - u_n) - (u_n - v_{n-1})\right], \tag{2.57}$$

$$M_2\ddot{v}_n = K\left[(u_{n+1} - v_n) - (v_n - u_n)\right]. \tag{2.58}$$

In Fig. 2.7, we show unit cells n and $n+1$. We assume solutions of (2.57) and (2.58) of the form

$$u_n = u_q \mathrm{e}^{\mathrm{i}q2an - \mathrm{i}\omega_q t}, \tag{2.59}$$

$$v_n = v_q \mathrm{e}^{\mathrm{i}q(2an+a) - \mathrm{i}\omega_q t}. \tag{2.60}$$

where u_q and v_q are constants. Substituting (2.59) and (2.60) into the equations of motion gives the following matrix equation.

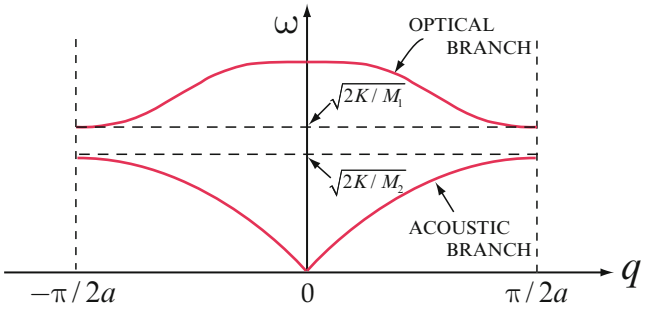


Fig. 2.8. Dispersion relations for the acoustical and optical modes of a diatomic linear chain

$$\begin{bmatrix} -M_1\omega^2 + 2K & -2K\cos qa \\ -2K\cos qa & -M_2\omega^2 + 2K \end{bmatrix} \begin{bmatrix} u_q \\ v_q \end{bmatrix} = 0. \tag{2.61}$$

The nontrivial solutions are obtained by setting the determinant of the 2×2 matrix multiplying the column vector $\begin{bmatrix} u_q \\ v_q \end{bmatrix}$ equal to zero. The roots are

$$\omega_{\pm}^2(q) = \frac{K}{M_1 M_2}\left\{M_1 + M_2 \mp \left[(M_1+M_2)^2 - 4M_1M_2\sin^2 qa\right]^{1/2}\right\}. \tag{2.62}$$

We shall assume that $M_1 < M_2$. Then at $q = \pm\frac{\pi}{2a}$, the two roots are $\omega_{\rm OP}^2(q = \frac{\pi}{2a}) = \frac{2K}{M_1}$ and $\omega_{\rm AC}^2(q = \frac{\pi}{2a}) = \frac{2K}{M_2}$. At $q \approx 0$, the two roots are given by $\omega_{\rm AC}^2(q) \simeq \frac{2Ka^2}{M_1+M_2}q^2$ and $\omega_{\rm OP}^2(q) = \frac{2K(M_1+M_2)}{M_1M_2}\left[1 - \frac{M_1M_2}{(M_1+M_2)^2}q^2a^2\right]$. The dispersion relations for both modes are sketched in Fig. 2.8.

2.5 Lattice Vibrations in Three Dimensions

Now let us consider a primitive unit cell in three dimensions defined by the translation vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. We will apply periodic boundary conditions such that N_i steps in the direction $\mathbf{a}_i$ will return us to the original lattice site. The Hamiltonian in the harmonic approximation can be written as

$$H = \sum_i \frac{\mathbf{P}_i^2}{2M} + \frac{1}{2}\sum_{i,j} \mathbf{u}_i \cdot \underline{C}_{ij} \cdot \mathbf{u}_j. \tag{2.63}$$

Here, the tensor $\underline{C}_{ij}$ (i and j refer to the ith and jth atoms and $\underline{C}_{ij}$ is a three-dimensional tensor for each value of i and j) is given by

$$\underline{C}_{ij} = \left[\nabla_{R_i}\nabla_{R_j}U(\mathbf{R}_1, \mathbf{R}_2, \ldots)\right]_{R_i^0 R_j^0}. \tag{2.64}$$

In obtaining (2.63) we have expanded $U(\mathbf{R}_1, \mathbf{R}_2, \ldots)$ in powers of $\mathbf{u}_i = \mathbf{R}_i - \mathbf{R}_i^0$, the deviation from the equilibrium position, and we have used the definition of equilibrium to eliminate the term that is linear in $\mathbf{u}_i$.

From Hamilton's equation we obtain the equation of motion

$$M\ddot{\mathbf{u}}_n = -\sum_j \underline{C}_{ij} \cdot \mathbf{u}_j. \tag{2.65}$$

We assume a solution to (2.65) of the form

$$\mathbf{u}_n = \boldsymbol{\xi}_{\mathbf{k}} \, \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}_n^0 - \mathrm{i}\omega_{\mathbf{k}} t}. \tag{2.66}$$

Here, $\xi_{\mathbf{k}}$ is a vector whose magnitude gives the size of the displacement associated with wave vector $\mathbf{k}$ and whose direction gives the direction of the displacement. It is convenient to write

$$\boldsymbol{\xi}_{\mathbf{k}} = \hat{\varepsilon}_{\mathbf{k}} \; q_{\mathbf{k}}, \tag{2.67}$$

where $\hat{\varepsilon}_{\mathbf{k}}$ is a unit polarization vector (a unit vector in the direction of $\xi_{\mathbf{k}}$) and $q_{\mathbf{k}}$ is the amplitude. Substituting the assumed solution into the equation of motion gives

$$M\omega_{\mathbf{k}}^2 \boldsymbol{\varepsilon}_{\mathbf{k}} = \sum_j \underline{C}_{ij} \cdot \hat{\varepsilon}_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\left(\mathbf{R}_j^0 - \mathbf{R}_i^0\right)}. \tag{2.68}$$

Because (2.68) is a vector equation, it must have three solutions for each value of $\mathbf{k}$. This is apparent if we define the tensor $\underline{F}(\mathbf{k})$ by

$$\underline{F}(\mathbf{k}) = -\sum_j \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\left(\mathbf{R}_j^0 - \mathbf{R}_i^0\right)} \underline{C}_{ij}. \tag{2.69}$$

Then, (2.68) can be written as a matrix equation

$$\begin{pmatrix} M\omega_{\mathbf{k}}^2 + F_{xx} & F_{xy} & F_{xz} \\ F_{yx} & M\omega_{\mathbf{k}}^2 + F_{yy} & F_{yz} \\ F_{zx} & F_{zy} & M\omega_{\mathbf{k}}^2 + F_{zz} \end{pmatrix} \begin{pmatrix} \hat{\varepsilon}_{\mathbf{k}x} \\ \hat{\varepsilon}_{\mathbf{k}y} \\ \hat{\varepsilon}_{\mathbf{k}z} \end{pmatrix} = 0. \tag{2.70}$$

The three solutions of the three by three secular equation for a given value of $\mathbf{k}$ can be labeled by a *polarization index* λ. The eigenvalues of (2.70) will be $\omega_{\mathbf{k}\lambda}^2$ and the eigenfunctions will be

$$\hat{\varepsilon}_{\mathbf{k}\lambda} = (\hat{\varepsilon}_{\mathbf{k}\lambda}^x, \hat{\varepsilon}_{\mathbf{k}\lambda}^y, \hat{\varepsilon}_{\mathbf{k}\lambda}^z)$$

with $\lambda = 1, 2, 3$.

When we apply periodic boundary conditions, then we must have the condition

$$\mathrm{e}^{\mathrm{i}k_i N_i a_i} = 1 \tag{2.71}$$

satisfied for $i = 1, 2, 3$ the three primitive translation directions. In (2.71), k_i is the component of $\mathbf{k}$ in the direction of $\mathbf{a}_i$ and N_i is the period associated

with the periodic boundary conditions in this direction. From the condition (2.71), it is clear that the allowed values of the wave vector $\mathbf{k}$ must be of the form

$$\mathbf{k} = 2\pi \left(\frac{n_1}{N_1}\mathbf{b}_1 + \frac{n_2}{N_2}\mathbf{b}_2 + \frac{n_3}{N_3}\mathbf{b}_3 \right), \tag{2.72}$$

where n_1, n_2, and n_3 are integers, and $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ are primitive translation vectors of the reciprocal lattice. As in the one-dimensional case, not all of the values of $\mathbf{k}$ given by (2.72) are independent. It is customary to chose as independent values of $\mathbf{k}$ those which satisfy (2.72) and the condition

$$-\frac{N_i}{2} \leq n_i \leq \frac{N_i}{2}. \tag{2.73}$$

This set of $\mathbf{k}$ values is restricted to the first Brillouin zone, the set of all values of $\mathbf{k}$ satisfying (2.72) that are closer to the origin in reciprocal space than to any other reciprocal lattice point. The total number of $\mathbf{k}$ values in the first Brillouin zone is $N = N_1N_2N_3$, and there are three normal modes (three polarizations λ) for each $\mathbf{k}$ value. This gives a total of 3N normal modes, the number required to describe a system of $N = N_1N_2N_3$ atoms each having three degrees of freedom. For $\mathbf{k}$ values that lie outside the Brillouin zone, one simply adds a reciprocal lattice vector $\mathbf{K}$ to obtain an equivalent $\mathbf{k}$ value inside the Brillouin zone.

2.5.1 Normal Modes

As we did in the one-dimensional case, we can define new coordinates $q_{\mathbf{k}\lambda}$ and $p_{\mathbf{k}\lambda}$ as

$$\mathbf{u}_n = N^{-1/2} \sum_{\mathbf{k}\lambda} \hat{\varepsilon}_{\mathbf{k}\lambda} q_{\mathbf{k}\lambda} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}_n^0}, \tag{2.74}$$

$$\mathbf{P}_n = N^{-1/2} \sum_{\mathbf{k}\lambda} \hat{\varepsilon}_{\mathbf{k}\lambda} p_{\mathbf{k}\lambda} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_n^0}. \tag{2.75}$$

The Hamiltonian becomes

$$H = \sum_{\mathbf{k}\lambda} H_{\mathbf{k}\lambda} = \sum_{\mathbf{k}\lambda} \left[\frac{1}{2M} p_{\mathbf{k}\lambda} p^*_{\mathbf{k}\lambda} + \frac{1}{2} M \omega^2_{\mathbf{k}\lambda} q_{\mathbf{k}\lambda} q^*_{\mathbf{k}\lambda} \right]. \tag{2.76}$$

It is customary to define the polarization vectors $\hat{\varepsilon}_{\mathbf{k}\lambda}$ to satisfy $\hat{\varepsilon}_{-\mathbf{k}\lambda} = -\hat{\varepsilon}_{\mathbf{k}\lambda}$ and $\hat{\varepsilon}_{\mathbf{k}\lambda} \cdot \hat{\varepsilon}_{\mathbf{k}\lambda'} = \delta_{\lambda\lambda'}$. Remembering that $\sum_n \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_n^0} = N\delta_{\mathbf{k},\mathbf{k}'}$, one can see immediately that

$$\sum_n \hat{\varepsilon}_{\mathbf{k}\lambda} \cdot \hat{\varepsilon}_{\mathbf{k}'\lambda'} \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_n^0} = N\delta_{\mathbf{k},\mathbf{k}'}\delta_{\lambda\lambda'}. \tag{2.77}$$

The conditions resulting from requiring $\mathbf{P}_n$ and $\mathbf{u}_n$ to be real are

$$\mathbf{p}^*_{\mathbf{k}\lambda} = \mathbf{p}_{-\mathbf{k}\lambda} \quad \text{and} \quad \mathbf{q}^*_{\mathbf{k}\lambda} = \mathbf{q}_{-\mathbf{k}\lambda} \tag{2.78}$$

where $\mathbf{p}_{\mathbf{k}\lambda} = \hat{\varepsilon}_{\mathbf{k}\lambda} p_{\mathbf{k}\lambda}$ and $\mathbf{q}_{\mathbf{k}\lambda} = \hat{\varepsilon}_{\mathbf{k}\lambda} q_{\mathbf{k}\lambda}$. The condition on the scalar quantities $p_{\mathbf{k}\lambda}$ and $q_{\mathbf{k}\lambda}$ differs by a minus sign from the vector relation (2.78) because $\hat{\varepsilon}_{\mathbf{k}\lambda}$ changes sign when $\mathbf{k}$ goes to $-\mathbf{k}$.

2.5.2 Quantization

To quantize, the dynamical variables $p_{\mathbf{k}\lambda}$ and $q_{\mathbf{k}\lambda}$ are replaced by quantum mechanical operators $\hat{p}_{\mathbf{k}\lambda}$ and $\hat{q}_{\mathbf{k}\lambda}$ which satisfy the commutation relations

$$[\hat{p}_{\mathbf{k}\lambda},\ \hat{q}_{\mathbf{k}'\lambda'}]_- = -\mathrm{i}\hbar\delta_{\mathbf{k}\mathbf{k}'}\delta_{\lambda\lambda'}. \tag{2.79}$$

It is again convenient to introduce creation and annihilation operators $a^\dagger_{\mathbf{k}\lambda}$ and $a_{\mathbf{k}\lambda}$ defined by

$$q_{\mathbf{k}\lambda} = \left(\frac{\hbar}{2M\omega_{\mathbf{k}\lambda}}\right)^{1/2}\left(a_{\mathbf{k}\lambda} - a^\dagger_{-\mathbf{k}\lambda}\right), \tag{2.80}$$

$$p_{\mathbf{k}\lambda} = \mathrm{i}\left(\frac{\hbar M\omega_{\mathbf{k}\lambda}}{2}\right)^{1/2}\left(a^\dagger_{\mathbf{k}\lambda} + a_{-\mathbf{k}\lambda}\right). \tag{2.81}$$

The differences in sign from one-dimensional case result from using scalar quantities $q_{\mathbf{k}\lambda}$ and $p_{\mathbf{k}\lambda}$ in defining $a_{\mathbf{k}\lambda}$ and $a^\dagger_{\mathbf{k}\lambda}$. The operators $a_{\mathbf{k}\lambda}$ and $a_{\mathbf{k}'\lambda'}$ satisfy the commutation relations

$$\left[a_{\mathbf{k}\lambda},\ a^\dagger_{\mathbf{k}'\lambda'}\right]_- = \delta_{\mathbf{k}\mathbf{k}'}\delta_{\lambda\lambda'}, \tag{2.82}$$

$$[a_{\mathbf{k}\lambda},\ a_{\mathbf{k}'\lambda'}]_- = \left[a^\dagger_{\mathbf{k}\lambda},\ a^\dagger_{\mathbf{k}'\lambda'}\right]_- = 0. \tag{2.83}$$

The Hamiltonian is given by

$$H = \sum_{\mathbf{k}\lambda}\hbar\omega_{\mathbf{k}\lambda}\left(a^\dagger_{\mathbf{k}\lambda}a_{\mathbf{k}\lambda} + \frac{1}{2}\right). \tag{2.84}$$

From this point on the analysis is essentially identical to that of the one-dimensional case which we have treated in detail already. In the three-dimensional case, we can write the displacement $\mathbf{u}_n$ and momentum $\mathbf{P}_n$ of the nth atom as the quantum mechanical operators given below:

$$\mathbf{u}_n = \sum_{\mathbf{k}\lambda}\left(\frac{\hbar}{2MN\omega_{\mathbf{k}\lambda}}\right)^{1/2}\hat{\varepsilon}_{\mathbf{k}\lambda}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}^0_n}\left(a_{\mathbf{k}\lambda} - a^\dagger_{-\mathbf{k}\lambda}\right), \tag{2.85}$$

$$\mathbf{P}_n = \sum_{\mathbf{k}\lambda}\mathrm{i}\left(\frac{\hbar M\omega_{\mathbf{k}\lambda}}{2N}\right)^{1/2}\hat{\varepsilon}_{\mathbf{k}\lambda}\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}^0_n}\left(a^\dagger_{\mathbf{k}\lambda} + a_{-\mathbf{k}\lambda}\right). \tag{2.86}$$

Mean Squared Displacement of an Atom

As an example of how to use the quantum mechanical eigenstates and the operator describing dynamical variables, let us evaluate the mean squared displacement of an atom from its equilibrium position in a three-dimensional crystal. We can write

$$\mathbf{u}_n \cdot \mathbf{u}_n = \sum_{\mathbf{k}\lambda,\mathbf{k}'\lambda'} \left(\frac{\hbar}{2MN}\right) \left(\omega_{\mathbf{k}\lambda}\omega_{\mathbf{k}'\lambda'}\right)^{-1/2} \hat{\varepsilon}_{\mathbf{k}\lambda} \cdot \hat{\varepsilon}_{\mathbf{k}'\lambda'} \left(a_{\mathbf{k}\lambda} + a^{\dagger}_{\mathbf{k}\lambda}\right) \left(a_{\mathbf{k}'\lambda'} + a^{\dagger}_{\mathbf{k}'\lambda'}\right). \tag{2.87}$$

Here, we have again chosen the origin at the equilibrium position of the n^{th} atom so that $R^0_n = 0$. Then, we can replace $\hat{\varepsilon}_{\mathbf{k}\lambda} a^{\dagger}_{-\mathbf{k}\lambda}$ by $-\hat{\varepsilon}_{\mathbf{k}\lambda} a^{\dagger}_{\mathbf{k}\lambda}$ in (2.85). This was done in obtaining (2.87). If we assume the eigenstate of the lattice is $|n_{\mathbf{k}_1\lambda_1}, n_{\mathbf{k}_2\lambda_2}, \ldots\rangle$, it is not difficult to see the that

$$\langle \mathbf{u}_n \rangle = \langle n_{\mathbf{k}_1\lambda_1}, n_{\mathbf{k}_2\lambda_2}, \ldots | \mathbf{u}_n | \, n_{\mathbf{k}_1\lambda_1}, n_{\mathbf{k}_2\lambda_2}, \ldots \rangle = 0, \tag{2.88}$$

and that

$$\langle \mathbf{u}_n \cdot \mathbf{u}_n \rangle = \sum_{\mathbf{k}\lambda} \left(\frac{\hbar}{2MN\omega_{\mathbf{k}\lambda}}\right) (2n_{\mathbf{k}\lambda} + 1)\,. \tag{2.89}$$

2.6 Heat Capacity of Solids

In the nineteenth century, it was known from experiment that at room temperature the specific heat of a solid was given by the Dulong–Petit law which said

$$C_{\text{v}} = 3R, \tag{2.90}$$

where $R = N_{\text{A}} k_{\text{B}}$, and N_{A} = Avogadro number ($=6.03 \times 10^{23}$ atoms/mole) and k_{B} = Boltzmann's constant ($=1.38 \times 10^{-16}$ ergs/°K). Recall that 1 calorie $= 4.18$ joule $= 4.18 \times 10^7$ ergs. Thus, (2.90) gave the result

$$C_{\text{v}} \simeq 6\,\text{cal/deg mole}. \tag{2.91}$$

The explanation of the Dulong–Petit law is based on the equipartition theorem of classical statistical mechanics. This theorem assumes that each atom oscillates harmonically about its equilibrium position, and that the energy of one atom is

$$E = \frac{p^2}{2m} + \frac{1}{2}kr^2 = \frac{1}{2m}\left(p_x^2 + p_y^2 + p_z^2\right) + \frac{1}{2}k\left(x^2 + y^2 + z^2\right). \tag{2.92}$$

The equipartition theorem states that for a classical system in equilibrium $\left\langle \frac{p_x^2}{2m} \right\rangle = \frac{1}{2}k_{\text{B}}T$. The same is true for the other terms in (2.92), so that the energy per atom at temperature T is $E = 3k_{\text{B}}T$. The energy of 1 mole is

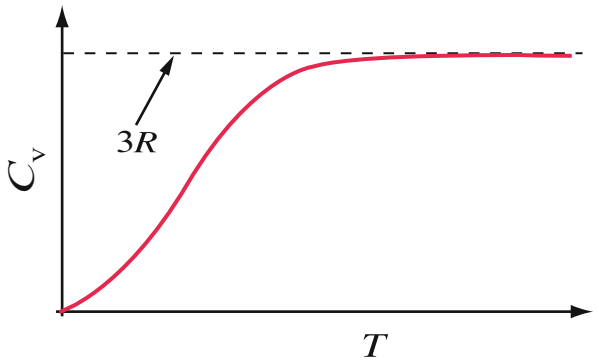


Fig. 2.9. Temperature dependence of the specific heat of a typical solid

$$U = 3N_A k_B T = 3RT. \tag{2.93}$$

It follows immediately that C_v, which is equal to $\left(\frac{\partial U}{\partial T}\right)_v$ is given by (2.90). It was later discovered that the Dulong–Petit law was valid only at sufficiently high temperature. The temperature dependence of C_v for a typical solid was found to behave as shown in Fig. 2.9.

2.6.1 Einstein Model

To explain why the specific heat decreased as the temperature was lowered, Einstein made the assumption that the atomic vibrations were quantized. By this we mean that if one assumes that the motion of each atom is described by a harmonic oscillator, then the allowed energy values are given by $\varepsilon_n = \left(n + \frac{1}{2}\right)\hbar\omega$, where $n = 0, 1, 2, \ldots$, and ω is the oscillator frequency.[2] Einstein used a very simple model in which each atom vibrated with the same frequency ω. The probability p_n that an oscillator has energy ε_n is proportional to $e^{-\varepsilon_n/k_B T}$. Because p_n is a probability and $\sum_{n=0}^{\infty} p_n = 1$, we find that it is convenient to write

$$p_n = Z^{-1} e^{-\varepsilon_n/k_B T}, \tag{2.94}$$

and to determine the constant Z from the condition $\sum_{n=0}^{\infty} p_n = 1$. Doing so gives

$$Z = e^{-\hbar\omega/2k_B T} \sum_{n=0}^{\infty} \left(e^{-\hbar\omega/k_B T}\right)^n. \tag{2.95}$$

The power series expansion of $(1 - x)^{-1}$ is equal to $\sum_{n=0}^{\infty} x^n$. Making use of this result in (2.95) gives

[2] See Appendix A for a quantum mechanical solution of a harmonic oscillator problem.

$$Z = \frac{\mathrm{e}^{-\hbar\omega/2k_\mathrm{B}T}}{1-\mathrm{e}^{-\hbar\omega/k_\mathrm{B}T}} = \frac{\mathrm{e}^{\hbar\omega/2k_\mathrm{B}T}}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T}-1}. \tag{2.96}$$

The mean value of the energy of one oscillator at temperature T is given by $\bar{\varepsilon} = \sum_n \varepsilon_n p_n$. Making use of (2.94) and (2.95) and the formula $\sum_n n\,\mathrm{e}^{-nx} = -\frac{\partial}{\partial x}\sum_n \mathrm{e}^{-nx}$ gives

$$\bar{\varepsilon} = \frac{\hbar\omega}{2} + \bar{n}\hbar\omega. \tag{2.97}$$

Here, $\bar{n}$ is the thermal average of n; it is given by

$$\bar{n} = \frac{1}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T}-1}, \tag{2.98}$$

and is called the *Bose–Einstein distribution function*. The internal energy of a lattice containing N atoms is simply $U = 3N\hbar\omega\left(\bar{n}+\frac{1}{2}\right)$, where $\bar{n}$ is given by (2.98). If N is the Avogadro number, then the specific heat is given by

$$C_\mathrm{v} = \left(\frac{\partial U}{\partial T}\right)_\mathrm{v} = 3Nk_\mathrm{B}F_\mathrm{E}\left(\frac{\hbar\omega}{k_\mathrm{B}T}\right), \tag{2.99}$$

where the *Einstein function* $F_\mathrm{E}(x)$ is defined by

$$F_\mathrm{E}(x) = \frac{x^2}{(\mathrm{e}^x-1)(1-\mathrm{e}^{-x})}. \tag{2.100}$$

It is useful to define the *Einstein temperature* T_E by $\hbar\omega = k_\mathrm{B}T_\mathrm{E}$. Then the x appearing in $F_\mathrm{E}(x)$ is $\frac{T_\mathrm{E}}{T}$.

In the high-temperature limit ($T \gg T_\mathrm{E}$), x is very small compared to unity. Expanding $F_\mathrm{E}(x)$ for small x gives

$$F_\mathrm{E}(x) = 1 - \frac{1}{12}x^2 + \cdots, \tag{2.101}$$

and

$$C_\mathrm{v} = 3Nk_\mathrm{B}\left[1 - \frac{1}{12}\left(\frac{T_\mathrm{E}}{T}\right)^2 + \cdots\right]. \tag{2.102}$$

This agrees with the classical Dulong–Petit law at very high temperature and it falls off with decreasing T.

In the low temperature limit ($T \ll T_\mathrm{E}$), x is very large compared to unity. In this limit,

$$F_\mathrm{E}(x) \simeq x^2\mathrm{e}^{-x}, \tag{2.103}$$

and

$$C_\mathrm{v} = 3Nk_\mathrm{B}\left(\frac{T_\mathrm{E}}{T}\right)^2 \mathrm{e}^{-T_\mathrm{E}/T}. \tag{2.104}$$

The Einstein temperature was treated as a parameter to be determined by comparison with experiment. The Einstein model reproduced the Dulong–Petit law at high temperature and showed that C_v decreased as the temperature was lowered. Careful comparison of experimental data with the model showed that the low temperature behavior was not quite correct. The experimental data fit a T^3 law at low temperature (i.e., $C_\mathrm{v} \propto T^3$) instead of decreasing exponentially as predicted by the simple Einstein model.

2.6.2 Modern Theory of the Specific Heat of Solids

We know from our study of lattice vibrations that Einstein's assumption that each atom in the crystal oscillated at a single frequency ω is too great a simplification. In fact, the normal modes of vibration have a spectrum $\omega_{\mathbf{q}\lambda}$, where $\mathbf{q}$ is a wave vector restricted to the first Brillouin zone and λ is a label that defines the polarization of the mode. The energy of the crystal at temperature T is given by

$$U = \sum_{\mathbf{q}\lambda} \left(\bar{n}_{\mathbf{q}\lambda} + \frac{1}{2} \right) \hbar \omega_{\mathbf{q}\lambda}. \tag{2.105}$$

In (2.105), $\bar{n}_{\mathbf{q}\lambda}$ is given by

$$\bar{n}_{\mathbf{q}\lambda} = \frac{1}{\mathrm{e}^{\hbar\omega_{\mathbf{q}\lambda}/k_\mathrm{B}T} - 1}. \tag{2.106}$$

From (2.105), the specific heat can be obtained; it is given by

$$C_\mathrm{v} = \left(\frac{\partial U}{\partial T} \right)_\mathrm{v} = k_\mathrm{B} \sum_{\mathbf{q}\lambda} \left(\frac{\hbar\omega_{\mathbf{q}\lambda}}{k_\mathrm{B}T} \right)^2 \left(\mathrm{e}^{\frac{\hbar\omega_{\mathbf{q}\lambda}}{k_\mathrm{B}T}} - 1 \right)^{-1} \left(1 - \mathrm{e}^{-\frac{\hbar\omega_{\mathbf{q}\lambda}}{k_\mathrm{B}T}} \right)^{-1}. \tag{2.107}$$

To carry out the summation appearing in (2.107), we must have either more information or some model describing how $\omega_{\mathbf{q}\lambda}$ depends on $\mathbf{q}$ and λ is needed.

Density of States

Recall that the allowed values of $\mathbf{q}$ were given by

$$\mathbf{q} = 2\pi \left(\frac{n_1}{N_1}\mathbf{b}_1 + \frac{n_2}{N_2}\mathbf{b}_2 + \frac{n_3}{N_3}\mathbf{b}_3 \right), \tag{2.108}$$

where $\mathbf{b}_i$ were primitive translations of the reciprocal lattice, n_i were integers, and N_i were the number of steps in the direction i that were required before the periodic boundary conditions returned one to the initial lattice site. For simplicity, let us consider a simple cubic lattice. Then $\mathbf{b}_i = a^{-1}\hat{x}_i$, where a is the lattice spacing and $\hat{x}_i$ is a unit vector (in the x, y, or z direction). The allowed (independent) values of $\mathbf{q}$ are restricted to the first Brillouin zone. In this case, that implies that $-\frac{1}{2}N_i \leq n_i \leq \frac{1}{2}N_i$. Then, the summations over

q_x, q_y, and q_z can be converted to integrals as follows:

$$\sum_{q_x} \Rightarrow \frac{\int \mathrm{d}q_x}{2\pi/N_x a} \Rightarrow \frac{L_x}{2\pi}\int \mathrm{d}q_x. \tag{2.109}$$

Therefore, the three-dimensional sum $\sum_{\mathbf{q}}$ becomes

$$\sum_{\mathbf{q}} = \frac{L_x L_y L_z}{(2\pi)^3}\int \mathrm{d}^3 q = \frac{V}{(2\pi)^2}\int \mathrm{d}^3 q. \tag{2.110}$$

In these equations L_x, L_y, and L_z are equal to the length of the crystal in the x, y, and z directions, and $V = L_x L_y L_z$ is the crystal volume. For any function $f(\mathbf{q})$, we can write

$$\sum_{\mathbf{q}} f(\mathbf{q}) = \frac{V}{(2\pi)^3}\int \mathrm{d}^3 q\, f(\mathbf{q}). \tag{2.111}$$

Now, it is convenient to introduce the *density of states* $g(\omega)$ defined by

$$g(\omega)\mathrm{d}\omega = \left\{ \begin{array}{c} \text{The number of normal modes per unit volume} \\ \text{whose frequency } \omega_{\mathbf{q}\lambda} \text{ satisfies } \omega < \omega_{\mathbf{q}\lambda} < \omega + \mathrm{d}\omega. \end{array} \right. \tag{2.112}$$

From this definition, it follows that

$$g(\omega)\mathrm{d}\omega = \frac{1}{V}\sum_{\substack{\mathbf{q}\lambda \\ \omega < \omega_{\mathbf{q}\lambda} < \omega + \mathrm{d}\omega}} 1 = \frac{1}{(2\pi)^3}\sum_{\lambda}\int_{\omega<\omega_{\mathbf{q}\lambda}<\omega+\mathrm{d}\omega} \mathrm{d}^3 q. \tag{2.113}$$

Let $S_\lambda(\omega)$ be the surface in three-dimensional wave vector space on which $\omega_{\mathbf{q}\lambda}$ has the value ω. Then $dS_\lambda(\omega)$ is an infinitesimal element of this surface of constant frequency (see Fig. 2.10). The frequency change $d\omega$ in going from the surface $S_\lambda(\omega)$ to the surface $S_\lambda(\omega + \mathrm{d}\omega)$ can be expressed in terms of $d\mathbf{q}$, an infinitesimal displacement in $\mathbf{q}$ space as

$$\mathrm{d}\omega = \mathrm{d}\mathbf{q}\cdot[\nabla_{\mathbf{q}}\omega_{\mathbf{q}\lambda}]_{\omega_{\mathbf{q}\lambda}=\omega} \quad \text{or} \quad \mathrm{d}\omega = \mathrm{d}q_\perp\,|\nabla_{\mathbf{q}}\omega_{\mathbf{q}\lambda}|_{\omega_{\mathbf{q}\lambda}=\omega}\,. \tag{2.114}$$

Here, $dq_\perp$ is the component of $d\mathbf{q}$ normal to the surface of constant frequency $S_\lambda(\omega)$. The volume element $\mathrm{d}^3 q$ in wave vector space can be written $\mathrm{d}^3 q = \mathrm{d}q_\perp \mathrm{d}S_\lambda(\omega)$, and using (2.114) allows us to write

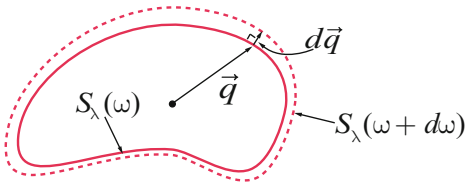


Fig. 2.10. Constant frequency surfaces in three-dimensional wave vector space

$$\mathrm{d}^3q = \frac{\mathrm{d}\omega}{|\nabla_{\mathbf{q}}\omega_{\mathbf{q}\lambda}|_{\omega_{\mathbf{q}\lambda}=\omega}}\mathrm{d}S_\lambda(\omega). \tag{2.115}$$

With this result, we can express the density of states as

$$g(\omega) = \frac{1}{(2\pi)^3}\sum_\lambda \int \frac{\mathrm{d}S_\lambda(\omega)}{|\nabla_{\mathbf{q}}\omega_{\mathbf{q}\lambda}|_\omega}. \tag{2.116}$$

In (2.116) the integration is performed over the surface of constant frequency $S_\lambda(\omega)$. The denominator contains the magnitude of the gradient of $\omega_{\mathbf{q}\lambda}$ (with respect to $\mathbf{q}$) evaluated at $\omega_{\mathbf{q}\lambda} = \omega$.

2.6.3 Debye Model

To evaluate (2.107) and obtain the specific heat, Debye[3] introduced a simple assumption about the phonon spectrum. He took $\omega_{\mathbf{q}\lambda} = s_\lambda\,|\mathbf{q}|$ for all values of $\mathbf{q}$ in the first Brillouin zone. Then, the surfaces of constant energy are spheres (i.e., $S_\lambda(\omega)$ is a sphere in $\mathbf{q}$ space of radius $q = \frac{\omega}{s_\lambda}$). In addition, Debye replaced the Brillouin zone by a sphere of the same volume. Since $\sum_{q\in 1^{\mathrm{st}}\mathrm{BZ}} 1 = N$, we can write

$$N = \left(\frac{L}{2\pi}\right)^3 \int_{|q|<q_\mathrm{D}} \mathrm{d}^3q = \frac{V}{(2\pi)^3}\,\frac{4}{3}\pi q_\mathrm{D}^3. \tag{2.117}$$

In (2.117) we have introduced q_D, the *Debye wave vector*. A sphere of radius q_D contains the N independent values of $\mathbf{q}$ associated with a crystal containing N atoms. From (2.117), $q_\mathrm{D}^3 = 6\pi^2 N/V$, where V is the volume of the crystal.

The density of states for the Debye model is very simple since $|\nabla_q\,\omega_{q\lambda}| = s_\lambda$. Substituting this result into (2.116) gives

$$g(\omega) = \frac{1}{(2\pi)^3}\sum_\lambda \left[\frac{4\pi q^2}{s_\lambda}\right]_{q=\frac{\omega}{s_\lambda}\le q_\mathrm{D}}. \tag{2.118}$$

If we introduce the unit step function $\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ for $x < 0$, $g(\omega)$ can be expressed

$$g(\omega) = \frac{\omega^2}{2\pi^2}\left[\frac{\theta(s_l k_\mathrm{D} - \omega)}{s_l^3} + \frac{2\,\theta(s_t k_\mathrm{D} - \omega)}{s_t^3}\right]. \tag{2.119}$$

Here, of course, s_l and s_t are the speed of a longitudinal and of a transverse sound wave. Figure 2.11 shows the frequency dependence of the three-dimensional density of states in the Debye model. Any summation over allowed values of wave vector can be converted into an integral over frequency by using the relation

[3] P. Debye, Annalen der Physik **39**, 789 (1912).

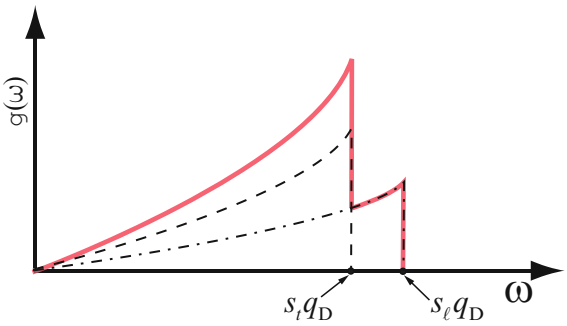


Fig. 2.11. Three-dimensional density of states in the Debye model

$$\sum_{\mathbf{q}\lambda} f(\omega_{\mathbf{q}\lambda}) = V \int \mathrm{d}\omega \, g(\omega) f(\omega). \tag{2.120}$$

Here, $f(\omega_{\mathbf{q}\lambda})$ is an arbitrary function of the normal mode frequencies $\omega_{\mathbf{q}\lambda}$. Making use of (2.120), the expression for the specific heat [(2.107)] can be written

$$C_\mathrm{v} = k_\mathrm{B} V \int \mathrm{d}\omega \left(\frac{\hbar\omega}{\Theta}\right)^2 \left(\mathrm{e}^{\hbar\omega/\Theta} - 1\right)^{-1} \left(1 - \mathrm{e}^{-\hbar\omega/\Theta}\right)^{-1} g(\omega). \tag{2.121}$$

Here, we have introduced $\Theta = k_\mathrm{B} T$. We define the *Debye temperature* T_D by $\Theta_\mathrm{D} = k_\mathrm{B} T_\mathrm{D} = \hbar s_l q_\mathrm{D}$. Remembering that $V = 6\pi^2 N q_\mathrm{D}^{-3}$ and that the integral $\int d\omega$ goes from $\omega = 0$ to $\omega = \omega_\mathrm{D} = s_l q_\mathrm{D}$ for longitudinal waves and from $\omega = 0$ to $\omega = s_t q_\mathrm{D} = \frac{s_t}{s_l}\omega_\mathrm{D}$ for transverse waves, it is not difficult to demonstrate that

$$C_\mathrm{v} = 3Nk_\mathrm{B} \left[\frac{1}{3} F_\mathrm{D}\left(\frac{\Theta_D}{\Theta}\right) + \frac{2}{3} F_\mathrm{D}\left(\frac{s_t \Theta_D}{s_l \Theta}\right)\right], \tag{2.122}$$

where the *Debye function* $F_\mathrm{D}(x)$ is defined by

$$F_\mathrm{D}(x) = \frac{3}{x^3} \int_0^x \frac{z^4 \, \mathrm{d}z}{(\mathrm{e}^z - 1)(1 - \mathrm{e}^{-z})}. \tag{2.123}$$

Behavior at $\Theta \gg \Theta_\mathrm{D}$

In this limit, x which equals $\frac{\Theta_\mathrm{D}}{\Theta}$ or $\frac{s_t \Theta_\mathrm{D}}{s_l \Theta}$ is much smaller than unity. Therefore, we can expand the exponentials for small argument to obtain

$$F_\mathrm{D}(x) \simeq \frac{3}{x^3} \int_0^x \frac{z^4 \, \mathrm{d}z}{z^2} \approx 1. \tag{2.124}$$

In this limit, $C_\mathrm{v} = 3Nk_\mathrm{B}$, in agreement with the classical Dulong–Petit law.

Behavior at $\Theta \ll \Theta_{\mathrm{D}}$

In this limit, x is much larger than unity, and because of the exponential in the denominator of the integral little error arises from replacing the upper limit by infinity. This gives

$$F_{\mathrm{D}}(x) \simeq \frac{3}{x^3} \int_0^\infty \frac{z^4 \, \mathrm{d}z}{(\mathrm{e}^z - 1)(1 - \mathrm{e}^{-z})}. \tag{2.125}$$

The integral is simply a constant. Its value can be obtained analytically

$$\int_0^\infty \frac{z^4 \, \mathrm{d}z}{(\mathrm{e}^z - 1)(1 - \mathrm{e}^{-z})} = \frac{4}{15}\pi^4. \tag{2.126}$$

The result for C_{v} at very low temperature is

$$C_{\mathrm{v}} = \frac{4}{5}\pi^4 N k_{\mathrm{B}} \left[1 + 2\left(\frac{s_l}{s_t}\right)^3\right] \left(\frac{\Theta}{\Theta_{\mathrm{D}}}\right)^3. \tag{2.127}$$

This agrees with the observed behavior of the specific heat at very low temperature, viz. $C_{\mathrm{v}} = AT^3$, where A is a constant.

2.6.4 Evaluation of Summations over Normal Modes for the Debye Model

In our calculation of the recoil free fraction in the Mössbauer effect (See (2.56)), and in the evaluation of (2.89), the mean square displacement $\langle \mathbf{u}_n \cdot \mathbf{u}_n \rangle$ of an atom from its equilibrium position, we encountered sums of the form

$$I = N^{-1} \sum_{\mathbf{q}\lambda} \frac{\bar{n}_{\mathbf{q}\lambda} + \frac{1}{2}}{\hbar\omega_{\mathbf{q}\lambda}}. \tag{2.128}$$

These sums can be performed by converting the sums to integrals through the standard prescription

$$\sum_{\mathbf{q}} f(\omega_{\mathbf{q}\lambda}) \to \frac{V}{(2\pi)^3} \int \mathrm{d}^3 q \, f(\omega_{\mathbf{q}\lambda}), \tag{2.129}$$

or by making use of the density of states $g(\omega)$ and the result that

$$\sum_{\mathbf{q}\lambda} f(\omega_{\mathbf{q}\lambda}) = V \int \mathrm{d}\omega \, g(\omega) f(\omega). \tag{2.130}$$

For simplicity, we will use a Debye model with the velocity of transverse and longitudinal waves both equal to s. Then

$$g(\omega) = \frac{3\omega^2}{2\pi^2 s^3} \theta(s k_{\mathrm{D}} - \omega). \tag{2.131}$$

The summation in (2.128) can then be written as

$$I = \frac{V}{N}\int_0^{\omega_{\rm D}} {\rm d}\omega \frac{3\omega^2}{2\pi^2 s^3}\frac{1}{\hbar\omega}\left[\frac{1}{2}+\frac{1}{{\rm e}^{\hbar\omega/\Theta}-1}\right]. \tag{2.132}$$

Let $z = \frac{\hbar\omega}{\Theta}$, and make use of $k_{\rm D}^3 = 6\pi^2\frac{N}{V}$. Then (2.132) can be rewritten

$$I = \frac{9}{\Theta_{\rm D}}\left(\frac{\Theta}{\Theta_{\rm D}}\right)^2\int_0^{\Theta_{\rm D}/\Theta} {\rm d}z\ z\left[\frac{1}{2}+\frac{1}{{\rm e}^z-1}\right]. \tag{2.133}$$

First, let us look at the high temperature limit of (2.133). If $\Theta \gg \Theta_{\rm D}$, then for values of z appearing in the integrand $\frac{1}{{\rm e}^z-1} \simeq \frac{1}{z}$. This corresponds to the classical equipartition of energy since the energy of a mode of frequency $\omega_{\mathbf{q}\lambda}$ is given by

$$\hbar\omega_{\mathbf{q}\lambda}\left[\frac{1}{{\rm e}^{\hbar\omega_{\mathbf{q}\lambda}/\Theta}-1}+\frac{1}{2}\right] \simeq \hbar\omega_{\mathbf{q}\lambda}\left[\frac{\Theta}{\hbar\omega_{\mathbf{q}\lambda}}+\frac{1}{2}\right],$$

and this is equal to Θ for every mode (the $\frac{1}{2}$ is negligible if $\Theta \gg \hbar\omega_{\mathbf{q}\lambda}$) as required by classical statistical mechanics. With this approximation

$$I \simeq \frac{9}{\Theta_{\rm D}}\left(\frac{\Theta}{\Theta_{\rm D}}\right)^2\int_0^{\Theta_{\rm D}/\Theta} {\rm d}z = \frac{9\Theta}{\Theta_{\rm D}^2}. \tag{2.134}$$

At very low temperature, $\Theta \ll \Theta_{\rm D}$, we can approximate the upper limit by ∞ in the term proportional to $({\rm e}^z-1)^{-1}$, since the contribution from very large values of z is very small. This gives

$$I = \frac{9}{\Theta_{\rm D}}\left(\frac{\Theta}{\Theta_{\rm D}}\right)^2\left[\int_0^{\infty}\frac{{\rm d}z\ z}{{\rm e}^z-1}+\int_0^{\Theta_{\rm D}/\Theta}{\rm d}z\ \frac{z}{2}\right]. \tag{2.135}$$

The first integral in the square bracket is a constant, while the second is $\frac{1}{4}\left(\frac{\Theta_{\rm D}}{\Theta}\right)^2$. The second term is much larger than the first for $\Theta \ll \Theta_{\rm D}$, so it is a reasonable approximation to take

$$I = \frac{9}{4\Theta_{\rm D}}. \tag{2.136}$$

(see, for example, Fig. 2.12).

2.6.5 Estimate of Recoil Free Fraction in Mössbauer Effect

Equation (2.56) gave the probability of starting in a lattice state $|n_i> = |n_1, n_2, \ldots, n_N>$ and ending, after the γ-ray emission, in the same state. If we assume that the crystal is in thermal equilibrium at a temperature Θ, then (2.56) is simply

$$P(\bar{n}_i, \bar{n}_i) = {\rm e}^{-2E(K)I}, \tag{2.137}$$

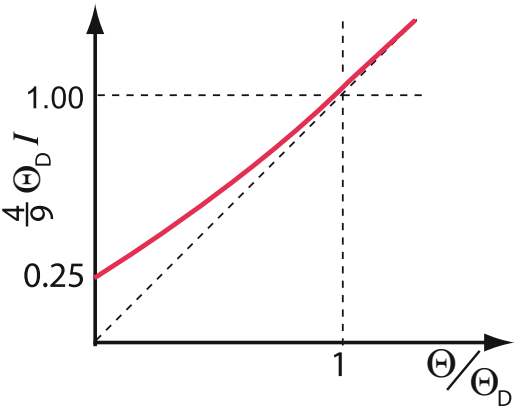


Fig. 2.12. Behavior of an integral I for $\Theta \leq \Theta_D$

where $\bar{n}_i$ is the *Bose–Einstein distribution function*, $E(K)$ is the recoil energy, and I is given by (2.132). We have just evaluated I using a simplified Debye model at both high ($\Theta \gg \Theta_D$) and low ($\Theta \ll \Theta_D$) temperatures. If we use (2.134) and (2.137), we find that at ($\Theta_D \gg \Theta$), $P \simeq e^{-\frac{9E(K)}{2\Theta_D}}$. Remember that $E(K) \simeq 2 \times 10^{-3}\,\text{eV}$. For a typical crystal $\Theta_D \simeq 300\text{K} \cdot k_B \approx 2.5 \times 10^{-2}\,\text{eV}$, giving for P, $P \simeq e^{-\frac{1}{3}} \approx 0.7$. This means that at very low temperature, 70% of the γ rays are emitted without any change in the number of phonons in the crystal.

At high temperature (let us take $\Theta = 400\,\text{K}$, larger than but not much larger than $\Theta_D \simeq 300\,\text{K}$) $I \simeq \frac{9\Theta}{\Theta_D^2}$ giving $P(\bar{n}_i, \bar{n}_i) \simeq e^{-\frac{9E(K)}{2\Theta_D}\frac{4\Theta}{\Theta_D}}$. This gives $P(\bar{n}_i, \bar{n}_i)$ at $\Theta = 400\,\text{K}$ of roughly 0.14, so that, even at room temperature the Mössbauer recoil free fraction is reasonably large.

2.6.6 Lindemann Melting Formula

The *Lindemann melting formula* is based on the idea that melting will occur when the amplitude of the atomic vibrations $\left(\text{i.e.,} \left\langle (\delta R)^2 \right\rangle^{1/2}\right)$ becomes equal to some fraction γ of the interatomic spacing. Recall that $\langle \mathbf{u}_n \cdot \mathbf{u}_n \rangle = \frac{\hbar^2}{M} I$ where I is given by (2.128) (see (2.89)). We can use the $\Theta \gg \Theta_D$ limit for I to write

$$\left\langle (\delta R)^2 \right\rangle \simeq \frac{9\hbar^2 \Theta}{M \Theta_D^2}. \tag{2.138}$$

The melting temperature is assumed to be given by $\Theta_{\text{Melting}} = \frac{M\Theta_D^2}{9\hbar^2}\gamma^2 r_0^2$, where r_0 is the atomic spacing and γ is a constant in the range ($0.2 \leq \gamma \leq 0.25$). This result is only very qualitative since it is based on a very much oversimplified model.

Some Remarks on the Debye Model

One can obtain an intuitive picture of the temperature dependence of the specific heat by applying the idea of classical equipartition of energy, but only to modes for which $\hbar\omega < \Theta$. By this we mean that only modes whose energy $\hbar\omega$ is smaller than $\Theta = k_{\mathrm{B}}T$ can be thermally excited at a temperature Θ and make a contribution to the internal energy U, and such modes contribute an energy Θ. Thus, we can write for U

$$U = \sum_{\mathbf{q}\lambda} \left(\bar{n}_{\mathbf{q}\lambda} + \frac{1}{2} \right) \hbar\omega_{\mathbf{q}\lambda} \simeq 3 \frac{V}{(2\pi)^3} \int_0^{\Theta/\hbar s} \Theta \; 4\pi q^2 \; \mathrm{d}q. \tag{2.139}$$

In writing (2.139) we have omitted the zero point energy since it does not depend on temperature and put $\hbar\omega[\bar{n}(\omega)] \simeq \Theta$ for all modes of energy less than Θ. This gives (using $V = \frac{6\pi^2 N}{k_{\mathrm{D}}^3}$ and $\hbar s k_{\mathrm{D}} = \Theta_{\mathrm{D}}$)

$$U = 3N \left(\frac{\Theta}{\Theta_{\mathrm{D}}} \right)^3 \Theta. \tag{2.140}$$

Differentiating with respect to T gives

$$C_{\mathrm{v}} = 12 N k_{\mathrm{B}} \left(\frac{\Theta}{\Theta_{\mathrm{D}}} \right)^3. \tag{2.141}$$

This rough approximation gives the correct T^3 temperature dependence, but the coefficient is not correct as might be expected from such a simple picture.

Experimental Data

Experimentalists measure the specific heat as a function of temperature over a wide range of temperatures. They often use the Debye model to fit their data, taking the Debye temperature as an adjustable parameter to be determined by fitting the data to (2.122) or some generalization of it. Thus, if you see a plot of Θ_{D} as a function of temperature, it only means that at that particular temperature T one needs to take $\Theta_{\mathrm{D}} = \Theta_{\mathrm{D}}(T)$ for that value of T to fit the data to a Debye model. It is always found that at very low T and at very high T the correct Debye temperature $\Theta_{\mathrm{D}} = \hbar s \left(\frac{6\pi^2 N}{V} \right)^{1/3}$ agrees with the experiment. At intermediate temperatures these might be fluctuations in Θ_{D} of the order of 10% from the correct value. The reason for this is that $g(\omega)$, the density of states, for the Debye model is a considerable simplification of the actual of $g(\omega)$ for real crystals. This can be illustrated by considering briefly the critical points in the phonon spectrum.

2.6.7 Critical Points in the Phonon Spectrum

Remember that the general expression for the density of states was given by (2.116). Points at which $\nabla_{\mathbf{q}}\omega_{\mathbf{q}\lambda} = 0$ are called *critical points*; the integrand in (2.116) becomes infinite at such points.

Suppose that $\mathbf{q}_c$ is a critical point in the phonon spectrum. Let $\boldsymbol{\xi} = \mathbf{q} - \mathbf{q}_c$; then for points in the neighborhood of $\mathbf{q}_c$ we can write

$$\omega_q = \omega_c + \alpha_1\xi_1^2 + \alpha_2\xi_2^2 + \alpha_3\xi_3^2, \tag{2.142}$$

where ξ_i are the components of $\boldsymbol{\xi}$, and $\omega_c = \omega(q_c)$. If α_1, α_2, and α_3 are all negative, by substituting into the expression for $g(\omega)$ and evaluating in the neighborhood of q_c, one obtains

$$g(\omega) = \begin{cases} 0 & \text{if } \omega > \omega_c, \\ \text{constant } (\omega_c - \omega)^{1/2} & \text{if } \omega < \omega_c. \end{cases} \tag{2.143}$$

Thus, although $g(\omega)$ is continuous at a critical point, its first derivative is discontinuous.

In three dimensions there are four kinds of critical points:

1. *Maxima*: Points at which all three α_i are negative.
2. *Minima*: Points at which all three α_i are positive.
3. *Saddle Points of the First Kind*: Points at which two α_i's are positive and one is negative.
4. *Saddle Points of the Second Kind*: Points at which one α_i is positive and the other two are negative.

The critical points all show up as points at which $\frac{dg(\omega)}{d\omega}$ is discontinuous. A rough sketch of $g(\omega)$ vs. ω showing several critical points is shown in Fig. 2.13. It is not too difficult to demonstrate that in three dimensions the phonon spectrum must have at least one maximum, one minimum, three saddle points of each kind. As an example, we look at the simpler case of two dimensions. Then

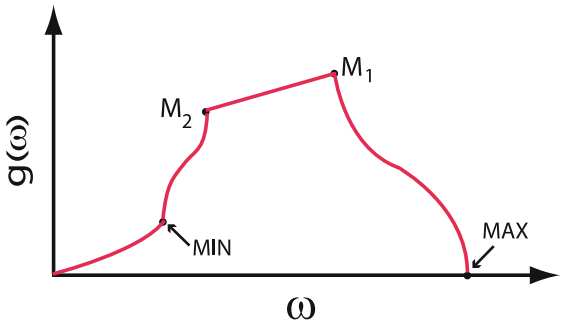


Fig. 2.13. Behavior of the density of states at various critical points

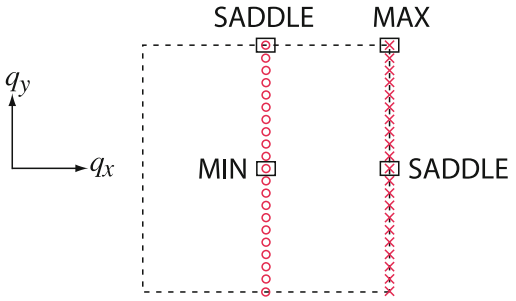


Fig. 2.14. Behavior of critical points in two dimensions

the phonon spectrum must have at least one maximum, one minimum, and two saddle points (there is only one kind of saddle point in two dimensions) (see Fig. 2.14). This can be demonstrated as follows:

1. We know ω_q is a periodic function of $\mathbf{q}$; values of $\mathbf{q}$ which differ by a reciprocal lattice vector $\mathbf{K}$ give the same ω_q.
2. For a Brillouin zone of a two-dimensional square, we can consider $\omega(q_x, q_y)$ as a function of q_x for a sequence of different fixed values of q_y. Because $\omega(q_x, q_y)$ is a periodic function of q_x there must be at least one maximum and one minimum on each line $q_y =$ constant.
3. Consider the locus of all maxima (represented by X's in Fig. 2.14). Along this locus $\omega(\mathbf{q})$ must have at least one maximum and one minimum as a function of q_y. These points will be an absolute maximum and a saddle point.
4. Doing the same for the locus of all minima (represented by O's in Fig. 2.14) gives one absolute minimum and another saddle point.

Because of the critical points, the phonon spectrum of a real solid looks quite different from that of the Debye model. However, the Debye model is constructed so that

1. The low frequency behavior of $g(\omega)$ is correct because for very small ω, $\omega_{\mathbf{q}\lambda} = s_\lambda |q|$ is a very accurate approximation.
2. The total area under the curve $g(\omega)$ is correct since k_D, the Debye wave vector is chosen so that there are exactly the correct total number of modes $3N$.

Because of this, the Debye model is good at very low temperature (where only very low frequency modes are important) and at very high temperature (where only the total number of modes and equipartition of energy are important). In Fig. 2.15 we compare $g(\omega)$ for a Debye model with that of a real crystal. We note that $\int g_\mathrm{Debye}(\omega)\mathrm{d}\omega \approx \int g_\mathrm{Actual}(\omega)\mathrm{d}\omega$.

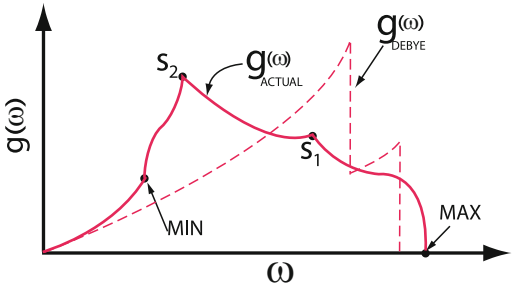


Fig. 2.15. Comparison of the density of states $g(\omega)$ for a Debye model and that of a real crystal

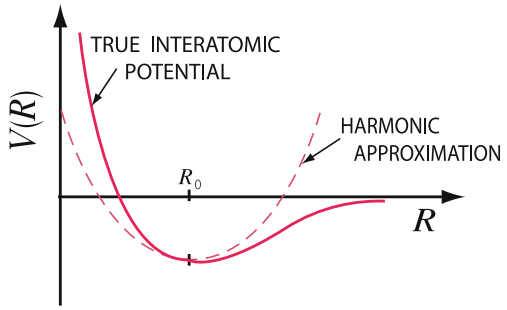


Fig. 2.16. Comparison of the potential felt by an atom and the harmonic approximation to it

2.7 Qualitative Description of Thermal Expansion

We have approximated the interatomic potential in a crystal by

$$V(R) = V(R_0) + \sum_{ij} c_{ij} u_i u_j + \text{higher terms.} \tag{2.144}$$

In Fig. 2.16 we show a sketch of the potential felt by one atom and the harmonic approximation to it. There are two main differences in the two potentials:

1. The true interatomic potential has a very strong repulsion at $u = R - R_0$ negative (i.e., close approach of the pair of atoms).
2. The true potential levels off as R becomes very large (i.e., for large positive u).

For a simple one-dimensional model we can write $x = x_0 + u$, where x_0 is the equilibrium separation between a pair of atoms and $u = x - x_0$ is the deviation from equilibrium. Then, we can model the behavior shown in Fig. 2.16 by assuming that

$$V(x) = V_0 + cu^2 - gu^3 - fu^4. \tag{2.145}$$

Here, g and f are positive constants. The fu^4 term simply accounts for the fact that the harmonic approximation rises too quickly for large u. The gu^3 term accounts for the asymmetry in the potential for u greater than or less than zero. When u is negative, $-gu^3$ is positive making the short range repulsion larger; when u is positive, $-gu^3$ is negative softening the interatomic repulsion for large R.

Now let us evaluate the expectation value of u at a temperature $k_{\mathrm{B}}T = \beta^{-1}$.

$$\langle u \rangle = \frac{\int_{-\infty}^{\infty} \mathrm{d}u\, u\, \mathrm{e}^{-\beta V}}{\int_{-\infty}^{\infty} \mathrm{d}u\, \mathrm{e}^{-\beta V}}. \tag{2.146}$$

But, $V = V_0 + cu^2 - gu^3 - fu^4$, and we can expand $\mathrm{e}^{\beta\left(gu^3 + fu^4\right)}$, for small values of u, to obtain

$$\mathrm{e}^{-\beta V} = \mathrm{e}^{-\beta\left(V_0 + cu^2\right)} \left(1 + \beta g u^3 + \beta f u^4\right). \tag{2.147}$$

The integrals in the numerator and denominator of (2.146) can be evaluated. Because of the factor $\mathrm{e}^{-\beta c u^2}$, we do not have to worry about the behavior of the integrand for very large values of $|u|$ so there is little error in taking the limit as $u = \pm\infty$. We can easily see that

$$\int_{-\infty}^{\infty} \mathrm{d}u\, \mathrm{e}^{-\beta V} = \mathrm{e}^{-\beta V_0} \int_{-\infty}^{\infty} \mathrm{d}u\, \mathrm{e}^{-\beta c u^2} \left(1 + \beta g u^3 + \beta f u^4\right). \tag{2.148}$$

The $\beta g u^3$ term vanishes because it is an odd function of u; the $\beta f u^4$ gives a small correction to the first term so it can be neglected. This results in

$$\int_{-\infty}^{\infty} \mathrm{d}u\, \mathrm{e}^{-\beta V} \simeq \mathrm{e}^{-\beta V_0} \left(\frac{\pi}{\beta c}\right)^{1/2}. \tag{2.149}$$

In writing down (2.149) we have made use of the result $\int_{-\infty}^{\infty} \mathrm{d}z\, \mathrm{e}^{-z^2} = \sqrt{\pi}$. The integral in the numerator of (2.146) becomes

$$\int_{-\infty}^{\infty} \mathrm{d}u\, u\, \mathrm{e}^{-\beta V} = \mathrm{e}^{-\beta V_0} \int_{-\infty}^{\infty} \mathrm{d}u\, u\, \mathrm{e}^{-\beta c u^2} \left(1 + \beta g u^3 + \beta f u^4\right). \tag{2.150}$$

Only the $\beta g u^3$ term in the square bracket contributes to the integral. The result is

$$\int_{-\infty}^{\infty} \mathrm{d}u\, u\, \mathrm{e}^{-\beta V} \simeq \mathrm{e}^{-\beta V_0} \frac{3\sqrt{\pi}}{4} \beta g \left(\beta c\right)^{-5/2}. \tag{2.151}$$

In obtaining (2.151) we have made use of the result $\int_{-\infty}^{\infty} \mathrm{d}z\, z^4\, \mathrm{e}^{-z^4} = \frac{3\sqrt{\pi}}{4}$. Substituting it back in (2.146) gives

$$\langle u \rangle = \frac{1}{\beta} \frac{3g}{4c^2} = \frac{3g}{4c^2} k_{\mathrm{B}} T. \tag{2.152}$$

The displacement from equilibrium is positive and increases with temperature. This suggests why a crystal expands with increasing temperature.

2.8 Anharmonic Effects

To get some idea about how one would go about treating *anharmonic effect*, let us go back to the simple one-dimensional model and include terms that we have neglected (up to this time) in the expansion of the potential energy. We can write $H = H_{\text{Harmonic}} + H'$, where H' is given by

$$H' = \frac{1}{3!}\sum_{lmn} d_{lmn} u_l u_m u_n + \frac{1}{4!}\sum_{lmnp} f_{lmnp} u_l u_m u_n u_p + \cdots . \tag{2.153}$$

As a first approximation, let us keep only the cubic anharmonic term and make use of

$$u_m = \sum_k \left(\frac{\hbar}{2MN\omega_k}\right)^{1/2} \left(a_k + a_{-k}^\dagger\right) \mathrm{e}^{\mathrm{i}kma}. \tag{2.154}$$

Substituting (2.154) in (2.153) gives

$$\begin{aligned} H_3' = {} & \tfrac{1}{3!}\sum_{lmn} d_{lmn} \sum_{kk'k''} \left(\frac{\hbar}{2MN}\right)^{3/2} (\omega_k \omega_k' \omega_k'')^{-1/2} \\ & \times \left(a_k + a_{-k}^\dagger\right)\left(a_k' + a_{-k'}^\dagger\right)\left(a_{k''} + a_{-k''}^\dagger\right) \mathrm{e}^{\mathrm{i}kna}\mathrm{e}^{\mathrm{i}k'ma}\mathrm{e}^{\mathrm{i}k''la}. \end{aligned} \tag{2.155}$$

As before, d_{lmn} does not depend on l, m, n individually, but on their relative positions. We can, therefore, write $d_{lmn} = d(n-m, n-l)$. Now introduce $g = n - m$ and $j = n - l$ and sum over all values of g, j, and n instead of l, m, and n. This gives for the cubic anharmonic correction to the Hamiltonian

$$\begin{aligned} H_3' = {} & \tfrac{1}{3!}\sum_{ngj} d(g,j) \sum_{kk'k''} \left(\frac{\hbar}{2MN}\right)^{3/2} (\omega_k \omega_{k'} \omega_{k''})^{-1/2} \\ & \times \left(a_k + a_{-k}^\dagger\right)\left(a_{k'} + a_{-k'}^\dagger\right)\left(a_{k''} + a_{-k''}^\dagger\right) \mathrm{e}^{\mathrm{i}kna}\mathrm{e}^{\mathrm{i}k'(n-g)a}\mathrm{e}^{\mathrm{i}k''(n-j)a}. \end{aligned} \tag{2.156}$$

The only factor depending on n is $\mathrm{e}^{\mathrm{i}(k+k'+k'')na}$, and

$$\sum_n \mathrm{e}^{\mathrm{i}(k+k'+k'')na} = N\delta\left(k + k' + k'', K\right). \tag{2.157}$$

Here, K is a reciprocal lattice vector; the value of K is uniquely determined since k, k', k'' must all lie within the first Brillouin zone. Eliminate k'' remembering that if $-(k + k')$ lies outside the first Brillouin zone, one must add a reciprocal lattice vector K to k'' to satisfy (2.157). With this H_3' becomes

$$\begin{aligned} H_3' = {} & N\sum_{kk'} \frac{1}{3!}\sum_{gj} d(g,j) \mathrm{e}^{-\mathrm{i}k'ga}\mathrm{e}^{\mathrm{i}(k+k')ja} \left(\frac{\hbar}{2MN}\right)^{3/2} \\ & \times (\omega_k \omega_{k'} \omega_{k+k'})^{-1/2} \left(a_k + a_{-k}^\dagger\right)\left(a_{k'} + a_{-k'}^\dagger\right)\left(a_{-(k+k')} + a_{k+k'}^\dagger\right). \end{aligned} \tag{2.158}$$

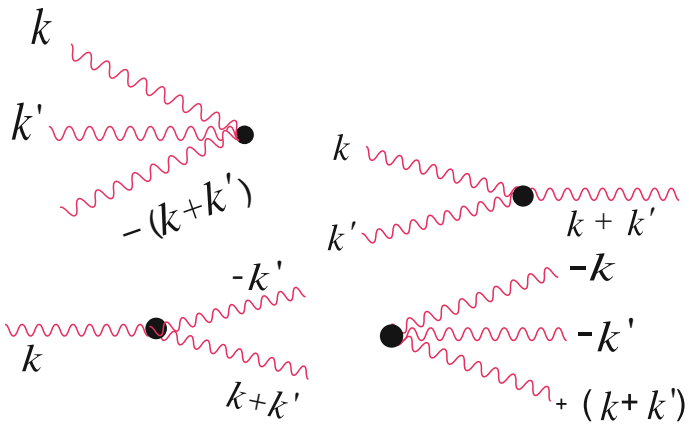


Fig. 2.17. Scattering of phonons: (**a**) annihilation of three phonons, (**b**) annihilation of two phonons and creation of a third phonon, (**c**) annihilation of one phonon and creation of two phonons, (**d**) creation of three phonons

Now define

$$G(k,k') = \frac{1}{3!}\sum_{gj} d(g,j)\mathrm{e}^{\mathrm{i}kja}\mathrm{e}^{\mathrm{i}k'(j-g)a}\left(\frac{\hbar^3}{2^3M^3N\omega_k\omega_{k'}\omega_{k+k'}}\right)^{1/2}. \quad (2.159)$$

Then, H_3' is simply

$$H_3' = \sum_{kk'} G(k,k')\left(a_k + a_{-k}^\dagger\right)\left(a_{k'} + a_{-k'}^\dagger\right)\left(a_{-(k+k')} + a_{k+k'}^\dagger\right). \quad (2.160)$$

Feynman Diagrams

In keeping track of the results obtained by applying H' to a state of the harmonic crystal, it is useful to use *Feynman diagrams*. A wavy line will represent a phonon propagating to the right (time increases to the right). The interaction (i.e., the result of applying H_3') is represented by a point into (or out of) which three wavy lines run. There are four fundamentally different kinds of diagrams (see Fig. 2.17):

1. $a_k a_{k'} a_{-(k+k')}$ annihilates three phonons (Fig. 2.17a).
2. $a_k a_{k'} a_{k+k'}^\dagger$ annihilates two phonons and creates a third phonon (Fig. 2.17b).
3. $a_k a_{-k'}^\dagger a_{k+k'}^\dagger$ annihilates a phonon but creates two phonons (Fig. 2.17c).
4. $a_{-k}^\dagger a_{-k'}^\dagger a_{k+k'}^\dagger$ creates three phonons (Fig. 2.17d).

Due to the existence of anharmonic terms (cubic, quartic, etc. in the displacements from equilibrium) the simple harmonic oscillators which describe

the normal modes in the harmonic approximation are coupled. This anharmonicity leads to a number of interesting results (e.g., thermal expansion, phonon–phonon scattering, phonon lifetime, etc.) We will not have space to take up these effects in this book. However, one should be aware that the harmonic approximation is an approximation. It ignores all the interesting effects resulting from anharmonicity.

2.9 Thermal Conductivity of an Insulator

When one part of a crystal is heated, a temperature gradient is set up. In the presence of the temperature gradient heat will flow from the hotter to the cooler region. The ratio of this heat current density to the magnitude of the temperature gradient is called the *thermal conductivity* κ_{T}.

In an insulating crystal (i.e., one whose electrical conductivity is very small at low temperatures as a result of the absence of nearly free electrons) the heat is transported by phonons. Let us define $u(x)$ as the internal energy per unit volume in a small region about the position x in the crystal. We assume that $u(x)$ depends on position because there is a temperature gradient $\frac{\partial T}{\partial x}$ in the x-direction. Because the temperature T depends on x, the local thermal equilibrium phonon density $\bar{n}_{q\lambda} = \left[\mathrm{e}^{\hbar\omega_{q\lambda}/\Theta} - 1\right]^{-1}$ will also depend on x. This takes a little explanation. In our discussion of phonons up until now, a phonon of wave vector $\mathbf{k}$ was not localized anywhere in the crystal. In fact, all of the atoms in the crystal vibrated with an amplitude u_k and different phases $\mathrm{e}^{\mathrm{i}kna-\mathrm{i}\omega_k t}$. In light of this, a phonon is everywhere in the crystal, and it seems difficult to think about difference in phonon density at different positions. In order to do so, we must construct wave packets with a spread in k values, Δk, chosen such that $(\Delta k)^{-1}$ is much larger than the atomic spacing but much smaller than the distance Δx over which the temperature changes appreciably. Then, by a phonon of wavenumber k we will mean a wavepacket centered at wavenumber k. The wavepacket can then be localized to a region Δx of the order $(\Delta k)^{-1}$. If the temperature at position x is different from that at some other position, the phonon will transport energy from the warmer to the cooler region. The thermal current density at position x can be written

$$j_{\mathrm{T}}(x) = \int \frac{\mathrm{d}\Omega}{4\pi} s\, \cos\theta\, u(x - l\cos\theta). \tag{2.161}$$

In this equation $u(x)$ is the internal energy per unit volume at position x, s is the sound velocity, l is the phonon mean free path ($l = s\tau$, where τ is the average time between phonon collisions), and θ is the angle between the direction of propagation of the phonon and the direction of the temperature gradient (see Fig. 2.18). A phonon reaching position x at angle θ (as shown in Fig. 2.18) had its last collision, on the average, at $x' = x - l\cos\theta$. But the phonons carry internal energy characteristic of the position where they had

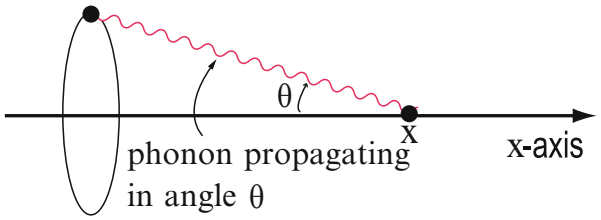


Fig. 2.18. Phonon propagation in the presence of a temperature gradient in the x-direction

their last collision, so such phonons carry internal energy $u(x-l\cos\theta)$. We can expand $u(x-l\cos\theta)$ as $u(x)-\frac{\partial u}{\partial x}\,l\cos\theta$, and integrate over $\mathrm{d}\Omega = 2\pi\sin\theta d\theta$. This gives the result

$$j_{\mathrm{T}}(x) = -\frac{1}{3}sl\frac{\partial u}{\partial x}. \tag{2.162}$$

Of course the internal energy depends on x because of the temperature gradient, so we can write $\frac{\partial u}{\partial x} = \frac{\partial u}{\partial T}\frac{\partial T}{\partial x}$. The result for the *thermal conductivity* $\kappa_{\mathrm{T}} = -j_{\mathrm{T}}\left(\frac{\partial T}{\partial x}\right)^{-1}$ is

$$\kappa_{\mathrm{T}} = \frac{1}{3}s^2\tau C_{\mathrm{v}}. \tag{2.163}$$

In (2.163), we have set $l = s\tau$ and $\frac{\partial u}{\partial T} = C_{\mathrm{v}}$, the specific heat of the solid.

2.10 Phonon Collision Rate

The collision rate τ^{-1} of phonons depends on

1. Anharmonic effects which cause phonon–phonon scattering
2. Defects and impurities which can scatter phonons and
3. The surfaces of the crystal which can also scatter phonons

Only the phonon–phonon collisions are very sensitive to temperature, since the phonon density available to scatter one phonon varies with temperature. For a perfect infinite crystal, defects, impurities, and surfaces can be ignored.

Phonon-phonon scattering can degrade the thermal current, but at very low temperature, where only low frequency ($\omega \ll \omega_{\mathrm{D}}$ or $k \ll k_{\mathrm{D}}$) phonons are excited, most phonon–phonon scattering conserves *crystal momentum*. By this we mean that in the real scattering processes shown in Fig. 2.19, no reciprocal lattice vector $\mathbf{K}$ is needed in the conservation of crystal momentum, and Fig. 2.19a would contain a delta function $\delta(\boldsymbol{k}_1+\boldsymbol{k}_2-\boldsymbol{k}_3)$, Fig. 2.19b a $\delta(\boldsymbol{k}_1-\boldsymbol{k}_2-\boldsymbol{k}_3)$, and Fig. 2.19c a $\delta(\boldsymbol{k}_1+k_2-\boldsymbol{k}_3-\boldsymbol{k}_4)$. This occurs because each k-value is very small compared to the smallest reciprocal lattice vector $\mathbf{K}$. These scattering processes are called *N-processes* (for normal scattering processes), and they do not degrade the thermal current.

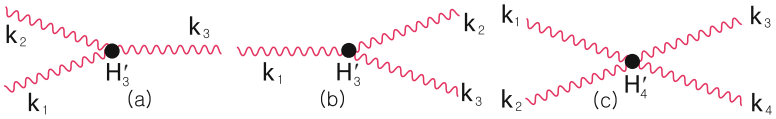


Fig. 2.19. Phonon–phonon scattering (**a**) Scattering of two phonons into one phonon, (**b**) Scattering of one phonon into two phonons, (**c**) Scattering of two phonons into two phonons

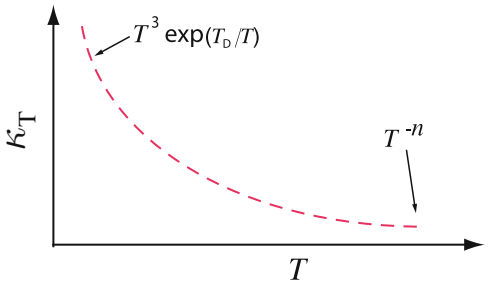


Fig. 2.20. Temperature dependence of the thermal conductivity of an insulator

At high temperatures phonons with k values close to a reciprocal lattice vector K will be thermally excited. In this case, the sum of $\mathbf{k}_1$ and $\mathbf{k}_2$ in Fig. 2.19a might be outside the first Brillouin zone so that $\mathbf{k}_3 = \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{K}$. It turns out that these processes, *U-processes* (for Umklapp processes) do degrade the thermal current. At high temperatures it is found that τ is proportional to temperature to the $-n$ power, where $1 \leq n \leq 2$. The high temperature specific heat is the constant Dulong–Petit value, so that according to (2.163) $\kappa_T \propto T^{-n}$ at high temperature.

At low temperature, only U-processes limit the thermal conductivity (or contribute to the thermal resistivity). But few phonons with $k \approx k_D$ are present at low temperature. A rough estimate would give $e^{-\hbar\omega_D/\Theta}$ for the probability of U-scattering at low temperature. Therefore, τ_U, the scattering time for U-processes is proportional to $e^{\Theta_D/\Theta}$. Since the low temperature specific heat varies as T^3, (2.163) would predict $\kappa_T \propto T^3 e^{T_D/T}$ for the thermal conductivity at low temperature. The result for the temperature dependence of thermal conductivity of an insulator is sketched in Fig. 2.20.

2.11 Phonon Gas

Landau introduced the concept of thinking of elementary excitations as particles. He suggested that it was possible to have a gas of phonons in a crystal whose properties were analogous to those of a classical gas. Both the atoms or molecules of a classical gas and the phonons in a crystal undergo collisions. For the former, the collisions are molecule–molecule collisions or molecule–wall of container collisions. For the latter they are phonon–phonon,

phonon–imperfection or phonon–surface collisions. Energy is conserved in these collisions. Momentum is conserved in molecule–molecule collisions in a classical gas and in N-process phonon–phonon collisions in a *phonon gas*. Of course, the number of particles is conserved in the molecule–molecule collisions of a classical gas, but phonons can be created or annihilated in phonon–phonon collisions, so their number is not a conserved quantity.

The sound waves of a classical gas are oscillations of the particle density. They occur if $\omega\tau \ll 1$, so that thermal equilibrium is established very quickly compared to the period of the sound wave. They also require that momentum be conserved in the collision process.

Landau[4] called normal sound waves in a gas *first sound*. He proposed an oscillation of the phonon density in a phonon gas that named *second sound*. This oscillation of the phonon density (or energy density) occurred in a crystal if $\omega\tau_{\mathrm{N}} \ll 1$ (as in first sound) but $\omega\tau_{\mathrm{U}} \gg 1$ so that crystal momentum is conserved. Second sound has been observed in He^4 and in a few crystals.

[4] L. Landau, J. Phys. U.S.S.R. **5**, 71 (1941).

Problems

2.1. Consider a three-dimensional Einstein model in which each degree of freedom of each atom has a vibrational frequency ω_0.

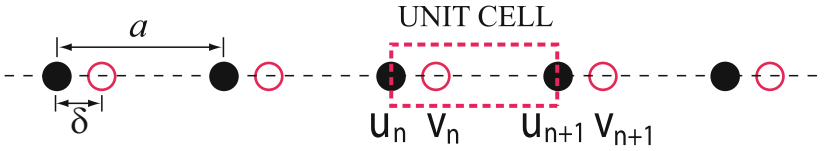
(a) Evaluate $G(\omega)$, the number of modes per unit volume whose frequency is less than ω.
(b) Evaluate $g(\omega) = \frac{\mathrm{d}G(\omega)}{\mathrm{d}\omega}$.
(c) Make a rough sketch of both $G(\omega)$ and $g(\omega)$ as a function of ω.

2.2. For a one-dimensional lattice a phonon of wave number k has frequency $\omega_k = \omega_0 \sin \frac{|k|a}{2}$ for a nearest neighbor coupling model. Now approximate this model by a Debye model with $\omega = s|k|$.

(a) Determine the value of s, the sound speed, and k_D, the Debye wave vector.
(b) Sketch ω as a function of k for each model over the entire Brillouin zone.
(c) Evaluate $g(\omega)$ for each model and make a sketch of $g(\omega)$ vs. ω for each.

2.3. Consider a diatomic linear chain. Evaluate u_q/v_q for the acoustic and optical modes at $q = 0$ and at $q = \frac{\pi}{2a}$.

2.4. Consider a linear chain with two atoms per unit cell (each of mass M) located at 0 and δ, where $\delta < \frac{a}{2}$, a being the primitive translation vector. Let C_1 be the force constant between nearest neighbors and C_2 the force constant between next nearest neighbors. Determine $\omega_\pm(k = 0)$ and $\omega_\pm(k = \frac{\pi}{a})$.



2.5. Show that the normal mode density (for samll ω) in a d-dimensional harmonic crystal varies as ω^{d-1}. Use this result to determine the temperature dependence of the specific heat.

2.6. In a linear chain with nearest neighbor interactions $\omega_k = \omega_0 \sin \frac{|k|a}{2}$. Show that $g(\omega) \simeq \left(\frac{2}{\pi a}\right) \frac{1}{\sqrt{\omega_0^2 - \omega^2}}$.

2.7. For a certain three-dimensional simple cubic lattice the phonon spectrum is independent of polarization λ and is given by

$$\omega(k_x, k_y, k_z) = \omega_0 \left[\sin^2\left(\frac{k_x a}{2}\right) + \sin^2\left(\frac{k_y a}{2}\right) + \sin^2\left(\frac{k_z a}{2}\right) \right]^{1/2} .$$

(a) Sketch a graph of ω vs. k for
   1. $k_y = k_z = 0$ and $0 \leq k_x \leq \frac{\pi}{a}$ (i.e., along $\Gamma \rightarrow$ X),
   2. $k_z = 0$ and $k_x = k_y = \frac{k}{\sqrt{2}}$ for $0 \leq k \leq \frac{\sqrt{2}\pi}{a}$ (i.e., along $\Gamma \rightarrow$ K),
   3. $k_x = k_y = k_z = \frac{k}{\sqrt{3}}$ for $0 \leq k \leq \frac{\sqrt{3}\pi}{a}$ (i.e., along $\Gamma \rightarrow$ L).

(b) Draw the ω vs. k curve for the Debye approximation to these dispersion curves as dashes lines on the diagram used in part (a).

(c) What are the critical points of this phonon spectrum? How many are there?

(d) Make a rough sketch of the Debye density of states $g(\omega)$. How will the actual density of states differ from the Debye approximation?

(e) Using this example, discuss the shortcomings and the successes of the Debye model in predicting the thermodynamic properties (like specific heat) of solids.

2.8. For a two-dimensional crystal a simple Debye model takes $\omega = sk$ for the longitudinal and the single transverse modes for all allowed k values up to the Debye wave number k_D.

(a) Determine k_D as a function of $\frac{N}{L^2}$, where N is the number of atoms and L^2 is the area of the crystal.

(b) Determine $g(\omega)$, the density of normal modes per unit area.

(c) Find the expression for the internal energy at a temperature T as an integral over the density of states times an appropriate function of frequency and temperature.

(d) From the result of part (c) determine the specific heat c_v.

(e) Evaluate c_v for $k_\mathrm{B}T \ll \hbar\omega_\mathrm{D} = \hbar s k_\mathrm{D}$.

Summary

In this chapter, we discussed the vibrations of the atoms in solids. Quantum mechanical treatment of lattice dynamics and dispersion curves of the normal modes are described.

The Hamiltonian of a linear chain is written, in the harmonic approximation, as $H = \sum_i \frac{P_i^2}{2M} + \frac{1}{2}\sum_{i,j} c_{ij} u_i u_j$, where P_i is the momentum and $u_i = R_i - R_i^0$ is the deviation of the ith atom from its equilibrium position. A general dispersion relation of the normal modes is $M\omega_q^2 = \sum_{l=1}^{N} c(l)\mathrm{e}^{\mathrm{i}qla}$. The normal coordinates are given by

$$q_k = N^{-1/2}\sum_n u_n \mathrm{e}^{-\mathrm{i}kna};\ p_k = N^{-1/2}\sum_n P_n \mathrm{e}^{+\mathrm{i}kna}.$$

The inverse of q_k and p_k are $u_n = N^{-1/2}\sum_k q_k \mathrm{e}^{\mathrm{i}kna};\ P_n = N^{-1/2}\sum_k p_k \mathrm{e}^{-\mathrm{i}kna}$.

The quantum mechanical Hamiltonian is given by $H = \sum_k H_k$, where

$$H_k = \frac{\hat{p}_k \hat{p}_k^\dagger}{2M} + \frac{1}{2}M\omega_k^2 \hat{q}_k \hat{q}_k^\dagger.$$

The dynamical variables q_k and p_k are replaced by quantum mechanical operators $\hat{q}_k$ and $\hat{p}_k$ which satisfy the commutation relation $[p_k,\ q_{k'}] = -\mathrm{i}\hbar\delta_{k,k'}$. It is convenient to rewrite $\hat{q}_k$ and $\hat{p}_k$ in terms of the operators a_k and $a_k^\dagger$, which are defined by

$$q_k = \left(\frac{\hbar}{2M\omega_k}\right)^{1/2}\left(a_k + a_{-k}^\dagger\right);\ p_k = \imath\left(\frac{\hbar M\omega_k}{2}\right)^{1/2}\left(a_k^\dagger - a_{-k}\right).$$

The a_k's and $a_k^\dagger$'s satisfy $\left[a_k,\ a_{k'}^\dagger\right]_- = \delta_{k,k'}$ and $[a_k,\ a_{k'}]_- = \left[a_k^\dagger,\ a_{k'}^\dagger\right]_- = 0$. The displacement of the nth atom and its momentum can be written

$$u_n = \sum_k \left(\tfrac{\hbar}{2MN\omega_k}\right)^{1/2} \mathrm{e}^{\mathrm{i}kna}\left(a_k + a_{-k}^\dagger\right),$$
$$P_n = \sum_k \imath\left(\tfrac{\hbar\omega_k M}{2N}\right)^{1/2} \mathrm{e}^{-\mathrm{i}kna}\left(a_k^\dagger - a_{-k}\right).$$

The Hamiltonian of the linear chain of atoms can be written

$$H = \sum_k \hbar\omega_k\left(a_k^\dagger a_k + \frac{1}{2}\right),$$

and its eigenfunctions and eigenvalues are

$$|n_1, n_2, \ldots, n_N> = \frac{\left(a_{k_1}^\dagger\right)^{n_1}}{\sqrt{n_1!}} \cdots \frac{\left(a_{k_N}^\dagger\right)^{n_N}}{\sqrt{n_N!}}|0>$$

and $E_{n_1,n_2,\ldots,n_N} = \sum_i \hbar\omega_{k_i}(n_i + \frac{1}{2})$.

In the three-dimensional case, the Hamiltonian is given by

$$H = \sum_{\mathbf{k}\lambda} \hbar\omega_{\mathbf{k}\lambda} \left(a^{\dagger}_{\mathbf{k}\lambda} a_{\mathbf{k}\lambda} + \frac{1}{2} \right).$$

The allowed values of $\mathbf{k}$ are given by $\mathbf{k} = 2\pi \left(\frac{n_1}{N_1}\mathbf{b}_1 + \frac{n_2}{N_2}\mathbf{b}_2 + \frac{n_3}{N_3}\mathbf{b}_3 \right)$. The displacement $\mathbf{u}_n$ and momentum $\mathbf{P}_n$ of the nth atom are written, respectively, as

$$\mathbf{u}_n = \sum_{\mathbf{k}\lambda} \left(\frac{\hbar}{2MN\omega_{\mathbf{k}\lambda}} \right)^{1/2} \hat{\varepsilon}_{\mathbf{k}\lambda} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}^0_n} \left(a_{\mathbf{k}\lambda} - a^{\dagger}_{-\mathbf{k}\lambda} \right)$$

$$\mathbf{P}_n = \sum_{\mathbf{k}\lambda} \mathrm{i} \left(\frac{\hbar M\omega_{\mathbf{k}\lambda}}{2N} \right)^{1/2} \hat{\varepsilon}_{\mathbf{k}\lambda} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}^0_n} \left(a^{\dagger}_{\mathbf{k}\lambda} + a_{-\mathbf{k}\lambda} \right).$$

The energy of the crystal is given by $U = \sum_{\mathbf{q}\lambda} \left(\bar{n}_{\mathbf{q}\lambda} + \frac{1}{2} \right) \hbar\omega_{\mathbf{q}\lambda}$, where $\bar{n}_{\mathbf{q}\lambda}$ is given by $\bar{n}_{\mathbf{q}\lambda} = \frac{1}{\mathrm{e}^{\hbar\omega_{\mathbf{q}\lambda}/k_{\mathrm{B}}T} - 1}$. The lattice heat capacity is written as

$$C_{\mathrm{v}} = \left(\frac{\partial U}{\partial T} \right)_{\mathrm{v}} = k_{\mathrm{B}} \sum_{\mathbf{q}\lambda} \left(\frac{\hbar\omega_{\mathbf{q}\lambda}}{k_{\mathrm{B}}T} \right)^2 \left(\mathrm{e}^{\frac{\hbar\omega_{\mathbf{q}\lambda}}{k_{\mathrm{B}}T}} - 1 \right)^{-1} \left(1 - \mathrm{e}^{-\frac{\hbar\omega_{\mathbf{q}\lambda}}{k_{\mathrm{B}}T}} \right)^{-1}.$$

The *density of states* $g(\omega)$ defined by

$$g(\omega)\mathrm{d}\omega = \begin{cases} \text{The number of normal modes per unit volume} \\ \text{whose frequency } \omega_{\mathbf{q}\lambda} \text{ satisfies } \omega < \omega_{\mathbf{q}\lambda} < \omega + \mathrm{d}\omega. \end{cases}$$

Then we have $g(\omega) = \frac{1}{(2\pi)^3} \sum_{\lambda} \int \frac{\mathrm{d}S_{\lambda}(\omega)}{|\nabla_{\mathbf{q}}\omega_{\mathbf{q}\lambda}|_{\omega}}$. Here, $\mathrm{d}S_{\lambda}(\omega)$ is an infinitesimal element of the surface of constant frequency in three-dimensional wave vector space on which $\omega_{\mathbf{q}\lambda}$ has the value ω. Near a critical point $\mathbf{q}_c$, at which $\nabla_{\mathbf{q}}\omega_{\mathbf{q}\lambda} = 0$, in the phonon spectrum, we can write

$$\omega_q = \omega_c + \alpha_1\xi_1^2 + \alpha_2\xi_2^2 + \alpha_3\xi_3^2,$$

where ξ_i are the components of $\xi = \mathbf{q} - \mathbf{q}_c$, and $\omega_c = \omega(q_c)$. In three dimensions, there are four kinds of critical points:

1. *Maxima*: points at which all three α_i are negative.
2. *Minima*: points at which all three α_i are positive.
3. *Saddle Points of the First Kind*: Points at which two α_i's are positive and one is negative.
4. *Saddle Points of the Second Kind*: Points at which one α_i is positive and the other two are negative.

The density of states for the Debye model is expressed as

$$g(\omega) = \frac{\omega^2}{2\pi^2} \left[\frac{\theta(s_l k_{\mathrm{D}} - \omega)}{s_l^3} + \frac{2\,\theta(s_t k_{\mathrm{D}} - \omega)}{s_t^3} \right].$$

Here, s_l and s_t are the speed of a longitudinal and of a transverse sound wave.

3

Free Electron Theory of Metals

3.1 Drude Model

The most important characteristic of a metal is its high electrical conductivity. Around 1900, shortly after J.J. Thompson's discovery of the electron, people became interested in understanding more about the mechanism of metallic conduction. The first work by E. Riecke in 1898 was quickly superseded by that of Drude in 1900. Drude[1] proposed an exceedingly simple model that explained a well-known empirical law, the Wiedemann–Franz law (1853). This law states that at a given temperature the ratio of the thermal conductivity to the electrical conductivity is the same for all metals. The assumptions of the *Drude model* are:

1. A metal contains free electrons which form an electron gas.
2. the electrons have some average thermal energy $\left\langle \frac{1}{2} m v_{\mathrm{T}}^2 \right\rangle$, but they pursue random motions through the metal so that $\langle \mathbf{v}_{\mathrm{T}} \rangle = 0$ even though $\left\langle v_{\mathrm{T}}^2 \right\rangle \neq 0$. The random motions result from collisions with the ions.
3. Because the ions have a very large mass, they are essentially immovable.

3.2 Electrical Conductivity

In the presence of an electric field $\mathbf{E}$, the electrons acquire a drift velocity $\mathbf{v}_{\mathrm{D}}$ which is superimposed on the thermal motion. Drude assumed that the probability that an electron collides with an ion during a time interval $\mathrm{d}t$ is simply proportional to $\frac{\mathrm{d}t}{\tau}$, where τ is called the *collision time* or *relaxation time*. Then Newton's law gives

$$m \left(\frac{\mathrm{d}\mathbf{v}_{\mathrm{D}}}{\mathrm{d}t} + \frac{\mathbf{v}_{\mathrm{D}}}{\tau} \right) = -e\mathbf{E}, \tag{3.1}$$

[1] P. Drude, Annalen der Physik **1**, 566 (1900); ibid., **3**, 369 (1900); ibid., **7**, 687 (1902).

where $-e$ is the charge on an electron. Some appreciation of the term of relaxation time $\frac{\mathrm{d}t}{\tau}$ can be obtained by assuming that the system acquires a drift velocity $\mathbf{v}_\mathrm{D}$ in the presence of an electric field $\mathbf{E}$, and then, at time $t = 0$, the electric field is turned off. The behavior of $\mathbf{v}_\mathrm{D}(t)$ as a function of time is given by

$$\mathbf{v}_\mathrm{D}(t) = \mathbf{v}_\mathrm{D}(0)\mathrm{e}^{-t/\tau}, \tag{3.2}$$

showing that $\mathbf{v}_\mathrm{D}$ relaxes from $\mathbf{v}_\mathrm{D}(0)$ toward zero with a relaxation time τ.

In the steady-state (where $\dot{\mathbf{v}}_\mathrm{D} = 0$), $\mathbf{v}_\mathrm{D}$ is given by

$$\mathbf{v}_\mathrm{D} = -\frac{e\mathbf{E}\tau}{m}. \tag{3.3}$$

The quantity $\frac{e\tau}{m}$, the drift velocity per unit electric field, is called μ, the *drift mobility*. The velocity of an electron including both thermal and drift components is

$$\mathbf{v} = \mathbf{v}_\mathrm{T} - \frac{e\tau\mathbf{E}}{m}. \tag{3.4}$$

The current density caused by the electric field $\mathbf{E}$ is simply

$$\mathbf{j} = V^{-1} \sum_{\substack{\text{all} \\ \text{electrons}}} (-e)\mathbf{v}. \tag{3.5}$$

But $\sum_{\substack{\text{all} \\ \text{electrons}}} \mathbf{v}_\mathrm{T} = 0$, so that

$$\mathbf{j} = V^{-1}N(-e)\left(-\frac{e\tau\mathbf{E}}{m}\right) = \sigma\mathbf{E}. \tag{3.6}$$

Here σ, the *electrical conductivity*, is equal to $\frac{n_0 e^2 \tau}{m}$ where $n_0 = \frac{N}{V}$ is the electron concentration.

3.3 Thermal Conductivity

The thermal conductivity is the ratio of the thermal current (i.e., the energy current) to the magnitude of the temperature gradient. In the presence of a temperature gradient $\frac{\partial T}{\partial x}$, the average thermal energy $\left\langle \frac{1}{2}mv_\mathrm{T}^2 \right\rangle$ will depend on the local temperature $T(x)$. The electrons sense the local temperature through collisions with the lattice. Thus, the thermal energy of a given electron will depend on where it made its last collision. If we choose an electron at random, the mean time back to its last collision is τ. Therefore, an electron crossing the plane $x = x_0$ at an angle θ to the x-axis had its last collision at $x = x_0 - v_\mathrm{T}\tau\cos\theta$. (See Fig. 3.1.) The energy of such an electron is $E(x) = E\left(x_0 - v_\mathrm{T}\tau\cos\theta\right)$. The number of electrons per unit volume whose direction of motion is in the solid angle $\mathrm{d}\Omega$ is simply $n_0\frac{\mathrm{d}\Omega}{4\pi}$. (See Fig. 3.2.)

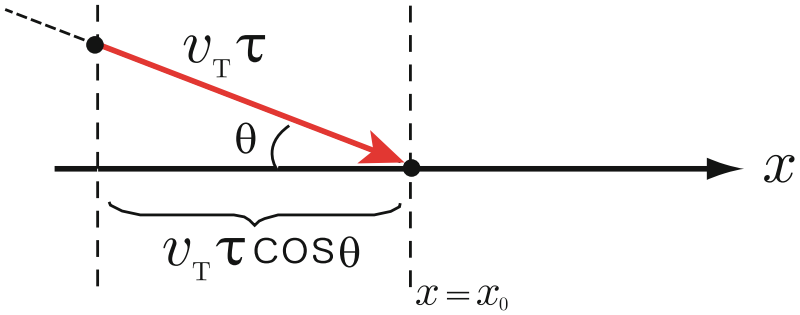


Fig. 3.1. An electron crossing the plane $x = x_0$ at an angle θ to the x-axis

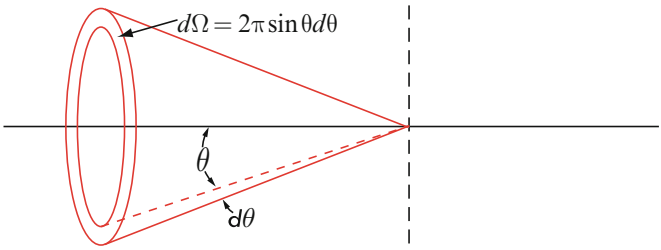


Fig. 3.2. Solid angle dΩ in which electrons moving to cross the plane $x = x_0$ at an angle θ to the x-axis

The number of such electrons crossing a unit area at x_0 is $n_0 \frac{\mathrm{d}\Omega}{4\pi} v_\mathrm{T} \cos\theta$, giving for the energy flux through a unit area at x_0

$$w(x_0) = \int E\left(x_0 - v_\mathrm{T}\tau\cos\theta\right) n_0 v_\mathrm{T} \cos\theta \frac{\mathrm{d}\Omega}{4\pi}. \tag{3.7}$$

Just as we did for the thermal conductivity due to phonons we expand $E(x_0 - v_\mathrm{T}\tau\cos\theta)$ and perform the integral over θ from 0 to π. This gives

$$w(x) = -\frac{1}{3} n_0 v_\mathrm{T}^2 \tau \left(\frac{\partial E}{\partial x}\right). \tag{3.8}$$

But $\frac{\partial E}{\partial x} = \frac{\partial E}{\partial T}\frac{\partial T}{\partial x}$, so the thermal conductivity κ is given by

$$\kappa = \frac{w}{-\partial T/\partial x} = \frac{1}{3} n_0 v_\mathrm{T}^2 \tau \frac{\mathrm{d}E}{\mathrm{d}T} = \frac{1}{3} v_\mathrm{T}^2 \tau C_\mathrm{v}, \tag{3.9}$$

where $C_\mathrm{v} = n_0 \frac{\mathrm{d}E}{\mathrm{d}T}$ is the heat capacity per unit volume (or the specific heat).

3.4 Wiedemann–Franz Law

The ratio of κ to σ is given by

$$\frac{\kappa}{\sigma} = \frac{\frac{1}{3} v_{\mathrm{T}}^2 \tau C_{\mathrm{v}}}{\frac{n_0 e^2 \tau}{m}}. \tag{3.10}$$

Now Drude applied the classical gas laws to evaluation of v_{T}^2 and C_{v}, viz., $\left\langle \frac{1}{2} m v_{\mathrm{T}}^2 \right\rangle = \frac{3}{2} k_{\mathrm{B}} T$ and $C_{\mathrm{v}} = n_0 \left(\frac{3}{2}\right) k_{\mathrm{B}}$. This gave

$$\frac{\kappa}{\sigma} = \frac{3}{2} \left(\frac{k_{\mathrm{B}}}{e} \right)^2 T. \tag{3.11}$$

In addition to agreeing with the Wiedemann–Franz law, the ratio $\mathcal{L} = \frac{\kappa}{\sigma T}$ had the value $\frac{3}{2} \left(\frac{k_{\mathrm{B}}}{e}\right)^2$ which was equal to 1.24×10^{-13} esu. The observed values for $\mathcal{L}$, called the *Lorenz number*[2], averaged to roughly 2.72×10^{-13} esu. Drude made an error of a factor of 2 in his original paper and found that $\mathcal{L} \sim 2.48 \times 10^{-13}$ esu, remarkably close to the experimental value.

3.5 Criticisms of Drude Model

1. If $\left\langle \frac{1}{2} m v_{\mathrm{T}}^2 \right\rangle = \frac{3}{2} k_{\mathrm{B}} T$, then the electronic contribution to C_{v} had to be $C_{\mathrm{v}} = \frac{3}{2} N k_{\mathrm{B}} = \frac{3}{2} R$. This is half as big as the lattice contribution and was simply not observed.
2. Experimentally σ varies as T^{-1}. This implies that $n_0 \tau \propto T^{-1}$ since e^2 and m are constants. In Drude's picture, the mean free path $l \simeq v_{\mathrm{T}} \tau$ was thought to be of the order of the atomic spacing and therefore independent of T. Since $v_{\mathrm{T}} \propto T^{1/2}$ this would imply that $\tau \propto T^{-1/2}$ and, to satisfy $n_0 \tau \propto T^{-1}$, that $n_0 \propto T^{-1/2}$. This did not make any sense.

3.6 Lorentz Theory

Since Drude's simple model gave some results that agree fairly well with experiment, Lorentz[3] decided to use the full apparatus of kinetic theory to investigate the model more carefully. He did not succeed in improving on Drude's model, but he did make use of the Boltzmann distribution function and Boltzmann equation which we would like to describe.

[2] Ludvig Valentin Lorenz (1829–1891).
[3] Hendrik Antoon Lorentz (1853–1928).

3.6.1 Boltzmann Distribution Function

The Boltzmann distribution function $f(\mathbf{v}, \mathbf{r}, t)$ is defined by

$f(\mathbf{v}, \mathbf{r}, t)\mathrm{d}^3r\mathrm{d}^3v =$ the number of electrons in the volume element d^3r centered at $\mathbf{r}$ whose velocity is between $\mathbf{v}$ and $\mathbf{v} + \mathrm{d}\mathbf{v}$ at time t.

Boltzmann's equation says that the total time rate of change of $f(\mathbf{v}, \mathbf{r}, t)$ must be balanced by its time rate of change due to collisions, i.e.,

$$\frac{\mathrm{d}f(\mathbf{v}, \mathbf{r}, t)}{\mathrm{d}t} = \left(\frac{\partial f}{\partial t}\right)_{\mathrm{c}}. \tag{3.12}$$

Here, $\left(\frac{\partial f}{\partial t}\right)_{\mathrm{c}} \mathrm{d}^3r\,\mathrm{d}^3v\,\mathrm{d}t$ is the net number of electrons forced into the volume element $\mathrm{d}^3r\mathrm{d}^3v$ (in phase space) by collisions in the time interval $\mathrm{d}t$.

3.6.2 Relaxation Time Approximation

The simplest form of the collision term is

$$\left(\frac{\partial f}{\partial t}\right)_{\mathrm{c}} = -\frac{f - f_0}{\tau}, \tag{3.13}$$

where f_0 is the thermal equilibrium distribution function, f the actual nonequilibrium distribution function (which differs from f_0 due to some external disturbance), and τ is a relaxation time. Once again if $f - f_0$ is nonzero due to some external disturbance, and if at time $t = 0$ the disturbance is turned off, one can simply write

$$(f - f_0)_t = (f - f_0)_{t=0}\,\mathrm{e}^{-t/\tau}. \tag{3.14}$$

3.6.3 Solution of Boltzmann Equation

We are frequently interested in small perturbations away from equilibrium and can linearize the Boltzmann equation. For example, suppose the external perturbation is a small electric field $\mathbf{E}$ in the x-direction, and a temperature gradient $\frac{\partial T}{\partial x}$. The steady-state Boltzmann equation ($\frac{\partial f}{\partial t} = 0$) is

$$\frac{\partial f}{\partial v_x}\left(-\frac{eE}{m}\right) + \frac{\partial f}{\partial x}v_x = -\frac{f - f_0}{\tau}. \tag{3.15}$$

If $f - f_0$ is small we can approximate f on the left-hand side by f_0 and obtain

$$f \simeq f_0 + \tau\left[\frac{eE}{m}\frac{\partial f_0}{\partial v_x} - v_x\frac{\partial f_0}{\partial x}\right]. \tag{3.16}$$

This is *linear response* since E and $\frac{\partial f_0}{\partial x}$ are already linear in E or $\frac{\partial T}{\partial x}$. The electrical current density and thermal current density are given, respectively, by

$$\mathbf{j}(\mathbf{r},t) = \int (-e)\mathbf{v}\, f(\mathbf{r},\mathbf{v},t)\, \mathrm{d}^3 v, \tag{3.17}$$

and

$$\mathbf{w}(\mathbf{r},t) = \int \varepsilon \mathbf{v}\, f(\mathbf{r},\mathbf{v},t)\, \mathrm{d}^3 v. \tag{3.18}$$

In (3.18) $\varepsilon = \frac{1}{2}mv^2$ is the kinetic energy of the electron of velocity $\mathbf{v}$. We substitute the solution for f given by (3.16) into (3.17) and (3.18) to calculate $\mathbf{j}$ and $\mathbf{w}$.

3.6.4 Maxwell–Boltzmann Distribution

To evaluate $\mathbf{j}$ and $\mathbf{w}$ it is necessary to know $f_0(v)$. Lorentz used the following expression:

$$f_0(v) = n_0 \left(\frac{m}{2\pi\Theta}\right)^{3/2} \mathrm{e}^{-\varepsilon/\Theta}. \tag{3.19}$$

Here $n_0 = N/V$, $\Theta = k_\mathrm{B}T$, and $\varepsilon = \frac{1}{2}mv^2$. The normalization constant has been chosen so that $\int f_0(v)\mathrm{d}^3 v = n_0$. The reader should check this. (Use $\int_0^\infty x^{1/2}\mathrm{e}^{-x}\mathrm{d}x = \Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$.)

The use of classical statistical mechanics and the Maxwell–Boltzmann distribution function is the source of the difficulty with the Lorentz theory. In 1925 Pauli[4] proposed the *exclusion principle*; in 1926 Fermi and Dirac[5] proposed the Fermi–Dirac statistics, and in 1928 Sommerfeld published the *Sommerfeld Theory of Metals*. The Sommerfeld theory was simply the Lorentz theory with the Fermi–Dirac distribution function replacing the Boltzmann–Maxwell distribution function.

3.7 Sommerfeld Theory of Metals

Sommerfeld[6] treated the Drude electron gas quantum mechanically. We can assume that the electron gas is contained in a cubic box of edge L, and that the potential inside the box is constant. The Schrödinger equation is

$$-\frac{\hbar^2}{2m}\nabla^2\Psi(\mathbf{r}) = E\Psi(\mathbf{r}), \tag{3.20}$$

[4] W. Pauli, Z. Physik **31**, 765 (1925).
[5] E. Fermi, Z. Physik **36**, 902 (1926); P.A.M. Dirac, Proc. Roy. Soc. London, A **112**, 661 (1926).
[6] A. Sommerfeld, Zeits. fur Physik **47**, 1 (1928).

and its solution is

$$\Psi_k(\mathbf{r}) = V^{-1/2}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \qquad (3.21)$$
$$E_{\mathbf{k}} = \tfrac{\hbar^2 k^2}{2m}.$$

To avoid difficulties with boundaries, we can assume periodic boundary conditions so that $x = 0$ and $x = L$ are the same point. Then the allowed values of k_x (and k_y and k_z) satisfy $k_x = \frac{2\pi}{L}\, n_x$, where $n_x = 0, \pm 1, \pm 2, \ldots$, and

$$E_{\mathbf{k}} = \frac{\hbar^2}{2m}\left(\frac{2\pi}{L}\right)^2 \left(n_x^2 + n_y^2 + n_z^2\right). \qquad (3.22)$$

The functions $|\mathbf{k}\rangle$ form a complete orthonormal set with

$$\langle \mathbf{k}|\mathbf{k}'\rangle = \int \mathrm{d}^3 r \Psi^*_{\mathbf{k}}(\mathbf{r})\Psi_{\mathbf{k}'}(\mathbf{r}) = \delta_{\mathbf{k}\mathbf{k}'}. \qquad (3.23)$$

$$\sum_{\mathbf{k}} |\mathbf{k}\rangle \langle \mathbf{k}| = 1 \text{ or } \sum_{\mathbf{k}} \Psi^*_{\mathbf{k}}(\mathbf{r}')\Psi_{\mathbf{k}}(\mathbf{r}) = \delta(\mathbf{r}' - \mathbf{r}). \qquad (3.24)$$

Fermi Energy

The Pauli principle states that only one electron can occupy a given quantum state. In the Sommerfeld model, states are labeled by $\{\mathbf{k}, \sigma\} = (k_x, k_y, k_z)$ and σ, where σ is a spin index which takes on the two values $\uparrow$ or $\downarrow$. At $T = 0$, only the lowest N energy states will be occupied by the N electrons in the system. Define k_{F} as the value of k for the highest energy occupied state. Then the number of particles is given by

$$N = \sum_{\substack{k < k_{\mathrm{F}} \\ \sigma}} 1 = \frac{V}{(2\pi)^3} 2 \int_{k<k_{\mathrm{F}}} \mathrm{d}^3 k. \qquad (3.25)$$

The factor of 2 comes from summing over spin. The integration simply gives $\frac{4}{3}\pi k_{\mathrm{F}}^3$, resulting in the relation

$$k_{\mathrm{F}}^3 = 3\pi^2 n_0. \qquad (3.26)$$

The *Fermi energy* $\varepsilon_{\mathrm{F}}(\equiv\Theta_{\mathrm{F}})$, *Fermi velocity* v_{F}, and *Fermi temperature* $T_{\mathrm{F}}\left(= \frac{\Theta_{\mathrm{F}}}{k_{\mathrm{B}}}\right)$ are defined, respectively, by

$$\varepsilon_{\mathrm{F}} = \frac{\hbar^2 k_{\mathrm{F}}^2}{2m} = \frac{1}{2} m v_{\mathrm{F}}^2 = \Theta_{\mathrm{F}}. \qquad (3.27)$$

For a typical metal, we have $n_0 = 10^{23}\,\mathrm{cm}^{-3}$ giving $\varepsilon_{\mathrm{F}} \simeq 5\,\mathrm{eV}$, $v_{\mathrm{F}} \simeq 10^8\,\mathrm{cm/s}$, and $T_{\mathrm{F}} \simeq 10^5\,\mathrm{K}$.

3.8 Review of Elementary Statistical Mechanics

Suppose that the states of an ideal Fermi gas are labeled $\phi_1, \phi_2, \ldots, \phi_i, \ldots$ and that they have energies $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i, \ldots$. Then, if N is the total number of Fermions

$$\sum_i n_i = N, \tag{3.28}$$

where $n_i = 1$ if the state ϕ_i is occupied and $n_i = 0$ if it is not. The *partition function* Z_N for this N particle system is defined by

$$Z_N = \sum_{\{n_i\}}{}' \mathrm{e}^{-\beta \sum_i n_i \varepsilon_i}. \tag{3.29}$$

In (3.29) $\beta = (k_\mathrm{B} T)^{-1}$ and the symbol $\sum'_{\{n_i\}}$ means a summation over all sets of values $\{n_i\} = \{n_1, n_2, \ldots, n_i, \ldots\}$ which satisfy the condition $\sum_i n_i = N$. This restriction makes performing the sum to obtain Z_N difficult. One can avoid this difficulty by using the *grand canonical ensemble* instead of the *canonical ensemble*. The *grand partition function* Q is defined by

$$Q = \sum_{N=0}^{\infty} \mathrm{e}^{\beta \zeta N} Z_N. \tag{3.30}$$

The symbol ζ is is called the *chemical potential*. When we substitute (3.29) into (3.30), the summation over N removes the restriction on the sets of values $\{n_i\}$ included in the sum appearing in (3.29). We can rewrite the grand partition function as follows:

$$\begin{aligned} Q &= \sum_{N=0}^{\infty} \sum_{\{n_i\}}{}' \mathrm{e}^{-\beta \sum_i n_i (\varepsilon_i - \zeta)} \\ &= \sum_{n_1=0}^{1} \sum_{n_2=0}^{1} \cdots \sum_{n_i=0}^{1} \cdots \mathrm{e}^{-\beta(\varepsilon_1 - \zeta) n_1} \mathrm{e}^{-\beta(\varepsilon_2 - \zeta) n_2} \cdots \mathrm{e}^{-\beta(\varepsilon_i - \zeta) n_i} \cdots . \end{aligned} \tag{3.31}$$

It is easy to see that

$$\sum_{n_i=0}^{1} \mathrm{e}^{-\beta(\varepsilon_i - \zeta) n_i} = 1 + \mathrm{e}^{-\beta(\varepsilon_i - \zeta)}$$

so that

$$Q = \prod_i \left[1 + \mathrm{e}^{-\beta(\varepsilon_i - \zeta)} \right]. \tag{3.32}$$

The average occupancy of some quantum state l is given by

$$\bar{n}_l = Q^{-1} \sum_{\{n_i\}} n_l \mathrm{e}^{-\beta \sum_i n_i (\varepsilon_i - \zeta)}. \tag{3.33}$$

For all $i \neq l$, the factor involving i in the numerator is exactly cancelled by the same factor in Q^{-1} leaving us

$$\bar{n}_l = \frac{\sum_{n_l} n_l \mathrm{e}^{-\beta(\varepsilon_l - \zeta)n_l}}{\sum_{n_l} \mathrm{e}^{-\beta(\varepsilon_l - \zeta)n_l}} = \frac{\mathrm{e}^{-\beta(\varepsilon_l - \zeta)}}{1 + \mathrm{e}^{-\beta(\varepsilon_l - \zeta)}}. \tag{3.34}$$

Thus we find the *Fermi–Dirac distribution function* of

$$\bar{n}_l = \frac{1}{\mathrm{e}^{(\varepsilon_l - \zeta)/\Theta} + 1}. \tag{3.35}$$

At $\Theta = 0$ all states whose energy is smaller than ε_F are occupied; all states of higher energy empty. Notice that (3.33) can be written

$$\bar{n}_l = -k_\mathrm{B} T \frac{\partial}{\partial \varepsilon_l} \ln Q, \tag{3.36}$$

a form that is sometimes useful.

3.8.1 Fermi–Dirac Distribution Function

At zero temperature the Fermi–Dirac distribution function can be written, as a function of energy ε, as

$$f_0(\varepsilon) = \begin{cases} 1 \text{ if } \varepsilon < \varepsilon_\mathrm{F}, \\ 0 \text{ if } \varepsilon > \varepsilon_\mathrm{F}. \end{cases} \tag{3.37}$$

At a finite temperature

$$f_0(\varepsilon) = \frac{1}{\mathrm{e}^{(\varepsilon - \zeta)/\Theta} + 1}. \tag{3.38}$$

Clearly at $\varepsilon = \zeta$, $f_0(\varepsilon = \zeta)$ is equal to $\frac{1}{2}$. (See Fig. 3.3.) The value of ζ is determined (as a function of T) by the condition

$$\sum_{\mathbf{k}\sigma} f_0\left(\varepsilon_{\mathbf{k}\sigma}\right) = N. \tag{3.39}$$

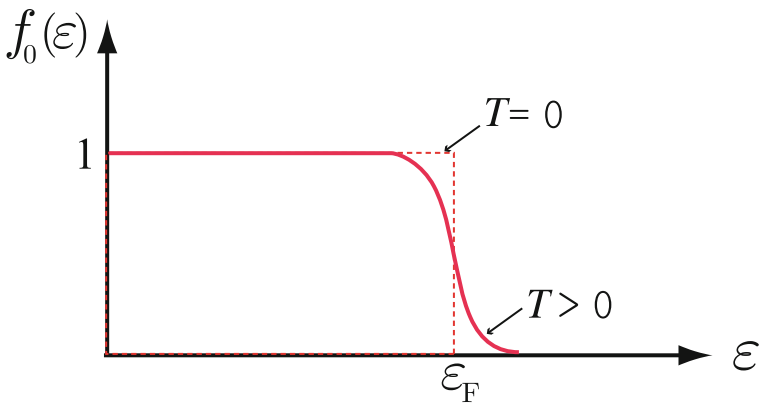


Fig. 3.3. Fermi–Dirac distribution function $f_0(\varepsilon)$ for two different temperatures

3.8.2 Density of States

It is easy to determine $G(\varepsilon)$, the total number of states per unit volume whose energy is less than ε, and then obtain $g(\varepsilon)$ from it.

$$G(\varepsilon + \mathrm{d}\varepsilon) - G(\varepsilon) = \frac{\mathrm{d}G}{\mathrm{d}\varepsilon}\mathrm{d}\varepsilon = g(\varepsilon)\mathrm{d}\varepsilon. \tag{3.40}$$

For free electrons we have

$$G(\varepsilon) = V^{-1} \sum_{\substack{\mathbf{k}\sigma \\ \varepsilon_{\mathbf{k}\sigma} \leq \varepsilon}} 1 = \frac{2}{(2\pi)^3}\frac{4\pi}{3}k^3, \tag{3.41}$$

where $\frac{\hbar^2 k^2}{2m} = \varepsilon$. It is easy to see that

$$G(\varepsilon) = n_0\left(\frac{k}{k_\mathrm{F}}\right)^3 = n_0\left(\frac{\varepsilon}{\varepsilon_\mathrm{F}}\right)^{3/2}. \tag{3.42}$$

Thus, from (3.40) we find

$$g(\varepsilon) = \frac{3}{2}\frac{n_0}{\varepsilon_\mathrm{F}}\left(\frac{\varepsilon}{\varepsilon_\mathrm{F}}\right)^{1/2} = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\varepsilon^{1/2}. \tag{3.43}$$

For electrons moving in a periodic potential, $g(\varepsilon)$ does not have such a simple form.

At a finite temperature Θ, the number of electrons per unit volume having energies between ε and $\varepsilon + \mathrm{d}\varepsilon$ is simply the product of $g(\varepsilon)\mathrm{d}\varepsilon$ and $f_0(\varepsilon)$: $n(\varepsilon)\mathrm{d}\varepsilon = g(\varepsilon)f_0(\varepsilon)\mathrm{d}\varepsilon$. (See Fig. 3.4.) The chemical potential ζ is determined from

$$N = V\int_0^\infty g(\varepsilon)f_0(\varepsilon)\mathrm{d}\varepsilon. \tag{3.44}$$

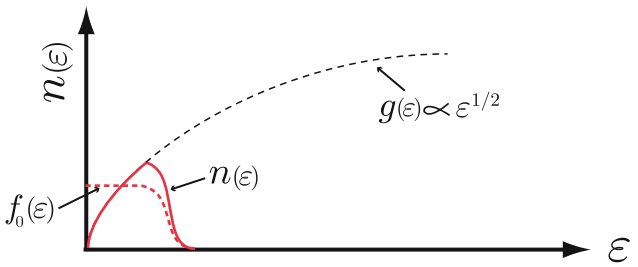


Fig. 3.4. Particle density $n(\varepsilon)$ and the density of states $g(\varepsilon)$

3.8.3 Thermodynamic Potential

The thermodynamic potential Ω is defined by

$$\Omega = -\Theta \ln Q = -\Theta \sum_i \ln\left(1 + \mathrm{e}^{(\zeta-\varepsilon_i)/\Theta}\right). \tag{3.45}$$

Functions that are commonly used in statistical mechanics are:

$$\begin{aligned}
&\textit{Internal energy } U,\\
&\textit{Helmholtz free energy } F = U - TS,\\
&\textit{Thermodynamic potential } \Omega = U - TS - \zeta N = -PV,\\
&\textit{enthalpy } \mathcal{H} = U + PV = TS + \zeta N,\\
&\textit{Gibbs free energy } G = U - TS + PV = \zeta N.
\end{aligned} \tag{3.46}$$

These definitions together with Euler's relation

$$U = TS - PV + \zeta N \tag{3.47}$$

and the second law of thermodynamics

$$\mathrm{d}U = T\mathrm{d}S - P\mathrm{d}V + \zeta \mathrm{d}N \tag{3.48}$$

are very useful to remember. By using (3.47) and (3.48) and $\Omega = -PV$, one can obtain

$$\mathrm{d}\Omega = -S\mathrm{d}T - P\mathrm{d}V - N\mathrm{d}\zeta. \tag{3.49}$$

From (3.49) one can see that the entropy S, pressure P, and particle number N can be obtained from the thermodynamic potential Ω

$$\begin{aligned}
S &= -\left(\frac{\partial \Omega}{\partial T}\right)_{V,\zeta},\\
P &= -\left(\frac{\partial \Omega}{\partial V}\right)_{T,\zeta},\\
N &= -\left(\frac{\partial \Omega}{\partial \zeta}\right)_{V,T}.
\end{aligned} \tag{3.50}$$

3.8.4 Entropy

We know that

$$\Omega = -\Theta \sum_i \ln\left(1 + \mathrm{e}^{(\zeta-\varepsilon_i)/\Theta}\right). \tag{3.51}$$

But we can write

$$1 - \bar{n}_i = 1 - \frac{1}{\mathrm{e}^{(\varepsilon_i-\zeta)/\Theta} + 1} = \frac{1}{\mathrm{e}^{(\zeta-\varepsilon_i)/\Theta} + 1}, \tag{3.52}$$

so that

$$\ln\left(1-\bar{n}_i\right) = -\ln\left[1+\mathrm{e}^{(\zeta-\varepsilon_i)/\Theta}\right]. \tag{3.53}$$

We can express (3.51) as

$$\Omega = \Theta\sum_i \ln\left(1-\bar{n}_i\right). \tag{3.54}$$

Since the entropy is given by $S=-\frac{\partial\Omega}{\partial\Theta}$, we can obtain

$$S = -\sum_i \ln\left(1-\bar{n}_i\right) + \Theta\sum_i \left(1-\bar{n}_i\right)^{-1}\frac{\partial\bar{n}_i}{\partial\Theta}. \tag{3.55}$$

Evaluating $\frac{\partial\bar{n}_i}{\partial\Theta}$ and multiplying by $\Theta\left(1-\bar{n}_i\right)^{-1}$ gives

$$\frac{\Theta}{1-\bar{n}_i}\frac{\partial\bar{n}_i}{\partial\Theta} = \bar{n}_i\ln\left(\frac{1-\bar{n}_i}{\bar{n}_i}\right). \tag{3.56}$$

Sustituting this result into (3.55) gives

$$S = -k_\mathrm{B}\sum_i\left[\left(1-\bar{n}_i\right)\ln\left(1-\bar{n}_i\right)+\bar{n}_i\ln\bar{n}_i\right]. \tag{3.57}$$

We have inserted the factor k_B into (3.57); in the derivation we had essentially set it equal to unity. Notice that the expression for S goes to zero as T goes to zero because $\bar{n}_i$ takes on the values 0 or 1 in this limit. In addition, we can write that

$$\begin{aligned}
\Theta S &= -\Theta\sum_i\left[\ln\left(1-\bar{n}_i\right)+\bar{n}_i\ln\left(\frac{\bar{n}_i}{1-\bar{n}_i}\right)\right],\\
&= -\Theta\sum_i\ln\left(1-\bar{n}_i\right) - \Theta\sum_i\bar{n}_i\left(\frac{\zeta-\varepsilon}{\Theta}\right),\\
&= -\Theta\sum_i\ln\left(1-\bar{n}_i\right) - \zeta\sum_i\bar{n}_i + \sum_i\bar{n}_i\varepsilon_i,\\
&= -\Theta\sum_i\ln\left(1-\bar{n}_i\right) - \zeta N + U.
\end{aligned} \tag{3.58}$$

If we write $F=U-TS$ we have

$$\begin{aligned}
F &= N\zeta + \Theta\sum_i\ln\left(1-\bar{n}_i\right),\\
&= N\zeta - \Theta\sum_i\ln\left(1+\mathrm{e}^{\frac{\zeta-\varepsilon_i}{\Theta}}\right).
\end{aligned} \tag{3.59}$$

If we hold V and T constant, the energy levels ε_i are unchanged and

$$\left(\frac{\partial F}{\partial N}\right)_{T,V} = \zeta + N\left(\frac{\partial \zeta}{\partial N}\right)_{T,V} - \Theta\frac{\partial}{\partial N}\sum_i \ln\left(1+\mathrm{e}^{\frac{\zeta-\varepsilon_i}{\Theta}}\right). \tag{3.60}$$

It is not difficult to show that (since $\ln\left[1+\mathrm{e}^{\frac{\zeta-\varepsilon_i}{\Theta}}\right]$ depends on N through ζ) the last two terms cancel and hence

$$\left(\frac{\partial F}{\partial N}\right)_{T,V} = \zeta. \tag{3.61}$$

3.9 Fermi Function Integration Formula

To study how the chemical potential ζ and internal energy U vary with temperature, we must evaluate the integrals

$$\frac{N}{V} = n_0 = \int_0^\infty \mathrm{d}\varepsilon\, g(\varepsilon) f_0(\varepsilon) \tag{3.62}$$

and

$$\frac{U}{V} = u = \int_0^\infty \mathrm{d}\varepsilon\, \varepsilon g(\varepsilon) f_0(\varepsilon). \tag{3.63}$$

In evaluating integrals of this type there is a very useful integration formula which we will now derive. Let us define an integral I as follows:

$$I = \int_0^\infty \mathrm{d}\varepsilon\, f_0(\varepsilon)\frac{\mathrm{d}F(\varepsilon)}{\mathrm{d}\varepsilon}. \tag{3.64}$$

Integrating by parts gives

$$I = [f_0(\varepsilon)F(\varepsilon)]_0^\infty - \int_0^\infty \mathrm{d}\varepsilon\, \frac{\partial f_0}{\partial \varepsilon}F(\varepsilon). \tag{3.65}$$

For many functions $F(\varepsilon)$, $F(0) = 0$ and $\lim_{\varepsilon\to\infty} f_0(\varepsilon)F(\varepsilon) \to 0$. For such functions we can write (3.65) simply as

$$I = -\int_0^\infty \mathrm{d}\varepsilon\, \frac{\partial f_0}{\partial \varepsilon}F(\varepsilon). \tag{3.66}$$

The functions f_0 changes rather quickly in an interval of width of the order of $k_\mathrm{B}T$ about $\varepsilon = \zeta$. It is obvious that

$$\int_0^\infty \left(-\frac{\partial f_0}{\partial \varepsilon}\right)\mathrm{d}\varepsilon = 1.$$

If the function $F(\varepsilon)$ is slowly varying compared to $\frac{\partial f_0}{\partial \varepsilon}$ in the region $\varepsilon \simeq \zeta$, we can expand $F(\varepsilon)$ in Taylor series as follows:

$$F(\varepsilon) = F(\zeta) + (\varepsilon - \zeta)F'(\zeta) + \frac{1}{2!}(\varepsilon - \zeta)^2 F''(\zeta) + \cdots . \tag{3.67}$$

Then, we can write I as

$$I = F(\zeta)\int_0^\infty \mathrm{d}\varepsilon \left(-\frac{\partial f_0}{\partial \varepsilon}\right) + F'(\zeta)\int_0^\infty \mathrm{d}\varepsilon\,(\varepsilon - \zeta)\left(-\frac{\partial f_0}{\partial \varepsilon}\right) + \frac{1}{2!}F''(\zeta)\int_0^\infty \mathrm{d}\varepsilon\,(\varepsilon - \zeta)^2\left(-\frac{\partial f_0}{\partial \varepsilon}\right) + \cdots .$$

But we note that

$$-\frac{\partial f_0}{\partial \varepsilon} = \beta\frac{\mathrm{e}^{\beta(\varepsilon-\zeta)}}{\left[\mathrm{e}^{\beta(\varepsilon-\zeta)}+1\right]^2}.$$

Introduce the parameter $z = \beta(\varepsilon - \zeta)$ and note that

$$\int_0^\infty \mathrm{d}\varepsilon\,(\varepsilon - \zeta)^n\left(-\frac{\partial f_0}{\partial \varepsilon}\right) = \Theta^n\int_{-\zeta/\Theta}^\infty \mathrm{d}z\,\frac{z^n}{(\mathrm{e}^z+1)(\mathrm{e}^{-z}+1)}.$$

If ζ is much larger than Θ (this is certainly true in metals) the lower limit on the integral over z can be replaced by $-\infty$. Since $\frac{z^n}{(\mathrm{e}^z+1)(\mathrm{e}^{-z}+1)}$ is an odd function of z for n odd, we obtain

$$I \simeq F(\zeta) + \frac{1}{2!}\Theta^2 F''(\zeta)\int_{-\infty}^\infty \mathrm{d}z\,\frac{z^2}{(\mathrm{e}^z+1)(\mathrm{e}^{-z}+1)} + \cdots + \frac{1}{(2n)!}\Theta^{2n}F^{(2n)}(\zeta)\int_{-\infty}^\infty \mathrm{d}z\,\frac{z^{2n}}{(\mathrm{e}^z+1)(\mathrm{e}^{-z}+1)}. \tag{3.68}$$

The first few integrals are

$$\int_{-\infty}^\infty \mathrm{d}z\,\frac{z^2}{(\mathrm{e}^z+1)(\mathrm{e}^{-z}+1)} = \frac{\pi^2}{3},$$

$$\int_{-\infty}^\infty \mathrm{d}z\,\frac{z^4}{(\mathrm{e}^z+1)(\mathrm{e}^{-z}+1)} = \frac{7\pi^4}{15}.$$

To order Θ^2 we have

$$I = F(\zeta) + \frac{\pi^2}{6}\Theta^2 F''(\zeta). \tag{3.69}$$

To evaluate the integral given in (3.62), we note that $F(\zeta)$ is just $G(\varepsilon)$, the total number of states per unit volume whose energy is less than ε. Then using (3.69) gives us

$$n_0 = G(\zeta) + \frac{\pi^2}{6}\Theta^2 G''(\zeta). \tag{3.70}$$

Define ζ_0 as the chemical potential at $T = 0$. Then $n_0 = G(\zeta_0)$ and

$$G(\zeta) = G(\zeta_0) - \frac{\pi^2}{6}\Theta^2 g'(\zeta). \tag{3.71}$$

Here, we have used $G'(\varepsilon) = g(\varepsilon)$ and set $n_0 = G(\zeta_0)$. Write $G(\zeta)$ as $G(\zeta_0) + g(\zeta_0)(\zeta - \zeta_0)$ and substitute into (3.71) to obtain

$$\zeta = \zeta_0 - \frac{\pi^2}{6}\Theta^2 \frac{g'(\zeta_0)}{g(\zeta_0)}.$$

But for free electrons $g(\zeta) = \frac{3}{2}\frac{n_0}{\zeta_0}\left(\frac{\varepsilon}{\zeta_0}\right)^{1/2}$ so that

$$\zeta = \zeta_0 \left[1 - \frac{\pi^2}{12}\left(\frac{\Theta}{\zeta_0}\right)^2 + \cdots\right]. \tag{3.72}$$

Applying the integration formula to the integral for $\frac{U}{V}$, $F(\varepsilon)$ is simply $\int_0^\varepsilon \varepsilon' g(\varepsilon')\,\mathrm{d}\varepsilon'$; therefore we have

$$\frac{U}{V} = \int_0^\zeta \varepsilon g(\varepsilon)\,\mathrm{d}\varepsilon + \frac{\pi^2}{6}\Theta^2 \left[\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\varepsilon\, g(\varepsilon)\right)\right]_{\varepsilon=\zeta}. \tag{3.73}$$

Define $U_0 = V\int_0^{\zeta_0} \varepsilon g(\varepsilon)\,\mathrm{d}\varepsilon$ and use the expression for $g(\varepsilon)$ given above for free electrons. One can find that

$$\frac{U}{V} = \frac{U_0}{V} + \frac{\pi^2}{6}\Theta^2 g(\zeta_0). \tag{3.74}$$

3.10 Heat Capacity of a Fermi Gas

The heat capacity $C_\mathrm{v} = \left(\frac{\partial U}{\partial T}\right)_V$ is given, using (3.74), by

$$C_\mathrm{v} = V\frac{\pi^2}{3}k_\mathrm{B}^2 g(\zeta_0)T = \gamma T. \tag{3.75}$$

For free electrons we have $\gamma = \frac{\pi^2 k_\mathrm{B}^2}{2\zeta_0}N$. It is interesting to compare the quantum mechanical Sommerfeld result $C_\mathrm{v}^\mathrm{QM} = \gamma T$ with the classical Drude result $C_\mathrm{v}^\mathrm{CM} = \frac{3}{2}Nk_\mathrm{B}$:

$$\frac{C_\mathrm{v}^\mathrm{QM}}{C_\mathrm{v}^\mathrm{CM}} = \frac{\pi^2}{3}\,\frac{T}{T_\mathrm{F}}. \tag{3.76}$$

For a typical metal, $T_\mathrm{F} \simeq 10^5\,\mathrm{K}$, while at room temperature $T \simeq 300\,\mathrm{K}$. This solves the problem that perplexed Drude concerning why the classical specific heat C_v^CM was not observed. The correct quantum mechanical specific heat is so small (because $\frac{T}{T_\mathrm{F}} \ll 1$) that it is difficult to observe even at room temperature.

One can obtain a rough estimate of the specific heat by saying that only quantum states within $k_\mathrm{B}T$ of the Fermi energy contribute to the classical estimate of the specific heat. This means that

$$N_{\text{eff}} = V\left[G(\varepsilon_{\text{F}}) - G(\varepsilon_{\text{F}} - k_{\text{B}}T)\right].$$

This gives

$$U \approx \left(\frac{3}{2}k_{\text{B}}T\right)\ N_{\text{eff}} = \left[\frac{3}{2}k_{\text{B}}T\right]\ [Vg(\varepsilon_{\text{F}})k_{\text{B}}T], \tag{3.77}$$

and hence

$$C_{\text{v}} = \frac{\partial U}{\partial T} \approx V3k_{\text{B}}^2 g(\varepsilon_{\text{F}})T. \tag{3.78}$$

3.11 Equation of State of a Fermi Gas

The equation of state relates the variables P, V, and T. For the Fermi gas we know that

$$P = -\left(\frac{\partial \Omega}{\partial V}\right)_{T,\zeta}. \tag{3.79}$$

But $\Omega = -\Theta\sum_i \ln\left(1 + \mathrm{e}^{(\zeta-\varepsilon_i)/\Theta}\right)$. At constant values of $\Theta = k_{\text{B}}T$ and ζ, Ω depends on V through ε_i:

$$\varepsilon_i = \frac{\hbar^2}{2m}\left(\frac{2\pi}{L}\right)^2\left(n_{ix}^2 + n_{iy}^2 + n_{iz}^2\right). \tag{3.80}$$

We can write $\frac{\partial \varepsilon_i}{\partial V} = \frac{\partial \varepsilon_i}{\partial L}\left(\frac{\partial V}{\partial L}\right)^{-1}$. Since $\varepsilon_i \propto L^{-2}$ and $V \propto L^3$ this gives $\frac{\partial \varepsilon_i}{\partial V} = -\frac{2}{3}\frac{\varepsilon_i}{V}$. Using this result in (3.79) gives

$$P = \Theta\sum_i \left(\frac{\mathrm{e}^{(\zeta-\varepsilon_i)/\Theta}}{1 + \mathrm{e}^{(\zeta-\varepsilon_i)/\Theta}}\right)\left(-\Theta^{-1}\right)\frac{\partial \varepsilon_i}{\partial V}. \tag{3.81}$$

From this we find (since $G(\varepsilon) = \frac{2}{3}\varepsilon g(\varepsilon)$ for a free Fermi gas) that

$$P = \frac{2}{3}\frac{U}{V} = \int_0^\infty \mathrm{d}\varepsilon G(\varepsilon) f_0(\varepsilon). \tag{3.82}$$

If we keep terms to order Θ^2 we obtain

$$P = P_0 + \frac{\pi^2}{6}\Theta^2\left\{g(\varepsilon_{\text{F}}) - \frac{g'(\varepsilon_{\text{F}})}{g(\varepsilon_{\text{F}})}G(\varepsilon_{\text{F}})\right\}. \tag{3.83}$$

3.12 Compressibility

The compressibility κ_T is defined by

$$\begin{aligned}\kappa_T^{-1} &= -V\left(\frac{\partial P}{\partial V}\right)_{T,\zeta} \\ &= -V\frac{\partial}{\partial V}\int_0^\infty G(\varepsilon) f_0(\varepsilon)\mathrm{d}\varepsilon.\end{aligned} \tag{3.84}$$

If we define $H(\varepsilon) = \int_0^{\varepsilon} G(\varepsilon)\,\mathrm{d}\varepsilon$, then the integral can be evaluated by integrating by parts to get (at $T = 0$)

$$\int_0^{\infty} G(\varepsilon) f_0(\varepsilon)\ \mathrm{d}\varepsilon = H\left(\varepsilon_{\mathrm{F}}\right).$$

But we know that $G(\varepsilon) = A\varepsilon^{3/2}$, therefore $H(\varepsilon) = \frac{2}{5}A\varepsilon^{5/2} = \frac{2}{5}\varepsilon G(\varepsilon)$ to have

$$\kappa_T^{-1} = -V\frac{\partial}{\partial V}\left(\frac{2}{5}\varepsilon_{\mathrm{F}} n_0\right),$$

since $G(\varepsilon_{\mathrm{F}}) = n_0$. ε_{F} is proportional to L^{-2}, and n_0 is proportional to L^{-3} so $\varepsilon_{\mathrm{F}} n_0$ is proportional to $L^{-5} = V^{-5/3}$. This gives

$$\kappa_T^{-1} = -V\frac{2}{5}\left(-\frac{5}{3}\right)\frac{n_0\varepsilon_{\mathrm{F}}}{V} = \frac{2}{3}n_0\varepsilon_{\mathrm{F}}.$$

Using $g(\varepsilon_{\mathrm{F}}) = \frac{3}{2}\frac{n_0}{\varepsilon_{\mathrm{F}}}$ allows us to write

$$\kappa_T^{-1} = \frac{n_0^2}{g(\varepsilon_{\mathrm{F}})}. \tag{3.85}$$

For free electrons $g(\zeta_0) = \frac{3n_0}{2\varepsilon_{\mathrm{F}}}$ and $\kappa_T = \frac{3}{2n_0\zeta_0}$. The velocity of sound in a solid is given by

$$s = \left(\kappa_T \rho\right)^{-1/2} \tag{3.86}$$

where ρ is the mass density. κ_T is the compressibility of the material in (3.86), and it includes the ion core repulsion as well as the pressure due to compressing the electron gas. The ionic contribution is small in simple metals like the alkali metals. If we neglect it and put $\rho = \frac{n_0 M}{z}$, where M is the ionic mass and z the number of electrons per atom, we find

$$s = \left(\frac{zm}{3M}\right)^{1/2} v_{\mathrm{F}}. \tag{3.87}$$

This result was first obtained by Bohm and Staver in a somewhat different way.

3.13 Electrical and Thermal Conductivities

Assume that there is an electric field $\mathbf{E} = E\hat{x}$ and a temperature gradient $\nabla T = \frac{\partial T}{\partial x}\hat{x}$. In discussing the Lorentz model, we wrote down the solution to the linearized Boltzmann equation

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{r}} f + \dot{\mathbf{v}}\cdot\nabla_{\mathbf{v}} f = -\left(\frac{f - f_0}{\tau}\right) \tag{3.88}$$

in the form

$$f = f_0 - \tau \left[-\frac{eE}{m} \frac{\partial f_0}{\partial v_x} + v_x \frac{\partial f_0}{\partial x} \right]. \tag{3.89}$$

The equilibrium distribution function is the Fermi–Dirac distribution function

$$f_0 = \frac{1}{1 + \mathrm{e}^{(\varepsilon - \zeta)/\Theta}}. \tag{3.90}$$

Because of the temperature gradient, both T and ζ depend on the coordinate x, but the energy ε does not. We can write

$$\frac{\partial f_0}{\partial x} = \frac{\partial f_0}{\partial \alpha} \frac{\partial \alpha}{\partial x}, \tag{3.91}$$

where $\alpha = \frac{\varepsilon - \zeta}{\Theta}$. This can be rewritten

$$\begin{aligned} \frac{\partial f_0}{\partial x} &= \Theta \frac{\partial f_0}{\partial \varepsilon} \frac{\partial}{\partial x} \left(\frac{\varepsilon - \zeta}{\Theta} \right) \\ &= -\frac{\partial f_0}{\partial \varepsilon} \left[\frac{\varepsilon}{\Theta} \frac{\partial \Theta}{\partial x} + \Theta \frac{\partial}{\partial x} \left(\frac{\zeta}{\Theta} \right) \right]. \end{aligned} \tag{3.92}$$

Because $\varepsilon = \frac{1}{2} m v^2$ we can write

$$\frac{\partial f_0}{\partial v_x} = \frac{\partial f_0}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial v_x} = m v_x \frac{\partial f_0}{\partial \varepsilon}. \tag{3.93}$$

We now substitute (3.90), (3.92), and (3.93) into (3.89) and use the resulting expression for $f(\varepsilon)$ in the equations for the electrical current density j_x and the thermal current density w_x:

$$j_x = \int_0^\infty \mathrm{d}\varepsilon \; (-e v_x) g(\varepsilon) f(\varepsilon), \tag{3.94}$$

$$w_x = \int_0^\infty \mathrm{d}\varepsilon \; (\varepsilon v_x) g(\varepsilon) f(\varepsilon). \tag{3.95}$$

For the electrical current density we obtain

$$j_x = e \int_0^\infty \mathrm{d}\varepsilon \; v_x g(\varepsilon) \tau v_x \left(-\frac{\partial f_0}{\partial \varepsilon} \right) \left[eE + \frac{\varepsilon}{\Theta} \frac{\partial \Theta}{\partial x} + \Theta \frac{\partial}{\partial x} \left(\frac{\zeta}{\Theta} \right) \right]. \tag{3.96}$$

Factoring all quantities that are independent of ε out of the integral gives

$$\begin{aligned} j_x = &\left[e^2 E + e\Theta \frac{\partial}{\partial x} \left(\frac{\zeta}{\Theta} \right) \right] \int_0^\infty \mathrm{d}\varepsilon \; v_x^2 g(\varepsilon) \tau \left(-\frac{\partial f_0}{\partial \varepsilon} \right) \\ &+ \frac{e}{\Theta} \frac{\partial \Theta}{\partial x} \int_0^\infty \mathrm{d}\varepsilon \; v_x^2 \varepsilon g(\varepsilon) \tau \left(-\frac{\partial f_0}{\partial \varepsilon} \right). \end{aligned} \tag{3.97}$$

Now substitute $v_x^2 = \frac{2}{3}\frac{\varepsilon}{m}$ and $g(\varepsilon) = \frac{3}{2}\frac{n_0}{\zeta_0^{3/2}}\varepsilon^{1/2}$ into (3.97) to have

$$j_x = \left[e^2 E + e\Theta \frac{\partial}{\partial x}\left(\frac{\zeta}{\Theta}\right)\right]\mathcal{K}_1 + \frac{e}{\Theta}\frac{\partial \Theta}{\partial x}\mathcal{K}_2, \tag{3.98}$$

where we have introduced the symbol $\mathcal{K}_n$ defined by

$$\mathcal{K}_n = \frac{n_0}{m\zeta_0^{3/2}} \int_0^\infty \mathrm{d}\varepsilon \ \left(-\frac{\partial f_0}{\partial \varepsilon}\right) \varepsilon^{n+1/2}\tau. \tag{3.99}$$

In the calculation of w_x, a factor of ε replaces $(-e)$; this gives

$$w_x = -\left[eE + \Theta \frac{\partial}{\partial x}\left(\frac{\zeta}{\Theta}\right)\right]\mathcal{K}_2 - \frac{1}{\Theta}\frac{\partial \Theta}{\partial x}\mathcal{K}_3. \tag{3.100}$$

The function $\mathcal{K}_n$ can be evaluated using the integration formula (3.69). We obtain

$$\mathcal{K}_n = \frac{n_0}{m\zeta_0^{3/2}}\left[\zeta^{n+1/2}\tau(\zeta) + \frac{\pi^2}{6}\Theta^2 \frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2}\left(\varepsilon^{n+1/2}\tau(\varepsilon)\right)\,|_{\varepsilon=\zeta}\right]. \tag{3.101}$$

At $T = 0$ we have

$$\mathcal{K}_n = \frac{n_0}{m}\zeta_0^{n-1}\tau(\zeta_0). \tag{3.102}$$

3.13.1 Electrical Conductivity

If we set $\frac{\partial T}{\partial x} = 0$, then j_x is given by $j_x = e^2 E\mathcal{K}_1$, and at $T = 0$ we have

$$j_x = \frac{n_0 e^2 \tau(\zeta_0)}{m} E = \sigma E. \tag{3.103}$$

This is exactly the Drude result for the conductivity σ with $\tau(\varepsilon)$ evaluated on the Fermi surface so that $\tau = \tau(\zeta_0)$.

3.13.2 Thermal Conductivity

The thermal conductivity is defined as the ratio of the thermal current w_x to $\left(-\frac{\partial T}{\partial x}\right)$ under conditions of zero electrical current. Therefore, we must set $j_x = 0$ in (3.98) and solve for E. This gives

$$j_x = \left[e^2 E + e\Theta \frac{\partial}{\partial x}\left(\frac{\zeta}{\Theta}\right)\right]\mathcal{K}_1 + \frac{e}{\Theta}\frac{\partial \Theta}{\partial x}\mathcal{K}_2 = 0$$

or

$$-\left[eE + \Theta \frac{\partial}{\partial x}\left(\frac{\zeta}{\Theta}\right)\right]\mathcal{K}_1 = \frac{1}{\Theta}\frac{\partial \Theta}{\partial x}\mathcal{K}_2. \tag{3.104}$$

Substitute this into (3.100) to obtain w_x; the result is

$$\begin{aligned} w_x &= \frac{1}{\Theta}\frac{\partial\Theta}{\partial x}\frac{\mathcal{K}_2}{\mathcal{K}_1}\mathcal{K}_2 - \frac{1}{\Theta}\frac{\partial\Theta}{\partial x}\mathcal{K}_3 \\ &= \frac{\mathcal{K}_3\mathcal{K}_1 - \mathcal{K}_2^2}{\mathcal{K}_1\Theta}\left(-\frac{\partial\Theta}{\partial x}\right). \end{aligned} \tag{3.105}$$

Thus the thermal conductivity $\kappa_T = k_{\mathrm{B}} w_x \left(-\frac{\partial\Theta}{\partial x}\right)^{-1}$ is

$$\kappa_T = k_{\mathrm{B}}\frac{\mathcal{K}_3\mathcal{K}_1 - \mathcal{K}_2^2}{\mathcal{K}_1\Theta}. \tag{3.106}$$

If we evaluate $\mathcal{K}_n$ as a function of Θ using (3.101) and the result

$$\left(\frac{\zeta}{\zeta_0}\right)^l \simeq 1 - \frac{\pi^2}{12}\left(\frac{\Theta}{\zeta_0}\right)^2 l, \tag{3.107}$$

we find (for τ independent of energy)

$$\begin{aligned} \mathcal{K}_1 &\simeq \frac{n_0\tau}{m}, \\ \mathcal{K}_2 &\simeq \frac{n_0\tau\zeta_0}{m}\left[1 + \frac{5}{12}\pi^2\left(\frac{\Theta}{\zeta_0}\right)^2\right], \\ \mathcal{K}_3 &\simeq \frac{n_0\tau\zeta_0^2}{m}\left[1 + \frac{7}{6}\pi^2\left(\frac{\Theta}{\zeta_0}\right)^2\right]. \end{aligned} \tag{3.108}$$

Now, substitute these results into (3.106) to obtain

$$\kappa_T = k_{\mathrm{B}}\frac{\pi^2}{3}\frac{n_0\tau}{m}\Theta.$$

Thus the Sommerfeld expression for κ_T can be written

$$\kappa_T = \frac{\pi^2}{3}k_{\mathrm{B}}^2\frac{n_0\tau}{m}T. \tag{3.109}$$

The Lorenz ratio for the Sommerfeld model $\mathcal{L}_{\mathrm{S}}$ is given by

$$\mathcal{L}_{\mathrm{S}} = \frac{\kappa_T}{\sigma T} = \frac{\pi^2}{3}\left(\frac{k_{\mathrm{B}}}{e}\right)^2 \simeq 2.71\times10^{-13}\mathrm{esu}. \tag{3.110}$$

Recall that for the Drude model

$$\mathcal{L}_{\mathrm{D}} = \frac{3}{2}\left(\frac{k_{\mathrm{B}}}{e}\right)^2 \simeq 1.24\times10^{-13}\,\mathrm{esu}, \tag{3.111}$$

and the average experimental result is $\mathcal{L} \approx 2.72\times10^{-13}\,\mathrm{esu}$.

3.14 Critique of Sommerfeld Model

The main achievements of the Sommerfeld model were as follows:

1. It explained the specific heat dilemma by showing that $C_{\rm v}$ for the electrons was very small.
2. It showed that even though there was one free electron per atom, the Pauli principle made only those in an energy range $k_{\rm B}T$ about ζ_0 effectively free.
3. It not only explained the Wiedemann–Franz law but it gave a very accurate value for the Lorenz ratio.
4. It correctly predicted the Pauli spin paramagnetism of metals.
5. It predicted Fermi energies that agreed with observed X-ray band widths.

The main shortcomings of the Sommerfeld model were:

1. It said nothing about the relaxation time $\tau(\varepsilon)$, τ appeared only as a phenomenological parameter. To agree with observed conductivities, the mean free path $l = v_{\rm F}\tau$ had to be of the order of a 100 atomic spacings ($\sim 5 \times 10^{-6}$ cm), and had to vary as T^{-1} at room temperature. These requirements were difficult to understand in 1928.
2. The model ignored the interaction of the free electrons with the fixed ions and with one another. These interactions were surely large. How could one achieve such excellent agreement with experiment when they were ignored. Furthermore, attempts to include these interaction ran into great difficulties.

3.15 Magnetoconductivity

In the presence of a large dc magnetic field $\mathbf{B}$, the conductivity of a metal displays some new effects. These can be understood very simply using the Drude model (the Sommerfeld model gives exactly the same result but involves much more mathematics).

In the presence of an electric field $\mathbf{E}$ and a dc magnetic field $\mathbf{B}$, the Drude model would predict a drift velocity $\mathbf{v}_{\rm D}$ which was a solution of the equation

$$m\left(\frac{{\rm d}\mathbf{v}_{\rm D}}{{\rm d}t} + \frac{\mathbf{v}_{\rm D}}{\tau}\right) = -e\mathbf{E} - \frac{e}{c}\mathbf{v}_{\rm D} \times \mathbf{B}. \tag{3.112}$$

Let us choose the z-axis along $\mathbf{B}$ and assume that $\mathbf{E}$ is spatially uniform but varies in time as $\mathrm{e}^{\mathrm{i}\omega t}$. Then (3.112) can be rewritten (we drop the subscript D of $\mathbf{v}_{\rm D}$ in the rest of this section) as follows:

$$\begin{aligned}
(1 + \mathrm{i}\omega\tau)v_x &= -\frac{e\tau}{m}E_x - \frac{e\tau}{mc}Bv_y, \\
(1 + \mathrm{i}\omega\tau)v_y &= -\frac{e\tau}{m}E_y + \frac{e\tau}{mc}Bv_x, \\
(1 + \mathrm{i}\omega\tau)v_z &= -\frac{e\tau}{m}E_z.
\end{aligned} \tag{3.113}$$

Define the cyclotron frequency $\omega_c = \frac{eB}{mc}$ and solve for $\mathbf{v}$. The result is

$$\begin{aligned} v_x &= -\left(\frac{e\tau}{m}\right)\frac{[(1+\mathrm{i}\omega\tau)E_x - \omega_c\tau E_y]}{(1+\mathrm{i}\omega\tau)^2+(\omega_c\tau)^2}, \\ v_y &= -\left(\frac{e\tau}{m}\right)\frac{[\omega_c\tau E_x + (1+\mathrm{i}\omega\tau)E_y]}{(1+\mathrm{i}\omega\tau)^2+(\omega_c\tau)^2}, \\ v_z &= -\left(\frac{e\tau}{m}\right)\frac{E_z}{1+\mathrm{i}\omega\tau}. \end{aligned} \tag{3.114}$$

The current density is given by $\mathbf{j} = -en_0\mathbf{v}$. This can be written $\mathbf{j} = \overline{\overline{\sigma}} \cdot \mathbf{E}$, where $\overline{\overline{\sigma}}$ is called the *magnetoconductivity tensor*. Its components are $\sigma_{xz} = \sigma_{zx} = \sigma_{yz} = \sigma_{zy} = 0$, and

$$\begin{aligned} \sigma_{xx} &= \sigma_{yy} = \frac{\sigma_0(1+\mathrm{i}\omega\tau)}{(1+\mathrm{i}\omega\tau)^2+(\omega_c\tau)^2}, \\ \sigma_{xy} &= -\sigma_{yx} = \frac{\sigma_0(-\omega_c\tau)}{(1+\mathrm{i}\omega\tau)^2+(\omega_c\tau)^2}, \\ \sigma_{zz} &= \frac{\sigma_0}{1+\mathrm{i}\omega\tau}. \end{aligned} \tag{3.115}$$

Here $\sigma_0 = \frac{n_0e^2\tau}{m}$ is just the Drude's dc conductivity.

3.16 Hall Effect and Magnetoresistance

If we apply an electric field $\mathbf{E}$ in the x-direction, the Lorentz force, $-\frac{e}{c}\mathbf{v}\times\mathbf{B}$ causes a drift velocity in the y-direction. If $\omega = 0$ charge will accumulate on the surfaces normal to the y-direction until a field E_y builds up that exactly cancels the Lorentz force. (See Fig. 3.5.) The condition $j_y = 0$ gives

$$j_y = \sigma_{xx}E_y - \sigma_{xy}E_x = 0,$$

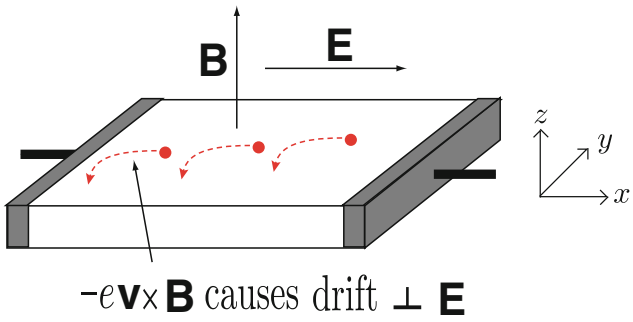


Fig. 3.5. Schematics of the Hall effect experiment. The initial drift of the negatively charged electron is illustrated

or

$$E_y = \frac{\sigma_{xy}}{\sigma_{xx}} E_x.$$

The *Hall coefficient* $\mathcal{R}$ is defined as the ratio of E_y to $j_x B$:

$$\mathcal{R} = \frac{E_y}{j_x B}. \tag{3.116}$$

But $j_x = \sigma_{xx} E_x + \sigma_{xy} E_y = (\sigma_{xx}^2 + \sigma_{xy}^2)/\sigma_{xx}\ E_x$. If we substitute it into the expression for the Hall coefficient, we find

$$\begin{aligned} \mathcal{R} = \frac{E_y}{j_x B} &= \frac{(\sigma_{xy}/\sigma_{xx})\ E_x}{\left[\left(\sigma_{xx}^2 + \sigma_{xy}^2\right)/\sigma_{xx}\right] E_x B} \\ &= \frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2}\ \frac{1}{B}. \end{aligned}$$

Making use of (3.115) in the limit $\omega \longrightarrow 0$ gives

$$\mathcal{R} = \frac{1}{n_0(-e)c}. \tag{3.117}$$

Because $\mathcal{R}$ depends on the carrier concentration, the Hall effect is often used to measure n_0.

Magnetoresistance

When $B = 0$, $j_x = \sigma_0 E_x$. In the presence of the magnetic field B, we have

$$j_x = \frac{\sigma_{xx}^2 + \sigma_{xy}^2}{\sigma_{xx}}\ E_x. \tag{3.118}$$

For the free electron model $\frac{\sigma_{xx}^2+\sigma_{xy}^2}{\sigma_{xx}} = \sigma_0$ (one can check this relation as an exercise). Therefore even in the presence of the B-field $j_x = \sigma_0 E_x$. The magnetic field causes no change in the ratio $\frac{E_x}{j_x} = \rho$, the *resistivity*, and we would say that

$$\Delta\rho = \rho(B) - \rho(0) = 0,$$

or that the magnetoresistance vanishes. This does not occur in more general cases than the simple free electron model as we shall see later.

3.17 Dielectric Function

The electrical current density $\mathbf{j}$ can be thought of as the time rate of change of the polarization $\mathbf{P}$. Assume $\mathbf{D}$, $\mathbf{P}$, and $\mathbf{E}$ vary as $\mathrm{e}^{\mathrm{i}\omega t}$. Then $\mathbf{j} = \dot{\mathbf{P}} = \mathrm{i}\omega\mathbf{P}$ and $\mathbf{D} = \varepsilon\mathbf{E} = \mathbf{E} + 4\pi\mathbf{P}$ where ε is *the dielectric function*. This gives us the relation

$$\varepsilon(\omega) = 1 - \frac{4\pi\mathrm{i}}{\omega}\sigma(\omega), \tag{3.119}$$

where $\sigma(\omega)$ is the frequency dependent Drude conductivity. In the presence of a dc magnetic field, the dielectric function and conductivity become tensor quantities: $\underline{\sigma}(\omega)$ and $\underline{\varepsilon}(\omega)$, whose off-diagonal components result from the Lorentz force.

The dielectric function $\varepsilon(\omega)$ or conductivity $\sigma(\omega)$ appear in Maxwell's equation for $\nabla \times \mathbf{B}$:

$$\nabla \times \mathbf{B} = \frac{1}{c}\dot{\mathbf{E}} + \frac{4\pi}{c}\mathbf{j} = \frac{\mathrm{i}\omega}{c}\left[\mathbf{1} - \frac{4\pi\mathrm{i}}{\omega}\underline{\sigma}(\omega)\right] \cdot \mathbf{E} = \frac{1}{c}\underline{\varepsilon}(\omega) \cdot \dot{\mathbf{E}}. \tag{3.120}$$

In the Drude model

$$\begin{aligned}\varepsilon(\omega) &= 1 - \frac{4\pi\mathrm{i}}{\omega}\sigma(\omega), \\ &= 1 - \frac{4\pi\mathrm{i}}{\omega}\frac{n_0 e^2 \tau/m}{1 + \mathrm{i}\omega\tau}.\end{aligned}$$

Define the *plasma frequency* ω_p by $\omega_p^2 = \frac{4\pi n_0 e^2}{m}$; then we have

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega - \mathrm{i}/\tau)}. \tag{3.121}$$

$\varepsilon(\omega)$ has real and imaginary parts, ε_1 and ε_2, respectively, as

$$\begin{aligned}\varepsilon_1(\omega) &= 1 - \frac{\omega_p^2}{\omega^2 + 1/\tau^2}, \\ \varepsilon_2(\omega) &= -\frac{\omega_p^2/\omega\tau}{\omega^2 + 1/\tau^2}.\end{aligned} \tag{3.122}$$

In general, we can ask how electromagnetic waves propagate in a medium described by a dielectric tentor $\underline{\varepsilon}(\omega)$. The wave equation can be obtained from the two Maxwell equations:

$$\begin{aligned}\nabla \times \mathbf{E} &= -\frac{1}{c}\dot{\mathbf{B}}, \\ \nabla \times \mathbf{B} &= \frac{1}{c}\underline{\varepsilon} \cdot \dot{\mathbf{E}}.\end{aligned} \tag{3.123}$$

Assume $\mathbf{E}$ and $\mathbf{B}$ vary as $\mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}\mathbf{q}\cdot\mathbf{r}}$. The two Maxwell equations can then be combined, by eliminating $\mathbf{B}$, to give

$$\mathbf{q}(\mathbf{q} \cdot \mathbf{E}) - q^2\mathbf{E} + \frac{\omega^2}{c^2}\underline{\varepsilon} \cdot \mathbf{E} = 0. \tag{3.124}$$

This can be applied to a case in which a dc magnetic field $\mathbf{B}_0$ is present and oriented in the z-direction. Then without loss of generality we can choose $\mathbf{q} = (0, q_y, q_z)$ and write (3.124) as

$$\begin{pmatrix} \frac{\omega^2}{c^2}\varepsilon_{xx} - q^2 & \frac{\omega^2}{c^2}\varepsilon_{xy} & 0 \\ -\frac{\omega^2}{c^2}\varepsilon_{xy} & \frac{\omega^2}{c^2}\varepsilon_{xx} - q_z^2 & q_y q_z \\ 0 & q_y q_z & \frac{\omega^2}{c^2}\varepsilon_{zz} - q_y^2 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = 0. \tag{3.125}$$

The description of the bulk modes are given by setting the determinant of the matrix multiplying the column vector equal to zero.

For surface modes at a metal–dielectric interface, think of ω and q_y (wave vector along the surface) as given and determine the allowed values of q_z in the solid (with a given $\underline{\varepsilon}(\omega)$) and in the dielectric. One can get modes from applying standard boundary conditions.

Problems

3.1. A two-dimensional electron gas is contained within a square box of side L.

(a) Apply periodic boundary conditions and determine $E(k_x, k_y)$ and $\Psi_{\mathbf{k}}(x, y)$ for the free electron Hamiltonian $H = \frac{1}{2m}\left(p_x^2 + p_y^2\right)$.
(b) Determine the Fermi wave number k_{F} in terms of the density $n_0 = \frac{N}{L^2}$.
(c) Evaluate $G(\varepsilon)$ and $g(\varepsilon)$ for this system.
(d) Use $n_0 = \int_0^\infty \mathrm{d}\varepsilon\, g(\varepsilon) f_0(\varepsilon)$ together with the Fermi function integration formula to determine how the chemical potential ζ depends on T.

3.2. In the simple Drude model it is assumed that the probability that an electron has a collision in a time interval dt is given by $\frac{dt}{\tau}$.

(a) Show that an electron picked at random at a given instant had no collision during the past t seconds with probability $\mathrm{e}^{-t/\tau}$. Demonstrate that it will have no collision during the next t seconds with the same probability,
(b) Show that the probability that the time interval between two successive collisions falls in the range t to $t + \mathrm{d}t$ is $\left(\frac{\mathrm{d}t}{\tau}\right)\mathrm{e}^{-t/\tau}$,
(c) as a consequence of (a), the mean time (at any given instant) back to the last collision (or up to the next collision) averaged over all electrons is τ,
(d) and, as a consequence of (b), the mean time between collisions is τ.

3.3. Think about the following argument, and use physical intuition (not formal mathematics) to say why it its wrong.
Argument: Both (c) and (d) in Problem 2 cannot be simultaneously correct for the following reason. If I catch a bus each morning to go to work, and the buses arrive at random times except that the average time between buses is 15 min, then (c) would imply that, on the average, I have to wait 15 min. However, if a friend got off each bus at my stop, and then got on the next bus with me, then the average time waited by my friend boarding the bus with me would be 15 min also. Clearly the friend boarding with me is waiting at the bus stop when I arrive. How can it be that we wait the same amount of time?

3.4. A metal is described by the conductivity tensor $\sigma_{xx}=\sigma_{yy}=\frac{\sigma_0(1+\mathrm{i}\omega\tau)}{(1+\mathrm{i}\omega\tau)^2+(\omega_{\mathrm{c}}\tau)^2}$, $\sigma_{xy}=-\sigma_{yx}=\frac{\sigma_0(-\omega_{\mathrm{c}}\tau)}{(1+\mathrm{i}\omega\tau)^2+(\omega_{\mathrm{c}}\tau)^2}$, and $\sigma_{zz} = \frac{\sigma_0}{1+\mathrm{i}\omega\tau}$ in the presence of a dc magnetic field $\mathbf{B} = B\hat{z}$. Consider the propagation of an electromagnetic wave $E_\pm = (E_x \pm \mathrm{i}E_y)\,\mathrm{e}^{\mathrm{i}\omega t}$ parallel to the z-axis. Use Maxwell's equations to obtain the wave equation, and show that $c^2k^2 = \omega^2\varepsilon_\pm(\omega)$, where $\underline{\varepsilon} = \underline{\mathbf{1}} - \frac{4\pi i}{\omega}\underline{\sigma}(\omega)$. Consider the cases $\omega_{\mathrm{c}}\tau \gg 1$ and $\omega_{\mathrm{c}} \gg \omega$ and show that $\omega = \frac{c^2k^2\omega_{\mathrm{c}}}{\omega_{\mathrm{p}}^2}$ for one circular polarization.

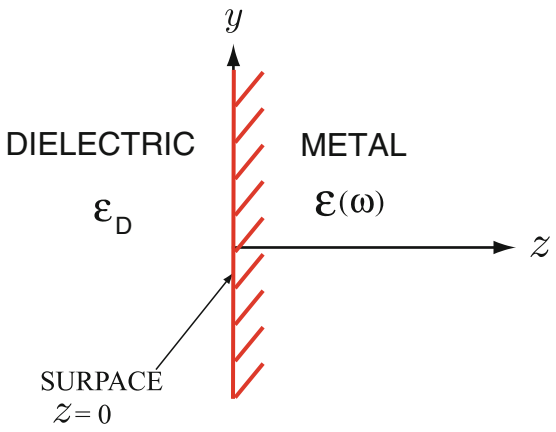


Fig. 3.6. Interface between a dielectric and a metal

3.5. Let us consider the interface between a dielectric of dielectric constant ε_{D} and a metal of dielectric function $\varepsilon(\omega) = 1 - \frac{\omega_{\mathrm{p}}^2}{\omega^2}$, where $\omega_{\mathrm{p}}^2 = \frac{4\pi n e^2}{m\omega^2}$. It is illustrated in Fig. 3.6.

If the normal to the surface is in the z direction and the wave vector $\mathbf{q} = (0, q_y, q_z)$, consider the region of $\omega - q_y$ space in which q_z is imaginary (i.e. $q_z^2 < 0$) both in the dielectric and in the metal. Impose the appropriate boundary conditions at $z = 0$ and at $\mid z \mid \longrightarrow \infty$, and determine the dispersion relation (ω as a function of q_y) for these surface plasma modes.

3.6. At a temperature T a semiconductor contains n_{e} electrons and n_{h} holes per unit volume in parabolic energy bands (i.e. $E \propto k^2$). The mass, charge, and collision time of the electrons and holes are m_{e}, $-e$, τ_{e}, and m_{h}, e, τ_{h}, respectively.

(a) Use the equations of motion of charged particles in the presence of a dc magnetic field $\mathbf{B} = B\hat{z}$ and an ac electric field $\mathbf{E}e^{i\omega t}$ to determine $\underline{\sigma}_{\mathrm{e}}(\omega)$ and $\underline{\sigma}_{\mathrm{h}}(\omega)$, the electron and hole contributions to the frequency dependent magnetoconductivity tensor.
(b) Consider $\omega_{\mathrm{ce}} = \frac{eB}{m_{\mathrm{e}}c}$ and $\omega_{\mathrm{ch}} = \frac{eB}{m_{\mathrm{h}}c}$ to be large compared to τ_{e}^{-1} and τ_{h}^{-1}, respectively. Determine the Hall coefficient for $\omega = 0$.
(c) Under the conditions of part (b), determine the magnetoresistance.

Summary

In this chapter, first we have briefly reviewed classical kinetic theories of an electron gas both by Drude and by Lorentz as simple models of metals. Then Sommerfeld's elementary quantum mechanical theory of metals is discussed.

In the Drude model, the electrical conductivity $\sigma = \frac{n_0 e^2 \tau}{m}$ is determined by the Newton's law of motion given by

$$m\left(\frac{\mathrm{d}\mathbf{v}_\mathrm{D}}{\mathrm{d}t} + \frac{\mathbf{v}_\mathrm{D}}{\tau}\right) = -e\mathbf{E}.$$

Here, $n_0 = \frac{N}{V}$ and $-e$ are the electron concentration and the charge on an electron. The thermal conductivity is given by

$$\kappa = \frac{w}{-\partial T/\partial x} = \frac{1}{3} n_0 v_\mathrm{T}^2 \tau \frac{dE}{dT} = \frac{1}{3} v_\mathrm{T}^2 \tau C_\mathrm{v},$$

where $C_\mathrm{v} = n_0 \frac{\mathrm{d}E}{\mathrm{d}T}$ is the electronic specific heat.

The electrical current density $\mathbf{j}$ and thermal current density $\mathbf{w}$ are given, in terms of distribution function f, by

$$\mathbf{j}(\mathbf{r},t) = \int (-\mathrm{e})\mathbf{v}\, f(\mathbf{r},\mathbf{v},t)\,\mathrm{d}^3 v \quad \text{and} \quad \mathbf{w}(\mathbf{r},t) = \int \varepsilon \mathbf{v}\, f(\mathbf{r},\mathbf{v},t)\,\mathrm{d}^3 v.$$

In the Sommerfeld model, states are labeled by $\{\mathbf{k},\sigma\} = (k_x, k_y, k_z)$ and σ, where σ is a spin index. The *Fermi energy* $\varepsilon_\mathrm{F}\,(\equiv \Theta_\mathrm{F})$, *Fermi velocity* v_F, and *Fermi temperature* $T_\mathrm{F}\left(= \frac{\Theta_\mathrm{F}}{k_\mathrm{B}}\right)$ are defined, respectively, by

$$\varepsilon_\mathrm{F} = \frac{\hbar^2 k_\mathrm{F}^2}{2m} = \frac{1}{2} m v_\mathrm{F}^2 = \Theta_\mathrm{F},$$

where the Fermi wave number k_F is related to the carrier concentration n_0 by $k_\mathrm{F}^3 = 3\pi^2 n_0$. The density of states of an electron gas is

$$g(\varepsilon) = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} \varepsilon^{1/2}.$$

For electrons moving in a periodic potential, $g(\varepsilon)$ does not have such a simple form. At a finite temperature, the chemical potential ζ is determined from

$$N = V \int_0^\infty g(\varepsilon) f_0(\varepsilon) \mathrm{d}\varepsilon.$$

The internal energy U is given by

$$\frac{U}{V} = u = \int_0^\infty \mathrm{d}\varepsilon\, \varepsilon g(\varepsilon) f_0(\varepsilon).$$

These integrals are of the form $I = \int_0^\infty \mathrm{d}\varepsilon\, f_0(\varepsilon) \frac{\mathrm{d}F(\varepsilon)}{\mathrm{d}\varepsilon}$. At low temperatures, we have, to order Θ^2,

$$I = F(\zeta) + \frac{\pi^2}{6}\Theta^2 F''(\zeta).$$

The electronic heat capacity $C_\mathrm{v} = \left(\frac{\partial U}{\partial T}\right)_V$ is given, at low temperature, by $C_\mathrm{v} = \gamma T$, where $\gamma = \frac{\pi^2 k_\mathrm{B}^2}{2\zeta_0} N$ for free electrons.

The electrical and thermal current densities j_x and w_x are, respectively, written as

$$j_x = \left[e^2 E + e\Theta \frac{\partial}{\partial x}\left(\frac{\zeta}{\Theta}\right)\right] \mathcal{K}_1 + \frac{e}{\Theta}\frac{\partial \Theta}{\partial x}\mathcal{K}_2$$

and

$$w_x = -\left[eE + \Theta \frac{\partial}{\partial x}\left(\frac{\zeta}{\Theta}\right)\right] \mathcal{K}_2 - \frac{1}{\Theta}\frac{\partial \Theta}{\partial x}\mathcal{K}_3.$$

where

$$\mathcal{K}_n = \frac{n_0}{m\zeta_0^{3/2}} \int_0^\infty \mathrm{d}\varepsilon \, \left(-\frac{\partial f_0}{\partial \varepsilon}\right) \varepsilon^{n+1/2}\tau.$$

The function $\mathcal{K}_n$ is given by

$$\mathcal{K}_n = \frac{n_0}{m\zeta_0^{3/2}} \left[\zeta^{n+1/2}\tau(\zeta) + \frac{\pi^2}{6}\Theta^2 \frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2}\left(\varepsilon^{n+1/2}\tau(\varepsilon)\right) \mid_{\varepsilon=\zeta}\right].$$

The electrical and thermal conductivities are given, in terms of $\mathcal{K}_1$, by $\sigma = e\mathcal{K}_1$ and $\kappa_T = k_\mathrm{B}\frac{\mathcal{K}_3\mathcal{K}_1 - \mathcal{K}_2^2}{\mathcal{K}_1\Theta}$. The Sommerfeld expression for κ_T is $\kappa_T = \frac{\pi^2}{3} k_\mathrm{B}^2 \frac{n_0\tau}{m} T$.

In the presence of an electric field $\mathbf{E}$ and a dc magnetic field $\mathbf{B}$, the magnetoconductivity tensor has nonzero components, for the case $\mathbf{B}$ along the z-axis, as follows: $\sigma_{xx} = \sigma_{yy} = \frac{\sigma_0(1+\mathrm{i}\omega\tau)}{(1+\mathrm{i}\omega\tau)^2 + (\omega_\mathrm{c}\tau)^2}$, $\sigma_{xy} = -\sigma_{yx} = \frac{\sigma_0(-\omega_\mathrm{c}\tau)}{(1+\mathrm{i}\omega\tau)^2 + (\omega_\mathrm{c}\tau)^2}$, $\sigma_{zz} = \frac{\sigma_0}{1+\mathrm{i}\omega\tau}$. Here $\omega_\mathrm{c} = \frac{eB}{mc}$ and $\sigma_0 = \frac{n_0 e^2 \tau}{m}$ is just the Drude's dc conductivity.

The electrical current density $\mathbf{j}$ can be thought of as the time rate of change of the polarization $\mathbf{P}$, that is, $\mathbf{j} = \dot{\mathbf{P}} = \mathrm{i}\omega\mathbf{P}$, where $\mathbf{P} = \frac{\varepsilon - 1}{4\pi}\mathbf{E}$ and $\mathbf{D}$, $\mathbf{P}$, and $\mathbf{E}$ are assumed to vary as $\mathrm{e}^{\mathrm{i}\omega t}$. Hence, we have the relation

$$\varepsilon(\omega) = 1 - \frac{4\pi\mathrm{i}}{\omega}\sigma(\omega).$$

$\varepsilon(\omega)$ has real and imaginary parts, ε_1 and ε_2, respectively, and in the Drude model, we have $\varepsilon_1(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + 1/\tau^2}$ and $\varepsilon_2(\omega) = -\frac{\omega_p^2/\omega\tau}{\omega^2 + 1/\tau^2}$. The two Maxwell equations $\nabla \times \mathbf{E} = -\frac{1}{c}\dot{\mathbf{B}}$ and $\nabla \times \mathbf{B} = \frac{1}{c}\underline{\varepsilon} \cdot \dot{\mathbf{E}}$ can be combined to obtain the wave equation

$$\mathbf{q}(\mathbf{q} \cdot \mathbf{E}) - q^2\mathbf{E} + \frac{\omega^2}{c^2}\underline{\varepsilon} \cdot \mathbf{E} = 0.$$

4

Elements of Band Theory

4.1 Energy Band Formation

So far, we have completely neglected the effects of the ion cores on the motion of the valence electrons. We consider "valence electrons" to be those outside of a closed shell configuration, so that

1. Na has a single $3s$ valence electron outside a [Ne] core.
2. Mg has two $3s$ electrons outside a [Ne] core.
3. Ga has ten $3d$ electrons, two $4s$ electrons, and one $3p$ electrons outside an [Ar] core.

The s and p electrons are usually considered as the "valence" electrons, since they are responsible for "bonding". Sometimes the mixing of d-electron atomic states with "valence" electron states is important.

To get some idea about the potential due to the ion cores let's consider the simple case of an isolated Na^+ ion. This ion has charge $+e$ and attracts an electron via the Coulomb potential. (See Fig. 4.1.)

$$V(r) = -\frac{e^2}{r} \quad \text{if } r > \text{ion radius.} \tag{4.1}$$

For a pair of Na atoms separated by a large distance, each "conduction electron" ($3s$-electron in Na) has a well-defined atomic energy level. As the two atoms are brought closer together, the atomic potentials $V(r)$ begin to overlap. Then, each electron can feel the potential of both ions. This gives rise a splitting of the degeneracy of atomic levels.

For a large number of atoms, the same effect occur. Think of a crystal structure with a nearest neighbor separation of 1 cm. The energy levels of the system will be atomic in character. However, as we decrease the nearest neighbor separation the atomic energy levels will begin to broaden into bands. (See Fig. 4.2.) The equilibrium separation of the crystal is the position at which the total energy of the system is a minimum. In all crystalline solids the

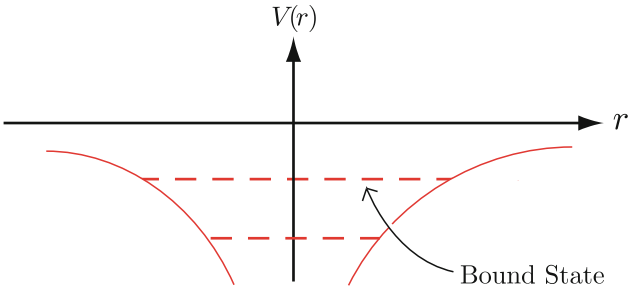


Fig. 4.1. Coulomb potential

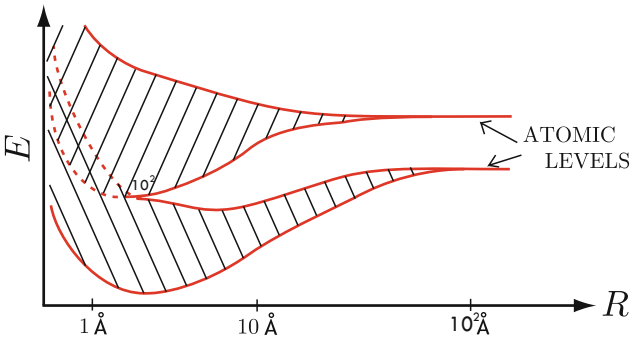


Fig. 4.2. Schematic illustration of energy band formation

electronic energies form bands of allowed energy values separated by energy gaps (bands of forbidden energy values). These energy bands determine the electrical properties of the solid.

4.2 Translation Operator

Because the crystalline potential seen by a single electron in a solid is a periodic function of position, with the period of the lattice, it is useful to introduce a translation operator T defined by

$$Tf(x) = f(x + a), \tag{4.2}$$

where $f(x)$ is an arbitrary function of position and a is the period of the lattice. It is clear that T commutes with the single particle Hamiltonian H

$$H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x), \tag{4.3}$$

because if we let $x' = x + a$, we can see that $\partial/\partial x' = \partial/\partial x$ and $V(x') = V(x)$.

One of the most useful theorems of linear algebra for the study of quantum systems states that if two operators commute, one can find common eigenfunctions for them (i.e, they can both be diagonal in the same representation). Let Ψ be an eigenfunction of H and of T

$$H\Psi = E\Psi \quad \text{and} \quad T\Psi = \lambda\Psi. \tag{4.4}$$

Here, E and λ are eigenvalues. Clearly applying T to Ψ N times gives

$$T^N\Psi(x) = \lambda^N\Psi(x) = \Psi(x + Na). \tag{4.5}$$

If we apply periodic boundary conditions with period N, then $\Psi(x + Na) = \Psi(x)$. This implies that

$$T^N\Psi(x) = \Psi(x), \tag{4.6}$$

or that $\lambda^N = 1$. Thus, λ itself must be the Nth root of unity

$$\lambda = \mathrm{e}^{\mathrm{i}\frac{2\pi}{N}n}, \tag{4.7}$$

where $n = 0, \ \pm 1, \ldots$. We can write λ as

$$\lambda = \mathrm{e}^{\mathrm{i}ka}, \tag{4.8}$$

where $k = \frac{2\pi}{Na} \times n$. Then, it is apparent that two values of k which differ by $2\pi/a$ times an integer give identical values of λ. As usual, we choose the N independent values of k to lie in the range $-\pi/a < k \leq \pi/a$, the first Brillouin zone of a one dimensional crystal.

For more than one dimension, $T_{\mathbf{R}}$ translates through a lattice vector $\mathbf{R}$

$$T_{\mathbf{R}}\Psi(r) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}}\Psi(r), \tag{4.9}$$

where $\mathbf{k} = \frac{2\pi}{N}(n_1\mathbf{b}_1 + n_2\mathbf{b}_2 + n_3\mathbf{b}_3)$. Here n_1, n_2, n_3 are integers and $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ are primitive translations of the reciprocal lattice. We have assumed a period N for periodic boundary condition with $L_1 = N\mathbf{a}_1$, $L_2 = N\mathbf{a}_2$, $L_3 = N\mathbf{a}_3$, and values of n_1, n_2, n_3 are chosen to restrict k to the first Brillouin zone.

4.3 Bloch's Theorem

We have just demonstrated that for a one-dimensional crystal with N-atoms and periodic boundary conditions

$$T\Psi(x) \equiv \Psi(x + a) = \mathrm{e}^{\mathrm{i}ka}\Psi(x), \tag{4.10}$$

where the N independent values of k are restricted to the first Brillouin zone. We can define a function

$$u_k(x) = \mathrm{e}^{-\mathrm{i}kx}\Psi(x). \tag{4.11}$$

It is apparent that

$$Tu_k(x) = T\{\mathrm{e}^{-\mathrm{i}kx}\Psi(x)\} = \mathrm{e}^{-\mathrm{i}k(x+a)}\Psi(x+a) = \mathrm{e}^{-\mathrm{i}kx}\Psi(x) = u_k(x).$$

Therefore, we can write

$$\Psi(x) = \mathrm{e}^{\mathrm{i}kx}u_k(x), \tag{4.12}$$

where $u_k(x)$ is a periodic function i.e. $u_k(x+a) = u_k(x)$. This is known to physicists as Bloch's theorem, although it had been proven sometime earlier than Bloch[1] and is known to mathematicians as Floquet's theorem.

4.4 Calculation of Energy Bands

There are two very different starting points from which one can approach energy bands in solids. The first approach is to start with atomic orbitals and to form linear combinations which satisfy Bloch's theorem. The second is to start with a Sommerfeld free electron gas picture (for the electrons outside a closed shell core) and to see how the periodic potential of ions changes the $\varepsilon_k = k^2\hbar^2/2m$ free electron dispersion. The first approach works well for systems of rather tightly bound electrons, while the second works well for weakly bound electrons. We will spend a good deal of time on the "nearly free electron" model and how group theory helps to make the calculations easier. Before doing that, we begin with the first approach called the "tight-binding method" or the LCAO (linear combination of atomic orbitals).

4.4.1 Tight-Binding Method

Suppose that a free atom has a potential $V_a(r)$, so that a "conduction electron," like the $3s$ electron in sodium, satisfies the Schrödinger equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_a(r) - E_a\right)\phi(r) = 0. \tag{4.13}$$

Here, E_a is the atomic energy level of this conduction electron. When atoms form a crystal, the potentials of the individual atoms overlap, as indicated schematically in Fig. 4.3.

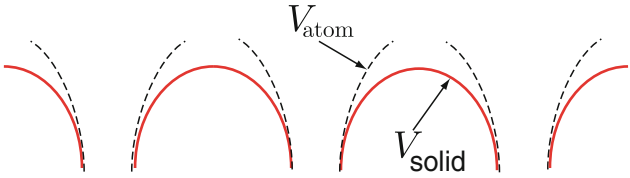


Fig. 4.3. Tight-Binding Potential

[1] Felix Bloch, Z. Physik **52**, 555 (1928).

In the tight-binding approximation, one assumes that the electron in the unit cell about R_j is only slightly influenced by atoms other than the one located at R_j. Its wave function in that cell will be close to $\phi(r - R_j)$, the atomic wave function, and its energy close to E_a. One can make a linear combination of atomic orbitals $\phi(r - R_j)$ as a trial function for the electronic wave function in the solid.

To satisfy Bloch's theorem, we can write

$$\Psi_k(r) = \frac{1}{\sqrt{N}} \sum_j \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}_j} \phi(r - R_j). \tag{4.14}$$

Clearly, the translation operator operating on $\Psi_k(r)$ gives

$$\begin{aligned} T_{R_n}\Psi_k(r) &= \Psi_k(r + R_n) \\ &= \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}_n} \frac{1}{\sqrt{N}} \sum_j \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{R}_j - \mathbf{R}_n)} \phi(r - R_j + R_n) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}_n} \Psi_k(r). \end{aligned}$$

The energy of a state $\Psi_k(r)$ is given by

$$E_k = \frac{\langle \Psi_k | H | \Psi_k \rangle}{\langle \Psi_k | \Psi_k \rangle}, \tag{4.15}$$

where H is the Hamiltonian for an electron in the crystal, and

$$\langle \Psi_k | \Psi_k \rangle = \frac{1}{N} \sum_{j,m} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{R}_j - \mathbf{R}_m)} \int \mathrm{d}^3 r \phi^*(r - R_m) \phi(r - R_j). \tag{4.16}$$

If we neglect overlap between $\phi(r - R_j)$ and $\phi(r - R_m)$, the $\mathrm{d}^3 r$ integration gives $\delta_{j,m}$ and the sum over j simply gives a factor N, the number of atoms in the crystal.

The Hamiltonian for an electron in the solid contains the potential $V(r)$. Let's write

$$V(r) = \tilde{V}(r - R_j) + V_a(r - R_j). \tag{4.17}$$

In other words, $\tilde{V}(r - R_j)$ is the full potential of the solid minus the potential of an atom located at R_j. It's clear from Fig. 4.3 that V_a is larger than $V(r)$ in the cell containing R_j so that $\tilde{V}(r - R_j)$ is negative. Since

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V_a(r - R_j) - E_a \right] \phi(r - R_j) = 0, \tag{4.18}$$

$$E_k = \frac{1}{N} \sum_j \sum_m \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(R_j - R_m)} \int \mathrm{d}^3 r \phi^*(r - R_m) \left[E_a + \tilde{V}(r - R_j) \right] \phi(r - R_j). \tag{4.19}$$

In the first term, E_a is the constant value of the atomic energy level and it can be taken out of the integration. All that remains in the integral is $\langle \Psi_k | \Psi_k \rangle$ which is 1, so the first term is just E_a. We can define

$$\alpha = -\int \mathrm{d}^3 r \phi^*(r - R_j)\tilde{V}(r - R_j)\phi(r - R_j), \tag{4.20}$$

and

$$\gamma = -\int \mathrm{d}^3 r \phi^*(r - R_m)\tilde{V}(r - R_j)\phi(r - R_j). \tag{4.21}$$

In the definition of γ we assume that the only terms that are not negligible are terms in which R_m is a nearest neighbor of R_j. Then, we have

$$E_k = E_a - \alpha - \gamma \sum_m{}' \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{R}_j - \mathbf{R}_m)} \tag{4.22}$$

where the sum is over all nearest neighbors of $\mathbf{R}_j$. We chose minus signs in the definition of α and γ to make α and γ positive (since $\tilde{V}(r - R_j)$ is negative).

Now look at what happens for a simple cubic lattice. There are six nearest neighbors of $\mathbf{R}_j$ located at $\mathbf{R}_j \pm a\hat{x}$, $\mathbf{R}_j \pm a\hat{y}$, and $\mathbf{R}_j \pm a\hat{z}$. Substituting into (4.22) gives

$$E_k = E_a - \alpha - 2\gamma\left(\cos k_x a + \cos k_y a + \cos k_z a\right). \tag{4.23}$$

Because γ is positive

$$E_k^{\mathrm{Min}} = E_a - \alpha - 6\gamma, \tag{4.24}$$

and

$$E_k^{\mathrm{Max}} = E_a - \alpha + 6\gamma. \tag{4.25}$$

The result is sketched in Fig. 4.4.

For $k \ll \pi/a$

$$\begin{aligned} E_k &\simeq E_a - \alpha - 6\gamma + \gamma a^2 k^2; \quad k^2 = k_x^2 + k_y^2 + k_z^2 \\ &= E_k^{\mathrm{Min}} + \frac{\hbar^2 k^2}{2m^*}. \end{aligned}$$

The effective mass $m^* = \frac{\hbar^2}{2\gamma a^2}$. As γ decreases, the band width ΔE gets smaller and the effective mass near $E = E_k^{\mathrm{Min}}$ increases.

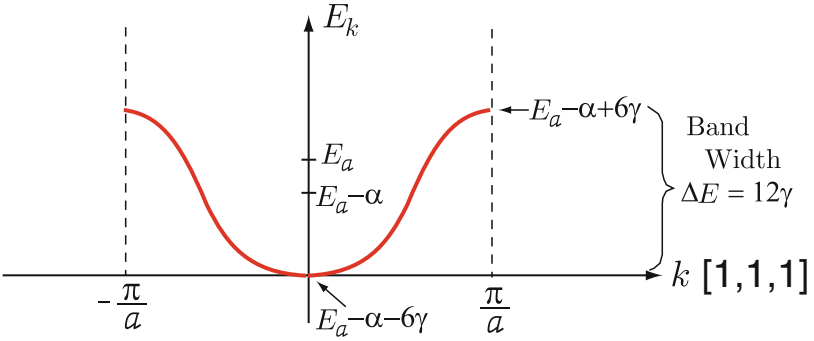


Fig. 4.4. Tight-binding dispersion for simple cubic lattice

4.4.2 Tight Binding in Second Quantization Representation

Suppose a system of free electrons is described by the Hamiltonian

$$H_0 = \sum_k \varepsilon_k c_k^\dagger c_k, \tag{4.26}$$

where $\varepsilon_k = \frac{\hbar^2 k^2}{2m}$ is the kinetic energy. In the presence of a periodic potential $V(\mathbf{r}) = \sum_{\mathbf{K}} V_{\mathbf{K}} \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}}$, we can write the potential energy of the electrons as

$$H' = \sum_{k,K} V_K c_{k+K}^\dagger c_k. \tag{4.27}$$

Now, introduce the operators c_n and $c_n^\dagger$ which annihilate or create electrons at site R_n^0.

$$c_n = \frac{1}{\sqrt{N}} \sum_k c_k \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}_n^0}. \tag{4.28}$$

The inverse transformation is

$$c_k = \frac{1}{\sqrt{N}} \sum_n c_n \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_n^0}. \tag{4.29}$$

Substitute the latter equation into H_0 to obtain

$$H_0 = \sum_k \varepsilon_k \sum_{nm} \frac{1}{N} c_n^\dagger c_m \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{R}_n^0 - \mathbf{R}_m^0)}. \tag{4.30}$$

Define

$$T_{nm} = \frac{1}{N} \sum_k \varepsilon_k \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{R}_n^0 - \mathbf{R}_m^0)}. \tag{4.31}$$

Then

$$H_0 = \sum_{nm} T_{nm} c_n^\dagger c_m. \tag{4.32}$$

Now, look at H'

$$\begin{aligned} H' &= \sum_{\mathbf{k},\mathbf{K}} V_K \frac{1}{N} \sum_{nm} c_n^\dagger \mathrm{e}^{\mathrm{i}(\mathbf{k}+\mathbf{K})\cdot\mathbf{R}_n^0} c_m \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_m^0} \\ &= \sum_{\mathbf{K}nm} \left[\sum_k \frac{1}{N} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{R}_n^0 - \mathbf{R}_m^0)} \right] V_K \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{R}_n^0} c_n^\dagger c_m. \end{aligned}$$

Since $\frac{1}{N} \sum_k \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{R}_n^0 - \mathbf{R}_m^0)} = \delta_{nm}$, we have

$$H' = \sum_{\mathbf{K},n} V_K \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{R}_n^0} c_n^\dagger c_n. \tag{4.33}$$

But we note that $\sum_K V_K e^{i\mathbf{K}\cdot\mathbf{R}_n^0} = V(R_n^0)$, and hence H' becomes

$$H' = \sum_n V(R_n^0) c_n^\dagger c_n. \tag{4.34}$$

Adding H_0 and H' gives

$$\begin{aligned} H &= \sum_n \left(T_{nn} + V(R_n^0)\right) c_n^\dagger c_n + \sum_{n\neq m} T_{nm} c_n^\dagger c_m \\ &= \sum_n \varepsilon_n c_n^\dagger c_n + \sum_{n\neq m} T_{nm} c_n^\dagger c_m, \end{aligned} \tag{4.35}$$

where $\varepsilon_n = T_{nn} + V(R_n^0)$ represents an energy on site n and T_{nm} denotes the amplitude of hopping from site m to site n. Starting with atomic levels ε_n and allowing hopping to neighboring sites results in energy bands, and the band width depends on the hopping amplitude T_{nm}. Later we will see that starting with free electrons and adding a periodic potential $V(r) = \sum_K V_K e^{i\mathbf{K}\cdot\mathbf{r}}$ also results in energy bands. The band gaps between bands depend on the Fourier components V_K of the periodic potential.

4.5 Periodic Potential

Because the potential experienced by an electron is periodic with the period of the lattice, it can be expanded in a Fourier series

$$V(\mathbf{r}) = \sum_{\mathbf{K}} V_{\mathbf{K}} e^{i\mathbf{K}\cdot\mathbf{r}}, \tag{4.36}$$

where the sum is over all vectors $\mathbf{K}$ of the reciprocal lattice, and

$$V_{\mathbf{K}} = \frac{1}{\Omega} \int d^3r\, V(\mathbf{r})\, e^{-i\mathbf{K}\cdot\mathbf{r}}.$$

For any reciprocal lattice vector $\mathbf{K}$

$$\mathbf{K}\cdot\mathbf{R} = 2\pi \times \text{integer},$$

if $\mathbf{R}$ is any translation vector of the lattice. Thus

$$\begin{aligned} V(\mathbf{r}+\mathbf{R}) &= \sum_{\mathbf{K}} V_{\mathbf{K}} e^{i\mathbf{K}\cdot(\mathbf{r}+\mathbf{R})}, \\ &= \sum_{\mathbf{K}} V_{\mathbf{K}} e^{i\mathbf{K}\cdot\mathbf{r}} = V(\mathbf{r}). \end{aligned}$$

The periodic part of the Bloch function can also be expanded in Fourier series. We can write

$$u_n(\mathbf{k},\mathbf{r}) = \sum_{\mathbf{K}} C_{\mathbf{K}}(n,\mathbf{k}) \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}}, \tag{4.37}$$

For the moment, let us omit the band index n and wave number $\mathbf{k}$ and simply write

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} u(\mathbf{r}) = \sum_{\mathbf{K}} C_{\mathbf{K}} \mathrm{e}^{\mathrm{i}(\mathbf{k}+\mathbf{K})\cdot\mathbf{r}}. \tag{4.38}$$

Use the Fourier expansion of $V(\mathbf{r})$ and $u(\mathbf{r})$ in the Schrödinger equation; this gives

$$\sum_{\mathbf{K}'} \Big[\frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{K}')^2 + \sum_{\mathbf{K}''} V_{\mathbf{K}''} \mathrm{e}^{i\mathbf{K}''\cdot\mathbf{r}}\Big] C_{K'} \mathrm{e}^{i(\mathbf{k}+\mathbf{K}')\cdot\mathbf{r}} = E \sum_{\mathbf{K}'} C_{\mathbf{K}'} \mathrm{e}^{i(\mathbf{k}+\mathbf{K}')\cdot\mathbf{r}}. \tag{4.39}$$

We multiply by $\mathrm{e}^{-\mathrm{i}(\mathbf{k}+\mathbf{K})\cdot\mathbf{r}}$ and integrate recalling that $\int \mathrm{d}^3 r \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}} = \Omega\delta(\mathbf{K})$ where Ω is the volume. This gives

$$\left[E - V_{\mathbf{0}} - \frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{K})^2\right] C_{\mathbf{K}} = \sum_{\mathbf{H}\neq 0} V_{\mathbf{H}} C_{\mathbf{K}-\mathbf{H}} \tag{4.40}$$

Here, we have set $\mathbf{K}'' = \mathbf{H}$ and have separated the $\mathbf{H} = 0$ term from the other terms in the potential. This is an infinite set of linear equations for the coefficients $C_{\mathbf{K}}$. The non-trivial solutions are obtained by setting the determinant of the matrix multiplying the column vector

$$\begin{pmatrix} \vdots \\ C_{\mathbf{K}} \\ \vdots \end{pmatrix}$$

equal to zero. The roots give the energy levels (an infinite number – one for each value of $\mathbf{K}$) in the periodic potential of a crystal. We can express the infinite set of linear equations in the following matrix notation:

$$\begin{pmatrix} \varepsilon_{\mathbf{K}_1} + \langle \mathbf{K}_1 | V | \mathbf{K}_1 \rangle - E & \langle \mathbf{K}_1 | V | \mathbf{K}_2 \rangle \cdots & \langle \mathbf{K}_1 | V | \mathbf{K}_n \rangle \cdots \\ \langle \mathbf{K}_2 | V | \mathbf{K}_1 \rangle & \varepsilon_{\mathbf{K}_2} + \langle \mathbf{K}_2 | V | \mathbf{K}_2 \rangle - E \cdots & \langle \mathbf{K}_2 | V | \mathbf{K}_n \rangle \cdots \\ \vdots & \vdots & \vdots \\ \langle \mathbf{K}_n | V | \mathbf{K}_1 \rangle & \langle \mathbf{K}_n | V | \mathbf{K}_2 \rangle \cdots & \varepsilon_{\mathbf{K}_n} + \langle \mathbf{K}_n | V | \mathbf{K}_n \rangle - E \cdots \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} C_{\mathbf{K}_1} \\ C_{\mathbf{K}_2} \\ \vdots \\ C_{\mathbf{K}_n} \\ \vdots \end{pmatrix} = 0 \tag{4.41}$$

Here, $\varepsilon_{\mathbf{K}} = \frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{K})^2$ and $\langle \mathbf{K} | V | \mathbf{K}' \rangle = V_{\mathbf{K}-\mathbf{K}'}$, where $|\mathbf{K}\rangle = \frac{1}{\sqrt{\Omega}} \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}}$. The object of energy band theory is to obtain a good approximation to $V(r)$ and to solve this infinite set of equations in an approximate way.

4.6 Free Electron Model

If $V_{\mathbf{K}} = 0$ for all $\mathbf{K} \neq 0$, then in the notation used earlier

$$(E - V_0 - \varepsilon_{\mathbf{k}+\mathbf{K}})\, C_{\mathbf{K}} = 0. \tag{4.42}$$

This is exactly the Sommerfeld model of free electrons in a constant potential V_0. We can write

$$\begin{aligned} E^{(0)}_{\mathbf{K}} &= V_0 + \varepsilon_{\mathbf{k}+\mathbf{K}} \\ C^{(0)}_{\mathbf{K}} &= \begin{cases} 1 & \text{For the band } \mathbf{K} \\ 0 & \text{For all other bands} \end{cases} \end{aligned} \tag{4.43}$$

Let us discuss this in detail by considering a simple one-dimensional case, for example, as shown in Fig. 4.5. The allowed values of the Bloch wave vector $\mathbf{k}$ are restricted to the first Brillouin zone. Values of $\mathbf{k}$ outside the first Brillouin zone are obtained by adding a reciprocal lattice vector $\mathbf{K}$ to $\mathbf{k}$. The labels C_n refer to the non-zero coefficients for that particular band; e.g. for C_{+2} we have

$$\left.\begin{aligned} E^{(0)}_2 &= V_0 + \tfrac{\hbar^2}{2m}\left[k + 2\left(\tfrac{2\pi}{a}\right)\right]^2 \\ \Psi^{(0)}_{n\mathbf{k}}(\mathbf{r}) &= \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} u^{(0)}_{n\mathbf{k}}(\mathbf{r}). \end{aligned}\right\} \tag{4.44}$$

But $u_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{K}} C_{\mathbf{K}}(\mathbf{k})\mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}}$; with $C^{(0)}_{\mathbf{K}}(\mathbf{k}) = 1$ for $|\ \mathbf{K}\ | = K_n = \left(\frac{2\pi}{a}\right)\cdot n$, we have $\Psi^{(0)}_{n\mathbf{k}}(r) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \cdot \mathrm{e}^{\mathrm{i}\mathbf{K}_n\cdot\mathbf{r}}$. All of this is simply a restatement of the free electron model in the "reduced" zone scheme (i.e. all Bloch $\mathbf{k}-$vectors are in the first Brillouin zone, but energies of higher bands are obtained by adding reciprocal lattice vectors $\mathbf{K}$ to $\mathbf{k}$; the periodic part of the Bloch function is $u_K = \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}}$).

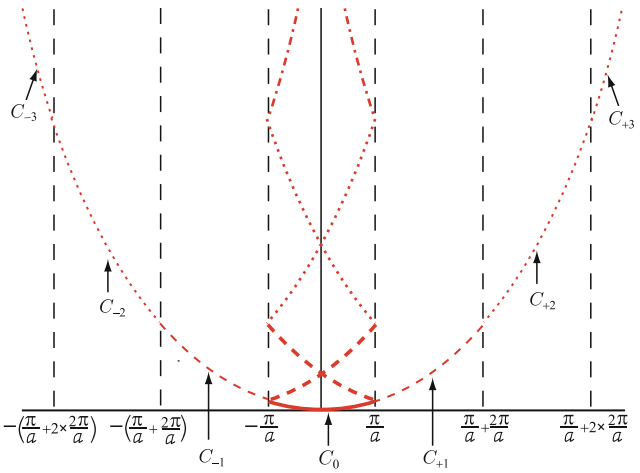


Fig. 4.5. One dimensional free electron band

4.7 Nearly Free Electron Model

If we take $V_{\mathbf{K}}$ for $\mid \mathbf{K} \mid \neq 0$ to be very small but non-zero, we can use "perturbation theory" to solve the infinite set of coupled equations approximately.

For the lowest band (the one with $C_0^{(0)} = 1$) we know that in zeroth order (i.e. with $V_{\mathbf{K}} = 0$ for $\mid \mathbf{K} \mid \neq 0$)

$$E^{(0)} = V_0 + \frac{\hbar^2 k^2}{2m},$$
$$C_0^{(0)} = 1 \text{ and } C_{\mathbf{K}}^{(0)} = 0 \text{ for } \mid \mathbf{K} \mid \neq 0.$$

Let us look at other values of $C_{\mathbf{K}}$ (i.e. not $\mid \mathbf{K} \mid = 0$ value) for $V_{\mathbf{K}} \neq 0$ (but very small) when $\mid \mathbf{K} \mid \neq 0$. The first order correction to $C_{\mathbf{K}}^{(0)}$ is given by

$$[E - V_0 - \frac{\hbar^2}{2m}(\mathbf{k} + \mathbf{K})^2] C_{\mathbf{K}}^{(1)} = \sum_{|H| \neq 0} C_{\mathbf{K}-\mathbf{H}} V_{|H|}. \tag{4.45}$$

On the right-hand side all the $V_{|\mathbf{H}|}$ appearing are small; therefore to first order we can use for $C_{\mathbf{K}-\mathbf{H}}$ the value $C_{\mathbf{K}-\mathbf{H}}^{(0)}$ which equals unity for $\mathbf{K} - \mathbf{H} = 0$ and zero otherwise. Solving for $C_{\mathbf{K}}^{(1)}$ gives

$$C_{\mathbf{K}}^{(1)} = \frac{V_{\mathbf{K}}}{\frac{\hbar^2}{2m}[k^2 - (\mathbf{k} + \mathbf{K})^2]} \tag{4.46}$$

Here, we have used $E \simeq \frac{\hbar^2 k^2}{2m} + V_0$ for the zeroth order approximation to the energy of the lowest band (the one we are investigating). We substitute this result back into the equation for C_0, which is approximately equal to unity.

$$\left(E - V_{\mathbf{0}} - \frac{\hbar^2 k^2}{2m}\right) C_0 \simeq \sum_{|H| \neq 0} C_{\mathbf{0}-\mathbf{H}} V_{\mathbf{H}} = \frac{V_{-\mathbf{H}} V_{\mathbf{H}}}{\frac{\hbar^2}{2m}[k^2 - (\mathbf{k} - \mathbf{H})^2]} \tag{4.47}$$

We have that $C_0 = 1 + C_0^{(1)}$, but $C_0^{(1)}$ can be neglected since the right-hand side is already small. Setting $C_0 \simeq 1$ and solving for E gives

$$E = V_{000} + \varepsilon_{\mathbf{k}} - \sum_{|\mathbf{K}| \neq 0} \frac{|V_{\mathbf{K}}|^2}{\varepsilon_{\mathbf{k}+\mathbf{K}} - \varepsilon_{\mathbf{k}}} \tag{4.48}$$

In this equation, we have used $V_{-\mathbf{H}} = V_{\mathbf{H}}^*$ and let $-\mathbf{H} = \mathbf{K}$. As long as $\mid \varepsilon_{\mathbf{k}+\mathbf{K}} - \varepsilon_{\mathbf{k}} \mid \gg \mid V_{\mathbf{K}} \mid$, this perturbation expansion is rather good. It clearly breaks down when

$$\varepsilon_{\mathbf{k}+\mathbf{K}} = \varepsilon_{\mathbf{k}}$$

or

$$|\mathbf{k} + \mathbf{K}| = |\mathbf{k}|\,.$$

This is exactly the condition for a Bragg reflection; when $\mathbf{k}' - \mathbf{k} = \mathbf{K}$ we get Bragg reflection.

4.7.1 Degenerate Perturbation Theory

Suppose that for some particular reciprocal lattice vector $\mathbf{K}$

$$|\mathbf{k}+\mathbf{K}| \simeq |\mathbf{k}| \tag{4.49}$$

Our simple perturbation theory result gave

$$C_{\mathbf{K}}^{(1)} = \frac{V_{\mathbf{K}}}{\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{K}}}.$$

It is clear that this result is inconsistent with our starting assumption that $C_{\mathbf{K}}$ was very small except for $\mathbf{K} = 0$. To remedy this shortcoming we assume that both C_0 and $C_{\mathbf{K}}$ (for the particular value satisfying $|\mathbf{k}| = |\mathbf{k}+\mathbf{K}|$) are important. This assumption gives us a pair of equations

$$\begin{aligned} (E - V_0 - \varepsilon_{\mathbf{k}})\, C_0 &= C_{\mathbf{K}} V_{-\mathbf{K}} \\ (E - V_0 - \varepsilon_{\mathbf{k}+\mathbf{K}})\, C_{\mathbf{K}} &= C_0 V_{\mathbf{K}}. \end{aligned} \tag{4.50}$$

The solutions are obtained by setting the determinant of the matrix multiplying the column vector $\begin{pmatrix} C_0 \\ C_{\mathbf{K}} \end{pmatrix}$ equal to zero. The two roots are given by

$$E_{\pm}(k) = V_0 + \frac{1}{2}[\varepsilon_{\mathbf{k}} + \varepsilon_{\mathbf{k}+\mathbf{K}}] \pm \left\{ |V_{\mathbf{K}}|^2 + \left(\frac{\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{K}}}{2} \right)^2 \right\}^{1/2} \tag{4.51}$$

For $|\mathbf{k}| = |\mathbf{k}+\mathbf{K}|$, $\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{K}} = 0$ and the solutions become

$$E_{\pm}(k) = V_0 + \varepsilon_{\mathbf{k}} \pm |V_{\mathbf{K}}| \tag{4.52}$$

This behavior is shown in Fig. 4.6. If we introduce $\mathbf{q} = \frac{\mathbf{K}}{2} + \mathbf{k}$, where $q \ll \frac{K}{2}$, we can expand the roots for small q and obtain

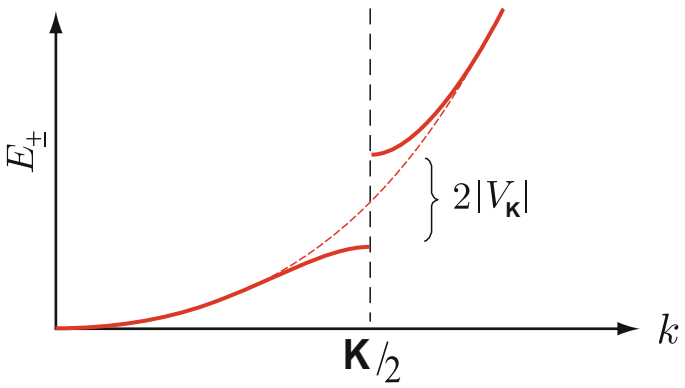


Fig. 4.6. Bandgap formation at the zone boundary $\mathbf{k} = \frac{\mathbf{K}}{2}$

$$E_{\pm} = V_0 + \varepsilon_{\frac{\mathbf{K}}{2}} + \varepsilon_q \pm |V_{\mathbf{K}}| \left\{ 1 + \frac{1}{2} \frac{\varepsilon_{\mathbf{K}}}{|V_{\mathbf{K}}|^2} \varepsilon_{\mathbf{q}} \right\}.$$

If we define

$$E_0 = V_0 + \varepsilon_{\frac{\mathbf{K}}{2}} - |V_{\mathbf{K}}|,$$

and choose zero of energy at E_0, the two roots can be written (for small q)

$$\begin{aligned} E_- &= \tfrac{\hbar^2 q^2}{2m_-^*} \\ E_+ &= E_{\mathrm{G}} + \tfrac{\hbar^2 q^2}{2m_+^*}. \end{aligned} \tag{4.53}$$

Here, the energy gap E_{G} is equal to $2\,|V_{\mathbf{K}}|$ and the *effective masses* $m_{\pm}^*$ are given by

$$m_{\pm}^* = \frac{m}{1 \mp \frac{\varepsilon_{\mathbf{K}}}{2|V_{\mathbf{K}}|}}. \tag{4.54}$$

It is common for $\varepsilon_{\mathbf{K}}$ to be larger than $2 \mid V_{\mathbf{K}} \mid$ so that m_-^* is negative. Then the two roots are commonly expressed as

$$\begin{aligned} E_{\mathrm{v}}(k) &= -\tfrac{\hbar^2 k^2}{2m_{\mathrm{v}}}, \\ E_{\mathrm{c}}(k) &= E_{\mathrm{G}} + \tfrac{\hbar^2 k^2}{2m_{\mathrm{c}}}, \end{aligned}$$

where $m_{\mathrm{v}} = -m_-^*$. These results are frequently used to describe the valence band and conduction band in semiconductors. The results are only valid near $q = 0$ since we expanded the original equations for small deviations q from the extrema. This result is called the *effective mass approximation*.

4.8 Metals–Semimetals–Semiconductors–Insulators

The very simple Bloch picture of energy bands and energy gaps allows us to understand in a qualitative way why some crystals are metallic, some insulating, and some in between. For a one-dimensional crystal there will be a gap separating every band (assuming that V_K is non-zero for all values of K). We know that the gaps occur when $|\mathbf{k}| = |\mathbf{k} + \mathbf{K}|$. This defines the first Brillouin zone's boundaries.

In more than one dimension, the highest energy levels in a lower band can exceed the energy at the bottom of a higher band. This is referred to as band overlap. It is illustrated for a two-dimensional square lattice in Fig. 4.7. The square represents the first Brillouin zone bounded by $|k_x| = \pi/a$ and $|k_y| = \pi/a$. The point Γ indicates the zone center of $\mathbf{k} = 0$. The points $\mathrm{X} = \left(\frac{\pi}{a}, 0\right)$ and $\mathrm{M} = \left(\frac{\pi}{a}, \frac{\pi}{a}\right)$ are particular $\mathbf{k}$-values lying on the zone boundary. The Δ and Σ are arbitrary points on the lines connecting $\Gamma \to \mathrm{X}$ and $\Gamma \to \mathrm{M}$, respectively. If we plot the energy along these lines, we obtain the result illustrated in Fig. 4.8. It is clear that if the gaps are not too large, the maximum energy in

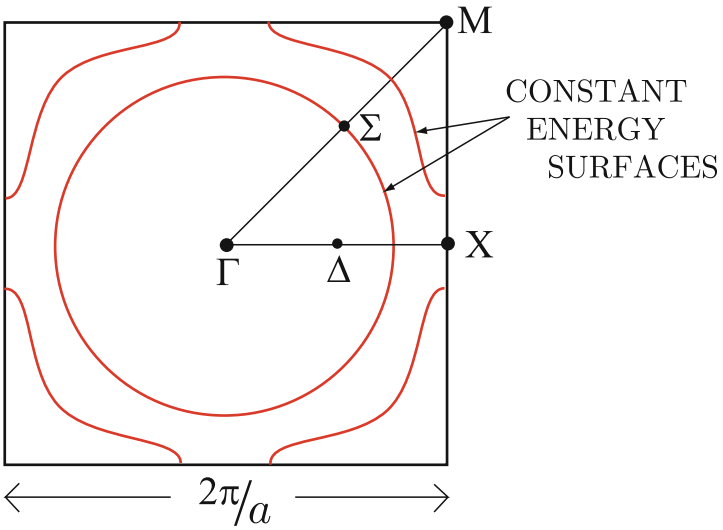


Fig. 4.7. Constant energy surface in two-dimensional square lattice

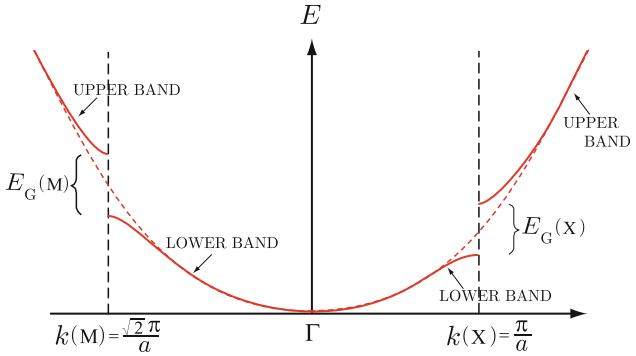


Fig. 4.8. Energy dispersion along the line M – Γ – X

the lower band $E_{\mathrm{LB}}(\mathrm{M})$ is higher than the minimum energy in the upper band $E_{\mathrm{UB}}(\mathrm{X})$. If we fill all the lowest energy states with electrons, being mindful of the exclusion principle, then it is clear that there will be some empty states in the lower band as we begin to fill the low-energy states of the upper band. The band overlap can be large (when the energy gaps are very small) or non-existent (when the energy gaps are very large).

The existence of

1. Band gaps in the energy spectrum at Brillouin zone boundaries
2. Band overlap in more than one dimension when energy gaps are small and
3. The Pauli exclusion principle allows us to classify solids as follows:

- Metal
  1. *Monovalent metal.* A material which contains one electron (outside a closed core) on each atom and one atom per unit cell. Na, K, Rb, Cs are good examples of monovalent metals. Because the total number of allowed **k** values in the first Brillouin zone is equal to N, the number of unit cells in the crystal, and because each **k**-state can accommodate one spin up and one spin down electron, each band can accommodate $2N$ electrons. A monovalent metal has N electrons, so the conduction band will be half filled. The Fermi energy is far (in comparison to $k_{\mathrm{B}}T$) from the band edges and band gaps. Therefore, the crystal acts very much like a Sommerfeld free electron model. The same picture holds for any odd valency solid containing 1, 3, 5, ... electrons per unit cell.
  2. Even-valency metals When the band gaps are very small, there can be a very large overlap between neighboring bands. The resultant solid will have a large number n_{h} of empty states in the lower band and an equal number n_{e} ($n_{\mathrm{e}} = n_{\mathrm{h}}$) of electrons in the higher band. If n_{e} and n_{h} are of the order of N, the number of unit cells, the material is metallic.
- *Insulator.* For a material with an even number of electrons per unit cell and a large gap ($\geq 4\mathrm{eV}$) between the highest filled state and the lowest empty state, an insulating crystal results. The application of a modest electric field cannot alter the electron distribution function because to do so would require a large energy E_{G}.
- *Semiconductor.* A material which is insulating at low temperature, but whose band gap E_{G} is small ($0.1\,\mathrm{eV} \leq E_{\mathrm{G}} \leq 2\,\mathrm{eV}$) is called a semiconductor. At finite temperature a few electrons will be excited from the filled valence band to the empty conduction band. These electrons and holes (empty states in the valence band) can carry current. Because the concentration of electrons in the conduction band varies like $\mathrm{e}^{-E_{\mathrm{G}}/2k_{\mathrm{B}}T}$, the conductivity increases with increasing temperature.
- *Semimetal.* These materials are even-valency materials with small band overlap. The number of electrons n_{e} equals the number of holes n_{h} but both are small compared to N, the number of unit cells in the crystal. Typically n_{e} and n_{h} might be 10^{-3} or 10^{-4} times the number of unit cells.

Examples

Monovalent metals Li, Na, K, Rb, Cs, Cu, Ag, Au
Divalent metals Zn, Cd, Ca, Mg, Ba
Polyvalent metals Al, Ga, In, Tl
Semimetals As, Sb, Bi, Sn, graphite
Insulators Al_2O_3, diamond
Semiconductors Ge, Si, InSb, GaAs, AlSb, InAs, InP

Problems

4.1. In an infinite linear chain of A and B atoms (...ABABAB...) with equal spacings R between each atom, the energies of electrons in the system are given by $E_k = \pm(\varepsilon^2 + 4\beta^2 \cos^2 kR)^{1/2}$, where k is the wavevector of the electron state. What is the band gap in the electronic band structure for this system? How would you expect the electrical and optical properties of this structure to depend on ε and β?

4.2. Consider a crystal of sodium with a volume $0.10\,\mathrm{cm}^3$, estimate the average spacing between the energy levels in the $3s$ band given that the $3s$ electron eneregies span a range of $3.20\,\mathrm{eV}$. The electron concentration of a sodium crystal is approximately $2.65 \times 10^{28}\,\mathrm{m}^{-3}$. (You can estimate this value by yourself.)

4.3. A BCC lattice has eight nearest neighbors at $\mathbf{r} = \frac{a}{2}(\pm\hat{x} \pm \hat{y} \pm \hat{z})$. Use the s-band tight-binding formula, (4.22), to evaluate $E(\mathbf{k})$.

4.4. A single graphite sheet (called graphene) has a honeycomb structure. Let us assume that there is one p_z orbital, which is oriented perpendicular to the sheet, on each carbon atom and forms the active valence and conduction bands of this two-dimensional crystal.

(a) Using the tight-binding method and only nearest-neighbor interactions, calculate and sketch the p-electron band structure $E_n(k)$ for graphene. You may assume the overlap matrix is the unity matrix.
(b) Show that this is a zero-gap semiconductor or a zero-overlap semimetal. Note that there is one p electron per carbon atom.
(c) Locate the position where the zero gap occurs in the momentum space.

4.5. A one-dimensional attractive potential is given by $V(z) = -\lambda\delta(z)$.

(a) Show that the lowest energy state occurs at $\varepsilon_0 = -\frac{\hbar^2 K^2}{2m}$, where $K = \frac{m\lambda}{\hbar^2}$.
(b) Determine the normalized wave function $\phi_0(z)$.

4.6. Consider a superlattice with $V(z) = -\lambda \sum_{l=-\infty}^{\infty} \delta(z - la)$.

(a) Obtain $\varepsilon_0(k_z)$, the energy of the lowest band as a function of k_z by using the tight-binding approximation (i.e. overlap between only neighboring sites is kept).
(b) Use the tight binding form of the wave function $\Psi_0(k_z, z)$ and show that it can be expressed as

$$\Psi_0(k_z, z) = \mathrm{e}^{\mathrm{i}k_z z}\, u(k_z, z),$$

where $u(k_z, z)$ is periodic with period a. Determine an expression for $u(k_z, z)$.

4.7. Let us consider electrons in a one-dimensional Bravais lattice described by the wave function and potential written as $\Psi(x) = \alpha e^{ikx} + \beta e^{i(k+G)x}$ and $V(x) = 2V_1 \cos(Gx)$. The zone boundary is at $k = G/2 = \pi/a$, where a is the lattice constant.

(a) Obtain the band structure by solving the Schrödinger equation, and sketch the band for $0 \leq k \leq G$ for $V_1 = 0$ and $V_1 = 0.2\hbar^2 G^2/2m$.
(b) What kind of material do we have when $V_1 = 0$?
(c) If $V_1 = 0.2\hbar^2 G^2/2m$ and each atom has three electrons, what kind of material do we have? Explain the reason.

Summary

In this chapter, we studied the electronic states from the consideration of the periodicity of the crystal structure. In the presence of periodic potential the electronic energies form bands of allowed energy values separated by bands of forbidden energy values. Bloch functions were introduced as a consequence of the translational symmetry of the lattice. Energy bands obtained by a simplest tight-binding method and nearly free electron model are discussed.

The eigenfunction of the Hamiltonian H can be written as

$$\Psi(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}} u_{\mathbf{k}}(\mathbf{r}),$$

where $u_{\mathbf{k}}(\mathbf{r})$ is a lattice periodic function i.e. $u_{\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{\mathbf{k}}(\mathbf{r})$. This is the Bloch's theorem. For an electron in a crystalline potential, we have

$$T_{\mathbf{R}}\Psi(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}}\Psi(\mathbf{r}),$$

where $T_{\mathbf{R}}$ is a lattice translation operator and $\mathbf{k} = \frac{2\pi}{N}(n_1\mathbf{b}_1 + n_2\mathbf{b}_2 + n_3\mathbf{b}_3)$. Here n_1, n_2, n_3 are integers and $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ are primitive translations of the reciprocal lattice.

In the tight-binding approximation one assumes that the electron in the unit cell about R_j is only slightly influenced by atoms other than the one located at R_j. Its wave function in that cell will be close to $\phi(r - R_j)$, the atomic wave function, and its energy close to E_a. One can make a linear combination of atomic orbitals $\phi(r - R_j)$ as a trial function for the electronic wave function in the solid:

$$\Psi_k(r) = \frac{1}{\sqrt{N}} \sum_j \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}_j} \phi(r - R_j).$$

The energy of a state $\Psi_k(r)$ is given by

$$E_k = E_a - \alpha - \gamma {\sum_m}' \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{R}_j - \mathbf{R}_m)}$$

where the sum is over all nearest neighbors of $\mathbf{R}_j$ and

$$\alpha = -\int \mathrm{d}^3 r \phi^*(r - R_j)\tilde{V}(r - R_j)\phi(r - R_j)$$

and

$$\gamma = -\int \mathrm{d}^3 r \phi^*(r - R_m)\tilde{V}(r - R_j)\phi(r - R_j).$$

In second quantization representation, the tight-binding Hamiltonian is given by

$$H = \sum_n \varepsilon_n c_n^\dagger c_n + \sum_{n \neq m} T_{nm} c_n^\dagger c_m$$

where $\varepsilon_n = T_{nn} + V(R_n^0)$ represents an energy on site n and T_{nm} denotes the amplitude of hopping from site m to site n.

The periodic part of the Bloch function is expanded in Fourier series as

$$u_n(\mathbf{k},\mathbf{r}) = \sum_{\mathbf{K}} C_{\mathbf{K}}(n,\mathbf{k}) \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}}$$

and the energy eigenfunction is simply written as

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} u(\mathbf{r}) = \sum_{\mathbf{K}} C_{\mathbf{K}} \mathrm{e}^{\mathrm{i}(\mathbf{k}+\mathbf{K})\cdot\mathbf{r}}.$$

The Schrödinger equation of an electron in a lattice periodic potential is written as an infinite set of linear equations for the coefficients $C_{\mathbf{K}}$:

$$\left[E - V_{\mathbf{0}} - \frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{K})^2\right] C_{\mathbf{K}} = \sum_{\mathbf{H}\neq 0} V_{\mathbf{H}} C_{\mathbf{K}-\mathbf{H}}.$$

We can express the infinite set of linear equations in a matrix notation:

$$\begin{pmatrix} \varepsilon_{\mathbf{K}_1} + \langle \mathbf{K}_1 | V | \mathbf{K}_1 \rangle - E & \langle \mathbf{K}_1 | V | \mathbf{K}_2 \rangle \cdots & \langle \mathbf{K}_1 | V | \mathbf{K}_n \rangle \cdots \\ \langle \mathbf{K}_2 | V | \mathbf{K}_1 \rangle & \varepsilon_{\mathbf{K}_2} + \langle \mathbf{K}_2 | V | \mathbf{K}_2 \rangle - E \cdots & \langle \mathbf{K}_2 | V | \mathbf{K}_n \rangle \cdots \\ \vdots & \vdots & \vdots \\ \langle \mathbf{K}_n | V | \mathbf{K}_1 \rangle & \langle \mathbf{K}_n | V | \mathbf{K}_2 \rangle \cdots & \varepsilon_{\mathbf{K}_n} + \langle \mathbf{K}_n | V | \mathbf{K}_n \rangle - E \cdots \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} C_{\mathbf{K}_1} \\ C_{\mathbf{K}_2} \\ \vdots \\ C_{\mathbf{K}_n} \\ \vdots \end{pmatrix} = 0.$$

Here, $\varepsilon_{\mathbf{K}} = \frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{K})^2$ and $\langle \mathbf{K} | V | \mathbf{K}' \rangle = V_{\mathbf{K}-\mathbf{K}'}$, where $|\mathbf{K}\rangle = \frac{1}{\sqrt{\Omega}} \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}}$.

In the nearly free electron method, the energies near the zone boundary become

$$E_{\pm}(k) = V_0 + \varepsilon_{\mathbf{k}} \pm |V_{\mathbf{K}}|.$$

The two roots can be written, for small q, as

$$E_- = \frac{\hbar^2 q^2}{2m_-^*}; \quad E_+ = E_{\mathrm{G}} + \frac{\hbar^2 q^2}{2m_+^*},$$

where $\mathbf{q} = \frac{\mathbf{K}}{2} + \mathbf{k}$. Here, the energy gap E_{G} is equal to $2|V_{\mathbf{K}}|$ and the *effective masses* $m_{\pm}^*$ are given by

$$m_{\pm}^* = \frac{m}{1 \mp \frac{\varepsilon_{\mathbf{K}}}{2|V_{\mathbf{K}}|}}.$$

The results are only valid near $q = 0$ since we expanded the original equations for small deviations q from the extrema. This result is called the *effective mass approximation*. Crystalline solids are classified as metals, semimetals, semiconductors, and insulators according to the magnitudes and shapes of the energy gap of the material.

5

Use of Elementary Group Theory in Calculating Band Structure

5.1 Band Representation of Empty Lattice States

For a three-dimensional crystal, the free electron energies and wave functions can be expressed in the Bloch function form in the following way:

1. Write the plane-wave wave vector as a sum of a Bloch wave vector and a reciprocal lattice vector. The Bloch wave vector $\mathbf{k}$ is restricted to the first Brillouin zone; the reciprocal lattice vectors are given by

$$\mathbf{K}_{\boldsymbol{\ell}} = 2\pi[l_1\mathbf{b}_1 + l_2\mathbf{b}_2 + l_3\mathbf{b}_3] \tag{5.1}$$

   where $(l_1, l_2, l_3) = \boldsymbol{\ell}$ are integers and $\mathbf{b}_i$ are primitive translations of the reciprocal lattice. Then

$$\Psi_{\boldsymbol{\ell}}(\mathbf{k}, \mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\mathrm{e}^{\mathrm{i}\mathbf{K}_{\boldsymbol{\ell}}\cdot\mathbf{r}}. \tag{5.2}$$

   The second factor has the periodicity of the lattice since $\mathrm{e}^{\mathrm{i}\mathbf{K}_l\cdot\mathbf{R}} = 1$ for any translation vector $\mathbf{R}$.
2. The energy is given by

$$E_{\boldsymbol{\ell}}(\mathbf{k}) = \frac{\hbar^2}{2m}\left(\mathbf{k} + \mathbf{K}_{\boldsymbol{\ell}}\right)^2. \tag{5.3}$$

3. Each band is labeled by $\boldsymbol{\ell} = (l_1, l_2, l_3)$ and has $\Psi_{\boldsymbol{\ell}}(k, r)$ given by (5.2) and $E_{\boldsymbol{\ell}}(k)$ given by (5.3).

5.2 Review of Elementary Group Theory

In our brief discussion of translational and rotational symmetries of a lattice, we introduced a few elementary concepts of group theory. The object of this chapter is to show how group theory can be used in evaluating the band structure of a solid. We begin with a few definitions.

Order of a group: If a group $\mathcal{G}$ contains g elements, it is said to be of order g.
Abelian group: A group in which all elements commute.
Cyclic group: A group of g elements, in which the elements can be written

$$A, A^2, A^3, \ldots, A^{g-1}, A^g = E.$$

Class When an element R of a group is multiplied by A and A^{-1} to form $R' = ARA^{-1}$, where A and A^{-1} are elements of the group, the set of elements R' obtained by using every A belonging to $\mathcal{G}$ is said to form a class. Elements belong to the same class if they do essentially the same thing when viewed from different coordinate system. For example, for 4 mm there are five classes:
(1) E, (2)R_{90° and R_{-90°, (3) R_{180°, (4) m_x and m_y, (5) m_+ and m_-
Rearrangement theorem If $\mathcal{G} = \{E, A, B, \ldots\}$ is the set of elements of a group, $A\mathcal{G} = \{AE, AA, AB, \ldots\}$ is simply a rearrangement of this set. Therefore $\sum_{R\in\mathcal{G}} f(R) = \sum_{R\in\mathcal{G}} f(AR)$.
Generators If all the elements of a group can be expressed in form $A^m D^n$, where m and n are integers, then A and D are called generators of the group. For example, the four operators of 2 mm can all be expressed in terms of R and m_x such as $E = R^2 = m_x^2$, $R = R$, $m_x = m_x$, $m_y = R^1 m_x^1$.

5.2.1 Some Examples of Simple Groups

Cyclic Group of Order

n Consider an n-sided regular polygon. Rotation by $R_j = \frac{2\pi}{n} \times j$ with $j = 0, 1, 2, \ldots, n-1$ form a group of symmetry operations. The generator of this group is R_1 = rotation by $\frac{2\pi}{n}$.

$$\mathcal{G} = \{R_1, R_1^2, R_1^3, \ldots, R_1^n = E\} \tag{5.4}$$

Symmetry Operations of an Equilateral Triangle

$$\mathcal{G} = \{E, R_{120}, R_{-120}, J_A, J_B, J_C\} \tag{5.5}$$

Here, R_{120} and J_A are generators of $\mathcal{G}$. In this case, we have three classes of $\{E\}$, $\{R_{120}, R_{-120}\}$, and $\{J_A, J_B, J_C\}$. (See Fig. 5.1.)

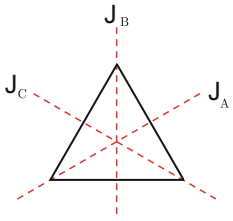


Fig. 5.1. Equilateral triangle

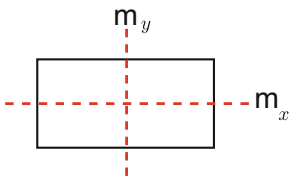


Fig. 5.2. Rectangle

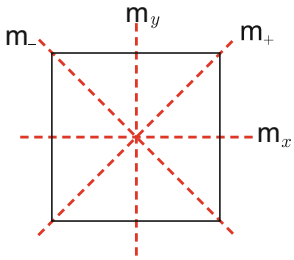


Fig. 5.3. Square

Symmetry Group of a Rectangle

$$\mathcal{G} = \{E, R, m_x, m_y\} \tag{5.6}$$

Here, R and m_x are generators of $\mathcal{G}$, and each element belongs to a different class since x and y directions differ. (See Fig. 5.2.)

Symmetry Group of a Square

$$\mathcal{G} = \{E; R_{90}\,(\equiv R_1)\,, R_{-90}\,(\equiv R_3)\,; R_{180}\,(\equiv R_2)\,; m_x, m_y; m_+, m_-\} \tag{5.7}$$

Here, R_{90} and m_x are generators, and classes are discussed earlier. (See Fig. 5.3.)

Other Examples

Groups of matrices, e.g., (1) $n \times n$ matrices with determinant equal to unity and (2) $n \times n$ orthogonal matrices.

5.2.2 Group Representation

A group of matrices that satisfy the same multiplication as the elements of the group is said to form a representation of the group. To be concrete, suppose a group $\mathcal{G} = \{E, A, B, \ldots\}$ of symmetry operations operates on some function $\Psi(x, y, z)$. These operations give us the set of functions.

$$E\Psi, A\Psi, B\Psi, \ldots. \tag{5.8}$$

which form a vector space that is invariant under the operations of the group. By this, we mean that the space of all functions of the form

$$\Phi = c_1 E\Psi + c_2 A\Psi + c_3 B\Psi + \cdots \tag{5.9}$$

where c_i are complex numbers is invariant under the operations of the group. We can choose a basis set Ψ_j with $j = 1, 2, \ldots, l \leq g$, where g is the order of the group. Then for any ϕ belonging to this vector space we can write

$$\phi = \sum_{j=1}^{l} c_j \Psi_j. \tag{5.10}$$

For any element $R \in \mathcal{G}$,

$$R\Psi_j = \sum_{k=1}^{l} D_{jk}(R)\Psi_k, \tag{5.11}$$

where $D(R)$ is a matrix. The set of matrices $D(R)$ (for each $R \in \mathcal{G}$) form a representation of $\mathcal{G}$.

5.2.3 Examples of Representations of the Group 4 mm

Under the operations of the eight elements of the group 4 mm, x always transforms into $\pm x$ or into $\pm y$ as shown in Table 5.1.

Representation Γ_1 Consider the function $\Psi_0 = x^2 + y^2$. It is obvious that under every operation belonging to $\mathcal{G}$, Ψ_0 is unchanged. For example,

$$R_{180}\Psi_0 = R_{180}(x^2 + y^2) = (-x)^2 + (-y)^2 = x^2 + y^2 = \Psi_0. \tag{5.12}$$

Thus, every operation of $\mathcal{G}$ can be represented by the unit matrix

$$D(E) = D(R_{180}) = D(R_{90}) = D(R_{-90}) = \cdots = D(m_-) = 1. \tag{5.13}$$

This set of matrices forms a representation of $\mathcal{G}$ that is called the "identity" representation and denoted by the symbol Γ_1 (i.e. the representation Γ_1). Any function $f(x, y)$ that transforms under the group operation R in exactly the same manner as multiplication by the matrix $D(R)$ representing R in the representation Γ_n is said to belong to the representation Γ_n.

Table 5.1. Operations of the group 4 mm on functions of x and y

Operation	E	R_{180}	R_{90}	R_{-90}	m_x	m_y	m_+	m_-
x	x	$-x$	y	$-y$	x	$-x$	y	$-y$
y	y	$-y$	$-x$	x	$-y$	y	x	$-x$

Representation Γ_4 Consider the function $\Psi_4 = xy$. It is obvious that E, R_{180}, m_+, and m_- operating on Ψ_4 leave it unchanged but that R_{90}, R_{-90}, m_x, and m_y operating on Ψ_4 change it to $-\Psi_4$. Therefore the matrices

$$D(E) = D(R_{180}) = D(m_+) = D(m_-) = 1$$
$$D(R_{90}) = D(R_{-90}) = D(m_x) = D(m_y) = -1$$

form a representation of $\mathcal{G}$. This representation is called the Γ_4 representation.

By writing $\Psi_3 = x^2 - y^2$, $\Psi_2 = xy(x^2 - y^2)$, and $\Psi_5 = \begin{pmatrix} x \\ y \end{pmatrix}$, it is easy to construct Table 5.2, which illustrates the sets of matrices $D_{\Gamma_j}(R)$ for each $R \in \mathcal{G}$ of the group 4 mm. The sets of matrices $\{D_{\Gamma_j}(R)\}$ for each $R \in \mathcal{G}$ form representations of the group 4 mm. Functions belonging to the representation Γ_j transform under the operations of 4 mm in exactly the same way as multiplying them by the appropriate $D_{\Gamma_j}(R)$.

Table 5.2. Group representation table of the group 4 mm

R	$\Psi_0 = x^2 + y^2$ $D_{\Gamma_1}(R)$	$\Psi_2 = xy(x^2 - y^2)$ $D_{\Gamma_2}(R)$	$\Psi_3 = x^2 - y^2$ $D_{\Gamma_3}(R)$	$\Psi_4 = xy$ $D_{\Gamma_4}(R)$	$\Psi_5 = \begin{pmatrix} x \\ y \end{pmatrix}$ $D_{\Gamma_5}(R)$
E	1	1	1	1	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
R_{180}	1	1	1	1	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$
R_{90}	1	1	−1	−1	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$
R_{-90}	1	1	−1	−1	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$
m_x	1	−1	1	−1	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
m_y	1	−1	1	−1	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$
m_+	1	−1	−1	1	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
m_-	1	−1	−1	1	$\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$
	⇑ Γ_1	⇑ Γ_2	⇑ Γ_3	⇑ Γ_4	⇑ Γ_5

5.2.4 Faithful Representation

You will notice that the set of matrices forming the representation Γ_1 of 4 mm to which the function x^2+y^2 belonged were all identical, i.e. $D(E) = D(\mathrm{R}_1) = D(R_2) = \cdots = D(m_x) = 1$. Such a representation is a homomorphism between the group of symmetry operators and the group of matrices, and it is said to be an *unfaithful representation*. A representation in which each operation $R \in \mathcal{G}$ is represented by a different matrix $D(R)$ is called a *faithful representation*. In this case, the group of symmetry operations and the group of matrices are isomorphic.

5.2.5 Regular Representation

If we construct a multiplication table for a finite group $\mathcal{G}$ as shown in Table 5.3 for 2 mm and we form 4×4 matrices $D(E)$, $D(R)$, $D(m_x)$, $D(m_y)$ by substituting 1 in the 4×4 array wherever the particular operation appears and 0 everywhere else, the set of matrices form a representation known as the *regular representation*. Thus, we have

$$D(E) = \begin{pmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{pmatrix}, \; D(R) = \begin{pmatrix} 0\,1\,0\,0 \\ 1\,0\,0\,0 \\ 0\,0\,0\,1 \\ 0\,0\,1\,0 \end{pmatrix}, \text{ etc.} \tag{5.14}$$

5.2.6 Reducible and Irreducible Representations

Suppose $D^{(1)}(R)$ and $D^{(2)}(R)$ are two representations of the same group, then $D(R)$ defined by

$$D(R) = \begin{pmatrix} D^{(1)}(R) & 0 \\ 0 & D^{(2)}(R) \end{pmatrix} \tag{5.15}$$

also forms a representation of $\mathcal{G}$. $D(R)$ is called the direct sum of the first two representations $D^{(1)}(R)$ and $D^{(2)}(R)$. A representation which can be written as the direct sum of two smaller representations is said to be reducible. Sometimes a representation is reducible, but it is not at all apparent. The reason for this is that if the matrices $D(R)$ form a representation of $\mathcal{G}$, then

Table 5.3. Multiplication table of the group 2 mm

	$E^{-1} = E$	$R^{-1} = R$	$m_x^{-1} = m_x$	$m_y^{-1} = m_y$
E	E	R	m_x	m_y
R	R	E	m_y	m_x
m_x	m_x	m_y	E	R
m_y	m_y	m_x	R	E

$$\tilde{D}(R) = S^{-1}D(R)S \tag{5.16}$$

also form a representation (corresponding to a change in the basis vectors of the vector space on which the matrices act). This similarity transformation can scramble the block diagonal form so that the resulting $\tilde{D}(R)$ do not look reducible. A representation is reducible if and only if it is possible to perform the same similarity transformation on all the matrices in the representation and reduce them to block diagonal form. Otherwise, the representation is irreducible. Clearly all one-dimensional representations (1×1 matrices) are irreducible.

5.2.7 Important Theorems of Representation Theory (without proof)

1. *Theorem one* The number of irreducible representations (IRs) is equal to the number of conjugate classes.
2. *Theorem two* If l_i is the dimension of the ith IR and g is the number of operations in the group $\mathcal{G}$

$$\sum_i l_i^2 = g. \tag{5.17}$$

Examples

1. 2 mm. There are four operations E, R, m_x, m_y and there are four classes (remember that because the x and y directions are distinct m_x and m_y belong to different classes). From Theorem one, there are four IRs; from Theorem two, they are all one dimensional so that

$$\sum_i l_i^2 = 1^2 + 1^2 + 1^2 + 1^2 = 4 = g.$$

2. 4 mm. There are eight operations falling into five conjugate classes: E; R_2; R_1 and R_3; m_x and m_y; m_+ and m_-. Therefore, there are five IRs and four IR's must be one dimensional and one must be two dimensional so that

$$\sum_i l_i^2 = 1^2 + 1^2 + 1^2 + 1^2 + 2^2 = 8 = g$$

5.2.8 Character of a Representation

The character χ of a representation $D(R)$ is defined as

$$\chi(R) = \sum_j D_{jj}(R) \; for \; each \; R \in \mathcal{G}. \tag{5.18}$$

Because the application of a similarity transformation does not change the trace of a matrix

1. $\chi(R)$ is independent of the basis used for the vector space.
2. $\chi(R)$ is the same for all elements R belonging to the same conjugate class.

Thus, we can define $\chi(C)$ to be the common value of $\chi(R)$ for all R belonging to conjugate class C.[1]

5.2.9 Orthogonality Theorem

In trying to determine the IRs and their characters certain orthogonality theorems are very useful. We state them without proof.

1.
$$\sum_{R\in\mathcal{G}} \chi^i(R)\chi^{j^*}(R) = g\delta ij, \tag{5.19}$$
where χ^i and χ^j are the characters of two representations. This can also be written
$$\sum_{C} n_C \chi^i(C)\chi^{j^*}(C) = g\delta_{ij}, \tag{5.20}$$
where n_C is the number of elements in the class C.

2.
$$\sum_{i} \chi^i(C)\chi^{i^*}(C') = \frac{g}{n_C}\delta_{C,C'}. \tag{5.21}$$

3. If $D^{(i)}_{\mu\nu}(R)$ is the $\mu\nu$ matrix element of the ith IR for the operation R, then
$$\sum_{R\in\mathcal{G}} D^{(i)}_{\mu\nu}(R)\left[D^{(j)}(R)\right]^{-1}_{\nu'\mu'} = \frac{1}{l_i}\delta_{ij}\delta_{\mu\mu'}\delta_{\nu\nu'}. \tag{5.22}$$
For a unitary representation $\left[D^{(j)}(R)\right]^{-1}_{\nu'\mu'} = D^{(j)^*}_{\mu'\nu'}(R)$ so that for unitary representation
$$\sum_{R\in\mathcal{G}} D^{(i)}_{\mu\nu}(R)D^{(j)^*}(R)_{\nu'\mu'} = \frac{1}{l_i}\delta_{ij}\delta_{\mu\mu'}\delta_{\nu\nu'}. \tag{5.23}$$

Some Examples

1. 2 mm. We know there are four IRs all of which are one dimensional. We label them Γ_1, Γ_2, Γ_3, Γ_4. The 1×1 matrices representing each element are given in Table 5.4.
2. 4 mm. There are five IRs; one is two dimensional and the rest are one dimensional as are shown in Table 5.5.

The reader should use these simple examples to demonstrate that the orthogonality theorems hold.

[1] In the regular representation, each IR Γ_i appears l_i times, where l_i is the dimension of the IR Γ_i.

Table 5.4. Irreducible representations of the group 2 mm

	Γ_1	Γ_2	Γ_3	Γ_4
E	1	1	1	1
R	1	1	−1	−1
m_x	1	−1	1	−1
m_y	1	−1	−1	1

Table 5.5. Irreducible representations of the group 4 mm

	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5
E	1	1	1	1	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \Rightarrow 2$
R_2	1	1	1	1	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \Rightarrow -2$
R_1, R_3	1	1	−1	−1	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \Rightarrow 0$
m_x, m_y	1	−1	1	−1	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \Rightarrow 0$
m_+, m_-	1	−1	−1	1	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \Rightarrow 0$

5.3 Empty Lattice Bands, Degeneracies and IRs at High-Symmetry Points

In Sect. 5.2 we determined the free electron energies and wave functions in the reduced zone scheme for a two-dimensional rectangular lattice. The starting point for many band structure calculations is this empty lattice band structure obtained by writing, as (5.3),

$$E_{\mathbf{l}}(\mathbf{k}) = \frac{\hbar^2}{2m} \left(\mathbf{k} + \mathbf{K}_{\mathbf{l}}\right)^2 ,$$

where $\mathbf{k}$ is restricted to lie in the first Brillouin zone and $\mathbf{K}_{\mathbf{l}}$ is a reciprocal lattice vector of (5.1)

$$\mathbf{K}_{\mathbf{l}} = 2\pi[l_1\mathbf{b}_1 + l_2\mathbf{b}_2 + l_3\mathbf{b}_3].$$

Here, $\mathbf{b}_i$ are the primitive translation vectors of the reciprocal lattice and $l_i = 0, \pm 1, \pm 2, \ldots$. We evaluate the energy at particular symmetry points, e.g. at $k = 0$, the Γ point, along $k_y = k_z = 0$, the line Δ, etc.

The group of symmetry operations which leave the lattice invariant also leaves the reciprocal lattice invariant. Suppose we know some wave function $\Psi_{\mathbf{k}}$. A rotation or reflection operation of the point group acting on $\Psi_{\mathbf{k}}$ will give the same result as the rotation or reflection of $\mathbf{k}$, that is

$$R\Psi_{\mathbf{k}}(x) = \Psi_{R\mathbf{k}}(x) = \Psi_{\mathbf{k}}\left(R^{-1}x\right). \tag{5.24}$$

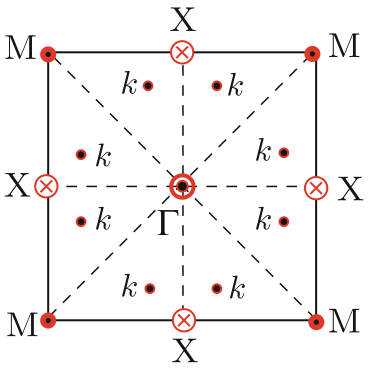


Fig. 5.4. STAR of square lattice

Here, we used the fact that applying the same orthogonal transformation to both vectors in a scalar product does not change the value of the product, for example $\mathbf{k} \cdot \mathrm{R}^{-1}\mathbf{r} = \mathrm{R}\mathbf{k} \cdot \mathrm{R}\mathrm{R}^{-1}\mathbf{r} = \mathrm{R}\mathbf{k} \cdot \mathbf{r}$. By applying every $R \in \mathcal{G}$ to a wave vector $\mathbf{k}$, we generate the *STAR* of $\mathbf{k}$. For a two-dimensional square lattice, all operations leave Γ invariant so Γ is its own STAR. (See Fig. 5.4) For a general point $\mathbf{k}$, there will be g (=8 for 4 mm) points in the STAR of $\mathbf{k}$. The symmetry point X has four points in its STAR; two of these lie along the x-axis and are equivalent because they are separated by a reciprocal lattice vector. The other two points in the STAR of X are not equivalent to the X-point along the x-axis because they are not separated from it by a reciprocal lattice vector. All four points in the STAR of M are equivalent since they are all separated by vectors of the reciprocal lattice.

5.3.1 Group of the Wave Vector k

The group (or subgroup of the original point group) of rotations and reflections that transform $\mathbf{k}$ into itself or into a new $\mathbf{k}$ vector separated from the original $\mathbf{k}$ point by a reciprocal lattice vector belong to the *group of the wave vector* $\mathbf{k}$.

Example
For two-dimensional square lattice (Fig. 5.5) we remember that Δ, Z, and Σ denote, respectively, any point on the line from $\Gamma \to X$, $X \to M$, and $\Gamma \to M$. We have the groups of the wave vectors as follows:

$$\begin{aligned} \mathcal{G}_\Gamma &= \mathcal{G}_M = \{E, R_1, R_2, R_3, m_x, m_y, m_+, m_-\} \\ \mathcal{G}_X &= \{E, R_2, m_x, m_y\} \\ \mathcal{G}_\Delta &= \{E, m_x\} \\ \mathcal{G}_\Sigma &= \{E, m_+\} \\ \mathcal{G}_Z &= \{E, m_y\} \end{aligned} \tag{5.25}$$

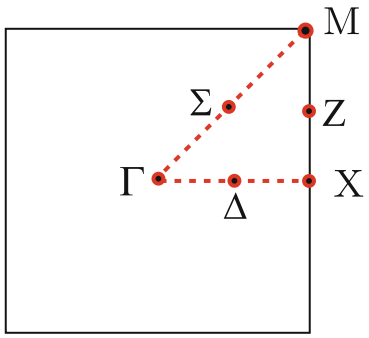


Fig. 5.5. First Brillouin zone of a square lattice

When the empty lattice bands are calculated, there are often a number of degenerate bands at points of high symmetry like the Γ-point, the M-point, and the X-point. Because the operations of the group of the wave vector Γ (or M or X) leave this point invariant, one can construct linear combinations of these degenerate states that belong to representations of the group of the wave vector Γ (or M or X, etc.).

Example
Empty lattice bands for a two-dimensional square lattice are written by

$$E_{l_1 l_2}(\mathbf{k}) = \frac{\hbar^2}{2m}\left(\mathbf{k} + \mathbf{K}_{l_1 l_2}\right)^2 \quad \text{and} \quad \Psi_{l_1 l_2}(\mathbf{k}, \mathbf{r}) = \mathrm{e}^{\mathrm{i}(\mathbf{k}+\mathbf{K}_{l_1 l_2})\cdot\mathbf{r}}. \tag{5.26}$$

Here

$$\mathbf{K}_{l_1 l_2} = \frac{2\pi}{a}\left(l_1 \hat{x} + l_2 \hat{y}\right) \tag{5.27}$$

with $l_1, l_2 = 0, \pm 1, \pm 2, \ldots$. The empty lattice bands are labelled by the pair of integers (l_1, l_2). (See Fig. 5.6.) By defining ξ and η by

$$\mathbf{k} = \frac{2\pi}{a}[\xi \hat{x} + \eta \hat{y}] \tag{5.28}$$

then we have

$$E_{l_1 l_2} = \frac{h^2}{2ma^2}\left[(\xi + l_1)^2 + (\eta + l_2)^2\right] \quad \text{and} \quad \Psi_{l_1 l_2} = \mathrm{e}^{\frac{2\pi \mathrm{i}}{a}[(\xi+l_1)x + (\eta+l_2)y]}. \tag{5.29}$$

The parameters ξ and η are restricted to the range $[-\frac{1}{2}, \frac{1}{2}]$.

Exercise
Evaluate the energies $E_{l_1 l_2}(\Gamma)$, $E_{l_1 l_2}(X)$ for energies up to $E = \frac{h^2}{2ma^2} \times 10$. Make a sketch of the empty square lattice bands going from $\Gamma \to \Delta \to X$. (Use straight lines to connect $E_{l_1 l_2}(\Gamma)$ to $E_{l_1 l_2}(X)$). List the degeneracies at Γ, Δ, and X. For example, see Table 5.6.

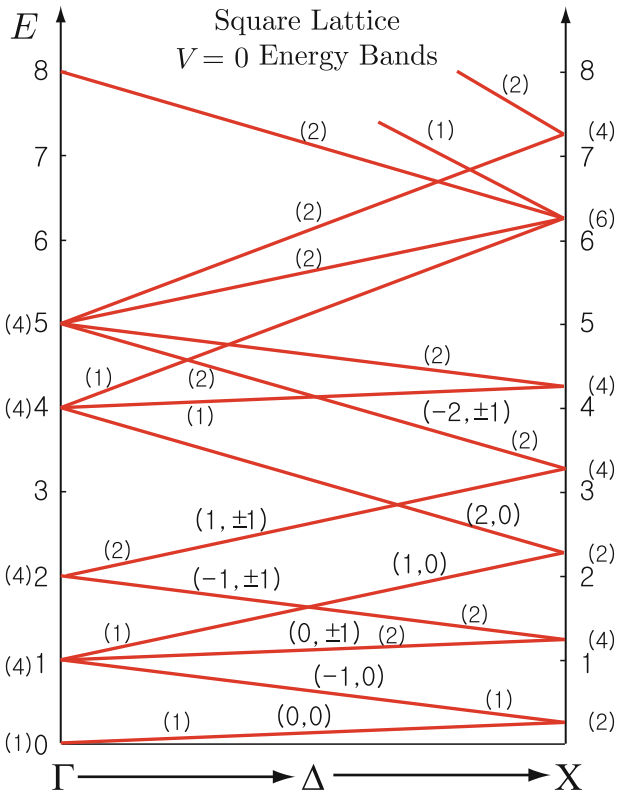


Fig. 5.6. Empty lattice bands of a square lattice along the line $\Gamma \to \Delta \to X$. Energy is measured in units of $\frac{h^2}{2ma^2}$. We have drawn *straight lines* connecting $E_{l_1 l_2}(\Gamma)$ to $E_{l_1 l_2}(X)$ for the sake of simplicity. The pair of integers (l_1, l_2) is indicated for each band, and the band degeneracy is given in the parenthesis. In fact the energy along Δ varies as $\xi^2 + 2l_1\xi + l_1^2 + l_2^2$

5.4 Use of Irreducible Representations

It is apparent from the $V(r) = 0$ empty lattice structure that at some points in the Brillouin zone more than one band has the same energy. If we refer to the two-dimensional square lattice, we find that at $E(X) = \frac{1}{4}\frac{h^2}{2ma^2}$ there are two degenerate bands, viz. $(l_1, l_2) = (0, 0)$ and $(-1,0)$. At $E(\Gamma) = \frac{h^2}{2ma^2}$ there are four degenerate bands $(-1,0)$, $(1,0)$, $(0,1)$, $(0,-1)$. The vector space formed by the degenerate bands at $E(\mathbf{k})$ is invariant under the operations of the group of the wave vector $\mathbf{k}$. This means that the space of degenerate states at a point $\mathbf{k}$ in the Brillouin zone provides a representation of the group of the wave vector $\mathbf{k}$. We can decompose this representation into its irreducible components and use the decomposition to label the states. This process does not change the empty lattice band structure that we have already

Table 5.6. Empty lattice bands of the group $4mm$

l_1	l_2	$\frac{2ma^2}{\hbar^2}E_{l_1 l_2}(\Gamma) = l_1^2 + l_2^2$	$\frac{2ma^2}{\hbar^2}E_{l_1 l_2}(X) = (l_1 + \frac{1}{2})^2 + l_2^2$
0	0	0	$\frac{1}{4}$
−1	0	1	$\frac{1}{4}$
1	0	1	$\frac{9}{4}$
−2	0	4	$\frac{9}{4}$
2	0	4	$\frac{25}{4}$
0	±1	1	$\frac{5}{4}$
0	±2	4	$\frac{17}{4}$
−1	±1	2	$\frac{5}{4}$
−1	±2	5	$\frac{17}{4}$

obtained. It is simply a convenient choice of basis functions for each of the spaces of degenerate energy states. However, once the periodic potential $V(r)$ is introduced, it is immediately seen that band gaps appear as a consequence of the decomposition of degenerate states into irreducible components. We shall see that states belonging to different IR's do not interact (i.e. they are not coupled together by the periodic potential).

The periodic potential $V(r)$ can be expressed as a Fourier series

$$V(r) = \sum_{\mathbf{K}} V_{\mathbf{K}} \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}}, \tag{5.30}$$

where $\mathbf{K}$ is a reciprocal lattice vector. For the two-dimensional square lattice we can write

$$V(r) = \sum_{l_1 l_2} V_{l_1,l_2} \mathrm{e}^{\frac{2\pi\mathrm{i}}{a}(l_1 x + l_2 y)}. \tag{5.31}$$

Because $V(r)$ is invariant under the operations of the point group, it is not difficult to see that

$$V_{l_1,l_2} = V_{-l_1,l_2} = V_{l_1,-l_2} = V_{-l_1,-l_2} = V_{l_2,l_1} = V_{-l_2,l_1} = V_{l_2,-l_1} = V_{-l_2,-l_1}. \tag{5.32}$$

In our previous discussion of the effect of the periodic potential, we were able to obtain an infinite set of coupled algebraic equations, (4.40), which could be written

$$\left[E - V_{\mathbf{0}} - \frac{\hbar^2}{2m}(\mathbf{k} + \mathbf{K})^2\right] C_{\mathbf{K}} = \sum_{\mathbf{H}\neq 0} V_{\mathbf{H}} C_{\mathbf{K}-\mathbf{H}}. \tag{5.33}$$

Here, $C_{\mathbf{K}}$ was the coefficient in the expansion of $u(r)$, the periodic part of the Bloch function, in Fourier series. This infinite set of equations could be expressed as a matrix equation, (4.41). The off-diagonal matrix elements are of the form $\langle \mathbf{K}_i | V(r) | \mathbf{K}_j \rangle = V_{\mathbf{K}_i - \mathbf{K}_j}$. When the degeneracy of a particular energy state becomes large (e.g. at $E(\Gamma) = 5\frac{h^2}{2ma^2}$ the degeneracy is 8), the degenerate states must be treated exactly and there is no reason to suppose that any off-diagonal matrix elements vanish. However, when we classify the degenerate states according to the IR's of the group of the wave vector, we are able to simplify the secular equation by virtue of a fundamental theorem on matrix elements.

Theorem on Matrix Elements (without proof)
Matrix elements of any operator which is invariant under all the operations of a group are zero between functions belonging to different IRs of the group. Matrix elements are also zero between functions *belonging to different rows* of the same representation.

When one classifies the degenerate states according to the IR's of the group of the wave vector, many of the degenerate states will belong to different IRs and therefore the off-diagonal matrix elements of $V(r)$ between them will vanish.

5.4.1 Determining the Linear Combinations of Plane Waves Belonging to Different IRs

Let us begin by considering the states at Γ. The plane-wave wave functions and energies are given, respectively, by

$$\begin{aligned} \Psi_{l_1 l_2}(\Gamma) &= e^{\frac{2\pi i}{a}(l_1 x + l_2 y)} \\ E_{l_1 l_2}(\Gamma) &= \tfrac{h^2}{2ma^2}(l_1^2 + l_2^2). \end{aligned} \tag{5.34}$$

Therefore, the energies at Γ, the bands corresponding to that energy, and the degeneracy are as given in Table 5.7.

At $E_\Gamma = 0$, there is a single state (let us measure E in units of $\frac{h^2}{2ma^2}$). The wave function is given by

$$\Psi_{00}(\Gamma) = 1 \tag{5.35}$$

It is unchanged by every operation of $\mathcal{G}_\Gamma$, the group of the wave vector Γ. Therefore, it belongs to the IR Γ_1 because every element of $\mathcal{G}_\Gamma$ operating on Ψ_{00} gives $+1 \times \Psi_{00}$ or every operation is represented by the 1×1 unit matrix $D = 1$. This, of course, is the Γ_1 representation. At $E_\Gamma = 1$, there are four states

$$\begin{aligned} \Psi_{\pm 1,0}(\Gamma) &= \mathrm{e}^{\pm \frac{2\pi i}{a} x}, \\ \Psi_{0,\pm 1}(\Gamma) &= \mathrm{e}^{\pm \frac{2\pi i}{a} y}. \end{aligned} \tag{5.36}$$

Table 5.7. Empty lattice energy at Γ and its degeneracy for a two dimensional square lattice

$\frac{2ma^2}{h^2}E(\Gamma)$	Bands $(l_1,\ l_2)$	Degeneracy
0	(0,0)	1
1	$(\pm1,0);(0,\pm1)$	4
2	$(-1,\pm1);(1,\pm1)$	4
4	$(\pm2,0);(0,\pm2)$	4
5	$(-1,\pm2);(1,\pm2);(-2,\pm1);(2,\pm1)$	8

The eight operations of $\mathcal{G}_\Gamma$ change x into $\pm x$ and y into $\pm y$ or x into $\pm y$ and y into $\pm x$. From the four function $\Psi_{\pm1,0}$ and $\Psi_{0,\pm1}$ we can form the following linear combinations

$$\Psi(\Gamma_1) = \cos\frac{2\pi x}{a} + \cos\frac{2\pi y}{a} \propto \Psi_{1,0} + \Psi_{-1,0} + \Psi_{0,1} + \Psi_{0,-1}, \tag{5.37}$$

$$\Psi(\Gamma_3) = \cos\frac{2\pi x}{a} - \cos\frac{2\pi y}{a} \propto \Psi_{1,0} + \Psi_{-1,0} - \Psi_{0,1} - \Psi_{0,-1}, \tag{5.38}$$

and

$$\Psi(\Gamma_5) = \begin{pmatrix} \Psi^{(1)}(\Gamma_5) \\ \Psi^{(2)}(\Gamma_5) \end{pmatrix} = \begin{pmatrix} \sin\frac{2\pi x}{a} \\ \sin\frac{2\pi y}{a} \end{pmatrix} \propto \begin{pmatrix} \Psi_{1,0} - \Psi_{-1,0} \\ \Psi_{0,1} - \Psi_{0,-1} \end{pmatrix}. \tag{5.39}$$

Because the cosine function is an even function of its argument, every operation of $\mathcal{G}_\Gamma$ leaves $(\cos\frac{2\pi x}{a} + \cos\frac{2\pi y}{a})$ unchanged, and this function transforms according to the IR Γ_1. The function $(\cos\frac{2\pi x}{a} - \cos\frac{2\pi y}{a})$ is left unchanged by operations $(E, R_2,\ m_x, m_y)$ which change $x \to \pm x$, but it changes to minus itself under operations $(R_1, R_3,\ m_+, m_-)$ which change $x \to \pm y$. Thus the operations of $\mathcal{G}_\Gamma$ operating on $(\cos\frac{2\pi x}{a} - \cos\frac{2\pi y}{a})$ do exactly the same thing as multiplying by the matrices belonging to the representation Γ_3. In a similar way one can show that the operations of $\mathcal{G}_\Gamma$ operating on the column vector $\begin{pmatrix} \sin\frac{2\pi x}{a} \\ \sin\frac{2\pi y}{a} \end{pmatrix}$ have exactly the same effect as multiplying by the set of matrices forming the 2×2 representation Γ_5.

Exercise
The reader should determine the linear combinations of plane waves at $E_\Gamma = 2$ and $E_\Gamma = 4$ belonging to the appropriate IRs of $\mathcal{G}_\Gamma$.

At $E_\Gamma = 5$ there are eight states

$$\begin{aligned} \Psi_{\pm1,\pm2}(\Gamma) &= \mathrm{e}^{\pm\frac{2\pi\mathrm{i}}{a}(\pm x \pm y)}, \\ \Psi_{\pm2,\pm1}(\Gamma) &= \mathrm{e}^{\pm\frac{2\pi\mathrm{i}}{a}(\pm 2x \pm y)}. \end{aligned} \tag{5.40}$$

The simplest way to determine the linear combinations belonging to IRs of $\mathcal{G}_\Gamma$ is first to form sine and cosine functions like

$$\Psi_1 = \cos\frac{2\pi}{a}x\cos\frac{2\pi}{a}2y,$$
$$\Psi_2 = \cos\frac{2\pi}{a}2x\cos\frac{2\pi}{a}y,$$
$$\Psi_3 = \sin\frac{2\pi}{a}x\sin\frac{2\pi}{a}2y,$$
$$\Psi_4 = \sin\frac{2\pi}{a}2x\sin\frac{2\pi}{a}y,$$
$$\Psi_5 = \cos\frac{2\pi}{a}x\sin\frac{2\pi}{a}2y,$$
$$\Psi_6 = \cos\frac{2\pi}{a}2x\sin\frac{2\pi}{a}y,$$
$$\Psi_7 = \sin\frac{2\pi}{a}x\cos\frac{2\pi}{a}2y,$$
$$\Psi_8 = \sin\frac{2\pi}{a}2x\cos\frac{2\pi}{a}y.$$

It is easy to see how these $\Psi_i's$ are transformed by the operations of $\mathcal{G}_\Gamma$. For example, all operations of $\mathcal{G}_\Gamma$ which transform $x \to \pm x$ transform Ψ_1 into itself; all operations which transform $x \to \pm y$ transform Ψ_1 into Ψ_2. Therefore, the linear combination $\Psi_1 + \Psi_2$ is unchanged by every operation of $\mathcal{G}_\Gamma$ and belongs to Γ_1. The linear combination $\Psi_1 - \Psi_2$ is unchanged by the operations which take $x \to \pm x$, but changed to $-(\Psi_1 - \Psi_2)$ by operations which take $x \to \pm y$. Thus $\Psi_1 - \Psi_2$ belongs to the IR Γ_3. We find, by similar analysis

$$\begin{aligned}
\Psi(\Gamma_1) &= \cos\tfrac{2\pi}{a}x\cos\tfrac{2\pi}{a}2y + \cos\tfrac{2\pi}{a}2x\cos\tfrac{2\pi}{a}y = \Psi_1 + \Psi_2,\\
\Psi(\Gamma_3) &= \cos\tfrac{2\pi}{a}x\cos\tfrac{2\pi}{a}2y - \cos\tfrac{2\pi}{a}2x\cos\tfrac{2\pi}{a}y = \Psi_1 - \Psi_2,\\
\Psi(\Gamma_2) &= \sin\tfrac{2\pi}{a}x\sin\tfrac{2\pi}{a}2y - \sin\tfrac{2\pi}{a}2x\sin\tfrac{2\pi}{a}y = \Psi_3 - \Psi_4,\\
\Psi(\Gamma_4) &= \sin\tfrac{2\pi}{a}x\sin\tfrac{2\pi}{a}2y + \sin\tfrac{2\pi}{a}2x\sin\tfrac{2\pi}{a}y = \Psi_3 + \Psi_4,\\
\Psi^{(1)}(\Gamma_5) &= \begin{pmatrix}\cos\frac{2\pi}{a}x\sin\frac{2\pi}{a}2y\\ \cos\frac{2\pi}{a}y\sin\frac{2\pi}{a}2x\end{pmatrix} = \begin{pmatrix}\Psi_5\\ \Psi_8\end{pmatrix},\\
\Psi^{(2)}(\Gamma_5) &= \begin{pmatrix}\sin\frac{2\pi}{a}x\cos\frac{2\pi}{a}2y\\ \sin\frac{2\pi}{a}y\cos\frac{2\pi}{a}2x\end{pmatrix} = \begin{pmatrix}\Psi_7\\ \Psi_6\end{pmatrix}.
\end{aligned} \tag{5.41}$$

5.4.2 Compatibility Relations

The character tables for $4mm$ and its principal subgroups are listed in Tables 5.8-5.11. Notice that under the operation m_x, functions belonging to the IR's

Table 5.8. Character table of the groups 4 mm and its principal subgroups

	$\Gamma_1 = M_1$	$\Gamma_2 = M_2$	$\Gamma_3 = M_3$	$\Gamma_4 = M_4$	$\Gamma_5 = M_5$
E	1	1	1	1	2
R_2	1	1	1	1	−2
R_1, R_3	1	1	−1	−1	0
m_x, m_y	1	−1	1	−1	0
m_+, m_-	1	−1	−1	1	0

Table 5.9. Character table of a principal subgroup $\mathcal{G}_X$

	X_1	X_2	X_3	X_4
E	1	1	1	1
R_2	1	1	−1	−1
m_x	1	−1	1	−1
m_y	1	−1	−1	1

Table 5.10. Character table of a principal subgroup $\mathcal{G}_\Delta$

	Δ_1	Δ_2
E	1	1
m_x	1	−1

Table 5.11. Character table of a principal subgroup $\mathcal{G}_\Sigma$

	Σ_1	Σ_2
E	1	1
m_+	1	−1

1. Γ_1 and Γ_3 are unchanged.
2. Γ_2 and Γ_4 change sign.
3. X_1 and X_3 are unchanged.
4. X_2 and X_4 change sign.
5. Δ_1 are unchanged.
6. Δ_2 change sign.

Because of this only a Δ_1 band can begin at an Γ_1 or Γ_3 band and only a Δ_1 band can end at an X_1 or X_3 band. We call such restrictions *compatibility relations*. For our purpose it is sufficient to know that

- A band Δ_1 can connect $\Gamma_1, \Gamma_3, \Gamma_5$ to X_1, X_3.
- A band Δ_2 can connect $\Gamma_2, \Gamma_4, \Gamma_5$ to X_2, X_4.
- A band Σ_1 can connect $\Gamma_1, \Gamma_4, \Gamma_5$ to M_1, M_4, M_5.
- A band Σ_2 can connect $\Gamma_2, \Gamma_3, \Gamma_5$ to M_2, M_3, M_5.

5.5 Using the Irreducible Representations in Evaluating Energy Bands

Instead of labelling energy bands at particular symmetry points or along particular symmetry lines by the integers (l_1, l_2) [or in three-dimensions (l_1, l_2, l_3)], it is possible to label the states by their energy and by the linear combination belonging to a particular IR of $\mathcal{G}_{\mathbf{k}}$. Thus, at the Γ point of $E = 1 \cdot \frac{h^2}{2ma^2}$ we may write the four states as

$$|l_1, l_2\rangle = \begin{cases} |1,0\rangle &= \mathrm{e}^{\frac{2\pi \mathrm{i}}{a}x} \\ |-1,0\rangle &= \mathrm{e}^{-\frac{2\pi \mathrm{i}}{a}x} \\ |1,0\rangle &= \mathrm{e}^{\frac{2\pi \mathrm{i}}{a}y} \\ |0,-1\rangle &= \mathrm{e}^{-\frac{2\pi \mathrm{i}}{a}y} \end{cases} \tag{5.42}$$

or we can write (in units of $\frac{h^2}{2ma^2} = 1$)

$$|E_\Gamma = 1, \Gamma_1\rangle = \cos\frac{2\pi x}{a} + \cos\frac{2\pi y}{a} \tag{5.43}$$

$$|E_\Gamma = 1, \Gamma_3\rangle = \cos\frac{2\pi x}{a} - \cos\frac{2\pi y}{a} \tag{5.44}$$

$$\begin{pmatrix} |E_\Gamma = 1, \Gamma_5\rangle_1 \\ |E_\Gamma = 1, \Gamma_5\rangle_2 \end{pmatrix} = \begin{pmatrix} \Psi^{(1)}(\Gamma_5) \\ \Psi^{(2)}(\Gamma_5) \end{pmatrix} = \begin{pmatrix} \sin\frac{2\pi x}{a} \\ \sin\frac{2\pi y}{a} \end{pmatrix}. \tag{5.45}$$

There is a distinct advantage to using the basis functions belonging to IRs of $\mathcal{G}_\Gamma$ that results from the theorem on matrix elements.

Any matrix elements of the periodic potential (i.e. an operator with the full symmetry of the point group) between states belonging to different IR's is zero. Thus, the secular equation becomes

$$\begin{vmatrix} \varepsilon_0(\Gamma_1) - E & \langle \Gamma_1 0 \,|V|\, 1\Gamma_1\rangle & \langle \Gamma_1 0 \,|V|\, 1\Gamma_3\rangle & \langle \Gamma_1 0 \,|V|\, 1\Gamma_5\rangle_1 & \cdots \\ \langle \Gamma_1 1 \,|V|\, 0\Gamma_1\rangle & \varepsilon_1(\Gamma_1) - E & \langle \Gamma_1 1 \,|V|\, 1\Gamma_3\rangle & \langle \Gamma_1 1 \,|V|\, 1\Gamma_5\rangle_1 & \cdots \\ \langle \Gamma_3 1 \,|V|\, 0\Gamma_1\rangle & \langle \Gamma_3 1 \,|V|\, 1\Gamma_1\rangle & \varepsilon_1(\Gamma_3) - E & \langle \Gamma_3 1 \,|V|\, 1\Gamma_5\rangle_1 & \cdots \\ {}_1\langle \Gamma_5 1 \,|V|\, 0\Gamma_1\rangle & {}_1\langle \Gamma_5 1 \,|V|\, 1\Gamma_1\rangle & {}_1\langle \Gamma_5 1 \,|V|\, 1\Gamma_3\rangle_1 & \varepsilon_1(\Gamma_5) - E & \cdots \\ {}_2\langle \Gamma_5 1 \,|V|\, 0\Gamma_1\rangle & {}_2\langle \Gamma_5 1 \,|V|\, 1\Gamma_1\rangle & \vdots & \vdots & \vdots \\ \vdots & \vdots & & & \end{vmatrix} = 0 \tag{5.46}$$

Equation (5.46) reduces to

$$\begin{vmatrix} \varepsilon_0(\Gamma_1) - E & \langle \Gamma_1 0 \,|V|\, 1\Gamma_1\rangle & 0 & 0 & \cdots \\ \langle \Gamma_1 1 \,|V|\, 0\Gamma_1\rangle & \varepsilon_1(\Gamma_1) - E & 0 & 0 & \cdots \\ 0 & 0 & \varepsilon_1(\Gamma_3) - E & 0 & \cdots \\ 0 & 0 & 0 & \varepsilon_1(\Gamma_5) - E & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{vmatrix} = 0 \tag{5.47}$$

Here, $\varepsilon_n(\Gamma_j) = \frac{h^2}{2ma^2}n^2 + \langle \Gamma_j n\,|V|\,\Gamma_j n\rangle$. There are two things to be noted:

1. The matrix elements of V between different IR's vanish, so many off-diagonal matrix elements are zero. This reduces the determinantal equation to a block diagonal form.
2. The diagonal matrix elements $\langle \Gamma_j n\,|V|\,\Gamma_j n\rangle$ are, in general, different for different IRs Γ_j. This lifts the degeneracy at the symmetry points and splits the fourfold degeneracy into nondegenerate states Γ_1 and Γ_3 and one doubly degenerate state Γ_5 at $E(\Gamma) \simeq 1 \cdot \frac{h^2}{2ma^2}$.

When the energy bands along Δ and along X are classified according to the IRs of the appropriate symmetry group, a band structure like that sketched in Fig. 5.7 results. The degeneracies at Γ and X are lifted by the diagonal matrix elements of the potential. The rare case when two different IRs have the same value of the diagonal matrix element of $V(r)$ is called as *accidental degeneracy*. Two Δ-bands belonging to Δ_1, or two belonging to Δ_2 cannot cross because they are coupled by the nonvanishing matrix elements of $V(r)$ between the two bands. However, a Δ_1 band can cross a Δ_2 band because

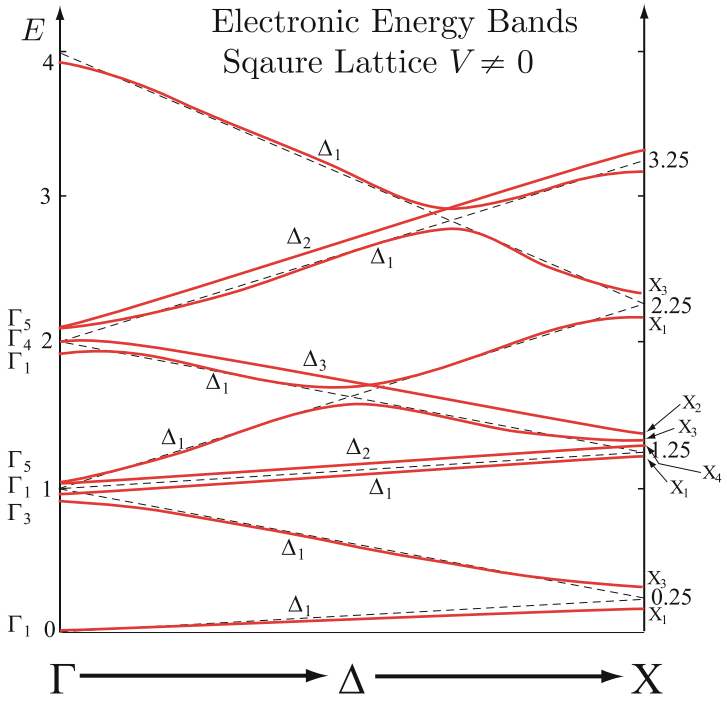


Fig. 5.7. Electronic energy bands of a square lattice with $V \neq 0$ along the line $\Gamma \rightarrow \Delta \rightarrow \mathrm{X}$. The bands are schematic, showing where splittings and anticrossings occur on the simplified diagram (Fig. 5.6) which connects $E(\Gamma)$ and $E(\mathrm{X})$ by *straight lines*

$\langle \Psi_{\Delta_1} | V(r) | \Psi_{\Delta_2} \rangle = 0$. Bands that are widely separated in energy (e.g. the bands at $E(\Gamma) = 1$ and $E(\Gamma) = 4$) can be treated by perturbation theory as was done in the nearly free electron model. One can observe that degeneracies do not occur frequently for bands belonging to the same IR's at Γ (or at X) until the energies become high.

5.6 Empty Lattice Bands for Cubic Structure

5.6.1 Point Group of a Cubic Structure

Every operation of the cubic group will turn x into $\pm x$, $\pm y$, $\pm z$. It is easy to see that there are 48 different operations that can be listed as follows:

1. $x \to \pm x$, $y \to \pm y$, $z \to \pm z$.
2. $x \to \pm x$, $y \to \pm z$, $z \to \pm y$.
3. $x \to \pm y$, $y \to \pm x$, $z \to \pm z$.
4. $x \to \pm y$, $y \to \pm z$, $z \to \pm x$
5. $x \to \pm z$, $y \to \pm y$, $z \to \pm x$
6. $x \to \pm z$, $y \to \pm x$, $z \to \pm y$

Since there are $\pm$ signs we have two possibilities at each step, giving $2^3 = 8$ operations on each line or 48 operations all together.

We can also think of the 48 operations in terms of 24 proper rotations and 24 improper rotations:

Proper Rotations
E: Identity $\to$ 1 operation
4: Rotation by $\pm 90°$ about x-, y-, or z-axis $\to$ six operations
4^2: Rotation by $\pm 180°$ about x-, y-, or z-axis $\to$ three operations
2: Rotation by $\pm 180°$ about the six $[110], [1\bar{1}0], [101], [10\bar{1}], [011], [01\bar{1}]$ axes $\to$ six operations
3: Rotation by $\pm 120°$ about the four $\langle 111 \rangle$ axes $\to$ eight operations Hence, we have 24 proper rotations in total.

Improper Rotations
Multiply each by J (inversion operator: $\mathbf{r} \to -\mathbf{r}$) to have 24 improper rotations.

The 24 improper rotations are obtained by multiplying each of the 24 proper rotations by J, the inversion operation ($\mathbf{r} \to -\mathbf{r}$). Clearly there are 10 classes and 48 operations. Using the theorem

$$\sum_{i=IR} l_i^2 = g$$

we can see that there are 10 IRs, four one-dimensional, and two two-dimensional, and four three-dimensional ones.

$$2\{1^2 + 1^2 + 2^2 + 3^2 + 3^2\} = 48.$$

It is probably worthwhile to sit down with a simple cube and work out the following rotations:

1. E: $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} x \\ y \\ z \end{pmatrix}$.
2. Rotation of $\pm\frac{\pi}{2}$ about x, y, z axes,

$$C_x(1): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} x \\ z \\ -y \end{pmatrix}, \; C_x(-1): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} x \\ -z \\ y \end{pmatrix}$$

$$C_y(1): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} -z \\ y \\ x \end{pmatrix}, \; C_y(-1): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} z \\ y \\ -x \end{pmatrix}$$

$$C_z(1): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} y \\ -x \\ z \end{pmatrix}, \; C_z(-1): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} -y \\ x \\ yz \end{pmatrix}$$

3. Rotation of π about x, y, z axes,

$$C_x(2): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} x \\ -y \\ -z \end{pmatrix}, \; C_y(2): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} -x \\ y \\ -z \end{pmatrix}$$

$$C_z(2): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} -x \\ -y \\ z \end{pmatrix}$$

4. Rotation of π about the six $\langle 110 \rangle$ axes,

$$C_{x,y}(2): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} y \\ x \\ -z \end{pmatrix}, \; C_{x,-y}(2): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} -y \\ -x \\ -z \end{pmatrix}$$

$$C_{y,z}(2): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} -x \\ z \\ y \end{pmatrix}, \; C_{y,-z}(2): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} -x \\ -z \\ -y \end{pmatrix}$$

$$C_{z,x}(2): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} z \\ -y \\ -x \end{pmatrix}, \; C_{z,-x}(2): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} -z \\ -y \\ -x \end{pmatrix}$$

5. Rotation of $\pm\frac{2\pi}{3}$ about the six $\langle 111 \rangle$ axes,

$$C_{x,y,z}(+): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} y \\ z \\ x \end{pmatrix}, \; C_{x,y,z}(-): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} z \\ x \\ y \end{pmatrix}$$

$$C_{x,y,-z}(+): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} -z \\ x \\ -y \end{pmatrix}, \; C_{x,y,-z}(-): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} y \\ -z \\ -x \end{pmatrix}$$

$$C_{x,-y,z}(+): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} -y \\ -z \\ x \end{pmatrix}, \; C_{x,-y,z}(-): \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} z \\ -x \\ -y \end{pmatrix}$$

Table 5.12. Characters and IRs of cubic group

(number of operations)	(1)	(3)	(6)	(6)	(8)	(1)	(3)	(6)	(6)	(8)
class →	E	4^2	4	2	3	J	J4^2	J4	J2	J3
representation ↓										
Γ_1 (A_{1g})	1	1	1	1	1	1	1	1	1	1
Γ_2 (A_{2g})	1	1	−1	−1	1	1	1	−1	−1	1
Γ_{12} (E_g)	2	2	0	0	−1	2	2	0	0	−1
Γ'_{15} (T_{1g})	3	−1	1	−1	0	3	−1	1	−1	0
Γ'_{25} (T_{2g})	3	−1	−1	1	0	3	−1	−1	1	0
Γ'_1 (A_{1u})	1	1	1	1	1	−1	−1	−1	−1	−1
Γ'_2 (A_{2u})	1	1	−1	−1	1	−1	−1	1	1	−1
Γ'_{12} (E_u)	2	2	0	0	−1	−2	−2	0	0	1
Γ_{15} (T_{1u})	3	−1	1	−1	0	−3	1	−1	1	0
Γ_{25} (T_{2u})	3	−1	−1	1	0	−3	1	1	−1	0

$$C_{x,-y,-z}(+)\colon \begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} -y \\ z \\ -x \end{pmatrix},\ C_{x,-y,-z}(-)\colon \begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} -z \\ -x \\ y \end{pmatrix}$$

6. All of the above operations multiplied by the inversion **J**,

$$\mathbf{J}\begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} -x \\ -y \\ -z \end{pmatrix}$$

gives the 24 improper rotations.

Characters and irreducible representations of the cubic group are listed in Table 5.12.

5.6.2 Face Centered Cubic Lattice

The primitive translation vectors of a face centered cubic lattice are given by

$$\mathbf{a}_1 = \frac{a}{2}(\hat{x}+\hat{y}), \quad \mathbf{a}_2 = \frac{a}{2}(\hat{z}+\hat{x}), \quad \mathbf{a}_3 = \frac{a}{2}(\hat{y}+\hat{z}). \tag{5.48}$$

The primitive vectors of the reciprocal lattice (including the factor 2π) are

$$\mathbf{b}_1 = \frac{2\pi}{a}(-\hat{x}-\hat{y}+\hat{z}), \mathbf{b}_2 = \frac{2\pi}{a}(-\hat{x}+\hat{y}-\hat{z}), \mathbf{b}_3 = \frac{2\pi}{a}(\hat{x}-\hat{y}-\hat{z}). \tag{5.49}$$

An arbitrary vector of the reciprocal lattice can be written as

$$\begin{aligned} &\mathbf{G_k} = h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3, \\ &\mathbf{G_k} = \tfrac{2\pi}{a}\left[(-h_1-h_2+h_3)\hat{x} + (-h_1+h_2-h_3)\hat{y} + (h_1-h_2-h_3)\hat{z}\right]. \end{aligned} \tag{5.50}$$

Brillouin Zone

There are eight shortest and six next shortest reciprocal lattice vectors from the origin of reciprocal space to neighboring points (remember that the reciprocal lattice of an FCC is a BCC). They are given by

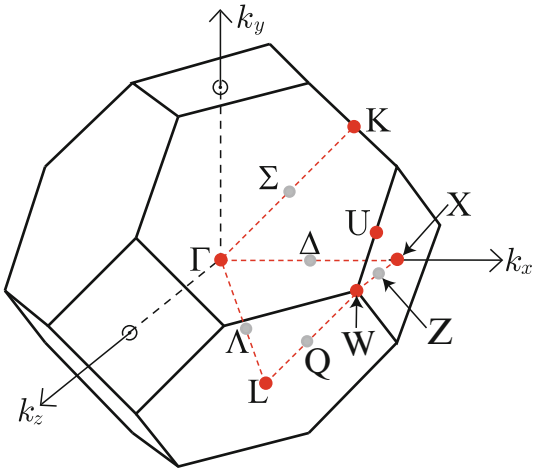


Fig. 5.8. The first Brillouin zone of FCC lattice

1. The eight vectors $\frac{2\pi}{a}[\pm\hat{x}, \pm\hat{y}, \pm\hat{z}]$ whose length is $|G| = \frac{2\pi}{a}\sqrt{3}$.
2. The six vectors $\frac{2\pi}{a}(\pm 2\hat{x}), \frac{2\pi}{a}(\pm 2\hat{y}), \frac{2\pi}{a}(\pm 2\hat{z})$ whose length is $|G| = \frac{2\pi}{a} \cdot 2$.

The first Brillouin zone is the volume enclosed by the planes which are the perpendicular bisectors of these 14 **G**-vectors. The first Brillouin zone of FCC lattice has six square faces perpendicular to (100) and eight hexagonal faces perpendicular to (111). (See Fig. 5.8.)

The names of high symmetry points are labelled in Fig. 5.8. Γ is the origin. Arbitrary points along (100), (110), and (111) directions are called Δ, Σ, and Λ, respectively. The special points X, L, K, and W are

$$X = \frac{2\pi}{a}(1,0,0), L = \frac{2\pi}{a}\left(\frac{1}{2},\frac{1}{2},\frac{1}{2}\right), K = \frac{2\pi}{a}\left(\frac{3}{4},\frac{3}{4},0\right), W = \frac{2\pi}{a}\left(\frac{1}{2},1,0\right).$$

The energy of a free electron is given by

$$E = \frac{h^2}{2ma^2}\left[(l_1+\xi)^2 + (l_2+\eta)^2 + (l_3+\zeta)^2\right] \tag{5.51}$$

where

$$l_1 = -h_1 - h_2 + h_3, l_2 = -h_1 + h_2 - h_3, l_3 = h_1 - h_2 - h_3$$

and h_i are integers. If we measure energy in units of $\frac{h^2}{2ma^2}$, then

$$\begin{aligned} E(\Gamma) &= l_1^2 + l_2^2 + l_3^2, \\ E(X) &= (l_1+1)^2 + l_2^2 + l_3^2, \\ E(L) &= (l_1+\tfrac{1}{2})^2 + (l_2+\tfrac{1}{2})^2 + (l_3+\tfrac{1}{2})^2. \end{aligned} \tag{5.52}$$

One should obtain a table similar to Table 5.13. From the table you constructed, you can draw the empty lattice band structure, showing the bands going from $\Gamma \to X$ and from $\Gamma \to L$.

Table 5.13. Energies for FCC empty lattice $E(\Gamma) \leq 8$. Energy is measured in units of $\frac{h^2}{2ma^2}$

h_1	h_2	h_3	l_1	l_2	l_3	$E(\Gamma)$	$E(X)$	$E(L)$
0	0	0	0	0	0	0	1	$\frac{3}{4}$
1	0	0	−1	−1	1	3	2	$\frac{11}{4}$
−1	0	0	1	1	−1	3	6	$\frac{19}{4}$
0	1	0	−1	1	−1	3	2	$\frac{11}{4}$
0	−1	0	1	−1	1	3	6	$\frac{19}{4}$
0	0	1	1	−1	−1	3	6	$\frac{11}{4}$
0	0	−1	−1	1	1	3	2	$\frac{19}{4}$
1	1	1	−1	−1	−1	3	2	$\frac{3}{4}$
−1	−1	−1	1	1	1	3	6	$\frac{27}{4}$
1	1	0	−2	0	0	4	1	$\frac{11}{4}$
−1	−1	0	2	0	0	4	9	$\frac{27}{4}$
1	0	1	0	−2	0	4	5	$\frac{11}{4}$
−1	0	−1	0	2	0	4	5	$\frac{27}{4}$
0	1	1	0	0	−2	4	5	$\frac{11}{4}$
0	−1	−1	0	0	2	4	5	$\frac{27}{4}$
1	−1	0	0	−2	2	8	9	$\frac{35}{4}$
−1	1	0	0	2	−2	8	9	$\frac{35}{4}$
1	0	−1	−2	0	2	8	5	$\frac{35}{4}$
−1	0	1	2	0	−2	8	13	$\frac{35}{4}$
0	1	−1	−2	2	0	8	5	$\frac{35}{4}$
0	−1	1	2	−2	0	8	13	$\frac{35}{4}$
2	1	1	−2	−2	0	8	5	$\frac{19}{4}$
−2	−1	−1	2	2	0	8	13	$\frac{51}{4}$
1	2	1	−2	0	−2	8	5	$\frac{19}{4}$
−1	−2	−1	2	0	2	8	13	$\frac{51}{4}$
1	1	2	0	−2	−2	8	9	$\frac{19}{4}$
−1	−1	−2	0	2	2	8	9	$\frac{51}{4}$

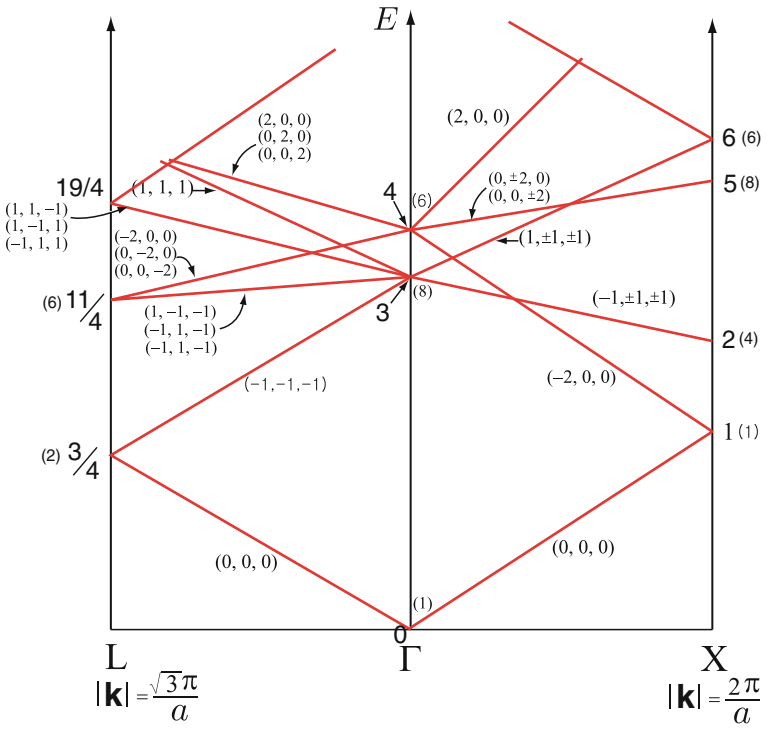


Fig. 5.9. Empty lattice band of the FCC lattice. The energy E is measured in units of $\frac{h^2}{2ma^2}$ and plotted as a function of k. Each band is schematically represented by a *straight line* going from $E_l(\Gamma)$ to $E_l(\mathrm{X})$ or $E_l(\mathrm{L})$ even though the bands really have a more complicated (quadratic form) dependence of the Bloch wave vector $\mathbf{k}$. The set of integers (l_1, l_2, l_3) is indicated for each band

This empty lattice lattice band structure is shown in Fig. 5.9. Note that $\mathbf{k} = \frac{2\pi}{a}\hat{x}$ at X and $\mathbf{k} = \frac{\pi}{a}(\hat{x} + \hat{y} + \hat{z})$ at L. The energy E is sketched as a function of k.

5.6.3 Body Centered Cubic Lattice

The primitive translations of the reciprocal lattice (including 2π) are

$$\mathbf{b}_1 = \frac{2\pi}{a}(\hat{x} + \hat{z}), \mathbf{b}_2 = \frac{2\pi}{a}(-\hat{x} + \hat{y}), \mathbf{b}_3 = \frac{2\pi}{a}(-\hat{y} + \hat{z}). \tag{5.53}$$

Therefore a general reciprocal lattice vector $\mathbf{G}_{h_1h_2h_3}$ is given by

$$\begin{aligned}\mathbf{G}_{h_1h_2h_3} &= \frac{2\pi}{a}\left[h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3\right] \\ &= \frac{2\pi}{a}\left[(h_1 - h_2)\hat{x} + (h_2 - h_3)\hat{y} + (h_3 + h_1)\hat{z}\right]. \end{aligned}\tag{5.54}$$

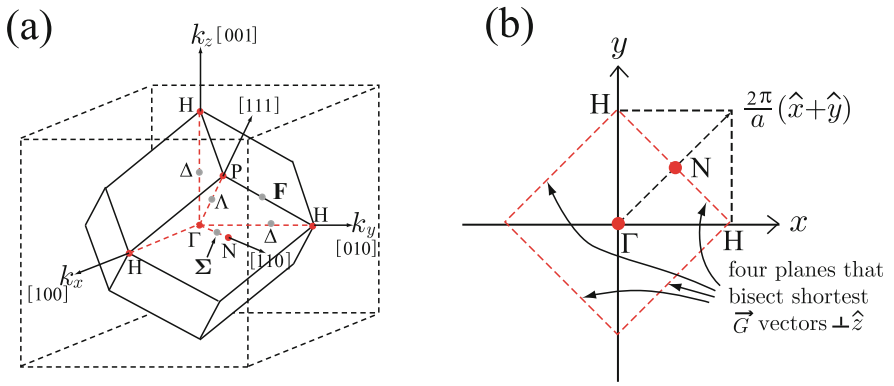


Fig. 5.10. (**a**)First Brillouin zone of the BCC lattice, (**b**) cross section of the four planes bisecting the $\mathbf{G} = \pm\frac{2\pi}{a}(\hat{x}\pm\hat{y})$ perpendicular to the z-axis of the first Brillouin zone of a BCC lattice

The 12 shortest reciprocal lattice vectors are

$$\pm\frac{2\pi}{a}\left(\hat{x}\pm\hat{y}\right), \pm\frac{2\pi}{a}\left(\hat{y}\pm\hat{z}\right), \pm\frac{2\pi}{a}\left(\hat{x}\pm\hat{z}\right).$$

They have length $\frac{2\pi}{a}\sqrt{2}$. The first Brillouin zone is formed by the 12 planes that bisect these 12 shortest reciprocal lattice vectors.(See Fig. 5.10a.)

Figure 5.10b shows the cross section of the four planes that bisect the shortest $\mathbf{G} = \pm\frac{2\pi}{a}(\hat{x}\pm\hat{y})$ perpendicular to the z-axis of the first Brillouin zone of the BCC lattice. The empty lattice energy bands can be written by

$$E_{l_1 l_2 l_3} = \frac{h^2}{2ma^2}\left[(l_1+\xi)^2+(l_2+\eta)^2+(l_3+\zeta)^2\right] \tag{5.55}$$

where

$$l_1 = h_1 - h_2,\ l_2 = h_2 - h_3,\ l_3 = h_3 + h_1,$$

and h_i are integers. We use the symbols of $\mathbf{k} = \frac{2\pi}{a}(\xi,\eta,\zeta)$, H=$\frac{2\pi}{a}(1,0,0)$, P=$\frac{2\pi}{a}(\frac{1}{2},\frac{1}{2},\frac{1}{2})$, and N=$\frac{2\pi}{a}(\frac{1}{2},\frac{1}{2},0)$. Thus, we have, in units of $\frac{h^2}{2ma^2}$

$$\begin{aligned}
&E_{l_1 l_2 l_3}(\Gamma) = l_1^2 + l_2^2 + l_3^2,\\
&E_{l_1 l_2 l_3}(H) = (l_1+1)^2 + l_2^2 + l_3^2,\\
&E_{l_1 l_2 l_3}(P) = (l_1+\tfrac{1}{2})^2 + (l_2+\tfrac{1}{2})^2 + (l_3+\tfrac{1}{2})^2,\\
&E_{l_1 l_2 l_3}(N) = (l_1+\tfrac{1}{2})^2 + (l_2+\tfrac{1}{2})^2 + l_3^2.
\end{aligned} \tag{5.56}$$

Since $l_1 = h_1 - h_2$, $l_2 = h_2 - h_3$, $l_3 = h_3 + h_1$, we can write

$$\begin{aligned}
&E(\Gamma) = (h_1-h_2)^2 + (h_2-h_3)^2 + (h_1+h_3)^2,\\
&E(H) = (h_1-h_2+1)^2 + (h_2-h_3)^2 + (h_1+h_3)^2,\\
&E(P) = (h_1-h_2+\tfrac{1}{2})^2 + (h_2-h_3+\tfrac{1}{2})^2 + (h_1+h_3+\tfrac{1}{2})^2,\\
&E(N) = (h_1-h_2+\tfrac{1}{2})^2 + (h_2-h_3+\tfrac{1}{2})^2 + (h_1+h_3)^2.
\end{aligned} \tag{5.57}$$

5.7 Energy Bands of Common Semiconductors

Many common semiconductors which crystallize in the cubic zincblende structure have valence–conduction band structures that are quite similar in gross features. This results from the fact that each atom (or ion) has four electrons outside a closed shell and there are two atoms per unit cell. For example, silicon has the electron configuration $[\mathrm{Ne}]3s^2 3p^2$, i.e., two $3s$ electrons and two $3p$ electrons outside a closed neon core. With two silicon atoms per unit cell, this gives eight electrons per unit cell. The empty lattice has a single Γ_1 band at $E_\Gamma = 0$ and 8-fold degenerate bands at $E_\Gamma = 3$. The eightfold degeneracy is lifted by the periodic potential, so the valence and conduction bands at Γ will arise from these eight bands. Germanium has the electron configuration of $[\mathrm{Ar}]3d^{10}4s^2 4p^2$, and III–V compounds like GaAs $\left\{\mathrm{Ga}\left([\mathrm{Ar}]3d^{10}4s^2 4p^1\right)\mathrm{As}\left([\mathrm{Ar}]3d^{10}4s^2 4p^3\right)\right\}$ look just like Ge if one $4p$ electron transfers from As to Ga leaving a somewhat ionic $\mathrm{Ga}^-\mathrm{As}^+$ molecule in the unit cell instead of two Ge atoms. The same is true if any III–V elements replace a pair of Si atoms or Ge atoms in a zincblende structure.

A nice example of the use of group concepts in studying energy band structure is a simple nearly free electron type model used to give a rather good description of the valence–conduction band semiconductors with zincblende structures. We will give a rough sketch of the calculation, referring the reader to an article by D. Brust[2]. To describe the band structures of Si and Ge, Brust use the following 15 plane-wave wave functions corresponding to the 15 bands at Γ which have energy $E \leq 4$. (See the 15 bands at Γ in Fig. 5.9.) We can write these 15 plane waves as w_i, with $i = 1, 2, 3, \ldots, 15$ defined by

$$
\begin{aligned}
w_0 &= 1 \; E_0(\Gamma) = 0, \\
w_1 &= w_5^* = \mathrm{e}^{\frac{2\pi \mathrm{i}}{a}(x+y+z)} \; E_1(\Gamma) = E_5(\Gamma) = 3, \\
w_2 &= w_6^* = \mathrm{e}^{\frac{2\pi \mathrm{i}}{a}(x-y-z)} \; E_2(\Gamma) = E_6(\Gamma) = 3, \\
w_3 &= w_7^* = \mathrm{e}^{\frac{2\pi \mathrm{i}}{a}(-x+y-z)} \; E_3(\Gamma) = E_7(\Gamma) = 3, \\
w_4 &= w_8^* = \mathrm{e}^{\frac{2\pi \mathrm{i}}{a}(-x-y+z)} \; E_4(\Gamma) = E_8(\Gamma) = 3, \\
w_9 &= w_{12}^* = \mathrm{e}^{\frac{2\pi \mathrm{i}}{a}x} \; E_9(\Gamma) = E_{12}(\Gamma) = 4, \\
w_{10} &= w_{13}^* = \mathrm{e}^{\frac{2\pi \mathrm{i}}{a}y} \; E_{10}(\Gamma) = E_{13}(\Gamma) = 4, \\
w_{11} &= w_{14}^* = \mathrm{e}^{\frac{2\pi \mathrm{i}}{a}z} \; E_{11}(\Gamma) = E_{14}(\Gamma) = 4,
\end{aligned}
$$

From these 15 functions $w_0, w_1, \ldots, w_{14}$, one can construct linear superpositions belonging to IRs of the group of the wave vector Γ, X, L, etc. Some examples are

[2] D. Brust, Phys. Rev. **134**, A1337 (1964).

$$
\begin{aligned}
&\Psi_1 = \tfrac{1}{\sqrt{V}} w_0 && \text{belongs to } \Gamma_1,\\
&\Psi_2 = \tfrac{1}{\sqrt{8V}}[w_1 - w_2 - w_3 - w_4 + w_5 - w_6 - w_7 - w_8] && \text{belongs to } \Gamma_1,\\
&\quad\vdots && \quad\vdots\\
&\Psi_9 = \tfrac{1}{\sqrt{8V}}[w_1 + w_2 - w_3 + w_4 - w_5 - w_6 + w_7 - w_8] && \text{belongs to } \Gamma_{15},\\
&\Psi_{10} = \tfrac{1}{\sqrt{8V}}[w_9 + w_{10} + w_{11} - w_{12} - w_{13} - w_{14}] && \text{belongs to } \Gamma'_2,\\
&\quad\vdots && \quad\vdots\\
&\Psi_{15} = \tfrac{1}{\sqrt{2V}}[w_{11} + w_{14}] && \text{belongs to } \Gamma_{25}.
\end{aligned}
$$

If you use these combinations of plane waves, the Schrödinger equation breaks up into a block diagonal 15×15 matrix as shown in (5.58).

$$
\begin{vmatrix}
\Gamma_1 & & & & & & & & & & & & & & \\
& \Gamma'_{25}[3\times 3] & & & & & & & & & & & & & \\
& & \Gamma_{15}[3\times 3] & & & & & & & & & & & & \\
& & & & \Gamma'_2 & & & & & & & & & & \\
& & & & & \Gamma_1 & & & & & & & & & \\
& & & & & & & \Gamma_1 & & & & & & & \\
& & & & & & & & \Gamma_{12}[2\times 2] & & & & & & \\
& & & & & & & & & & & \Gamma_{25}[3\times 3] & & & \\
& & & & & & & & & & & & & & \\
& & & & & & & & & & & & & & \\
& & & & & & & & & & & & & & \\
& & & & & & & & & & & & & & \\
& & & & & & & & & & & & & & \\
& & & & & & & & & & & & & &
\end{vmatrix} = 0. \qquad (5.58)
$$

Here Γ_1 and Γ'_2 are 1×1, Γ_{12} is a 2×2, and Γ_{15}, Γ_{25}, and Γ'_{25} are 3×3 matrices, respectively. Off diagonal elements and bands in higher empty lattice states are treated by standard nondegenerate perturbation theory.

Now the question arises "What do we use for the periodic potential $V(r) = \sum_{l_1 l_2 l_3} V_{l_1 l_2 l_3} \mathrm{e}^{\mathrm{i}\mathbf{K}_1 \cdot \mathbf{r}}$?" Brust simply treated the parameters $V_{l_1 l_2 l_3}$ as phenomenological coefficients to be obtained by fitting band gaps and effective masses measured experimentally or fitting more detailed first-principles band structure calculations. He found that he could obtain a reasonably satisfactory fit by keeping only three parameters:

$$
\begin{aligned}
V(3) &= V_{l_1 l_2 l_3} \text{ when } l_1^2 + l_2^2 + l_3^2 = 3,\\
V(8) &= V_{l_1 l_2 l_3} \text{ when } l_1^2 + l_2^2 + l_3^2 = 8,\\
V(11) &= V_{l_1 l_2 l_3} \text{ when } l_1^2 + l_2^2 + l_3^2 = 11.
\end{aligned} \qquad (5.59)
$$

Remember, by cubic symmetry, $V_{1,1,1} = V_{1,1,-1} = V_{1,-1,-1} = V_{-1,-1,-1}$ etc. For Si, Brust found that $V(3) \simeq -0.21$ Ry, $V(8) \simeq 0.04$ Ry, $V(11) \simeq 0.08$ Ry. For Ge he found that $V(3) \simeq -0.23$ Ry, $V(8) \simeq 0.00$ Ry, $V(11) \simeq 0.06$ Ry.

The band structure obtained for Si is illustrated in Fig. 5.11. This diagram shows 11 bands at Γ out of the 15 bands we put into the calculation. Since there are two atoms per unit cell and four valence electrons per atom, we have eight electrons per unit cell or enough to fill four bands. Thus the Γ'_{25} state is the top of the valence band. The conduction band is Γ_{15} at Γ, but the minimum is at near the X-point, so the conduction band minimum has six valleys, each very near one of the six X points.

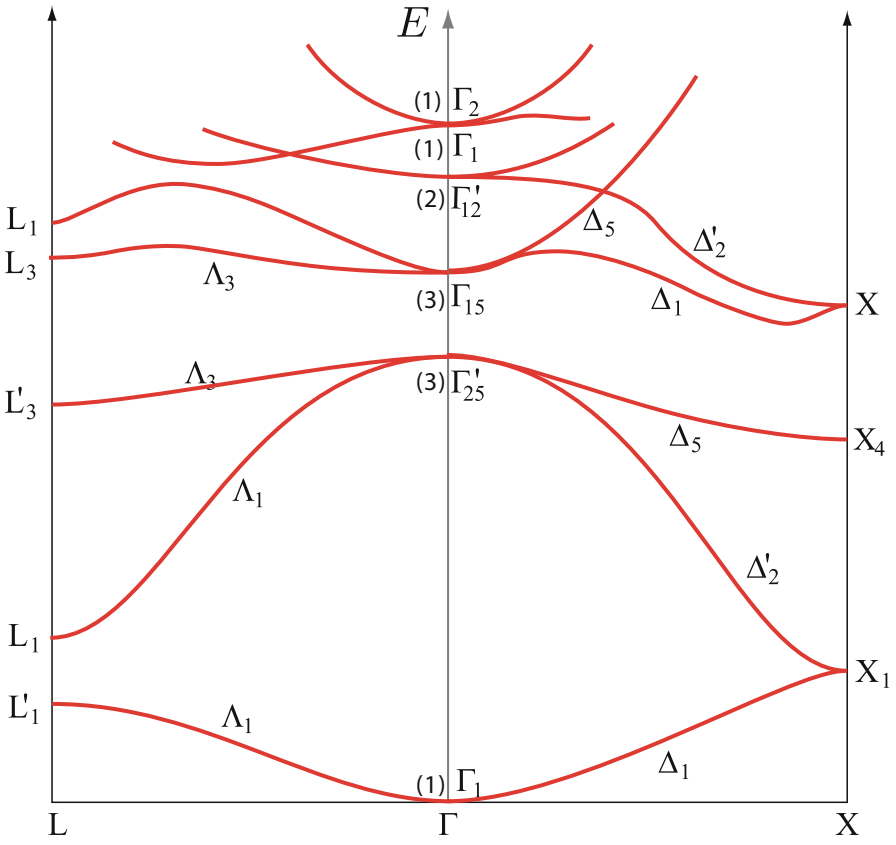


Fig. 5.11. Simple band structure of Si

Problems

5.1. At $E(\mathrm{X}) = 0.25\frac{h^2}{2ma^2}$ there are two degenerate bands. At $E(\mathrm{X}) = 1.25\frac{h^2}{2ma^2}$ there are four. Determine the linear combinations of degenerate states at these points belonging to IR's of $\mathcal{G}_X$. Do the same for $E(\Gamma) = 1\left[\frac{h^2}{2ma^2}\right]$ and $2\left[\frac{h^2}{2ma^2}\right]$

5.2. Make a table, with values of h_1, h_2, h_3, of the resulting l_1, l_2, l_3 and $E(\Gamma)$, $E(\mathrm{X})$, $E(L)$ for all bands with $E(\Gamma) \leq 4$ for an FCC lattice.

5.3. Use your knowledge of the irredicible representations at $E(\Gamma) = 0, 1, 2$ (in units of $\frac{h^2}{2ma^2}$) and at $E(X) = 0.25$ and 1.25, together with the compatibility relations to determine the irreducible representations for each of these bands along the line Δ.

5.4. Tabulate $E(\Gamma), E(H), E(P)$ for all bands that have $E(\Gamma) \leq 4$. Then sketch E vs. $\mathbf{k}$ along $\Delta(\Gamma \to H)$ and along $\Lambda(\Gamma \to P)$.

5.5. Do the same as above in part 4 for simple cubic lattice where

$$E_{\boldsymbol{l}} = \frac{h^2}{2ma^2}\left[(l_1+\xi)^2 + (l_2+\eta)^2 + (l_3+\zeta)^2\right]$$

for Γ, $X = \frac{\pi}{a}(1,0,0)$, and $R = \frac{\pi}{a}(1,1,1)$. Sketch E vs. $\mathbf{k}$ along $\Gamma \to X$ and along $\Gamma \to R$ for all bands having $E(\Gamma) \leq 4$.

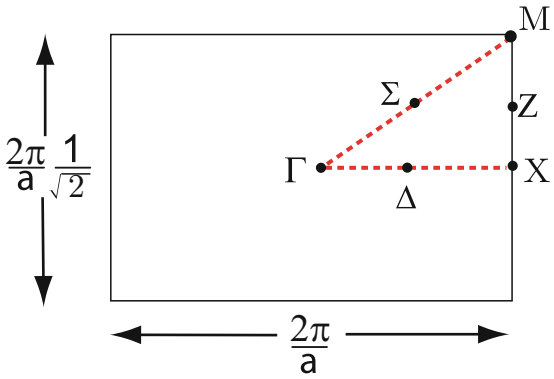
5.6. Use the irreducible representations at $E(\mathrm{X}) = 0.25$ to evaluate

$$V_{ij} = \left\langle \Psi_{\mathrm{X}_i}(0.25) \mid V(\mathbf{r}) \mid \Psi_{\mathrm{X}_j}(0.25) \right\rangle$$

where $\Psi_{\mathrm{X}_i}(0.25)$ is the wave function at $E(\mathrm{X}) = 0.25$ belonging to the irreducible representation X_i.

(a) Show that $V_{ij} = 0$ if $i \neq j$.
(b) Show that the diagonal matrix elements give the same energies (and band gap) as obtained by degenerate perturbation theory with the original plane waves.

5.7. A two dimensional rectangular lattice has a reciprocal lattice whose primitive translations, including the 2π, are $\mathbf{b}_1 = \frac{2\pi}{a}\hat{x}$ and $\mathbf{b}_2 = \frac{2\pi}{a}\frac{1}{\sqrt{2}}\hat{y}$.



(a) List the operations belonging to G_Γ.
(b) Do the same for G_X and G_Δ.
(c) For the empty lattice the wave functions and energies can be written $\Psi_{\boldsymbol{l}}(\mathbf{k},\mathbf{r}) = e^{i(\mathbf{k}+\mathbf{K}_{\boldsymbol{l}})\cdot\mathbf{r}}$ and $E_{\boldsymbol{l}}(\mathbf{k}) = \frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{K}_{\boldsymbol{l}})^2$. Here, $\mathbf{K}_{\boldsymbol{l}} = l_1\mathbf{b}_1 + l_2\mathbf{b}_2$, and l_1 and l_2 are integers. Tabulate the energies at Γ and at X for (l_1, l_2) =(0,0), $(0, \pm 1)$, $(-1,0)$, (1,0), and $(-1, \pm 1)$.
(d) Sketch (straight lines are OK) E vs. k along the line Δ (going from Γ to X) for these bands.
(e) Two degenerate bands at the point $E(\Gamma) = 0.5$ connect to $E(X) = 0.75$. Write down the wave functions for an arbitrary value of k_x for these two bands.
(f) From these wave functions, construct the linear combinations belonging to irreducible representations of G_Δ.

5.8. Graphene has a two-dimensional regular hexagonal reciprocal lattice whose primitive translations are $\boldsymbol{b}_1 = \frac{2\pi}{a}(1, -\frac{1}{\sqrt{3}})$ and $\boldsymbol{b}_2 = \frac{2\pi}{a}(0, \frac{2}{\sqrt{3}})$.

(a) List the operations belonging to G_Γ.
(b) Do the same for G_K and G_M. Note that $\mathbf{k}_K = \frac{2\pi}{a}(\frac{2}{3}, 0)$ and $\mathbf{k}_M = \frac{2\pi}{a}(\frac{1}{2}, \frac{1}{2\sqrt{3}})$.
(c) Write down the empty lattice wave functions and energies $\Psi_{\boldsymbol{l}}(\mathbf{k},\mathbf{r})$ and $E_{\boldsymbol{l}}(\mathbf{k})$ at Γ and K.
(d) Tabulate the energies at Γ and at K for (l_1, l_2) =(0,0), $(0, \pm 1)$, $(\pm 1, 0)$, $(-1, -1)$, and $(1, 1)$.
(e) Sketch E vs. k along the line going from Γ to K for these bands.
(f) Write down the wave functions for the three fold degenerate bands at the energy $E(K) = 4/9$.

Summary

In this chapter, we first reviewed elementary group theory and studied the electronic band structure in terms of elementary concepts of the group theory. We have shown that how group theory ideas can be used in obtaining the band structure of a solid. Group representations and characters of two-dimensional square lattice are discussed in depth and empty lattice bands of the square lattice are illustrated. Concepts of irreducible representations and compatibility relations are used in discussing the symmetry character of bands connecting different symmetry points and the removal of band degeneracies. We also discussed empty lattice bands of the cubic system and sketched the band calculation of common semiconductors.

The starting point for many band structure calculations is the empty lattice band structure. In the empty lattice band representation, each band is labeled by $\boldsymbol{\ell} = (l_1, l_2, l_3)$ where the reciprocal lattice vectors are given by

$$\mathbf{K}_{\boldsymbol{\ell}} = 2\pi[l_1\mathbf{b}_1 + l_2\mathbf{b}_2 + l_3\mathbf{b}_3]$$

where $(l_1, l_2, l_3) = \boldsymbol{\ell}$ are integers and $\mathbf{b}_i$ are primitive translations of the reciprocal lattice. Energy eigenvalues and eigenfunctions are written as

$$E_{\boldsymbol{\ell}}(\mathbf{k}) = \frac{\hbar^2}{2m}\left(\mathbf{k} + \mathbf{K}_\ell\right)^2$$

and

$$\Psi_{\boldsymbol{\ell}}(\mathbf{k}, \mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\mathrm{e}^{i\mathbf{K}_{\boldsymbol{\ell}}\cdot\mathbf{r}}.$$

The Bloch wave vector $\mathbf{k}$ is restricted to the first Brillouin zone.

The vector space formed by the degenerate bands at $E(\mathbf{k})$ is invariant under the operations of the group of the wave vector $\mathbf{k}$. That is, the space of degenerate states at a point $\mathbf{k}$ in the Brillouin zone provides a representation of the group of the wave vector $\mathbf{k}$. We can decompose this representation into its irreducible components and use the decomposition to label the states.

When we classify the degenerate states according to the IRs of the group of the wave vector, we are able to simplify the secular equation by virtue of a fundamental theorem on matrix elements:

1. The matrix elements of V between different IRs vanish, so many off-diagonal matrix elements are zero. This reduces the determinantal equation to a block diagonal form.
2. The diagonal matrix elements $\langle \Gamma_j n \,|V|\, \Gamma_j n\rangle$ are, in general, different for different IRs Γ_j. This lifts the degeneracy at the symmetry points.

Many common semiconductors which crystallize in the cubic zincblende structure have valence–conduction band structures that are quite similar in gross features. This results from the fact that each atom has four electrons outside a closed shell and there are two atoms per unit cell.

6

More Band Theory and the Semiclassical Approximation

6.1 Orthogonalized Plane Waves

So far, we have expanded the periodic part of the Bloch function $u_{\boldsymbol{\ell}}(\mathbf{k},\mathbf{r})$ in a plane wave basis, i.e.,

$$u_{\boldsymbol{\ell}}(\mathbf{k},\mathbf{r}) = \sum_{\mathbf{K}_{\boldsymbol{\ell}'}} C_{\boldsymbol{\ell}\boldsymbol{\ell}'}(k) \mathrm{e}^{\mathbf{K}_{\boldsymbol{\ell}'}\cdot\mathbf{r}} \tag{6.1}$$

It often occurs that the series for $u_{\boldsymbol{\ell}}(\mathbf{k},\mathbf{r})$ converges very slowly so that many different plane waves must be included in the expansion. The reason for this is that plane wave is not a very good description of the valence and conduction band states in the region of real space in which the core levels are of large amplitude. What are the core levels? They are the tightly bound atomic states associated with closed shell configurations. States outside the core are valence states that are responsible for the binding energy of the solid. For example, consider Table 6.1.

Let us define the eigenfunction

$$|cj\rangle = \Psi_{c_j}(r - R_j) \tag{6.2}$$

to be the core level c ($c = 1s, 2s, 2p, 3s, \ldots$) of the atom located at position R_j. The valence and conduction band states that we are interested in must be orthogonal to these core states. When we expand the periodic part of the Bloch function in plane waves, it takes a very large number of plane waves to give band wave functions with all the necessary wiggles needed to make them orthogonal to core states. For this reason, the *orthogonalized plane waves* (OPW) was introduced by Herring and Hill.[1] We define

$$|w_k\rangle = |k\rangle - \sum_{c',j'} \langle c'j'|k\rangle \, |c'j'\rangle \,. \tag{6.3}$$

[1] C. Herring, Phys. Rev. **57**, 1169 (1940) and C. Herring, A.G. Hill, Phys. Rev. **58**, 132 (1940).

Table 6.1. Electron configurations of core states and valence states of Na, Si, and Cu atoms

Atom	Core states	Valence states
Na	$1s^2, 2s^2, 2p^6$	$3s^1$ and higher
Si	$1s^2, 2s^2, 2p^6$	$3s^2, 3p^2$ and higher
Cu	$1s^2, 2s^2, 2p^6, 3s^2, 3p^6$	$3d^{10}, 4s^1$ and higher

Here, $|w_j\rangle$ is an OPW, $|k\rangle$ is a simple plane wave, and the sum is over all core levels on all atoms in the crystal. The core levels are solutions of the Schrödinger equation

$$\left[-\frac{\hbar^2\nabla^2}{2m} + V_j(r) - E_c\right]\Psi_{cj} = 0 \tag{6.4}$$

where $V_j(r)$ is the atomic potential for the atom located at r_j. Because the core levels are tightly bound, this potential is essentially identical to the value of the periodic crystalline potential in the unit cell centered at r_j.

It is clear from (6.3) that $|w_j\rangle$ is orthogonal to the core levels since

$$\langle cj|w_k\rangle = \langle cj|k\rangle - \sum_{c',j'} \langle c'j'|k\rangle \langle cj|c'j'\rangle\,, \tag{6.5}$$

but the core levels themselves satisfy

$$\langle cj|c'j'\rangle = \delta_{cc'}\delta_{jj'} \tag{6.6}$$

This gives $\langle cj|w_k\rangle = 0$. In an OPW calculation, the periodic part of the Bloch function is expanded in OPW's instead of in plane waves. This improves the convergence.

6.2 Pseudopotential Method

We can think of the operator P defined by

$$\mathrm{P} = \sum_{cj} |cj\rangle \langle cj| \tag{6.7}$$

as a projection operator. It gives the projection of any eigenfunction $|\phi\rangle$ onto the core states. If we expand the wave function Ψ_k in OPW's we can write

$$|\Psi_k\rangle = \sum_K a_K |w_{k+K}\rangle = (1-\mathrm{P})\sum_K a_K |(\mathbf{k}+\mathbf{K})\rangle\,. \tag{6.8}$$

Let us define $|\phi_k\rangle = \sum_K a_K |\mathbf{k}+\mathbf{K}\rangle$ as the *pseudo-wavefunction*. Clearly, we have

$$|\Psi_k\rangle = (1-\mathrm{P})\,|\phi_k\rangle\,. \tag{6.9}$$

We note that $|\phi_k\rangle$ is the plane wave part of the OPW expansion. Both $|cj\rangle$ and $|\Psi_k\rangle$ are solutions of the Schrödinger equation

$$\left[-\frac{\hbar^2\nabla^2}{2m} + V(r)\right]\Psi = E\Psi, \tag{6.10}$$

with eigenvalues E_c and $E(k)$, respectively. Let us substitute $|\Psi\rangle = (1-\mathrm{P})\,|\phi\rangle$ into (6.10). This gives

$$\left[-\frac{\hbar^2\nabla^2}{2m} + V(r) - E\right](1-\mathrm{P})\,|\phi\rangle = 0. \tag{6.11}$$

Recall that

$$\mathrm{P}\,|\phi\rangle = \sum_{cj} |cj\rangle\,\langle cj|\phi\rangle\,. \tag{6.12}$$

Therefore, we have

$$\begin{aligned} H\mathrm{P}\,|\phi\rangle &= \sum_{cj} H\,|cj\rangle\,\langle cj|\phi\rangle \\ &= \sum_{cj} E_{c_j}\,|cj\rangle\,\langle cj|\phi\rangle\,. \end{aligned} \tag{6.13}$$

We use this in the Schrödinger equation to obtain

$$\left[-\frac{\hbar^2\nabla^2}{2m} + V(r) - E\right]|\phi\rangle + \sum_{cj}(E-E_{c_j})\,|cj\rangle\,\langle cj|\phi\rangle = 0. \tag{6.14}$$

We define an *effective potential* or *pseudopotential* by

$$W(r) = V(r) + \sum_{cj}(E-E_{c_j})\,|cj\rangle\,\langle cj|\,. \tag{6.15}$$

The first term in the pseudopotential is just the usual periodic crystalline potential. The second term is a nonlocal repulsive potential.

$$\begin{aligned} W(r)\phi(r) &= V(r)\phi(r) + \sum_{cj}(E-E_{c_j})\Psi_{cj}(r)\,\langle cj|\phi(r')\rangle \\ &= \int \mathrm{d}^3r'[V(r')\delta(r-r') + \sum_{cj}(E-E_{c_j})\Psi_{cj}(r)\Psi^*_{cj}(r')]\phi(r'). \end{aligned} \tag{6.16}$$

It is clear that W is nonlocal since only the first term involving the periodic potential contains a δ-function. The second term

$$V_R = \sum_{cj} (E - E_{c_j}) \, |cj\rangle \, \langle cj| \tag{6.17}$$

is repulsive as opposed to an attractive potential like $V(r)$. We can see this by evaluating $\langle \phi | \, V_R \, | \phi \rangle$ for any function ϕ. We find that

$$\langle \phi | \, V_R \, | \phi \rangle = \sum_{cj} (E - E_{c_j}) \, |\langle \phi | cj \rangle|^2 \, . \tag{6.18}$$

Because $|\langle \phi | cj \rangle|^2$ is positive and the valence–conduction band energies E are, by definition, larger than core levels,

$$\langle \phi | \, V_R \, | \phi \rangle > 0. \tag{6.19}$$

Therefore, V_R cancels a portion of the attractive periodic potential V. The diagram shown in Fig. 6.1 is a sketch of what the periodic potential $V(r)$, the repulsive part of the pseudopotential V_R, and the full pseudopotential look like.

A number of people have used model pseudopotentials in which the potential is replaced by the one shown in Fig. 6.2. The pseudopotential $W(r)$ is taken to be a local potential which has (1) a constant value V_0 inside a core or radius d and (2) the actual potential $V(r)$ for $r > d$. Both V_0 and d are used as adjustable parameters to fit the energy bands to experimental observation.

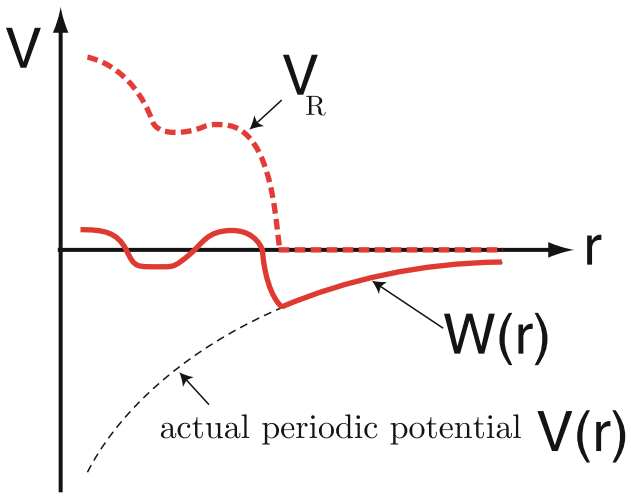


Fig. 6.1. A sketch of various potentials

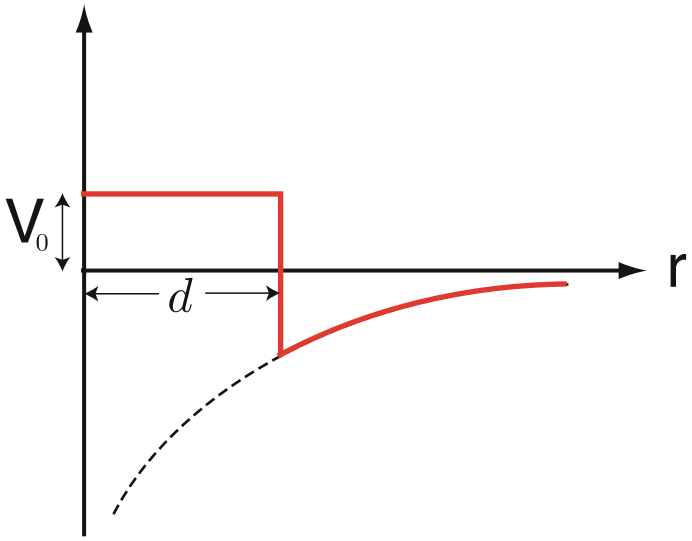


Fig. 6.2. A model pseudopotential

6.3 k · p Method and Effective Mass Theory

Often in discussing properties of semiconductors it is more important to have a single analytic description of the band structure very close to a conduction band minimum or valence band maximum than to have detailed numerical calculations of $E_n(k)$ and Ψ_{nk} throughout the Brillouin zone. One approach that has proven to be useful is called the $\mathbf{k}\cdot\mathbf{p}$ *method*. We know that $\Psi_k = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}u_k(r)$ is a solution of the Schrödinger equation

$$\left(\frac{p^2}{2m} + V(r) - E_k\right)\Psi_k(r) = 0. \tag{6.20}$$

By substituting the Bloch wave form for Ψ_k, it is easy to see that $u_k(r)$ satisfies the Schrödinger equation

$$\left[\frac{(\mathbf{p}+\hbar\mathbf{k})^2}{2m} + V(r) - E_k\right]u_k(r) = 0. \tag{6.21}$$

For $k = 0$ (i.e., at the Γ-point) this equation can be written as

$$\left(\frac{p^2}{2m} + V(r) - E_0\right)u_0(r) = 0. \tag{6.22}$$

There are an infinite number of solutions $u_0^{(1)}$, $u_0^{(2)}$, $u_0^{(3)}$, …, $u_0^{(n)}$, … with energies at the Γ-point $E_0^{(1)}, \ldots, E_0^{(n)}, \ldots$. Here, the superscript (n) is a band

index and the subscript 0 stands for $\mathbf{k} = 0$. For any fixed value of k the set of functions $u_k^{(n)}(r)$ form a complete orthonormal set in which any function with the periodicity of the lattice can be expanded. Therefore, we can use the set of function $u_0^{(n)}(r)$ as a basis set for a perturbation expansion of $u_k^{(m)}$ for $k \neq 0$ and for any band m. By this, we mean that we can write

$$u_k^{(m)} = \sum_n a_n^{(m)}(k) u_0^{(n)}. \tag{6.23}$$

The Schrödinger equation for $u_k^{(m)}$ can be written as

$$\left[\frac{p^2}{2m} + \frac{\hbar}{m}\mathbf{k}\cdot\mathbf{p} + \frac{\hbar^2 k^2}{2m} + V(r) - E_k\right] \sum_n a_n^{(m)}(k) u_0^{(n)}(r) = 0. \tag{6.24}$$

We omit the band superscript (m) for simplicity. We know that $\left[\frac{p^2}{2m} + V(r)\right]$ $u_0^{(n)} = E_0^{(n)} u_0^{(n)}$; therefore, we can write

$$\left(E_0^{(n)} + \frac{\hbar^2 k^2}{2m} - E_k\right) \sum_n a_n u_0^{(n)}(r) + \frac{\hbar}{m}\mathbf{k}\cdot\mathbf{p} \sum_n a_n u_0^{(n)}(r) = 0. \tag{6.25}$$

Take the scalar product with $u_0^{(m)}$ remembering that $\left\langle u_0^{(m)} | u_0^{(n)} \right\rangle = \delta_{mn}$. This gives

$$\left[E_0^{(m)} + \varepsilon_k - E_k\right] a_m + \sum_n \left\langle u_0^{(m)} \left| \frac{\hbar}{m}\mathbf{k}\cdot\mathbf{p} \right| u_0^{(n)} \right\rangle a_n = 0. \tag{6.26}$$

This is just a matrix equation of the form

$$\begin{pmatrix} E_0^{(1)} + \varepsilon_k - E_k & \left\langle u_0^{(1)} \left| H_1 \right| u_0^{(2)} \right\rangle & \cdots \\ \left\langle u_0^{(2)} \left| H_1 \right| u_0^{(1)} \right\rangle & E_0^{(2)} + \varepsilon_k - E_k & \cdots \\ \left\langle u_0^{(3)} \left| H_1 \right| u_0^{(1)} \right\rangle & \left\langle u_0^{(3)} \left| H_1 \right| u_0^{(2)} \right\rangle & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \end{pmatrix} = 0 \tag{6.27}$$

Here, $H_1 = \frac{\hbar}{m}\mathbf{k}\cdot\mathbf{p}$, where $\mathbf{p} = -\mathrm{i}\hbar\nabla$, $\varepsilon_k = \frac{\hbar^2 k^2}{2m}$, and we have put $\left\langle u_0^{(n)} \left| \mathbf{p} \right| u_0^{(n)} \right\rangle = 0$. This last result holds forcrystals with a center of symmetry because parity is a good quantum number and $\mathbf{p}$ is an operator that changes parity. If this matrix element does not vanish it must be added to ε_k.

If we consider k to be small (compared to $\frac{\pi}{a}$), then if $\left\langle u_0^{(m)} \left| \frac{\hbar}{m}\mathbf{k}\cdot\mathbf{p} \right| u_0^{(n)} \right\rangle$ does not vanish, it is usuallyquite small compared to $\left|E_0^{(n)} - E_0^{(m)}\right|$. When

the off-diagonal elements are treated by perturbation theory, the resulting expression for $E_k^{(n)}$ is written as

$$E_k^{(n)} = E_0^{(n)} + \frac{\hbar^2 k^2}{2m} + \frac{\hbar^2}{m^2} \sum_l \frac{\left| \mathbf{k} \cdot \left\langle u_0^{(n)} \left| \mathbf{p} \right| u_0^{(l)} \right\rangle \right|^2}{E_0^{(n)} - E_0^{(l)}}. \tag{6.28}$$

This can be rewritten as

$$E_k^{(n)} = E_0^{(n)} + \frac{\hbar^2}{2} \mathbf{k} \cdot \underline{m^{*-1}} \cdot \mathbf{k}, \tag{6.29}$$

where the *inverse effective mass tensor* (for the band n) is given by

$$m_{ij}^{*\,-1} = m^{-1} \delta_{ij} + \frac{2}{m^2} \sum_l \frac{\left\langle u_0^{(n)} \left| p_i \right| u_0^{(l)} \right\rangle \left\langle u_0^{(l)} \left| p_j \right| u_0^{(n)} \right\rangle}{E_0^{(n)} - E_0^{(l)}}. \tag{6.30}$$

Because $u_0^{(n)}$ is a periodic function with period a, the lattice spacing, the matrix element $\left\langle u_0^{(n)} \left| p_i \right| u_0^{(l)} \right\rangle$ is of the order of $\frac{\hbar}{a}$ if it does not vanish by symmetry considerations. Thus for two coupled bands separated by an energy gap ΔE

$$\frac{m}{m^*} \simeq 1 + 2 \, \frac{\hbar^2 / m a^2}{\Delta E} \tag{6.31}$$

Since $a \simeq 3 \times 10^{-8}\,\mathrm{cm}$, $\frac{\hbar^2}{ma^2} \simeq 10\,\mathrm{eV}$, but typical gaps in semiconductors can be as small as $10^{-1}\,\mathrm{eV}$. Thus, in small-gap semiconductors it is very possible to have

$$m^* = \frac{m}{1 + \frac{\hbar^2 / m a^2}{\Delta E}} \simeq 10^{-2} m.$$

Effective masses of $0.1 - 0.01\, m$ are not at all unusual in semiconductors.

The simple perturbation theory breaks down when there are a number of almost degenerate bands at the point in k-space about which the $\mathbf{k} \cdot \mathbf{p}$ expansion is being made. In that case, it is necessary to keep all the nearly degenerate states in the matrix Schrödinger equation and refrain from using second-order (nondegenerate) perturbation theory. One example of this is the Kane Model[2] used frequently in zincblende semiconductors (like InSb, InAs, GaSb, GaAs, etc.). In these materials, there are four bands that are rather close together (see Fig. 6.3). If spin–orbit coupling is included (it can be important in heavy atoms) one must add to the periodic potential $V(r)$ the *atomic spin–orbit coupling*

$$\frac{\hbar}{4m^2c^2} \left(\sigma \times \nabla_{\mathbf{r}} V \right) \cdot \mathbf{p}. \tag{6.32}$$

[2] E.O. Kane, J. Phys. Chem. Solids **1**, 82 (1956); ibid. **1**, 249 (1957); *Semiconductors and Semimetals*, Vol. 1, pp.75-100 (Academic, New York, 1966).

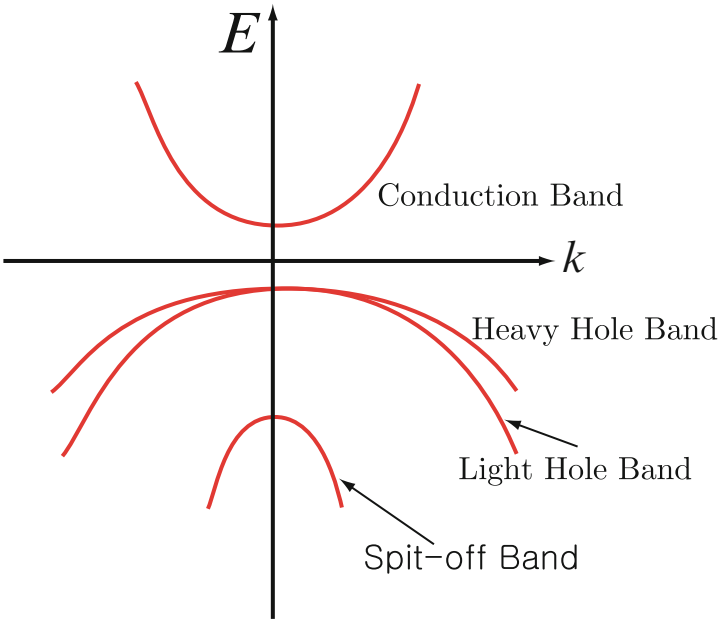


Fig. 6.3. Schematics of the band structure of zincblende semiconductors near the Γ point. Nonparabolicity of the valence band is neglected here

Then the four bands become eight (including the spin splitting) and $\hbar\mathbf{k}\cdot\mathbf{p}$ is replaced by $\mathbf{\Pi}\cdot\mathbf{p}$, where $\mathbf{\Pi}=\hbar\mathbf{k}+\frac{\hbar}{4mc^2}\sigma\times\nabla_{\mathbf{r}}V$. The 8×8 matrix must be diagonalized to obtain a good description of the conduction–valence band structure near the Γ point.

6.4 Semiclassical Approximation for Bloch Electrons

When we discussed the Sommerfeld model of free electrons, we discussed the motion of electrons in response to electric fields and temperature gradients which introduced **r**-dependence into the equilibrium distribution function

$$f_0(\varepsilon)=\frac{1}{\exp^{[\varepsilon-\zeta(\mathbf{r})]/k_{\mathrm{B}}T(\mathbf{r})}+1}. \tag{6.33}$$

The eigenfunctions of the Sommerfeld model were plane waves, so the probability that an electron was at a given position **r** was independent of **r**. Therefore, the r-dependence in $f_0(\varepsilon)$ only made sense if we introduced the idea of localized wave packets defined by

$$\Psi_{\mathbf{k}}(\mathbf{r},t)=\sum_{k'}g(\mathbf{k}-\mathbf{k}')\mathrm{e}^{\mathrm{i}(\mathbf{k}'\cdot\mathbf{r}-\omega_{\mathbf{k}'}t)}, \tag{6.34}$$

where $g(\mathbf{k}-\mathbf{k}') \simeq 0$ if $|\mathbf{k}-\mathbf{k}'|$ is larger than some value Δk. By the Heisenberg principle

$$\Delta k \Delta x \overset{\sim}{>} 1, \tag{6.35}$$

so that the electron can be localized in a region Δx of the order of $(\Delta k)^{-1}$. We must have

1. $\Delta x \gg a$, the atomic spacing.
2. $\Delta x \ll L$ the distance over which the potential $\phi(x) = eEx$ or the temperature $T(x)$ changes appreciably.

Thus, the semiclassical wave packet picture can be applied only to slowly varying (in space) perturbations on the free electrons.

In the presence of a periodic potential, we have Bloch states (or Bloch electrons) described by

$$\Psi_{n\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}'\cdot\mathbf{r}} u_{n\mathbf{k}}(r) \tag{6.36}$$

and

$$E = \varepsilon_n(\mathbf{k}) \tag{6.37}$$

Here, $\mathbf{k}$ is restricted to the first Brillouin zone, and there is a gap between different energy bands at the same value of $\mathbf{k}$, i.e., $\varepsilon_n(\mathbf{k}) - \varepsilon_{n'}(\mathbf{k}) = E_{\mathrm{Gap}}(\mathbf{k}) \neq 0$.

The semiclassical wave packet picture can be used to describe the motion of Bloch electrons in a given band in response to slowly varying perturbations by taking

$$\Psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{k'} g_n(\mathbf{k}-\mathbf{k}')\Psi_{n\mathbf{k}'}(\mathbf{r}) \tag{6.38}$$

with $g_n(\mathbf{k}-\mathbf{k}') \simeq 0$ if $|\mathbf{k}-\mathbf{k}'| > \Delta k$. Then, the standard expression for the group velocity of a wave packet gives

$$\mathbf{v}_n(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon_n(\mathbf{k}) \tag{6.39}$$

as the velocity of a Bloch electron of wave vector $\mathbf{k}$ in the nth band. In the presence of a force $\mathbf{F}$, the work done in moving an electron wavepacket a distance $\delta\mathbf{x}$ is written by

$$\delta W = \mathbf{F}\cdot\delta\mathbf{x} = \mathbf{F}\cdot\mathbf{v}_n\delta t \tag{6.40}$$

But this must equal the change in energy

$$\begin{aligned}\delta W &= E_n(\mathbf{k}+\delta\mathbf{k}) - E_n(\mathbf{k}) \\ &= \nabla_{\mathbf{k}}E_n(\mathbf{k})\cdot\delta\mathbf{k} = \hbar\mathbf{v}_n\cdot\dot{\mathbf{k}}\delta t\end{aligned} \tag{6.41}$$

Equating (6.40) and (6.41) gives

$$\dot{\mathbf{k}} = \hbar^{-1}\mathbf{F}. \tag{6.42}$$

The semiclassical description of Bloch electrons satisfies the following rules:

1. The band index n is a constant of the motion; no interband transitions are allowed.
2. $$\dot{\mathbf{r}} = \mathbf{v}_n(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon_n(\mathbf{k}). \tag{6.43}$$
3. $$\hbar\dot{\mathbf{k}} = -e\left(\mathbf{E} + \frac{1}{c}\mathbf{v}_n(\mathbf{k}) \times \mathbf{B}\right). \tag{6.44}$$
4. The contribution of the nth band to the electron density will be

$$2f_0\left(\varepsilon_n(k)\right)\frac{\mathrm{d}^3k}{(2\pi)^3} = \frac{\mathrm{d}^3k/4\pi^3}{1+\exp\left[\left(\varepsilon_n(k)-\mu\right)/k_{\mathrm{B}}T\right]}. \tag{6.45}$$

For free electrons (Sommerfeld model), electrons are not restricted to one band but move continuously in k-space according to $\hbar\dot{\mathbf{k}} =$ Force. For Bloch electrons $\mathbf{k}$ is restricted to the first Brillouin zone and $\mathbf{k} \equiv \mathbf{k} + \mathbf{K}$. Clearly the restriction to band n requirement must break down when the gap $E_{\mathrm{Gap}}(k)$ becomes very small. It can be shown (but not very easily) that the conditions

$$eEa \ll \frac{\left[E_{\mathrm{Gap}}(k)\right]^2}{E_{\mathrm{F}}} \tag{6.46}$$

and

$$\hbar\omega_c \ll \frac{\left[E_{\mathrm{Gap}}(k)\right]^2}{E_{\mathrm{F}}}. \tag{6.47}$$

must be satisfied for the semiclassical treatment of Bloch electrons to be valid. Here, E is the electric field and a the atomic spacing. ω_c is the electron cyclotron frequency and E_{F} the Fermi energy. The breakdown of the inequalities (6.46) and (6.47) lead to interband transitions; they are known as *electric breakdown* and *magnetic breakdown*, respectively.

6.4.1 Effective Mass

The acceleration of a Bloch electron in band n can be written as

$$\begin{aligned}\mathbf{a}_n \equiv \frac{\mathrm{d}\mathbf{v}_n}{\mathrm{d}t} &= \frac{1}{\hbar}\frac{\mathrm{d}}{\mathrm{d}t}\nabla_{\mathbf{k}}\varepsilon_n(\mathbf{k}) \\ &= \frac{1}{\hbar}\nabla_{\mathbf{k}}\nabla_{\mathbf{k}}\varepsilon_n(\mathbf{k}) \cdot \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t}\end{aligned} \tag{6.48}$$

If we write this tensor equation in terms of components, we have

$$\frac{\mathrm{d}v_i^{(n)}}{\mathrm{d}t} = \frac{1}{\hbar}\sum_j \frac{\partial}{\partial k_j}\frac{\partial}{\partial k_i}\varepsilon_n(\mathbf{k})\frac{\mathrm{d}k_j}{\mathrm{d}t}. \tag{6.49}$$

But $\hbar \frac{\mathrm{d}k_j}{\mathrm{d}t} = F_j$, the j-component of the force. Thus, we can write

$$\frac{\mathrm{d}\mathbf{v}_n}{\mathrm{d}t} = \underline{m}_n^{*\,-1} \cdot \mathbf{F} \tag{6.50}$$

where the *effective mass tensor* is defined by

$$\left(\underline{m}_n^{*\,-1}\right)_{ij} = \frac{1}{\hbar^2} \frac{\partial^2 \varepsilon_n(\mathbf{k})}{\partial k_i \partial k_j}. \tag{6.51}$$

Measured effective masses in different materials have widely different values. For example, in nickel there are electrons with $m^* \simeq 15\, m$ while in InSb there are electrons with $m^* \simeq 0.015\, m$.

6.4.2 Concept of a Hole

Symmetry requires that if a band has an energy $\varepsilon(k)$, then the solid must have an energy $\varepsilon(-k)$ satisfying $\varepsilon(-k) = \varepsilon(k)$. The group velocity of the states $\mathbf{k}$ and $-\mathbf{k}$ are equal in magnitude and opposite in direction. In equilibrium, if the state $\mathbf{k}$ is occupied, so is the state $-\mathbf{k}$. Since the velocities are equal in magnitude and opposite in direction, there is no current. A current is obtained by changing the probability of occupancy of the electron states.

A filled band cannot carry any current even in the presence of an electric field. Each electron is accelerated according to the equation:

$$\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = \frac{1}{\hbar}\mathbf{F}. \tag{6.52}$$

If $\mathbf{F}$ is in the x-direction, the electrons move in $\mathbf{k}$ space with $k_x(t) = k_x(0) + \frac{1}{\hbar}F_x t$. An electron arriving at $k_x = \frac{\pi}{a}$ (for a cubic crystal), the edge of the Brillouin zone, reappears at $k_x = -\frac{\pi}{a}$ (i.e., it is Bragg reflected through $k_x - k_x' = K = \frac{2\pi}{a}$. Thus at all times the band is filled; for each electron at $\mathbf{k}$ there is one at $-\mathbf{k}$ with equal but oppositely directed velocity. Therefore, the electrical current density $\mathbf{j} = 0$.

For a partially filled band we can write

$$\mathbf{j} = \frac{1}{V} \sum_{\mathbf{k}\ \text{occupied}} (-e\mathbf{v}_\mathbf{k})\,. \tag{6.53}$$

This can be rewritten as $\mathbf{j} = \frac{1}{V}\left[\sum_{\text{entire band}\mathbf{k}}(-e\mathbf{v}_\mathbf{k}) - \sum_{\mathbf{k}\ \text{unoccupied}}(-e\mathbf{v}_\mathbf{k})\right]$. The first term vanishes, so that

$$\mathbf{j} = \frac{1}{V} \sum_{\mathbf{k}\ \text{empty}} + e\mathbf{v}_\mathbf{k}. \tag{6.54}$$

Thus, for a nearly filled band, we can think of the current as being carried by *holes*, empty states in the almost filled band. These act as if they have a charge $+e$ instead of $-e$, the charge on an electron.

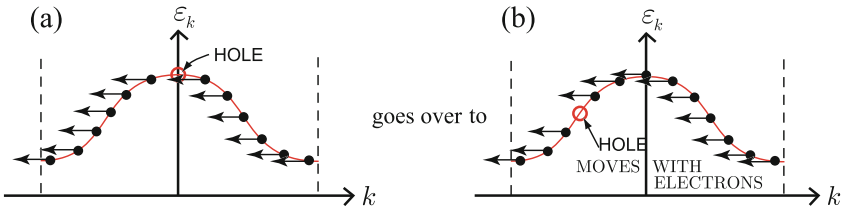


Fig. 6.4. Motion of an electron and a hole: panel (**a**) goes over to panel (**b**), and a hole moves with electrons

Because the equation of motion in **k** space is

$$\hbar \dot{k_x} = -eE_x \tag{6.55}$$

every electron in the nearly filled band moves in k space according to $k_x(t) = k_x(0) - \frac{eE_x}{\hbar}t$. Therefore, the hole moves in the same direction; Fig. 6.4a goes over to Fig. 6.4b. Of course, the effective mass m^* near the top of a valence band is negative since

$$\frac{1}{m^*} = \frac{1}{\hbar^2}\frac{\partial^2 \varepsilon_{\mathbf{k}}}{\partial k^2} < 0. \tag{6.56}$$

It is interesting to write down the following equations that describe the motion of a hole:

$$\hbar \dot{\mathbf{k}} = -e\left(\mathbf{E} + \frac{1}{c}\mathbf{v}_h \times \mathbf{B}\right). \tag{6.57}$$

$$\mathbf{v}_h = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon_{\mathbf{k}}. \tag{6.58}$$

$$m_k^* = -\frac{1}{\hbar^2}\frac{\partial^2 \varepsilon}{\partial k^2} > 0. \tag{6.59}$$

We can assume that a hole has a positive mass near the top of the band where an electron has a negative mass. Then,

$$\frac{\mathrm{d}\mathbf{v}_\mathrm{h}}{\mathrm{d}t} = \underline{m_k^*}^{-1} \cdot \left[e\mathbf{E} + \frac{e}{c}\mathbf{v}_\mathbf{k} \times \mathbf{B}\right]. \tag{6.60}$$

Here, we have used a positive mass $\underline{m_k^*}$ and a positive charge $+e$ to describe the hole. In the valence band of a semiconductor, a few holes can be thermally excited. They can be treated as particles having positive mass and positive charge.

6.4.3 Effective Hamiltonian of Bloch Electron

We know that for Bloch electrons we can write

1. $E = \varepsilon_n(\mathbf{k})$ for the energy of an electron in the nth band.
2. $\Psi_{n\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})$, where $u_{n\mathbf{k}}(\mathbf{r})$ is periodic with the lattice periodicity.

We have seen that close to a minimum (e.g. at $\mathbf{k} = 0$) we can write

$$\varepsilon_n(k) = \varepsilon_n(0) + \frac{\hbar^2}{2}\mathbf{k} \cdot \underline{m}^{*-1} \cdot \mathbf{k}. \tag{6.61}$$

The form of this equation might lead us to write an *effective Hamiltonian*

$$H_{\text{eff}} = \varepsilon_n(0) + \frac{\hbar^2}{2}\left(-i\nabla\right) \cdot \underline{m}^{*-1} \cdot \left(-\mathrm{i}\nabla\right), \tag{6.62}$$

and an effective Schrödinger equation

$$H_{\text{eff}}\varphi(\mathbf{r}) = E\varphi(\mathbf{r}). \tag{6.63}$$

The solution $\varphi(\mathbf{r})$ of (6.63) is not a true wave function for an electron. For example, if we set $\varphi(r) = V^{-1/2}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}$, we obtain $E = \varepsilon_n(0) + \frac{\hbar^2}{2}\mathbf{k} \cdot \underline{m}^{*-1} \cdot \mathbf{k}$. However, the true wave function Ψ is obtained from the *pseudo-wavefunction* φ by multiplying it by $u_{n\mathbf{k}}$, the periodic part of the Bloch function.

So far, we have not really done anything new. However, if we introduce a potential $W(\mathbf{r})$ which is very slowly varying on the atomic scale, we can take as the effective Hamiltonian

$$H_{\text{eff}} = \varepsilon_n(0) + \frac{\hbar^2}{2}\left(-\mathrm{i}\nabla\right) \cdot \underline{m}^{*-1} \cdot \left(-\mathrm{i}\nabla\right) + W(\mathbf{r}). \tag{6.64}$$

Then the solutions to $\left(H_{\text{eff}} - E\right)\varphi(\mathbf{r}) = 0$ will mix Bloch wave functions with different values of $\mathbf{k}$. The smooth function $\varphi(\mathbf{r})$ is called the *envelope function*. This approach can be justified rigorously if the perturbing potential $W(\mathbf{r})$ and the energy band $\varepsilon_n(\mathbf{k})$ satisfy certain conditions.

It turns out that the effective Hamiltonian approach works not only in the regime of the effective mass approximation. In fact, for a Bloch electron (in band n) in the presence of a time independent vector potential $\mathbf{A}(\mathbf{r})$ and a scalar potential $\varphi(\mathbf{r}, t)$, which is slowly varying in space and time, we may define an effective Hamiltonian

$$\mathcal{H}_{\text{eff}} = \varepsilon_n\left(-\mathrm{i}\nabla + \frac{e}{c}\mathbf{A}(\mathbf{r})\right) - e\phi(\mathbf{r}, t). \tag{6.65}$$

This effective Hamiltonian leads to the *semiclassical* equation of motion

$$\dot{\mathbf{r}} = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon_n(\mathbf{k}) = \mathbf{v}_n(\mathbf{k}) \tag{6.66}$$

$$\hbar\dot{\mathbf{k}} = -e\mathbf{E} - \frac{e}{c}\mathbf{v}_n(\mathbf{k}) \times \mathbf{B}, \tag{6.67}$$

where $\mathbf{E} = -\nabla\phi$ and $\mathbf{B} = \nabla \times \mathbf{A}$.

Problems

6.1. For a one-dimensional δ-function potential $V(z) = -\lambda\delta(z)$, show that the lowest energy state occurs at $\varepsilon_0 = -\frac{\hbar^2\kappa^2}{2m}$, where $\mid \kappa \mid = \frac{m\lambda}{\hbar^2}$. Determine the normalized wave function $\Psi_0(z)$.

6.2. Consider a periodic one-dimensional potential

$$V(z) = -\lambda \sum_{l=-\infty}^{\infty} \delta(z - la).$$

(a) Use the tight binding approximation (keeping overlap between nearest neighbors only) to determine the energy $\varepsilon_0(k)$ as a function of wave number k.

(b) Use the tight binding form of the wave function to express the Bloch function in the form $\Psi_0(k, z) = \mathrm{e}^{\mathrm{i}kz}u_0(k, z)$, where $u_0(k, z)$ is a periodic function of z. Write down an expression for $u_0(k, z)$ valid in the region $0 \leq z \leq a$.

6.3. A Wannier function for the nth band of a one-dimensional lattice can be written as

$$a_n(z - la) = \frac{1}{\sqrt{N}} \sum_k \mathrm{e}^{-\mathrm{i}kla}\Psi_{nk}(z).$$

Here, $\Psi_{nk}(z)$ is a Bloch function, $a_n(z-la)$ a Wannier function, and $k = \frac{2\pi}{Na}n$, where $-\frac{N}{2} \leq n \leq \frac{N}{2} - 1$.

(a) Use the orthogonality relation $\langle\Psi_{nk}|\Psi_{nk'}\rangle = \delta_{kk'}$ to show that Wannier functions on different sites are orthogonal, i.e.,

$$\langle a_n(z - l'a)|a_n(z - la)\rangle = \delta_{ll'}.$$

(b) For the model described in Problem 2, determine the Wannier function for the site at the origin, i.e., $l = 0$.

6.4. In the semiclassical treatment of Bloch electrons, one constructs wave packets by making linear combinations of Bloch functions within a single band, with a spread $\Delta\mathbf{k}$ in wave vectors about some particular value of $\mathbf{k}$. The wave packets are localized in coordinate space in a region Δx_i $(i = 1, 2, 3)$ centered on some point $\mathbf{r} = (x_1, x_2, x_3)$, and $\Delta x_i \Delta k_i \simeq 1$. The electron velocity is given by the group velocity $\mathbf{v}_n(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon_n(\mathbf{k})$. (Let us omit the band index n hereafter.) The time rate of change of the wave vector $\mathbf{k}$ is given by $\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = \frac{1}{\hbar}\mathbf{F}$, where $\mathbf{F}$ is the external force on the electron.

(a) If $\mathbf{F} = -eE\hat{x}$, show that

$$k_x(t) = k_x(0) - \frac{eE}{\hbar}t.$$

What happens when $k_x(t)$ reaches the Brillouin zone boundary?

(b) If $\mathbf{F} = -\frac{e}{c}\mathbf{v}_n \times B\hat{z}$, the Lorentz force in a magnetic field $\mathbf{B} = B\hat{z}$, show that the electron moves on a path in $\mathbf{k}$-space that is the intersection of a plane $k_z =$ constant and a surface of constant energy $\varepsilon(\mathbf{k}) =$ constant.

Summary

In this chapter, we studied more theories of band structure calculation and semiclassical description of Bloch electrons. We first introduced orthogonalized plane wave method for expanding the periodic part of the Bloch functions and discussed pseudopotential method and $\boldsymbol{k}\cdot\boldsymbol{p}$ effective mass theory as practical alternative ways of including the effects of periodic symmetry of crystal potential. Then, the semiclassical wave packet picture is discussed to describe the motion of the Bloch electrons in a given band. In addition, ideas of effective mass and hole are shown to be convenient in describing the behavior of band electrons.

It often occurs that the series for $u_{\boldsymbol{\ell}}(\mathbf{k},\mathbf{r}) = \sum_{\mathbf{K}_{\boldsymbol{\ell}'}} C_{\boldsymbol{\ell}\boldsymbol{\ell}'}(k)\mathrm{e}^{\mathbf{K}_{\boldsymbol{\ell}'}\cdot\mathbf{r}}$ converges very slowly so that many different plane waves must be included in the expansion. In an orthogonalized plane wave calculation, the periodic part of the Bloch function is expanded in orthogonalized plane waves instead of in plane waves. This improves the convergence. In many calculations, model pseudopotentials $W(r)$ are introduced in such a way that $W(r)$ is taken to be a local potential which has

1. A constant value V_0 inside a core or radius d and
2. The actual potential $V(r)$ for $r > d$.

Both V_0 and d are used as adjustable parameters to fit the energy bands to experimental observation.

In discussing properties of semiconductors, it is often more important to have a single analytic description of the band structure very close to a conduction band minimum or valence band maximum than to have detailed numerical calculations of $E_n(k)$ and Ψ_{nk} throughout the Brillouin zone. In a $\mathbf{k}\cdot\mathbf{p}$ *method*, energy eigenvalue $E_k^{(n)}$ is written as

$$E_k^{(n)} = E_0^{(n)} + \frac{\hbar^2}{2}\mathbf{k}\cdot\underline{m^{*-1}}\cdot\mathbf{k},$$

where the *inverse effective mass tensor* (for the band n) is given by

$$m_{ij}^{*\,-1} = m^{-1}\delta_{ij} + \frac{2}{m^2}\sum_l \frac{\left\langle u_0^{(n)}\left|p_i\right|u_0^{(l)}\right\rangle\left\langle u_0^{(l)}\left|p_j\right|u_0^{(n)}\right\rangle}{E_0^{(n)} - E_0^{(l)}}.$$

The semiclassical wave packet picture can be used to describe the motion of Bloch electrons in a given band in response to slowly varying perturbations, and the group velocity of a wave packet gives

$$\mathbf{v}_n(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon_n(\mathbf{k})$$

as the velocity of a Bloch electron of wave vector $\mathbf{k}$ in the nth band. In the presence of a force $\mathbf{F}$, we have $\dot{\mathbf{k}} = \hbar^{-1}\mathbf{F}$. The semiclassical description of Bloch electrons satisfies the following rules:

1. The band index n is a constant of the motion; no interband transitions are allowed.
2. $\dot{\mathbf{r}} = \mathbf{v}_n(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon_n(\mathbf{k})$.
3. $\hbar\dot{\mathbf{k}} = -e\left(\mathbf{E} + \frac{1}{c}\mathbf{v}_n(\mathbf{k}) \times \mathbf{B}\right)$.
4. The contribution of the nth band to the electron density is

$$2f_0\left(\varepsilon_n(k)\right)\frac{\mathrm{d}^3k}{(2\pi)^3} = \frac{\mathrm{d}^3k/4\pi^3}{1+\exp\left[\left(\varepsilon_n(k)-\mu\right)/k_{\mathrm{B}}T\right]}.$$

For free electrons, electrons are not restricted to one band but move continuously in k-space according to $\hbar\dot{\mathbf{k}} = \text{Force}$. For Bloch electrons $\mathbf{k}$ is restricted to the first Brillouin zone and $\mathbf{k} \equiv \mathbf{k} + \mathbf{K}$. Clearly the restriction to band n requirement must break down when the gap $E_{\mathrm{Gap}}(k)$ becomes very small. The conditions

$$eEa \ll \frac{\left[E_{\mathrm{Gap}}(k)\right]^2}{E_{\mathrm{F}}};\ \hbar\omega_c \ll \frac{\left[E_{\mathrm{Gap}}(k)\right]^2}{E_{\mathrm{F}}}.$$

must be satisfied for the semiclassical treatment of Bloch electrons to be valid. Here, E is the electric field and a the atomic spacing. ω_c is the electron cyclotron frequency and E_{F} the Fermi energy.

The equation of motion becomes

$$\frac{\mathrm{d}\mathbf{v}_n}{\mathrm{d}t} = \underline{m}_n^{*\,-1} \cdot \mathbf{F}$$

where the *effective mass tensor* is defined by $\left(\underline{m}_n^{*\,-1}\right)_{ij} = \frac{1}{\hbar^2}\frac{\partial^2\varepsilon_n(\mathbf{k})}{\partial k_i \partial k_j}$.

The motion of a hole is described by

$$\hbar\dot{\mathbf{k}} = -e\left(\mathbf{E} + \frac{1}{c}\mathbf{v}_h \times \mathbf{B}\right);\ \mathbf{v}_h = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon_{\mathbf{k}};\ m_k^* = -\frac{1}{\hbar^2}\frac{\partial^2\varepsilon}{\partial k^2} > 0.$$

Since a hole has a positive mass near the top of the band where an electron has a negative mass, we have

$$\frac{\mathrm{d}\mathbf{v}_{\mathrm{h}}}{\mathrm{d}t} = \underline{m}_k^{*\,-1} \cdot \left[e\mathbf{E} + \frac{e}{c}\mathbf{v}_{\mathbf{k}} \times \mathbf{B}\right].$$

The effective Hamiltonian of Bloch electron is written, in the presence of slowly varying potential $W(\mathbf{r})$, as

$$H_{\mathrm{eff}} = \varepsilon_n(0) + \frac{\hbar^2}{2}\left(-i\nabla\right) \cdot \underline{m}^{*\,-1} \cdot \left(-\mathrm{i}\nabla\right) + W(\mathbf{r}).$$

In the presence of a time-independent vector potential $\mathbf{A}(\mathbf{r})$ and a scalar potential $\varphi(\mathbf{r}, t)$, which is slowly varying in space and time, we have an effective Hamiltonian

$$\mathcal{H}_{\mathrm{eff}} = \varepsilon_n\left(-\mathrm{i}\nabla + \frac{e}{c}\mathbf{A}(\mathbf{r})\right) - e\phi(\mathbf{r}, t).$$

This effective Hamiltonian leads to the *semiclassical* equation of motion

$$\dot{\mathbf{r}} = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon_n(\mathbf{k}) = \mathbf{v}_n(\mathbf{k}) \text{ and } \hbar\dot{\mathbf{k}} = -e\mathbf{E} - \frac{e}{c}\mathbf{v}_n(\mathbf{k}) \times \mathbf{B},$$

where $\mathbf{E} = -\nabla\phi$ and $\mathbf{B} = \nabla \times \mathbf{A}$.

7

Semiconductors

7.1 General Properties of Semiconducting Material

In earlier sections, we have seen that a perfect crystal will be

1. An insulator at $T = 0\,\mathrm{K}$, if there is a gap separating the filled and empty energy bands.
2. A conductor at $T = 0\,\mathrm{K}$, if the conduction band is only partially occupied.

A special case of the insulating crystal is that of the semiconductor. In a semiconductor, the gap separating the filled and empty bands is very small, and at finite temperature some electrons from the filled valence band are thermally excited across the energy gap giving $n_\mathrm{e}(T)$ electrons per unit volume in the conduction band and $n_\mathrm{h}(T)$ holes per unit volume in the valence band (of course $n_\mathrm{e} = n_\mathrm{h}$).

If we recall the expression for the conductivity of a free electron model

$$\sigma = \frac{ne^2\tau}{m}, \tag{7.1}$$

where n is the number of carriers per unit volume, we find that different types of materials can be described by different values of n. For a metal $n \simeq 10^{22}$ to $10^{23}\,\mathrm{cm}^{-3}$ and is independent of temperature. For a semimetal $n \simeq 10^{18}$ to $10^{20}\,\mathrm{cm}^{-3}$ and is also roughly temperature independent. For an insulator or a semiconductor

$$n \simeq n_0 \mathrm{e}^{-\frac{E_\mathrm{Gap}}{2k_\mathrm{B}T}},$$

where $n_0 \simeq 10^{22}$ to $10^{23}\,\mathrm{cm}^{-3}$ and the energy gap E_Gap is large ($E_\mathrm{Gap} \geq 4\,\mathrm{eV}$) for an insulator and is small ($E_\mathrm{Gap} \leq 2\,\mathrm{eV}$) for a semiconductor.

At room temperature, $k_\mathrm{B}T \simeq 25\,\mathrm{meV}$, so that $\mathrm{e}^{-\frac{E_\mathrm{Gap}}{2k_\mathrm{B}T}} \leq \mathrm{e}^{-80} \simeq 10^{-35}$ for an insulator, while for a semiconductor $\mathrm{e}^{-\frac{E_\mathrm{Gap}}{2k_\mathrm{B}T}} \geq \mathrm{e}^{-20} \simeq 10^{-9}$. The factor 10^{-35} even when multiplied by $10^{23}\mathrm{cm}^{-3}$ gives $n \simeq 0$ for an insulator. With

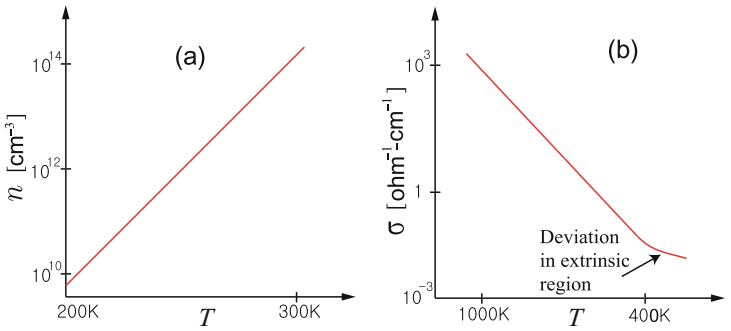


Fig. 7.1. Temperature dependence of carrier concentration (**a**) and electrical conductivity (**b**) of a typical semiconductor

$0.1\,\text{eV} \leq E_{\text{Gap}} < 2.0\,\text{eV}$ the carrier concentration satisfies $10^{22}\,\text{cm}^{-3} > n > 10^{13}\,\text{cm}^{-3}$. The relaxation time τ in the expression for the conductivity is associated with scattering events that dissipate current. These are scattering due to impurities, defects, and phonons. At room temperature, the relaxation time τ of a very pure material will be dominated by phonon scattering. For phonon scattering in this range of temperature $\tau \propto T^{-1}$. Therefore, in a metal the conductivity σ decreases as the temperature is increased. For a semiconductor, τ behaves the same as in a metal for the same temperature range. However, the carrier concentration n increases as the temperature increases. Since n increases exponentially with $\frac{1}{k_{\text{B}}T}$, this increase outweighs the decrease in relaxation time, which is a power law, and σ increases with increasing T.

Intrinsic Electrical Conductivity

In a very pure sample, the conductivity of a semiconductor is due to the excitation of electrons from the valence to the conduction band by thermal fluctuations. For a semiconductor at room temperature the resistivity is between $10^{-2}\,\Omega$-cm and $10^{9}\,\Omega$-cm depending on the band gap of the material. In contrast, a typical metal has a resistivity of $10^{-6}\,\Omega$-cm and a typical insulator satisfies $10^{14}\,\Omega\text{-cm} \leq \rho \leq 10^{22}\,\Omega$-cm. A plot of carrier concentration versus temperature and a plot of conductivity versus temperature is shown in Fig. 7.1a and b.

7.2 Typical Semiconductors

Silicon and germanium are the prototypical covalently bonded semiconductors. In our discussions of energy bands we stated that their valence band maxima were at the Γ-point. The valence band originates from atomic p-states and is threefold degenerate at Γ. Group theory tells us that this degeneracy gives rise to light hole and heavy hole bands, and that an additional splitting

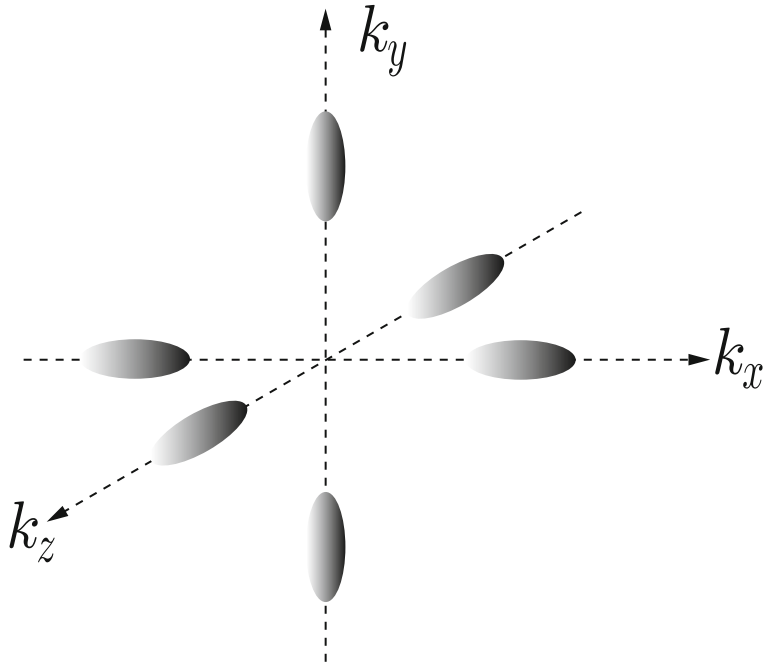


Fig. 7.2. Constant energy surfaces near the conduction band minima for Si

occurs if spin–orbit coupling is taken into account. The conduction band arises from an atomic s-state, but the minimum does not occur at the Γ-point. In Si, the conduction band minimum occurs along the line Δ, at about 90% of the way to the zone boundary. This gives six conduction band minima or *valleys*. (see Fig. 7.2). In the effective mass approximation, these valleys have a longitudinal mass $m_l \simeq 0.98\, m_\mathrm{e}$ along the axis and a transverse mass $m_\mathrm{t} \simeq 0.19\, m_\mathrm{e}$ perpendicular to it. Here, m_e is the mass of a free electron.

For Ge, the conduction band minimum is located at the L-point. This gives the Ge conduction band four minima (one half of each valley is at the zone boundary in the $\langle 111 \rangle$ directions). In Ge, $m_l \simeq 1.64\, m_\mathrm{e}$ and $m_t = 0.08\, m_\mathrm{e}$. Silicon and germanium are called *indirect gap* semiconductors because the valence band maximum and conduction band minimum are at different point in k-space. Materials like InSb, InAs, InP, GaAs, and GaSb are *direct gap* semiconductors because both conduction minimum and valence band maximum occur at the Γ-point. The band structures of many III–V compounds are similar; the sizes of energy gaps, effective masses, and spin splittings differ but the overall features are the same as those of Si and Ge (see Table 7.1). The energy gap is usually determined either by optical absorption or by measuring the temperature dependence of the conductivity. In optical absorption, the initial and final state must have the same wave vector $\mathbf{k}$ if no phonons

Table 7.1. Comparison of energy gaps of Si, Ge, diamond, and various III–V compound semiconductors

Crystal	Type of energy gap	E_{G}[eV] at 0K
Si	Indirect	1.2
Ge	Indirect	0.8
InSb	Direct	0.2
InAs	Direct	0.4
InP	Direct	1.3
GaP	Indirect	2.3
GaAs	Direct	1.5
GaSb	Direct	1.8
AlAs	Indirect	2.24
GaN	Direct	3.5
ZnO	Direct	3.4
Diamond	Indirect	5.48

are involved in the absorption process because the $\mathbf{k}_{\mathrm{ph}}$ vector of the photon is essentially zero on the scale of electron $\mathbf{k}$ vectors. This leads to a sharp increase in absorption at the energy gap of a direct band gap material. For an indirect gap semiconductor, the absorption process is phonon-assisted. It is less abrupt and shows a temperature dependence. The temperature dependence of the conductivity varies, aswe shall show, as $\mathrm{e}^{-\frac{E_{\mathrm{Gap}}}{2k_{\mathrm{B}}T}}$ where E_{Gap} is the minimum gap, the energy difference between the conduction band minimum and the valence band maximum.

7.3 Temperature Dependence of the Carrier Concentration

Let the conduction and valence band energies be given, respectively, by

$$\varepsilon_{\mathrm{c}}(k) = \varepsilon_{\mathrm{c}} + \frac{\hbar^2 k^2}{2m_{\mathrm{c}}} \tag{7.2}$$

and

$$\varepsilon_{\mathrm{v}}(k) = \varepsilon_{\mathrm{v}} - \frac{\hbar^2 k^2}{2m_{\mathrm{v}}}. \tag{7.3}$$

The minimum energy gap is $E_{\mathrm{Gap}} = \varepsilon_{\mathrm{c}} - \varepsilon_{\mathrm{v}}$. The density of states in the conduction band is given by

$$g_{\mathrm{c}}(\varepsilon)\mathrm{d}\varepsilon = \frac{2}{(2\pi)^3} \int_{\varepsilon < \varepsilon_{\mathrm{c}}(k) < \varepsilon + \mathrm{d}\varepsilon} \mathrm{d}^3 k. \tag{7.4}$$

Since, $\varepsilon_{\mathrm{c}}(k)$ is isotropic $\mathrm{d}^3k = 4\pi k^2 \mathrm{d}k$ and $\mathrm{d}\varepsilon = \frac{\hbar^2}{m_{\mathrm{c}}} k\mathrm{d}k$. Substituting into (7.4) gives

$$g_{\mathrm{c}}(\varepsilon) = \frac{\sqrt{2} m_{\mathrm{c}}^{3/2}}{\pi^2 \hbar^3} \left(\varepsilon - \varepsilon_{\mathrm{c}}\right)^{1/2} . \tag{7.5}$$

In a similar way, we have

$$g_{\mathrm{v}}(\varepsilon) = \frac{\sqrt{2} m_{\mathrm{v}}^{3/2}}{\pi^2 \hbar^3} \left(\varepsilon_{\mathrm{v}} - \varepsilon\right)^{1/2} . \tag{7.6}$$

The number of electrons per unit volume in the conduction band is given by

$$n_{\mathrm{c}}(T) = \int_{\varepsilon_{\mathrm{c}}}^{\infty} \mathrm{d}\varepsilon g_{\mathrm{c}}(\varepsilon) f_0(\varepsilon), \tag{7.7}$$

where $f_0(\varepsilon) = \frac{1}{\mathrm{e}^{\frac{\varepsilon-\zeta}{\Theta}}+1}$ is the Fermi distribution function. The concentration of holes in the valence band is written by

$$p_{\mathrm{v}}(T) = \int_{-\infty}^{\varepsilon_{\mathrm{v}}} \mathrm{d}\varepsilon g_{\mathrm{v}}(\varepsilon) \left[1 - f_0(\varepsilon)\right] . \tag{7.8}$$

Note that $1 - f_0(\varepsilon) = 1/\left[\mathrm{e}^{\frac{\zeta-\varepsilon}{\Theta}} + 1\right]$. Clearly $n_{\mathrm{c}}(T)$ and $p_{\mathrm{v}}(T)$ depend on the value of the chemical potential ζ. We will make the simplifying assumption that $\varepsilon_{\mathrm{c}} - \zeta \gg \Theta$ and $\zeta - \varepsilon_{\mathrm{v}} \gg \Theta$, where Θ is, of course, $k_{\mathrm{B}}T$. This nondegeneracy assumption makes the calculation much simpler, and we will evaluate ζ in the course of the calculation and check if the assumption is valid. With this assumption, we can write

$$\begin{aligned} f_0(\varepsilon) &\simeq \mathrm{e}^{-\frac{\varepsilon-\zeta}{\Theta}}, \\ 1 - f_0(\varepsilon) &\simeq \mathrm{e}^{-\frac{\zeta-\varepsilon}{\Theta}} \end{aligned} \tag{7.9}$$

The first line of (7.9) can be rewritten as $f_0(\varepsilon) \simeq \mathrm{e}^{-\frac{\varepsilon-\varepsilon_{\mathrm{c}}}{\Theta}} \mathrm{e}^{-\frac{\varepsilon_{\mathrm{c}}-\zeta}{\Theta}}$. The second factor is independent of ε and can be taken out of the integral in (7.7) to obtain

$$n_{\mathrm{c}}(T) = N_{\mathrm{c}}(T) \mathrm{e}^{-\frac{\varepsilon_{\mathrm{c}}-\zeta}{\Theta}}, \tag{7.10}$$

where

$$N_{\mathrm{c}}(T) = \int_{\varepsilon_{\mathrm{c}}}^{\infty} \mathrm{d}\varepsilon g_{\mathrm{c}}(\varepsilon) \mathrm{e}^{-\frac{\varepsilon-\varepsilon_{\mathrm{c}}}{\Theta}}. \tag{7.11}$$

In a similar manner, one can obtain

$$p_{\mathrm{v}}(T) = P_{\mathrm{v}}(T) \mathrm{e}^{-\frac{\zeta-\varepsilon_{\mathrm{v}}}{\Theta}}, \tag{7.12}$$

and

$$P_{\mathrm{v}}(T) = \int_{\infty}^{\varepsilon_{\mathrm{v}}} \mathrm{d}\varepsilon g_{\mathrm{v}}(\varepsilon) \mathrm{e}^{-\frac{\varepsilon_{\mathrm{v}}-\varepsilon}{\Theta}}. \tag{7.13}$$

Because the density of states varies as $g_c \propto (\varepsilon - \varepsilon_c)^{1/2}$ and $g_v \propto (\varepsilon_v - \varepsilon)^{1/2}$, the integral for $N_c(T)$ and $P_v(T)$ can be evaluated exactly by using the fact that $\int_0^\infty dx\ \sqrt{x}e^{-x} = \frac{1}{2}\sqrt{\pi}$. The results are

$$N_c(T) = \frac{1}{4}\left(\frac{2m_c\Theta}{\pi\hbar^2}\right)^{3/2}. \tag{7.14}$$

The result for $P_v(T)$ differs only in having m_v replace m_c. It is sometimes convenient to use the practical expression

$$N_c(T) \simeq 2.5\left(\frac{m_c}{m}\right)^{3/2}\left(\frac{T}{300\text{K}}\right)^{3/2} \times 10^{19}\,\text{cm}^{-3}. \tag{7.15}$$

Again for $P_v(T)$ we need only replace m_c by m_v. Note the very important fact that the product $n_c(T)p_v(T)$ is independent of ζ, so that

$$n_c(T)p_v(T) = N_c(T)P_v(T)e^{-E_{\text{Gap}}/\Theta}. \tag{7.16}$$

7.3.1 Carrier Concentration: Intrinsic Case

In the absence of impurities, the only carriers are thermally excited electron–hole pairs, so that $n_c(T) = p_v(T)$; this is defined as $n_i(T)$, where i stands for intrinsic. From (7.16), we have

$$n_i(T) = [N_c(T)P_v(T)]^{1/2}\, e^{-E_{\text{Gap}}/2\Theta}. \tag{7.17}$$

To obtain the value of ζ for this case (we will call it ζ_i, i for the intrinsic case) we note that $n_i(T) = n_c(T)$, or

$$[N_c(T)P_v(T)]^{1/2}\, e^{-E_{\text{Gap}}/2\Theta} = N_c(T)e^{-(\varepsilon_c - \zeta_i)\Theta}. \tag{7.18}$$

This can be rewritten as

$$[P_v(T)/N_c(T)]^{1/2} = e^{\frac{\zeta_i - \varepsilon_c + \frac{1}{2}E_{\text{Gap}}}{\Theta}}. \tag{7.19}$$

Solving for ζ_i gives

$$\zeta_i = \varepsilon_c - \frac{1}{2}E_{\text{Gap}} + \frac{3}{4}\Theta\ln\left(\frac{m_v}{m_c}\right). \tag{7.20}$$

In writing (7.20) we have used $[P_v(T)/N_c(T)] = (m_v/m_c)^{3/2}$. In terms of ε_v we can express (7.20) as

$$\zeta_i = \varepsilon_v + \frac{1}{2}E_{\text{Gap}} + \frac{3}{4}\Theta\ln\left(\frac{m_v}{m_c}\right). \tag{7.21}$$

If $m_v = m_c$, then ζ_i always sits in mid-gap. If $m_v \neq m_c$, ζ_i sits at mid-gap at $\Theta = 0$, but moves away from the higher density of states band as

Θ is increased. For $E_{\rm Gap} \simeq 1$ eV, the separations $\zeta_{\rm i} - \varepsilon_{\rm v}$ and $\varepsilon_{\rm c} - \zeta_{\rm i}$ are large compared to Θ for any reasonable temperature, so our assumption is justified.

7.4 Donor and Acceptor Impurities

Si and Ge have four valence electrons. If a small concentration of a column V element replaces some of the host atoms, then there is one electron more than necessary for the formation of the covalent bonds. The extra electron must be placed in the conduction band, and such atoms like As, Sb, and P are known as *donors*. For column III elements (Al, Ga, In, etc.) there is a shortage of one electron, thus, the valence band is not full and a hole exists for every *acceptor* atom.

Let us consider the case of donors (for acceptors, the same picture applies if electrons in the conduction band are replaced by holes in the valence band and, as an example, As^+ ions are replaced by Al^- ions). To a first approximation the extra electron of the As atom will go into the conduction band of the host material. This would give one conduction electron for each impurity from the column V. However, these conduction electron leaves behind an As^+ ion, and the As^+ ion acts as a center of attraction which can bind the conduction electron similar to the binding of an electron by a proton to form a hydrogen atom.

For a hydrogen atom, the Hamiltonian for an electron moving in the presence of a proton located at $r = 0$ is

$$H = \frac{p^2}{2m} - \frac{e^2}{r}. \tag{7.22}$$

The Schrödinger equation has, for its ground state eigenfunction and eigenvalue,

$$\Psi_0 = N_0 \mathrm{e}^{-r/a_{\rm B}} \quad \text{and} \quad E_0 = -\frac{e^2}{2a_{\rm B}}, \tag{7.23}$$

where $a_{\rm B} = \frac{\hbar^2}{me^2}$ is the Bohr radius ($a_{\rm B} \sim 0.5 \mathring{A}$).

For a conduction electron in the presence of a donor ion, we have

$$H = \frac{p^2}{2m_{\rm c}^*} - \frac{e^2}{\epsilon_{\rm s} r}. \tag{7.24}$$

Here, $m_{\rm c}^*$ is the conduction band effective mass and $\epsilon_{\rm s}$ is the background dielectric constant of the semiconductor. The ground state will have

$$\Psi_0 = N_0 \mathrm{e}^{-r/a_{\rm B}^*} \quad \text{and} \quad E_0 = -\frac{e^2}{2\epsilon_{\rm s} a_{\rm B}^*}. \tag{7.25}$$

The effective Bohr radius $a_{\rm B}^*$ is given by $a_{\rm B}^* = \frac{\hbar^2 \epsilon_{\rm s}}{m_{\rm c}^* e^2}$. For a typical semiconductor $m_{\rm c}^* \simeq 0.1m$ and $\epsilon_{\rm s} \simeq 10$. This gives $a_{\rm B}^* \approx 10^2 a_{\rm B} \simeq 50 \mathring{A}$ and $E_0 \approx -10^{-3} \frac{e^2}{2a_{\rm B}} \simeq -13\,\text{meV}$.

When donors are present, the chemical potential ζ will move from its intrinsic value ζ_i to a value near the conduction band edge. We know that $n_\mathrm{c}(T) = N_\mathrm{c}(T)\mathrm{e}^{-\frac{\varepsilon_\mathrm{c}-\zeta}{\Theta}}$; we can define the intrinsic carrier concentration $n_\mathrm{i}(T)$ by $n_\mathrm{i}(T) = N_\mathrm{c}(T)\mathrm{e}^{-\frac{\varepsilon_\mathrm{c}-\zeta_\mathrm{i}}{\Theta}}$. Then we can write for the general case

$$n_\mathrm{c}(T) = n_\mathrm{i}(T)\mathrm{e}^{\frac{\zeta-\zeta_\mathrm{i}}{\Theta}} \quad \text{and} \quad p_\mathrm{v}(T) = n_\mathrm{i}(T)\mathrm{e}^{-\frac{\zeta-\zeta_\mathrm{i}}{\Theta}}. \tag{7.26}$$

If $\zeta = \zeta_\mathrm{i}$, $n_\mathrm{c}(T) = p_\mathrm{v}(T) = n_\mathrm{i}(T)$. If $\zeta \neq \zeta_\mathrm{i}$, then $n_\mathrm{c}(T) \neq p_\mathrm{v}(T)$ and we can write

$$\Delta n \equiv n_\mathrm{c}(T) - p_\mathrm{v}(T) = 2\ n_\mathrm{i}\ \sinh\left(\frac{\zeta-\zeta_\mathrm{i}}{\Theta}\right) = n_\mathrm{c} - \frac{n_\mathrm{i}^2}{n_\mathrm{c}}. \tag{7.27}$$

The product $n_\mathrm{c}(T)p_\mathrm{v}(T)$ is still independent of ζ so we can write $n_\mathrm{c}(T)p_\mathrm{v}(T) = n_\mathrm{i}^2$. Using $p_\mathrm{v} = \frac{n_\mathrm{i}^2}{n_\mathrm{c}(T)}$, (7.27) gives a quadratic equation for n_c

$$n_\mathrm{c}^2 - \Delta n n_\mathrm{c} - n_\mathrm{i}^2 = 0$$

whose solution is

$$n_\mathrm{c} = \frac{\Delta n}{2} + \sqrt{\left(\frac{\Delta n}{2}\right)^2 + n_\mathrm{i}^2}. \tag{7.28}$$

We take the positive(+) root because donor impurities must increase the concentration $n_\mathrm{c}(T)$.

7.4.1 Population of Donor Levels

If the concentration of donors is sufficiently small ($N_\mathrm{d} \leq 10^{19}\mathrm{cm}^{-3}$) that interactions between donor electrons can be neglected, then the average occupancy of a single donor impurity state is given by

$$\langle n_\mathrm{d}\rangle = \frac{\sum_j N_j \mathrm{e}^{-\beta(E_j-\zeta N_j)}}{\sum_j \mathrm{e}^{-\beta(E_j-\zeta N_j)}}. \tag{7.29}$$

Here, $\beta = 1/\Theta$ and the possible values of N_j are

1. $N_j = 0$ when donor atom is empty.
2. $N_j = 1$ when donor atom is occupied by an electron of spin σ.
3. $N_j = 1$ when donor atom is occupied by an electron of spin $-\sigma$.
4. $N_j = 2$ when donor atom is occupied by two electrons of spin σ and $-\sigma$.

There is actually a large repulsion (repulsive energy U) between the electrons in case of $N_j = 2$, so that case of $N_j = 2$ does not actually occur. If we use the cases listed above in (7.29), we obtain

$$\langle n_\mathrm{d}\rangle = \frac{0 + 2\mathrm{e}^{-\beta(\varepsilon_\mathrm{d}-\zeta)} + 2\mathrm{e}^{-\beta(2\varepsilon_\mathrm{d}+U-2\zeta)}}{1 + 2\mathrm{e}^{-\beta(\varepsilon_\mathrm{d}-\zeta)} + \mathrm{e}^{-\beta(2\varepsilon_\mathrm{d}+U-2\zeta)}}. \tag{7.30}$$

If U is much larger than the other energies, then the terms involving U can be neglected; the following result is obtained:

$$\langle n_{\mathrm{d}} \rangle = \frac{1}{\frac{1}{2}\mathrm{e}^{\beta(\varepsilon_{\mathrm{d}}-\zeta)}+1}. \tag{7.31}$$

The numerical factor of $\frac{1}{2}$ in this expression comes from the fact that either spin up or spin down states can be occupied but not both.

7.4.2 Thermal Equilibrium in a Doped Semiconductor

Let us assume that we have N_{d} donors and N_{a} acceptors per unit volume, and let us take $N_{\mathrm{d}} \gg N_{\mathrm{a}}$. This material would be doped n-type since it has many more donors than acceptors. The energies of interest are shown in Fig. 7.3.

At zero temperature, there must be

- $n_{\mathrm{c}} = 0$, no electrons in the conduction band
- $p_{\mathrm{v}} = 0$, no holes in the valence band
- $p_{\mathrm{a}} = 0$, no holes bound to acceptors and
- $n_{\mathrm{d}} = N_{\mathrm{d}} - N_{\mathrm{a}}$, electrons bound to donor atoms.

The $(N_{\mathrm{d}} - N_{\mathrm{a}})$ donors with electrons bound to them are neutral. The remaining N_{a} donors have lost their electrons to the N_{a} acceptors. Thus we have N_{a} positively charged donor ions and N_{a} negatively charged acceptor ions per unit volume. The chemical potential must clearly be at the donor level since they are partially occupied, and only at the energy of $\varepsilon = \zeta$ can the Fermi function have a value different from unity or zero at $T = 0$.

At a finite temperature, we have

$$n_{\mathrm{c}}(T) = N_{\mathrm{c}}(T)\mathrm{e}^{-\beta(\varepsilon_{\mathrm{c}}-\zeta)}, \tag{7.32}$$

$$p_{\mathrm{v}}(T) = P_{\mathrm{v}}(T)\mathrm{e}^{-\beta(\zeta-\varepsilon_{\mathrm{v}})}, \tag{7.33}$$

$$n_{\mathrm{d}}(T) = \frac{N_{\mathrm{d}}}{\frac{1}{2}\mathrm{e}^{\beta(\varepsilon_{\mathrm{d}}-\zeta)}+1}, \tag{7.34}$$

Fig. 7.3. Impurity levels in semiconductors doped with N_{d} donors and N_{a} acceptors per unit volume

and

$$p_{\mathrm{a}}(T) = \frac{N_{\mathrm{a}}}{\frac{1}{2}\mathrm{e}^{\beta(\zeta-\varepsilon_{\mathrm{a}})}+1}. \tag{7.35}$$

In addition to these four equations, we must have charge neutrality so that

$$n_{\mathrm{c}} + n_{\mathrm{d}} = N_{\mathrm{d}} - N_{\mathrm{a}} + p_{\mathrm{v}} + p_{\mathrm{a}} \tag{7.36}$$

Here, $n_{\mathrm{c}}+n_{\mathrm{d}}$ is the number of electrons that are either in the conduction band or bound to a donor. If we forget about holes, $n_{\mathrm{c}}+n_{\mathrm{d}}$ must equal $N_{\mathrm{d}}-N_{\mathrm{a}}$, the excess number of electrons introduced by the impurities. For every hole, either bound to an acceptor or in the valence band, we must have an additional electron contributing to $n_{\mathrm{c}}+n_{\mathrm{d}}$. Equations (7.32)–(7.36) form a set of five equations in five unknowns. We know β, N_{d}, N_{a}, ε_{c}, ε_{v}, ε_{d}, and ε_{a}; the unknowns are $n_{\mathrm{c}}(T)$, $p_{\mathrm{v}}(T)$, $n_{\mathrm{d}}(T)$, $p_{\mathrm{a}}(T)$, and $\zeta(T)$. Although the equations can easily be solved numerically, it is worth looking at the simple case where $\varepsilon_{\mathrm{d}}-\zeta \gg \Theta$ and $\zeta-\varepsilon_{\mathrm{a}} \gg \Theta$. This does not occur at $T=0$ since $\zeta=\varepsilon_{\mathrm{d}}$ in that case; nor does it apparently occur at very high temperature. However, there is a range of temperature where the assumption is valid. With this assumption

$$n_{\mathrm{d}}(T) \simeq 2\ N_{\mathrm{d}}\mathrm{e}^{-\beta(\varepsilon_{\mathrm{d}}-\zeta)} \ll N_{\mathrm{d}}, \tag{7.37}$$

and

$$p_{\mathrm{a}}(T) \simeq 2\ N_{\mathrm{a}}\mathrm{e}^{-\beta(\zeta-\varepsilon_{\mathrm{a}})} \ll N_{\mathrm{a}}. \tag{7.38}$$

We know from (7.36)–(7.38) that

$$\Delta n \equiv n_{\mathrm{c}} - p_{\mathrm{v}} = N_{\mathrm{d}} - N_{\mathrm{a}} + p_{\mathrm{a}} - n_{\mathrm{d}} \approx N_{\mathrm{d}} - N_{\mathrm{a}}. \tag{7.39}$$

From (7.27) $\Delta n = 2\ n_{\mathrm{i}}\ \sinh\beta\left(\zeta-\zeta_{\mathrm{i}}\right)$, and for low concentrations of impurities at sufficiently high temperatures $\beta\left(\zeta-\zeta_{\mathrm{i}}\right)$ must be small. We can then approximate $\sin\mathrm{h}x$ by x and obtain

$$\Delta n \simeq 2\ n_{\mathrm{i}}\beta\left(\zeta-\zeta_{\mathrm{i}}\right). \tag{7.40}$$

We know that

$$n_{\mathrm{c}}(T) = n_{\mathrm{i}}(T)\mathrm{e}^{\beta(\zeta-\zeta_{\mathrm{i}})} \simeq n_{\mathrm{i}}\left[1+\beta\left(\zeta-\zeta_{\mathrm{i}}\right)\right]. \tag{7.41}$$

Using (7.39)–(7.41) gives

$$n_{\mathrm{c}} \simeq n_{\mathrm{i}} + \frac{1}{2}\left(N_{\mathrm{d}}-N_{\mathrm{a}}\right), \tag{7.42}$$

and

$$p_{\mathrm{v}} \simeq n_{\mathrm{i}} - \frac{1}{2}\left(N_{\mathrm{d}}-N_{\mathrm{a}}\right). \tag{7.43}$$

For low concentrations of donors and acceptors at reasonably high temperatures $\Delta n \ll n_{\mathrm{i}}(T)$, so that $2\beta\left(\zeta-\zeta_{\mathrm{i}}\right) \ll 1$ and ζ is relatively close to ζ_{i}. Because $\varepsilon_{\mathrm{d}}-\zeta_{\mathrm{i}}$ is an appreciable fraction of the band gap the assumptions $\beta\left(\varepsilon_{\mathrm{d}}-\zeta\right) \gg 1$ and $\beta\left(\zeta-\varepsilon_{\mathrm{a}}\right) \gg 1$ are valid.

7.4.3 High-Impurity Concentration

For high donor concentration $N_\mathrm{d} - N_\mathrm{a} \gg n_\mathrm{i}$; then $\beta\,(\zeta - \varepsilon_\mathrm{a}) \gg 1$ since the chemical potential moves from midgap closer to the conduction band edge. Because

$$p_\mathrm{a}(T) \simeq 2\ N_\mathrm{a} \mathrm{e}^{-\beta(\zeta-\varepsilon_\mathrm{a})}, \tag{7.44}$$

and

$$p_\mathrm{v}(T) = n_\mathrm{i} \mathrm{e}^{-\beta(\zeta-\zeta_\mathrm{i})} = P_\mathrm{v}(T) \mathrm{e}^{-\beta(\zeta-\varepsilon_\mathrm{v})}, \tag{7.45}$$

p_a must be very small compared to N_a and p_v must be very small compared to n_i which is, in turn, small compared to $N_\mathrm{d} - N_\mathrm{a}$. That is, $p_\mathrm{a} \ll N_\mathrm{a}$ and $p_\mathrm{v} \ll N_\mathrm{d} - N_\mathrm{a}$. Equation (7.36) then gives

$$n_\mathrm{c} + n_\mathrm{d} \simeq N_\mathrm{d} - N_\mathrm{a}. \tag{7.46}$$

But $n_\mathrm{d}(T) = \frac{N_\mathrm{d}}{\frac{1}{2}\mathrm{e}^{\beta(\varepsilon_\mathrm{d}-\zeta)}+1}$. If $\beta(\varepsilon_\mathrm{d} - \zeta) \gg 1$, then $n_\mathrm{d} \ll N_\mathrm{d}$, and we find

$$n_\mathrm{c} \simeq N_\mathrm{d} - N_\mathrm{a}, \tag{7.47}$$

$$p_\mathrm{v} \simeq \frac{n_\mathrm{i}^2}{N_\mathrm{d} - N_\mathrm{a}} \approx 0, \tag{7.48}$$

$$p_\mathrm{a} \simeq 2\ N_\mathrm{a} \mathrm{e}^{-\beta(\zeta-\varepsilon_\mathrm{a})} \approx 0, \tag{7.49}$$

$$n_\mathrm{d} \simeq 2\ N_\mathrm{d} \mathrm{e}^{-\beta(\varepsilon_\mathrm{d}-\zeta)} \approx 0. \tag{7.50}$$

7.5 p–n Junction

The p–n junction is of fundamental importance in understanding semiconductor devices, so we will spend a little time discussing the physics of p–n junctions. We consider a material with donor concentration $N_\mathrm{d}(z)$ and acceptor concentration $N_\mathrm{a}(z)$ given by

$$N_\mathrm{d}(z) = N_\mathrm{d}\Theta(z) \ \text{ and } \ N_\mathrm{a}(z) = N_\mathrm{a}\left[1 - \Theta(z)\right]. \tag{7.51}$$

We know that for $z \gg a$, where a is the atomic spacing the chemical potential must lie close to the donor levels and for $z \ll -a$ it must lie close to the acceptor levels. Since the chemical potential must be constant (independent of z) for the equilibrium case, we expect a picture like that sketched in Fig. 7.4. On the left we have a normal p-type material, and at low temperature, the chemical potential must sit very close to the acceptor levels which are shown by the dots at the chemical potential ζ. One the right, the chemical potential must be close to the donor levels (shown as dots at $\varepsilon = \zeta$) which are near the conduction band edge. In between, there must be a region in which there is a built-in potential $\phi(z)$ that results from the transfer of electrons from donors on the right to acceptors on the left in a region close to $z = 0$. We want to calculate this potential $\phi(z)$.

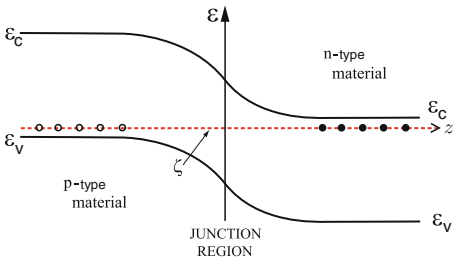


Fig. 7.4. Impurity levels and chemical potential across the p–n junction

7.5.1 Semiclassical Model

The effective Hamiltonian describing the conduction or valence band of a system containing a p-n junction can be written

$$H = \varepsilon\left(-\mathrm{i}\hbar\nabla\right) - e\phi(z), \tag{7.52}$$

where $\phi(z)$ is an electrostatic potential that must be slowly varying on the atomic scale in order for the semiclassical approximation to be valid. The energies of the conduction and valence band edges will be given by

$$\begin{aligned}\varepsilon_{\mathrm{c}}(z) &= \varepsilon_{\mathrm{c}} - e\phi(z),\\ \varepsilon_{\mathrm{v}}(z) &= \varepsilon_{\mathrm{v}} - e\phi(z).\end{aligned} \tag{7.53}$$

The concentration of electrons and holes will vary with position z as

$$\begin{aligned}n_{\mathrm{c}}(z) &= N_{\mathrm{c}}(T)\mathrm{e}^{-\beta[\varepsilon_{\mathrm{c}}-e\phi(z)-\zeta]},\\ p_{\mathrm{v}}(z) &= P_{\mathrm{v}}(T)\mathrm{e}^{-\beta[\zeta-\varepsilon_{\mathrm{v}}+e\phi(z)]}.\end{aligned} \tag{7.54}$$

The most important case to study is the high concentration limit where $N_{\mathrm{d}} \gg n_{\mathrm{i}}$ and $N_{\mathrm{a}} \gg n_{\mathrm{i}}$ on the right and left sides of the junction, respectively. In that case, the concentration of electrons and holes will vary with position z as

$$\begin{aligned}\lim_{z\to\infty} n_{\mathrm{c}}(z) &= N_{\mathrm{c}}(T)\mathrm{e}^{-\beta[\varepsilon_{\mathrm{c}}-e\phi(\infty)-\zeta]} \approx N_{\mathrm{d}},\\ \lim_{z\to-\infty} p_{\mathrm{v}}(z) &= P_{\mathrm{v}}(T)\mathrm{e}^{-\beta[\zeta-\varepsilon_{\mathrm{v}}+e\phi(-\infty)]} \approx N_{\mathrm{a}}.\end{aligned} \tag{7.55}$$

These two equations can be combined to give

$$e\Delta\phi = e\left[\phi(\infty) - \phi(-\infty)\right] = E_{\mathrm{Gap}} + \Theta\ln\left[\frac{N_{\mathrm{d}}N_{\mathrm{a}}}{N_{\mathrm{c}}(T)P_{\mathrm{v}}(T)}\right]. \tag{7.56}$$

The potential $\phi(z)$ must satisfy Poisson's equation given by

$$\frac{\partial^2\phi(z)}{\partial z^2} = -\frac{4\pi\rho(z)}{\epsilon_s}, \tag{7.57}$$

where the charge density $\rho(z)$ is given by

$$\rho(z) = e\left[N_d(z) - N_a(z) - n_c(z) + p_v(z)\right]. \tag{7.58}$$

In using (7.55), we are assuming that all donors and acceptors are ionized [since $n_c(\infty) = N_d$, all the donor electrons are in the conduction band so the donors must be positively charged]. Thus we have

$$N_d(z) = N_d\Theta(z), \tag{7.59}$$

$$N_a(z) = N_a\left[1 - \Theta(z)\right], \tag{7.60}$$

$$n_c(z) = N_d e^{-\beta[\phi(\infty)-\phi(z)]}, \tag{7.61}$$

$$p_v(z) = N_a e^{-\beta[\phi(z)-\phi(-\infty)]}. \tag{7.62}$$

Equations (7.57)–(7.62) form a complicated set of nonlinear equations. The solution is simple if we assume that the change in $\phi(z)$ occurs entirely over a relatively small region near the junction known as the *depletion region.*

We will assume that

$$\begin{aligned}
&\phi(z) = \phi(-\infty) \text{ for } z < -d_p; \text{ region I}\\
&\phi(z) = \phi(\infty) \text{ for } z > d_n; \text{ region II}\\
&\phi(z) \text{ varies with } z \text{ for } -d_p < z < d_n; \text{ region III}
\end{aligned} \tag{7.63}$$

The length d_p (or d_n) is called the *depletion length* of the p-type (or n-type) region. In region II the concentration of electron in the conduction band n_c is equal to the number of ionized donors N_d so that $\rho_{II}(z) = -en_c + eN_d = 0$. In region I, the concentration of holes p_v is equal to the number of ionized acceptors so that $\rho_I(z) = ep_v - eN_a = 0$. In region III, there are no electrons or holes (the built-in junction potential sweeps them out) so $p_v(z) = n_c(z) = 0$ in this region. Therefore, for $\rho(z)$ we have

$$\rho_{III}(z) = \begin{cases} +eN_d \text{ for } 0 < z < d_n, \\ -eN_a \text{ for } -d_p < z < 0. \end{cases} \tag{7.64}$$

We can integrate Poisson's equation. In the region $0 < z < d_n$, we have

$$\frac{\partial^2\phi(z)}{\partial z^2} = -\frac{4\pi e}{\epsilon_s}N_d, \tag{7.65}$$

and integration gives

$$\frac{\partial\phi(z)}{\partial z} = -\frac{4\pi e}{\epsilon_s}N_d z + C_1. \tag{7.66}$$

Here, C_1 is a constant of integration. Integrating (7.66) gives

$$\phi(z) = -\frac{2\pi e N_\mathrm{d}}{\epsilon_\mathrm{s}} z^2 + C_1 z + C_2. \tag{7.67}$$

We choose the constants so that $\phi(z)$ evaluated at $z = d_\mathrm{n}$ has the value $\phi(\infty)$ and $\frac{\partial\phi(z)}{\partial z} = 0$ at $z = d_n$. This gives

$$\phi(z) = \phi(\infty) - \frac{2\pi e N_\mathrm{d}}{\epsilon_\mathrm{s}} (z - d_\mathrm{n})^2, \text{ for } 0 < z < d_\mathrm{n}. \tag{7.68}$$

Doing exactly the same thing in the region $-d_\mathrm{p} < z < 0$ gives

$$\phi(z) = \phi(-\infty) + \frac{2\pi e N_\mathrm{a}}{\epsilon_\mathrm{s}} (z + d_\mathrm{p})^2, \text{ for } -d_\mathrm{p} < z < 0. \tag{7.69}$$

Of course, for $z > d_\mathrm{n}$, $\phi(z) = \phi(\infty)$ and for $z < -d_\mathrm{p}$, $\phi(z) = \phi(-\infty)$ (see Fig. 7.5).
Charge conservation requires that

$$N_\mathrm{d} d_\mathrm{n} = N_\mathrm{a} d_\mathrm{p}. \tag{7.70}$$

This condition insures the continuity of $\frac{\partial\phi}{\partial z}$ at $z = 0$. The continuity of $\phi(z)$ at $z = 0$ requires that

$$\phi(\infty) - \frac{2\pi e N_\mathrm{d}}{\epsilon_\mathrm{s}} d_\mathrm{n}^2 = \phi(-\infty) + \frac{2\pi e N_\mathrm{a}}{\epsilon_\mathrm{s}} d_\mathrm{p}^2. \tag{7.71}$$

We can solve (7.71) for $\Delta\phi \equiv \phi(\infty) - \phi(-\infty)$ to obtain

$$\Delta\phi = \frac{2\pi e}{\epsilon_\mathrm{s}} \left[N_\mathrm{d} d_\mathrm{n}^2 + N_\mathrm{a} d_\mathrm{p}^2 \right]. \tag{7.72}$$

Combining (7.70) and (7.72) allows us to determine d_n and d_p

$$d_\mathrm{n} = \left[\frac{(N_\mathrm{a}/N_\mathrm{d})\, \epsilon_\mathrm{s} \Delta\phi}{2\pi e (N_\mathrm{a} + N_\mathrm{d})} \right]^{1/2}. \tag{7.73}$$

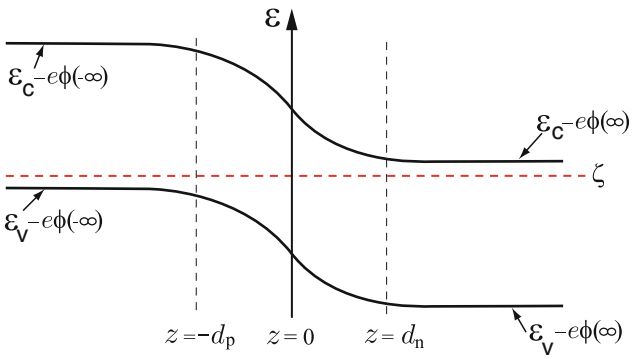


Fig. 7.5. Band bending across the p–n junction

The equation for $d_{\rm p}$ is obtained by interchanging $N_{\rm a}$ and $N_{\rm d}$. If $N_{\rm a}$ were equal to $N_{\rm d}$ then $d_{\rm n} = d_{\rm p} = d$ and is given by

$$d \simeq \left(\frac{\epsilon_{\rm s} e \Delta\phi}{4\pi e^2 N}\right)^{1/2} \approx \left(\frac{\epsilon_{\rm s} E_{\rm Gap}}{4\pi e^2 N}\right)^{1/2}, \tag{7.74}$$

where $N = N_{\rm d} = N_{\rm a}$. In the last result, we have simply put $e\Delta\phi \approx E_{\rm Gap}$.

7.5.2 Rectification of a p–n Junction

The region of the p–n junction is a high resistance region because the carrier concentration in the region $(-d_{\rm p} < z < d_{\rm n})$ is depleted. When a voltage V is applied, almost all of the voltage drop occurs across the high resistance junction region. We write $\Delta\phi$ in the presence of an applied voltage V as

$$\Delta\phi = (\Delta\phi)_0 - V. \tag{7.75}$$

Here, $(\Delta\phi)_0$ is, of course, the value of $\Delta\phi$ when $V = 0$. The sign of V is taken as positive (*forward bias*) when V decreases the voltage drop across the junction. The depletion layer width $d_{\rm n}$ changes with voltage

$$d_{\rm n}(V) = d_{\rm n}(0)\left[1 - \frac{V}{(\Delta\phi)_0}\right]^{1/2}. \tag{7.76}$$

A similar equation holds for $d_{\rm p}(V)$. When $V = 0$, there is no hole current $J_{\rm h}$ and no electron current $J_{\rm e}$. When V is finite both $J_{\rm e}$ and $J_{\rm h}$ are nonzero. Let us look at $J_{\rm h}$. It has two components:

Generation current This current results from the small concentration of holes on the n-side of the junction that are created to be in thermal equilibrium, i.e., to have ζ remain constant. These holes are immediately swept into the p-side of the junction by the electric field of the junction. This generation current is rather insensitive to applied voltage V, since the built-in potential $(\Delta\phi)_0$ is sufficient to sweep away all the carriers that are thermally generated.

Recombination current This current results from the diffusion of holes from the p-side to the n-side. On the p-side there is a very high concentration of holes. In order to make it cross the depletion layer (and recombine with an electron on the n-side), a hole must overcome the junction potential barrier $-e\left[(\Delta\phi)_0 - V\right]$. This recombination current does depend on V as

$$J_{\rm h}^{\rm rec} \propto {\rm e}^{-e[(\Delta\phi)_0 - V]/\Theta}. \tag{7.77}$$

Here, $J_{\rm h}^{\rm rec}$ indicates the number current density of holes from the p- to n-side.

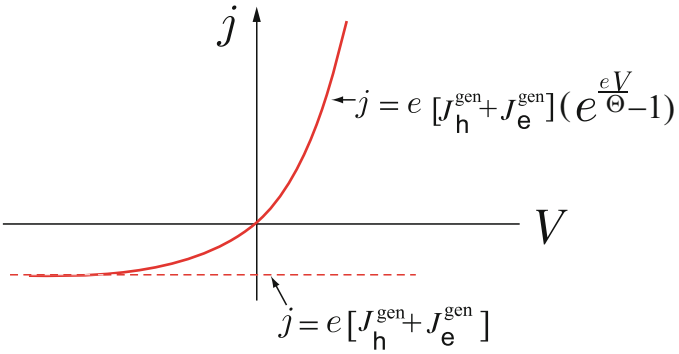


Fig. 7.6. Current–voltage characteristic across the p–n junction

Now, at $V = 0$ these two currents must cancel to give $J_{\mathrm{h}} = J_{\mathrm{h}}^{\mathrm{rec}} - J_{\mathrm{h}}^{\mathrm{gen}} = 0$ We can write

$$J_{\mathrm{h}} = J_{\mathrm{h}}^{\mathrm{gen}} \left[\mathrm{e}^{eV/\Theta} - 1\right]. \tag{7.78}$$

The electrical current density due to holes is $j_{\mathrm{h}} = eJ_{\mathrm{h}}$, and it vanishes at $V = 0$ and has the correct V dependence for $J_{\mathrm{h}}^{\mathrm{rec}}$. If we do the same for electrons, we obtain the current density $J_{\mathrm{e}} = J_{\mathrm{e}}^{\mathrm{gen}} \left[\mathrm{e}^{eV/\Theta} - 1\right]$, which flows oppositely to the J_{h}. The electrical current density of electrons j_{e} is parallel to the j_{h}. Therefore, the combined electrical current density becomes as follows:

$$j = e\,(J_{\mathrm{h}}^{\mathrm{gen}} + J_{\mathrm{e}}^{\mathrm{gen}}) \left(\mathrm{e}^{eV/\Theta} - 1\right). \tag{7.79}$$

A plot of j versus V looks as shown in Fig. 7.6. The applied-voltage behavior of an electrical current across the p–n junction is called *rectification* because a circuit can easily be arranged in which no current flows when V is negative (smaller than some value) but a substantial current flows for positive applied voltage.

7.5.3 Tunnel Diode

In the late 1950s, Leo Esaki[1] was studying the current voltage characteristics of very heavily doped p–n junctions. He found and explained the $j-V$ characteristic shown in Fig. 7.7. Esaki noted that, for very heavily doped materials, impurity band was formed and one would obtain degenerate n-type and p-type regions where the chemical potential ζ was actually in the conduction band on the n-side and in the valence band on the p-side as shown in Fig. 7.8. For a forward bias the electrons on the n-side can tunnel through the energy gap into the empty states (holes) in the valence band. This current occurs only for $V > 0$, and it cuts off when the voltage V exceeds the value at

[1] L. Esaki, Phys. Rev. **109**, 603–604 (1958).

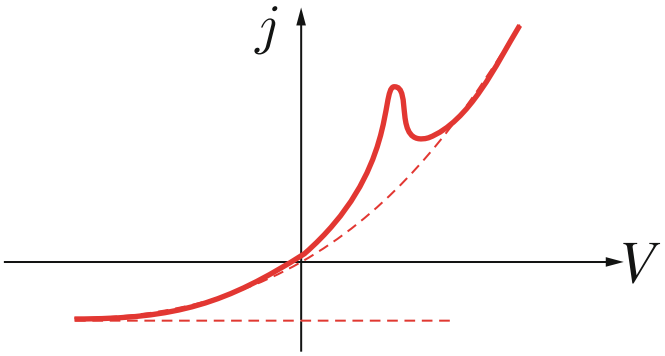


Fig. 7.7. Current–voltage characteristic across a heavily doped p–n junction

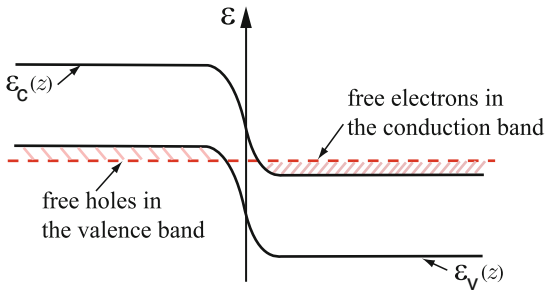


Fig. 7.8. Chemical potential across the heavily doped p–n junction

which $\varepsilon_c(\infty) = \varepsilon_v(-\infty)$. When the tunnel current is added to the normal p–n junction current, the negative resistance region shown in Fig. 7.7 occurs.

7.6 Surface Space Charge Layers

The metal–oxide–semiconductor (MOS) structure is the basis for all of current microelectronics. We will consider the *surface space charge layers* that can occur in an MOS structure. Assume a semiconductor surface is produced with a uniform and thin insulating layer (usually on oxide), and then on top of this oxide a metallic gate electrode is deposited as is shown in Fig. 7.9.

In the absence of any applied voltage, the bands line up as shown in Fig. 7.10. If a voltage is applied which lowers the Fermi level in the metal relative to that in the semiconductor, most of the voltage drop will occur across the insulator and the depletion layer of the semiconductor.

For a relatively small applied voltage, we obtain a band alignment as shown in Fig. 7.11. In the depletion layer all of the acceptors are ionized and the hole concentration is zero since the field in the depletion layer sweeps the holes into

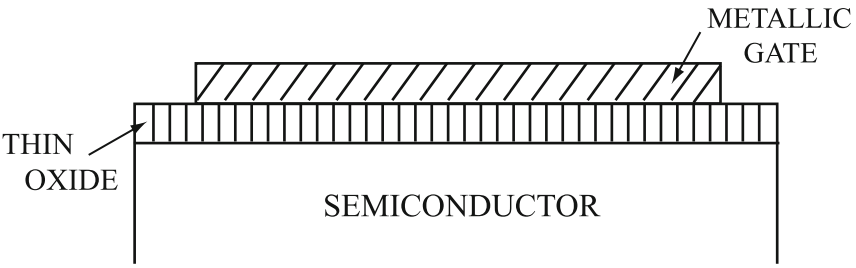


Fig. 7.9. Metal–oxide–semiconductor structure

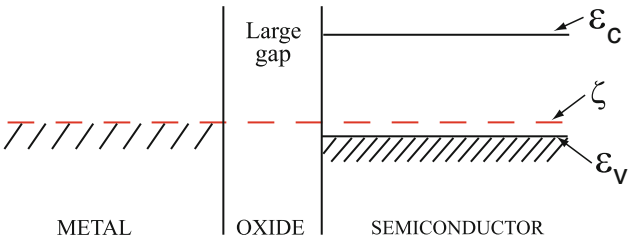


Fig. 7.10. Band edge alignment across an MOS structure

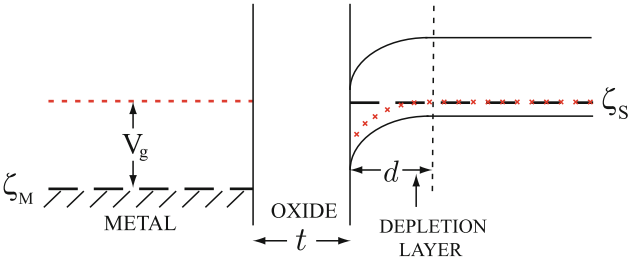


Fig. 7.11. Band alignment across an MOS structure in the presence of a small applied voltage

the bulk of the semiconductor. The normal component of the displacement field $D = \epsilon E$ must be continuous at the semiconductor–oxide interface, and the sum of the voltage drop V_d across the depletion layer and V_{ox} across the oxide must equal the applied voltage V_g. If we take the electrostatic potential to be $\phi(z)$, then

$$\phi(z) = \phi(\infty) \text{ for } z > d,$$
$$\phi(z) = \phi(\infty) + \frac{2\pi e N_a}{\epsilon_s}(z-d)^2 \text{ for } 0 < z < d. \tag{7.80}$$

The potential energy V is $-e\phi$. $V_{\rm ox}$ is simply $E_{\rm ox}t$, where $E_{\rm ox}$ and t are the electric field in the oxide and the thickness of the oxide layer, respectively. Equating $\epsilon_0 E_{\rm ox}$ to $-\varepsilon_{\rm s}\phi'(z=0)$ gives

$$e\frac{V_{\rm ox}}{t}\epsilon_0 = 4\pi e^2 N_{\rm a} d. \tag{7.81}$$

Equation (7.81) gives us d in terms of $V_{\rm ox}$. Adding $V_{\rm ox}$ to the voltage drop $\frac{2\pi e N_{\rm a}}{\epsilon_{\rm s}}d^2$ across the depletion layer gives

$$V_{\rm ox} = V_{\rm g} - \frac{2\pi e N_{\rm a}}{\epsilon_{\rm s}}d^2. \tag{7.82}$$

Note that d must grow as $V_{\rm g}$ increases since the voltage drop is divided between the oxide and the depletion layer. The only way that $V_{\rm d}$ can grow, since $N_{\rm a}$ is fixed, is by having d grow. The surface layer just discussed is called a *surface depletion layer* since the density of holes in the layer is depleted from its bulk value. For a gate voltage in the opposite direction the bands look as shown in Fig. 7.12. Here, the surface layer will have an excess of holes either bound to the acceptors or in the valence band. This is called an *accumulation layer* since the density of holes is increased at the surface. If the gate voltage Vg is increased to a large value in the direction of depletion, one can wind up with the conduction band edge at the interface below $\zeta_{\rm s}$, the chemical potential of the semiconductor. This is shown in Fig. 7.13.

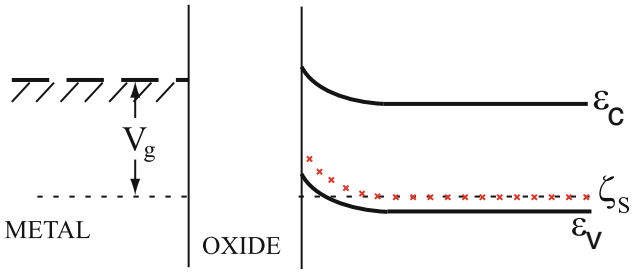


Fig. 7.12. Band alignment across an MOS structure in the presence of a small negative gate voltage

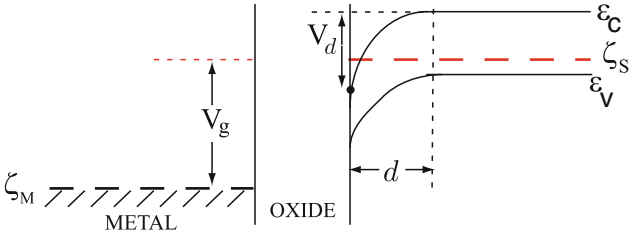


Fig. 7.13. Band alignment across an MOS structure in the presence of a large applied voltage

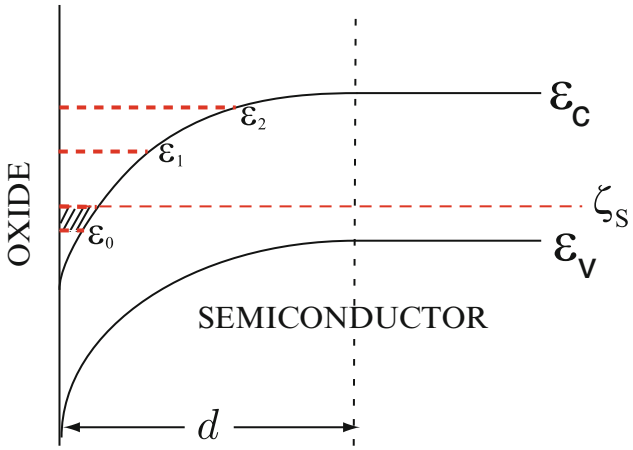


Fig. 7.14. Band edge near the interface of the semiconductor–insulator in the MOS structure in the presence of a large applied voltage

Now, there can be electrons in the conduction band because ζ_s is higher than $\varepsilon_c(z)$ evaluated at the semiconductor–oxide interface. The part of the diagram near this interface is enlarged in Fig. 7.14. This system is called a *semiconductor surface inversion layer* because in this surface layer we have trapped electrons (minority carriers in the bulk). The motion of the electrons in the direction normal to the interface is quantized, so there are discrete energy levels ε_0, ε_1, ... forming *subband structure*. If only ε_0 lies below the chemical potential and $\varepsilon_1 - \zeta \gg 0$, the electronic system behaves like a *two-dimensional electron gas* (2DEG)because

$$\varepsilon = \varepsilon_0 + \frac{\hbar^2}{2m_c^*}\left(k_x^2 + k_y^2\right), \tag{7.83}$$

and

$$\Psi_{n,k_x,k_y} = \frac{1}{L}\mathrm{e}^{\mathrm{i}(k_x x + k_y y)}\xi_n(z). \tag{7.84}$$

Here, $\xi_n(z)$ is the nth eigenfunction of a differential equation given by

$$\left[\frac{1}{2m_c^*}\left(-\mathrm{i}\hbar\frac{\partial}{\partial z}\right)^2 + V_{\mathrm{eff}}(z) - \varepsilon_n\right]\xi_n(z) = 0. \tag{7.85}$$

In (7.85) the effective potential $V_{\mathrm{eff}}(z)$ must contain contributions from the depletion layer charge, the *Hartree potential* of the electrons trapped in the inversion layer, an image potential if the dielectric constants of the oxide and semiconductor are different, and an *exchange–correlation potential* of the electrons with one another beyond the simple Hartree term. Because the electrons are completely free to move in the $x-y$ plane, but "frozen" into a single

quantized level ε_0 in the z-direction, the z-degree of freedom is frozen out of the problem, and in this sense the electrons behave as a two-dimensional electron gas. We fill up a circle in $k_x - k_y$ space up to k_F, and

$$2 \sum_{\substack{k_x, k_y \\ \varepsilon < \varepsilon_\mathrm{F}}} 1 = N, \tag{7.86}$$

giving $2\left(\frac{L}{2\pi}\right)^2 \pi k_\mathrm{F}^2 = N$. This means that $k_\mathrm{F}^2 = 2\pi n_\mathrm{s}$, where $n_\mathrm{s} \equiv N/L^2$ is the number of electrons per unit area of the inversion layer. Of course $\varepsilon_\mathrm{F} \equiv \frac{\hbar^2 k_\mathrm{F}^2}{2m_\mathrm{c}^*} = \zeta - \varepsilon_0$.

The potential due to the depletion charge is calculated exactly as before. The Hartree potential is a solution of Poisson's equation given below

$$\frac{\partial^2}{\partial z^2} V_\mathrm{H} = -\frac{4\pi e^2}{\epsilon_\mathrm{s}} \rho_e(z). \tag{7.87}$$

The electron density is given by

$$\rho_e(z) = \sum_{n,k} f_0(\varepsilon_{nk}) \left|\Psi_{nk}(z)\right|^2, \tag{7.88}$$

where $\Psi_{nk}(z) = L^{-1}\xi_n(z)\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}$ is the *envelope wave function* for the electrons in the effective potential. The exchange–correlation potential V_xc is a functional of the electron density $\rho_e(z)$. This surface inversion layer system is the basis of all large scale integrated circuit chips that we use every day. The basic unit is the MOS field effect transistor (MOSFET) shown in Fig. 7.15. The source–drain conductivity can be controlled by varying the applied gate voltage V_g. This allows one to make all kinds of electronic devices like oscillators, transistors *etc.* This was an extremely active field of semiconductor

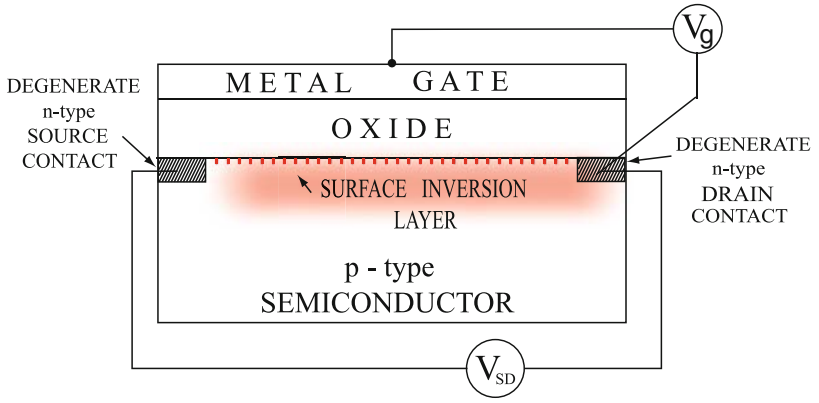


Fig. 7.15. Schematic diagram of the metal–oxide–semiconductor field effect transistor

physics from the late 60's till the present time. Some basic problems that were investigated include:

- *Transport along the layer*
  surface electron density n_s and relaxation time τ as a function of the gate voltage V_g, cyclotron resonance, localization, magnetoconductivity, and Hall effect.
- *Transport perpendicular to the layer*
  optical absorption, Raman scattering, coupling to optical phonons, intra and intersubband collective modes.
- *Many-body effects* on subband structure and on effective mass and effective g-value.

7.6.1 Superlattices

By novel growth techniques like *molecular beam epitaxy* (MBE) novel structures can be grown almost one atomic layer at a time. The requirements for such growth are

1. The lattice constants of the two materials must be rather close. Otherwise, large strains lead to many crystal imperfections.
2. The materials must form appropriate bonds with one another.

One very popular example is the GaAs–AlAs system shown in Fig. 7.16. A single layer of GaAs in an AlAs host would be called a *quantum well*. A periodic array of such layers is called a *superlattice*. It can be thought of as a new material with a *supercell* in real space that goes from one GaAs to AlAs interface to the next GaAs to AlAs interface.

7.6.2 Quantum Wells

If a quantum well is narrow, it will lead to quantized motion and subbands just as the MOS surface inversion layer did (see, for example, Fig. 7.17). For

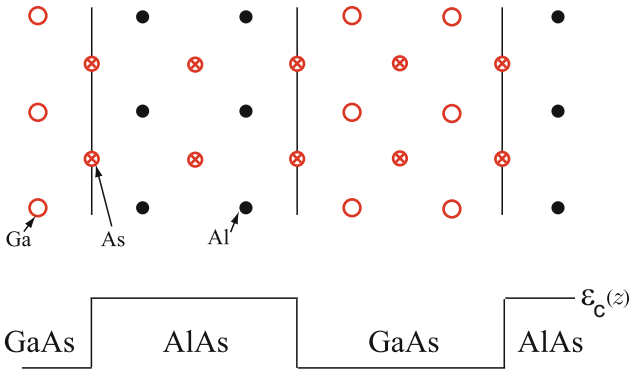


Fig. 7.16. Schematic diagram of the GaAs–AlAs superlattice system

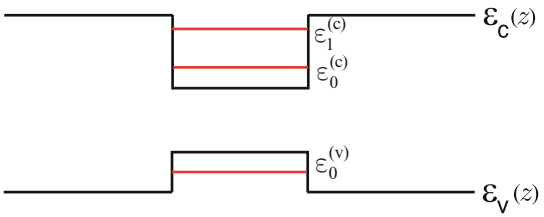


Fig. 7.17. Schematic diagram of the subbands formation in a quantum well of the GaAs–AlAs structure

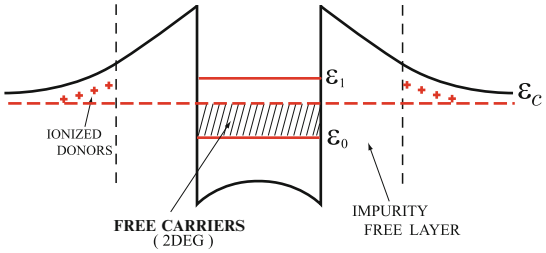


Fig. 7.18. Schematic diagram of a quantum well in a modulation doped $GaAs/Ga_{1-x}Al_xAs$ quantum well

the subbands in the conduction band we have

$$\varepsilon_n^{(c)}(\mathbf{k}) = \varepsilon_n^{(c)} + \frac{\hbar^2}{2m_c^*}\left(k_x^2 + k_y^2\right). \tag{7.89}$$

The band offsets are difficult to predict theoretically, but they can be measured.

7.6.3 Modulation Doping

The highest mobility materials have been obtained by growing modulation doped $GaAs/Ga_{1-x}Al_xAs$ quantum wells. In these materials, the donors are located in the GaAlAs barriers, but no closer than several hundred Angstroms to the quantum well. The bands look as shown in Fig. 7.18. A typical sample structure would look like GaAlAs with N_d donors/cm^3 $\|$ pure GaAlAs of 20 nm thick // GaAs of 10 nm thick // pure GaAlAs of 20 nm thick // GaAlAs with N_d donors/cm^3 (see, e.g., Fig. 7.19). Because the ionized impurities are rather far away from the quantum well electrons, ionized impurity scattering is minimized and very high mobilities can be attained.

7.6.4 Minibands

When the periodic array of quantum wells in a superlattice has very wide barriers, the subband levels in each quantum well are essentially unchanged

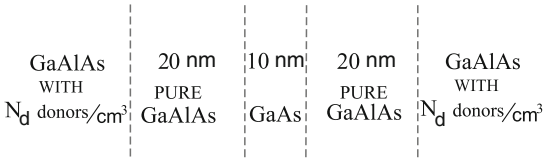


Fig. 7.19. Schematic layered structure of a typical modulation doped GaAs/$Ga_{1-x}Al_xAs$ quantum well system

(see, for example, Fig. 7.20). However, a new periodicity has been introduced, so we have a quantum number k_z that has to do with the eigenvalues of the translation operator.

$$T_a \Psi_{nk}(z) = \mathrm{e}^{\mathrm{i}k_z a} \Psi_{nk}(z) \tag{7.90}$$

This looks just like the problem of atomic energy levels that give rise to band structure when the atoms are brought together to form a crystal. For very large values of the barrier width, no tunneling occurs, and the *minibands* are essentially flat as is shown in Fig. 7.21. The supercell in real space extends from $z = 0$ to $z = a$. The first Brillouin zone in k-space extends $-\frac{\pi}{a} \leq k_z \leq \frac{\pi}{a}$. The minibands $\varepsilon_n(k_z)$ are flat if the barriers are so wide that no tunneling from one quantum well to its neighbor is possible. When the barriers are narrower and tunneling can take place, the flat bands become k_z-dependent. One can easily show that in tight binding calculation one would get bands with sinusoidal shape as shown in Fig. 7.22 below. Of course, the same band structure would result from taking free electrons moving in a periodic potential

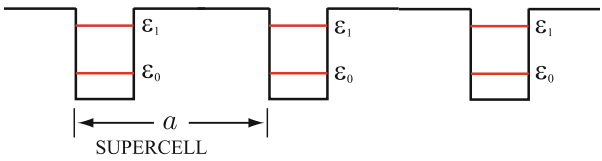


Fig. 7.20. Schematic subband alignments in a superlattice of supercell width a

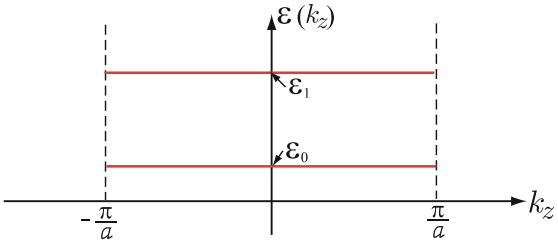


Fig. 7.21. Schematic miniband alignment in a superlattice of very large barrier width

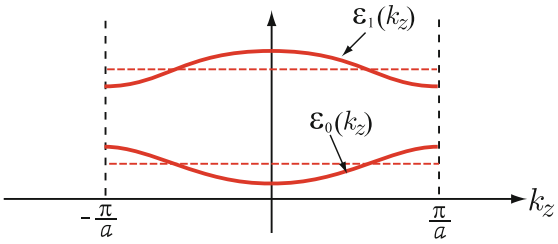


Fig. 7.22. Miniband structure in a superlattice of very narrow barrier width

$$V(r) = \sum_n V_n \, \mathrm{e}^{\mathrm{i}\frac{2\pi}{a}nz}. \tag{7.91}$$

During the past 20 years, there has been an enormous explosion in the study (both experimental and theoretical) of optical and transport properties of quantum wells, superlattices, quantum wires, and quantum dots. One of the most exciting developments was the observation by Klaus von Klitzing of the quantum Hall effect in a 2DEG in a strong magnetic field. Before we give a very brief description of this work, we must discuss the eigenstates of free electrons in two dimensions in the presence of a perpendicular magnetic field.

7.7 Electrons in a Magnetic Field

Consider a 2DEG with n_s electrons per unit area. In the presence of a dc magnetic field $\mathbf{B}$ applied normal to the plane of the 2DEG, the Hamiltonian of a single electron is written by

$$H = \frac{1}{2m}\left(\mathbf{p} + \frac{e}{c}\mathbf{A}\right)^2. \tag{7.92}$$

Here, $\mathbf{p} = (p_x, p_y)$ and $\mathbf{A}(\mathbf{r})$ is the vector potential whose curl gives $\mathbf{B} = (0, 0, B)$, i.e., $\mathbf{B} = \nabla \times \mathbf{A}$. There are a number of different possible choices for $\mathbf{A}(\mathbf{r})$ (different gauges) that give a constant magnetic field in the z-direction. For example, the *Landau gauge* chooses $\mathbf{A} = (0, Bx, 0)$ giving us $\left(\hat{x}\frac{\partial}{\partial x}\right) \times (\hat{y}Bx) = B\hat{z}$. Another common choice is $\mathbf{A} = \frac{B}{2}(-y, x, 0)$; this is called the *symmetric gauge*. Different gauges have different eigenstates, but the observables have to be the same.

Let us look at the Schrödinger equation in the Landau gauge. $(H - E)$ $\Psi = 0$ can be rewritten as

$$\left[\frac{p_x^2}{2m} + \frac{1}{2m}\left(p_y + \frac{e}{c}Bx\right)^2 - E\right]\Psi(\mathbf{r}) = 0. \tag{7.93}$$

Because H is independent of the coordinate y, we can write

$$\Psi(x, y, z) = \mathrm{e}^{\mathrm{i}ky}\varphi(x). \tag{7.94}$$

Substituting this into the Schrödinger equation gives

$$\left[\frac{p_x^2}{2m} + \frac{1}{2}m\omega_\mathrm{c}^2\left(x + \frac{\hbar k}{m\omega_\mathrm{c}}\right)^2 - E\right]\varphi(x) = 0. \tag{7.95}$$

Here, of course, $\omega_\mathrm{c} = \frac{eB}{mc}$ is the *cyclotron frequency*. If we define $\tilde{x} = x + \frac{\hbar k}{m\omega_\mathrm{c}}$, $\frac{\partial}{\partial x} = \frac{\partial}{\partial \tilde{x}}$ and the Schrödinger equation is just the simple harmonic oscillator equation. Its solutions are as follows:

$$\begin{aligned} E_{nk} &= \hbar\omega_\mathrm{c}\left(n + \frac{1}{2}\right), \ n = 0, 1, 2, \ldots \\ \Psi_{nk}(x, y, z) &= \mathrm{e}^{\mathrm{i}ky}\ u_n\left(x + \frac{\hbar k}{m\omega_\mathrm{c}}\right). \end{aligned} \tag{7.96}$$

The energy is independent of k, so the density of states (per unit length) is a series of δ-functions, as is shown in Fig. 7.23.

$$g(\varepsilon) \propto \sum_n \delta\left(\varepsilon - \hbar\omega_\mathrm{c}\left(n + \frac{1}{2}\right)\right). \tag{7.97}$$

The constant of proportionality for a finite size sample of area L^2 is $\frac{m\omega_\mathrm{c}L^2}{2\pi\hbar}$, so that the total number of states per Landau level is

$$N_\mathrm{L} = L\left(\frac{m\omega_\mathrm{c}L}{2\pi\hbar}\right) = \frac{BL^2}{hc/e}. \tag{7.98}$$

For a sample of area L^2, each Landau level can accommodate N_L electrons (we have omitted spin) and N_L is the *magnetic flux* through the sample divided

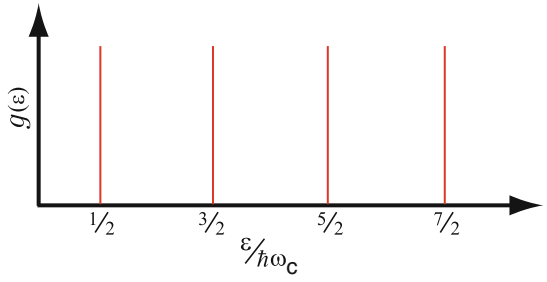


Fig. 7.23. Density of states for electrons in a dc magnetic field

by the *quantum of magnetic flux* $\frac{hc}{e}$. We note that the degeneracy of each Landau level can also be rewritten by $N_{\mathrm{L}} = \frac{L^2}{2\pi\lambda^2}$ in terms of the *magnetic length* $\lambda = \sqrt{\frac{\hbar c}{eB}}$.

7.7.1 Quantum Hall Effect

If we make contacts to the 2DEG, and send a current I in the x-direction, then we expect

$$\sigma_{xx} \propto \frac{I}{V_x} \quad \text{and} \quad \sigma_{xy} \propto \frac{I}{V_y}. \tag{7.99}$$

Here, V_x is the applied voltage in the direction of I and V_y is the Hall voltage. In the simple classical (Drude model) picture, we know that

$$\sigma_{xx} = \sigma_{yy} = \frac{\sigma_0}{1 + (\omega_{\mathrm{c}}\tau)^2} \quad \text{and} \quad \sigma_{xy} = -\sigma_{yx} = -\frac{\omega_{\mathrm{c}}\tau\sigma_0}{1 + (\omega_{\mathrm{c}}\tau)^2}, \tag{7.100}$$

where $\sigma_0 = \frac{n_{\mathrm{s}}e^2\tau}{m}$. In the limit as $\tau \to \infty$ we have $\sigma_0 \to \infty$, $\sigma_{xx} \to 0$, and $\sigma_{xy} \to -\frac{n_{\mathrm{s}}ec}{B}$.

In the absence of scattering, $g(\varepsilon)$ is a series of δ-functions. With scattering, the δ-functions are broadened as shown in Fig. 7.24. We know that, when one Landau level is completely filled and the one above it is completely empty, there can be no current, because to modify the distribution function $f_0(\varepsilon)$ would require promotion of electrons to the next Landau level. There is a gap for doing this, and at $T = 0$ there will be no current. If we plot σ_{xx} versus $n_{\mathrm{s}}/n_{\mathrm{L}} \equiv \nu$, the *filling factor* $\left(n_{\mathrm{L}} = N_{\mathrm{L}}/L^2\right)$ we expect σ_{xx} to go to zero at any integer values of ν as shown in Fig. 7.25.

Our understanding of the *integral quantum Hall effect* is based on the idea that within the broadened δ-functions representing the density of states, we have both *extended states* and *localized states* as shown in Fig. 7.26. The quantum Hall effect was very important because it led to

1. A resistance standard $\rho = \frac{h}{e^2}\,\frac{1}{n}$.
2. Better understanding of *Anderson localization*.
3. Discovery of the *fractional quantum Hall effect*.

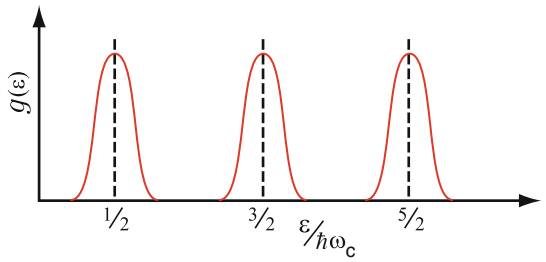


Fig. 7.24. Density of states for electrons in a dc magnetic field in the presence of scattering

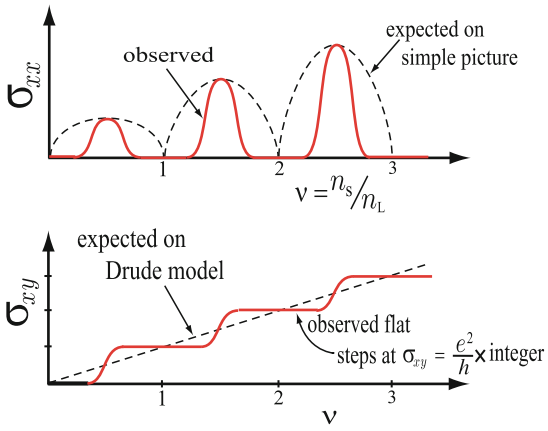


Fig. 7.25. Conductivities as a function of the Landau level filling factor. (**a**) Longitudinal conductivity σ_{xx}, (**b**) Hall conductivity σ_{xy}

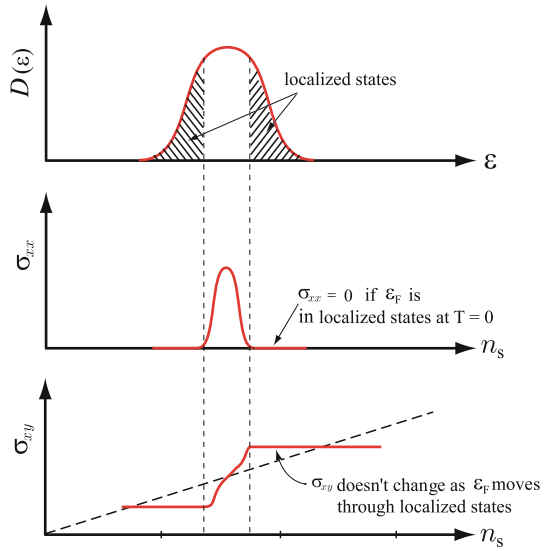


Fig. 7.26. Scattering effects on the density of states and conductivity components in an integral quantum Hall state

7.8 Amorphous Semiconductors

Except for introducing donors and acceptors in semiconductors, we have essentially restricted our consideration to ideal, defect-free infinite crystals. There are two important aspects of order that crystals display. The first is *short range order*. This has to do with the regular arrangement of atoms in the

vicinity of any particular atom. This short range order determines the local bonding and the crystalline fields acting on a given atom. The second aspect is *long range order*. This is responsible for the translational and rotational invariance that we used in discussing Bloch functions and band structure. It allowed us to use Bloch's theorem and to define the Bloch wave vector **k** within the first Brillouin zone.

In real crystals, there are always

- Surface effects associated with the finite size of the sample.
- Elementary excitations (dynamic perturbations like phonons, magnons, etc.)
- Imperfections and defects (static disorder).

For an *ordered solid*, one can start with the perfect crystal as the zeroth approximation and then treat static and dynamic perturbations by perturbation theory. For a *disordered solid* this type of approximation is not meaningful.

7.8.1 Types of Disorder

We can classify disorder by considering some simple examples in two dimensions that we can represent on a plane.
Perfect Crystalline Order Atoms in perfect crystalline array (see Fig. 7.27a).
Compositional Disorder Impurity atoms (e.g. in an alloy) are randomly distributed among crystalline lattice sites (See Fig. 7.27b.)
Positional Disorder Some separations and some bond angles are not perfect (See Fig. 7.27c.)
Topological Disorder Fig. 7.27d shows some topological disorder.
Because we cannot use translational invariance and energy band concepts, it is difficult to evaluate the eigenstates of a disordered system. What has been found is that in disordered systems, some of the electronic states can be extended states and some can be localized states. An extended state is one in which, if $|\Psi(0)|^2$ is finite, $|\Psi(r)|^2$ remains finite for r very large. A localized state is one in which $|\Psi(r)|^2$ falls off very quickly as r becomes large (usually exponentially). There is an enormous literature on disorder and localization (starting with a classic, but difficult, paper by P.W. Anderson[2] in the 1950s).

7.8.2 Anderson Model

The Anderson model described a system of atomic levels at different sites n and allowed for hopping from site n to m. The Hamiltonian is written by

$$H = \sum_n \varepsilon_n c_n^\dagger c_n + T \sum_{nm}{}' c_m^\dagger c_n \tag{7.101}$$

[2] P.W. Anderson, Phys. Rev. **109**, 1492 (1958).

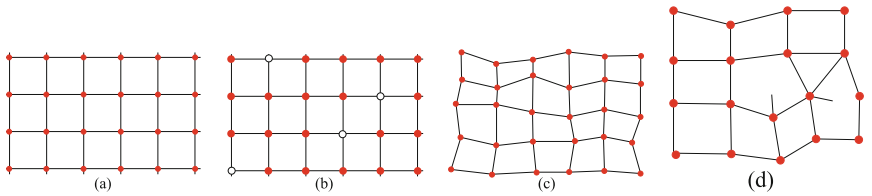


Fig. 7.27. Various types of disorder. (**a**) Atoms in perfect crystalline array. (**b**) Impurity atoms are randomly distributed among crystalline lattice sites. (**c**) Some separations and some bond angles are not perfect. (**d**) Not all fourfold rings, but some five and sixfold rings leaving dangling bonds represent topological disorder

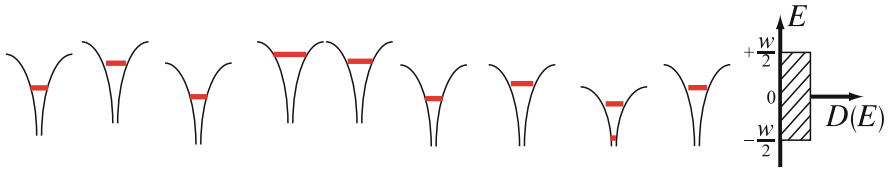


Fig. 7.28. Basic assumption of energy level distribution on different sites in the Anderson model

This is just the description of band structure in terms of an atomic level ε on site n where the periodic potential gives rise to the *hopping term*. In tight binding approximation, we would restrict T, the hopping term to nearest neighbor hops, and that is what the prime on $\sum'$ in the second-term means.

In Anderson model, it was assumed that ε_n the energy on site n was not a constant, but that it was randomly distributed over a range w (see, for example, Fig. 7.28). Anderson showed that if the parameter $\frac{w}{B}$, where B is the band width (caused by and proportional to T) satisfied $\frac{w}{B} \geq 5$, the state at $E = 0$ (the center of the band) is localized, while if $\frac{w}{B} \leq 5$ it is extended.

7.8.3 Impurity Bands

Impurity levels in semiconductors form Anderson-like systems. In these systems, the energy E is independent of n; it is equal to the donor energy ε_{d}. However, the hopping term T is randomly distributed between certain limits, since the impurities are randomly distributed. Sometimes (when two impurities are close together) it is easy to hop and T is large. Sometimes, when they are far apart, T is small.

7.8.4 Density of States

Although the eigenvalues of the Anderson Hamiltonian can not be calculated in a useful way, it is possible to make use of the idea of density of states. In

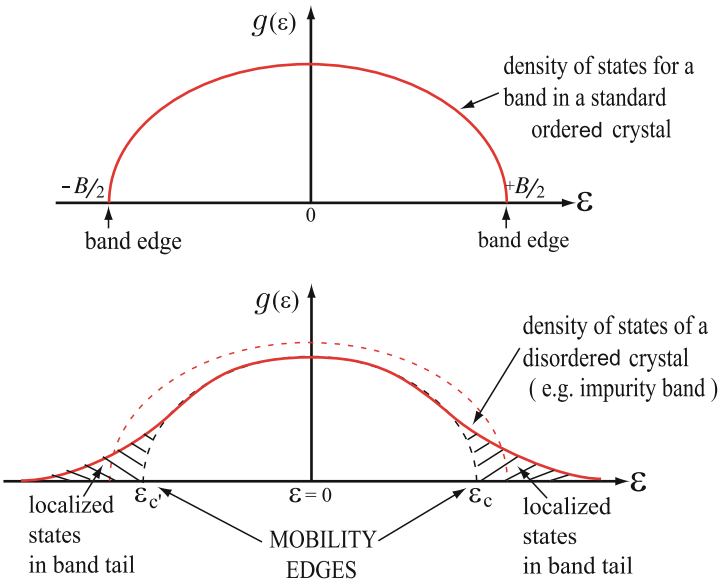


Fig. 7.29. Density of states of an ordinary crystal and that of a disordered material

Fig. 7.29, we sketch the density of states of an ordinary crystal and then the density of states of a disordered material. In the latter, the tails on the density of states usually contain localized states, while the states in the center of the band are extended. The energies $E_{c'}$ and E_c are called *mobility edges*. They separate localized and extended states. When E_F is in the localized states, there is no conduction at $T = 0$. The field of amorphous materials, Anderson localization, and mobility edges are of current research interest, but we do not have time to go into greater detail.

Problems

7.1. Intrinsic carrier concentration can be written as

$$n_i(T) = 2.5 \left(\frac{m_\mathrm{c}}{m}\frac{m_\mathrm{v}}{m}\right)^{3/4} \left(\frac{k_\mathrm{B}T}{E_\mathrm{Gap}}\right)^{3/2} \left(\frac{E_\mathrm{Gap}\text{ in eV}}{\frac{1}{40}\text{ eV}}\right)^{3/2} \mathrm{e}^{-\frac{E_\mathrm{Gap}}{2k_\mathrm{B}T}} \times 10^{19}/\mathrm{cm}^3.$$

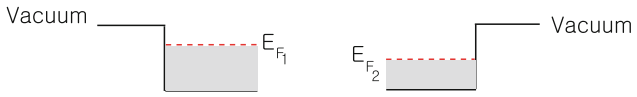
Take $E_\mathrm{Gap} = 1.5\mathrm{eV}$, $m_\mathrm{h} = 0.7m$, and $m_\mathrm{e} = 0.06m$ roughly those of GaAs, and plot $\ln n_i$ vs T in the range $T = 3\mathrm{K}$ to $300\mathrm{K}$.

7.2. Plot the chemical potential $\zeta(T)$ vs T in the range $T = 3 - 300\,\mathrm{K}$ for values of $E_\mathrm{Gap} = 1.5\,\mathrm{eV}$, $m_\mathrm{h} = 0.7\,\mathrm{m}$, and $m_\mathrm{e} = 0.06\,\mathrm{m}$.

7.3. For InSb, we have $E_\mathrm{Gap} \simeq 0.18\mathrm{eV}$, $\varepsilon_\mathrm{s} \simeq 17$, and $m_\mathrm{e} \simeq 0.014m$.

(a) Evaluate the binding energy of a donor.
(b) Evaluate the orbit radius of a conduction electron in the ground state.
(c) Evaluate the donor concentration at which overlap effects between neighboring impurities become significant.
(d) If $N_\mathrm{d} = 10^{14}\mathrm{cm}^{-3}$ in a sample of InSb, calculate n_c at $T = 4\mathrm{K}$.

7.4. Let us consider a case that the work function of two metals differ by 2 eV; $E_\mathrm{F1} - E_\mathrm{F2} = 2$ eV.



If the metals are brought into contact, electrons will flow from metal 1 to metal 2. Assume the transferred electrons are displaced by $3 \times 10^{-8}\,\mathrm{cm}$. How many electrons per square centimeter are transferred?

7.5. Consider a semiconductor quantum well consisting of a very thin layer of narrow gap semiconductor of $E_\mathrm{Gap} = \varepsilon_\mathrm{c} - \varepsilon_\mathrm{v}$ contained in a wide band gap host material of $E_\mathrm{Gap} = \varepsilon_\mathrm{c}^\mathrm{H} - \varepsilon_\mathrm{v}^\mathrm{H}$. The conduction and valence band edges are shown in the figure. The dashed lines indicate the positions of energy levels associated with the quantized motion of electrons (ε_0^c) and holes (ε_0^v) in this quantum well. We can write the electron and hole energies, respectively, as

$$\varepsilon_\mathrm{c}(k) = \tilde{\varepsilon}_\mathrm{c} + \frac{\hbar^2}{2m_\mathrm{c}}\left(k_x^2 + k_y^2\right)$$

and

$$\varepsilon_\mathrm{v}(k) = \tilde{\varepsilon}_\mathrm{v} - \frac{\hbar^2}{2m_\mathrm{v}}\left(k_x^2 + k_y^2\right)$$

where $\tilde{\varepsilon}_\mathrm{c} = \varepsilon_\mathrm{c} + \varepsilon_0^\mathrm{c}$ and $\tilde{\varepsilon}_\mathrm{v} = \varepsilon_\mathrm{v} + \varepsilon_0^\mathrm{v}$.

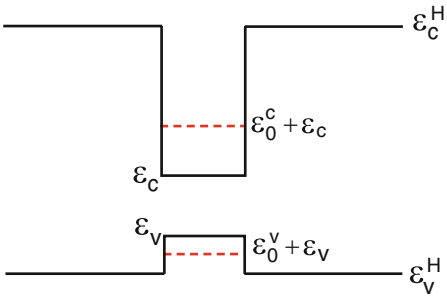
(a) Calculate the two-dimensional density of states for the electrons and holes assuming that other quantized levels can be ignored. Remember that

$$L^2 g_{\mathrm{c}}(\varepsilon) = \sum_{\substack{k_x, k_y, \sigma \\ \varepsilon < \varepsilon_k < \varepsilon + \mathrm{d}\varepsilon}} 1.$$

(b) Determine $N_{\mathrm{c}}(T)$ and $P_{\mathrm{v}}(T)$ for this two-dimensional system. Remember that

$$N_{\mathrm{c}}(T) = \int_{\tilde{\varepsilon}_{\mathrm{c}}}^{\infty} \mathrm{d}\varepsilon g_{\mathrm{c}}(\varepsilon) \mathrm{e}^{-\frac{\varepsilon - \tilde{\varepsilon}_{\mathrm{c}}}{\Theta}}.$$

(c) Determine $n_{\mathrm{c}}(T)$ and $p_{\mathrm{v}}(T)$ for the intrinsic case.
(d) Determine the value of the chemical potential for this case.



Summary

In this chapter, we studied the physics of semiconducting material and artificial structures made of semiconductors. General properties of typical semiconductors are reviewed and temperature dependence of carrier concentration is considered for both intrinsic and doped cases. Then basic physics of p-n junctions is covered in equilibrium and the current–voltage characteristic of the junction is described. The characteristics of two-dimensional electrons are discussed for the electrons in surface space charge layers formed in metal-oxide-semiconductor structures, semiconductor superlattices, and quantum wells. The fundamentals of the quantum Hall effects and the effects of disorders and modulation doping are also discussed.

The densities of states in the conduction and valence bands are given by

$$g_{\rm c}(\varepsilon) = \frac{\sqrt{2}m_{\rm c}^{3/2}}{\pi^2\hbar^3}\left(\varepsilon - \varepsilon_{\rm c}\right)^{1/2};\ g_{\rm v}(\varepsilon) = \frac{\sqrt{2}m_{\rm v}^{3/2}}{\pi^2\hbar^3}\left(\varepsilon_{\rm v} - \varepsilon\right)^{1/2}.$$

In the case of nondegenerate regime, we have $\varepsilon_{\rm c} - \zeta \gg \Theta$ and $\zeta - \varepsilon_{\rm v} \gg \Theta$, where Θ is $k_{\rm B}T$. Then the carrier concentrations become

$$n_{\rm c}(T) = N_{\rm c}(T){\rm e}^{-\frac{\varepsilon_{\rm c}-\zeta}{\Theta}};\ p_{\rm v}(T) = P_{\rm v}(T){\rm e}^{-\frac{\zeta-\varepsilon_{\rm v}}{\Theta}},$$

where

$$N_{\rm c}(T) = \int_{\varepsilon_{\rm c}}^{\infty} {\rm d}\varepsilon g_{\rm c}(\varepsilon){\rm e}^{-\frac{\varepsilon-\varepsilon_{\rm c}}{\Theta}};\ P_{\rm v}(T) = \int_{\infty}^{\varepsilon_{\rm v}} {\rm d}\varepsilon g_{\rm v}(\varepsilon).$$

The product $n_{\rm c}(T)p_{\rm v}(T)$ is independent of ζ such that

$$n_{\rm c}(T)p_{\rm v}(T) = N_{\rm c}(T)P_{\rm v}(T){\rm e}^{-E_{\rm Gap}/\Theta}.$$

In the absence of impurities, $n_{\rm c}(T) = p_{\rm v}(T)$ and we have

$$n_{\rm i}(T) = \left[N_{\rm c}(T)P_{\rm v}(T)\right]^{1/2}{\rm e}^{-E_{\rm Gap}/2\Theta}.$$

The chemical potential now becomes

$$\zeta_{\rm i} = \varepsilon_{\rm c} - \frac{1}{2}E_{\rm Gap} + \frac{3}{4}\Theta\ln\left(\frac{m_{\rm v}}{m_{\rm c}}\right);\ \zeta_{\rm i} = \varepsilon_{\rm v} + \frac{1}{2}E_{\rm Gap} + \frac{3}{4}\Theta\ln\left(\frac{m_{\rm v}}{m_{\rm c}}\right).$$

When donors are present, the chemical potential ζ will move from its intrinsic value $\zeta_{\rm i}$ to a value near the conduction band edge. If the concentration of donors is sufficiently small, the average occupancy of a single donor impurity state is given by

$$\langle n_{\rm d}\rangle = \frac{1}{\frac{1}{2}{\rm e}^{\beta(\varepsilon_{\rm d}-\zeta)}+1}.$$

The numerical factor of $\frac{1}{2}$ in $\langle n_{\rm d}\rangle$ comes from the fact that either spin up or spin down states can be occupied but not both.

At a finite temperature, we have

$$n_{\rm c}(T) = N_{\rm c}(T)\mathrm{e}^{-\beta(\varepsilon_{\rm c}-\zeta)},\ p_{\rm v}(T) = P_{\rm v}(T)\mathrm{e}^{-\beta(\zeta-\varepsilon_{\rm v})},$$
$$n_{\rm d}(T) = \frac{N_{\rm d}}{\frac{1}{2}\mathrm{e}^{\beta(\varepsilon_{\rm d}-\zeta)}+1},\ p_{\rm a}(T) = \frac{N_{\rm a}}{\frac{1}{2}\mathrm{e}^{\beta(\zeta-\varepsilon_{\rm a})}+1}.$$

In addition, we have charge neutrality condition given by

$$n_{\rm c} + n_{\rm d} = N_{\rm d} - N_{\rm a} + p_{\rm v} + p_{\rm a}.$$

The set of these five equations should be solved numerically in order to have self consistent result for five unknowns.

The region of the p–n junction is a high resistance region and the electrical current density becomes

$$j = e\left(J_{\rm h}^{\rm gen} + J_{\rm e}^{\rm gen}\right)\left(\mathrm{e}^{eV/\Theta} - 1\right),$$

where $J_{\rm h}^{\rm gen}$ and $J_{\rm e}^{\rm gen}$ are hole and electron generation current densities, respectively.

Near the interface of metal–oxide–semiconductor structure under a strong enough gate voltage, the motion of the electrons is characterized by

$$\varepsilon = \varepsilon_0 + \frac{\hbar^2}{2m_{\rm c}^*}\left(k_x^2 + k_y^2\right);\ \Psi_{n,k_x,k_y} = \frac{1}{L}\mathrm{e}^{\mathrm{i}(k_x x + k_y y)}\xi_n(z).$$

Here, $\xi_n(z)$ is the nth eigenfunction of a differential equation given by

$$\left[\frac{1}{2m_{\rm c}^*}\left(-\mathrm{i}\hbar\frac{\partial}{\partial z}\right)^2 + V_{\rm eff}(z) - \varepsilon_n\right]\xi_n(z) = 0.$$

If a quantum well is narrow, it leads to quantized motion and subbands:

$$\varepsilon_n^{(\rm c)}(\mathbf{k}) = \varepsilon_n^{(\rm c)} + \frac{\hbar^2}{2m_{\rm c}^*}\left(k_x^2 + k_y^2\right).$$

In the presence of a dc magnetic field $\mathbf{B}$ applied normal to the plane of the 2DEG, the Hamiltonian of a single electron is written by

$$H = \frac{1}{2m}\left(\mathbf{p} + \frac{e}{c}\mathbf{A}\right)^2.$$

Here, $\mathbf{p} = (p_x, p_y)$ and $\mathbf{A}(\mathbf{r})$ is the vector potential whose curl gives $\mathbf{B} = (0, 0, B)$. The electronic states are described by

$$E_{nk} = \hbar\omega_{\rm c}\left(n + \frac{1}{2}\right),\ \Psi_{nk}(x, y, z) = \mathrm{e}^{\mathrm{i}ky}\ u_n\left(x + \frac{\hbar k}{m\omega_{\rm c}}\right);\ n = 0, 1, 2, \ldots.$$

The density of states (per unit length) is given by $g(\varepsilon) \propto \sum_n \delta\left(\varepsilon - \hbar\omega_{\rm c}(n + \frac{1}{2})\right)$. The total number of states per Landau level is equal to the *magnetic flux* through the sample divided by the *flux quantum* $\frac{hc}{e}$:

$$N_{\rm L} = \frac{BL^2}{hc/e}.$$

8

Dielectric Properties of Solids

8.1 Review of Some Ideas of Electricity and Magnetism

When an external electromagnetic disturbance is introduced into a solid, it will produce induced charge density and induced current density. These induced densities produce induced electric and magnetic fields. We begin with a brief review of some elementary electricity and magnetism. In this chapter, we will neglect the magnetization produced by induced current density and concentrate on the electric polarization field produced by the induced charge density.

The potential $\phi(\mathbf{r})$ set up by a collection of charges q_i at positions $\mathbf{r}_i$ is given by

$$\phi(\mathbf{r}) = \sum_i \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|}. \tag{8.1}$$

The electric field $\mathbf{E}(\mathbf{r})$ is given by $\mathbf{E}(\mathbf{r}) = -\nabla\phi(\mathbf{r})$.

Now, consider a dipole at position $\mathbf{r}'$ (see Fig. 8.1).

$$\phi(\mathbf{r}) = \frac{q}{\left|\mathbf{r} - \mathbf{r}' - \frac{\mathbf{d}}{2}\right|} - \frac{q}{\left|\mathbf{r} - \mathbf{r}' + \frac{\mathbf{d}}{2}\right|}. \tag{8.2}$$

By a dipole we mean $\mathbf{p} = q\mathbf{d}$ is a constant, called the *dipole moment*, but $|\mathbf{d}| = d$ itself is vanishingly small. If we expand for $|\mathbf{r} - \mathbf{r}'| \gg |\ \mathbf{d}\ |$, we find

$$\phi(\mathbf{r}) = \frac{q\mathbf{d}\cdot(\mathbf{r} - \mathbf{r}')}{(\mathbf{r} - \mathbf{r}')^3} = \frac{\mathbf{p}\cdot(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}. \tag{8.3}$$

The potential produced by a collection of dipoles $\mathbf{p}_i$ located at $\mathbf{r}_i$ is simply

$$\phi(\mathbf{r}) = \sum_i \frac{\mathbf{p}_i\cdot(\mathbf{r} - \mathbf{r}_i)}{|\mathbf{r} - \mathbf{r}_i|^3}. \tag{8.4}$$

Again the electric field $\mathbf{E}(\mathbf{r}) = -\nabla\phi(\mathbf{r})$, so

$$\mathbf{E}(\mathbf{r}) = \sum_i \frac{3(\mathbf{r} - \mathbf{r}_i)\left[\mathbf{p}_i\cdot(\mathbf{r} - \mathbf{r}_i)\right] - (\mathbf{r} - \mathbf{r}_i)^2\mathbf{p}_i}{|\mathbf{r} - \mathbf{r}_i|^5}. \tag{8.5}$$

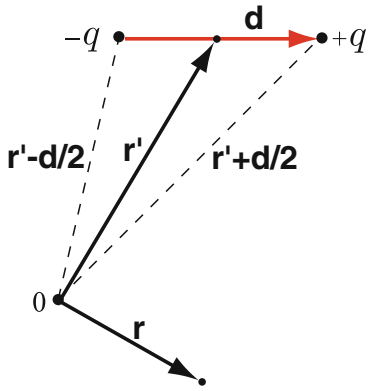


Fig. 8.1. Electric dipole of moment $\mathbf{p} = q\mathbf{d}$ located at $\mathbf{r}'$

8.2 Dipole Moment Per Unit Volume

Let us introduce the electric polarization $\mathbf{P}(\mathbf{r})$, which is the dipole moments per unit volume. Consider a volume V bounded by a surface S filled with a polarization $\mathbf{P}(\mathbf{r}')$ that depends on the position $\mathbf{r}'$. Then

$$\phi(\mathbf{r}) = \int \mathrm{d}^3 r' \, \frac{\mathbf{P}(\mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}. \tag{8.6}$$

If we look at the divergence of $\frac{\mathbf{P}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$ with respect to $\mathbf{r}'$, we note that

$$\nabla' \cdot \left[\frac{\mathbf{P}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \right] = \frac{1}{|\mathbf{r} - \mathbf{r}'|} \nabla' \cdot \mathbf{P}(\mathbf{r}') + \frac{\mathbf{P}(\mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}. \tag{8.7}$$

We can solve for $\frac{\mathbf{P}(\mathbf{r}')\cdot(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}$ and substitute into our expression for $\phi(\mathbf{r})$. The integral of the divergence term can be expressed as a surface integral using divergence theorem. This gives

$$\phi(\mathbf{r}) = \oint_S \mathrm{d}S' \, \frac{\mathbf{P}(\mathbf{r}') \cdot \hat{\mathbf{n}}'}{|\mathbf{r} - \mathbf{r}'|} + \int_V \mathrm{d}^3 r' \, \frac{[-\nabla' \cdot P(\mathbf{r}')]}{|\mathbf{r} - \mathbf{r}'|}. \tag{8.8}$$

The potential $\phi(\mathbf{r})$ can be associated with a potential produced by a volume distribution of charge density

$$\rho_P(\mathbf{r}) = -\nabla \cdot \mathbf{P}(\mathbf{r}) \tag{8.9}$$

and the potential produced by a surface charge density

$$\sigma_{\mathrm{P}}(\mathbf{r}) = \mathbf{P}(\mathbf{r}) \cdot \hat{\mathbf{n}}. \tag{8.10}$$

Here, of course, $\hat{\mathbf{n}}$ is a unit vector outward normal to the surface S bounding the volume V.

Poisson's equation tells us that

$$\nabla \cdot \mathbf{E} = 4\pi \left(\rho_0 + \rho_{\mathrm{P}}\right), \tag{8.11}$$

where ρ_0 is some *external charge density* and ρ_{P} is the *polarization charge density*. Since $\rho_{\mathrm{P}} = -\nabla \cdot \mathbf{P}$, we can write

$$\nabla \cdot \mathbf{E} = 4\pi\rho_0 - 4\pi\nabla \cdot \mathbf{P}. \tag{8.12}$$

If we define $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$, then

$$\nabla \cdot \mathbf{D} = 4\pi\rho_0. \tag{8.13}$$

Thus, $\mathbf{D}$ is the electric field that would be produced by the external charge density ρ_0 if a polarizable material were absent. $\mathbf{E}$ is the true electric field produced by all the charge densities including both ρ_0 and ρ_{P}.

In general, $\mathbf{P}$ and $\mathbf{E}$ need not be in the same direction. However, for sufficiently small value of $\mathbf{E}$, the relationship between $\mathbf{P}$ and $\mathbf{E}$ is linear. We can write

$$P_i = \sum_j \chi_{ij} E_j, \tag{8.14}$$

where $\underline{\chi}$ is called the *electrical susceptibility tensor*. We can write

$$\mathbf{D} = \underline{\varepsilon} \cdot \mathbf{E}, \tag{8.15}$$

where $\underline{\varepsilon} = \underline{1} + 4\pi\underline{\chi}$ is the *dielectric tensor*.

8.3 Atomic Polarizability

An atom in its ground state has no dipole moment. However, in the presence of an electric field $\mathbf{E}$, an induced dipole moment results from the relative displacements of the positive and negative charges within the atom. We can write

$$\mathbf{p}_{\mathrm{ind}} = \alpha\mathbf{E}, \tag{8.16}$$

and α is called the *atomic polarizability*.

8.4 Local Field in a Solid

In a dilute gas of atoms the electric field $\mathbf{E}$ that produces the induced dipole moment on an atom is simply the applied electric field. In a solid, however, all of the dipole moments produced on other atoms in the solid make a contribution to the field acting on a given atom. The value of this microscopic field at the position of the atom is called the *local field*. The local field $\mathbf{E}_{\mathrm{LF}}(\mathbf{r})$ is different from the applied electric field $\mathbf{E}_0$ and from the *macroscopic electric*

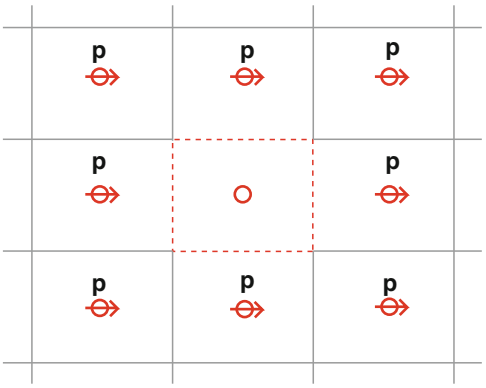


Fig. 8.2. Induced dipoles of moment **p** located on neighboring atoms

field **E** (which is the average of the microscopic field $\mathbf{E}_{\mathrm{LF}}(\mathbf{r})$ over a region that is large compared to a unit cell). Clearly, the contributions to the microscopic field from the induced dipoles on neighboring atoms vary considerably over the unit cell (see Fig. 8.2). The standard method of evaluating the local field $\mathbf{E}_{\mathrm{LF}}(\mathbf{r})$ in terms of the macroscopic field **E** is to make use of the *Lorentz sphere*. Before introducing the Lorentz field, let us review quickly the relation between the external field $\mathbf{E}_0$ and the macroscopic field **E** in the solid.

8.5 Macroscopic Field

Suppose the solid we are studying is shaped like an ellipsoid. It is a standard problem in electromagnetism to determine the electric field **E** inside the ellipsoid in terms of the external electric field $\mathbf{E}_0$ (see Fig. 8.3).

The applied field $\mathbf{E}_0$ is the value of the electric field very far away from the sample. The macroscopic field inside the ellipsoid is given by

$$\mathbf{E} = \mathbf{E}_0 - \lambda \mathbf{P} = \mathbf{E}_0 + \mathbf{E}_1. \tag{8.17}$$

The field $\mathbf{E}_1 = -\lambda \mathbf{P}$ is called the *depolarization field*, due to surface charge density $\hat{\mathbf{n}} \cdot \mathbf{P}$ on the outer surface of the specimen, and λ is called the *depolarization factor*.

8.5.1 Depolarization Factor

The standard electromagnetic theory problem of determining λ involves

1. Solving Laplace's equation $\nabla^2 \phi(\mathbf{r}) = 0$ in cylindrical coordinates so that

$$\phi(\mathbf{r}) = \left(ar^l + br^{-(l+1)}\right) P_l(\cos\theta)\,(C \sin m\phi + D \cos m\phi)$$

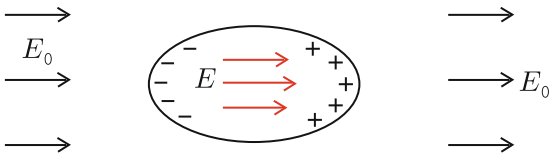


Fig. 8.3. The macroscopic electric field **E** inside an ellipsoid located in an external electric field $\mathbf{E}_0$ is the sum of $\mathbf{E}_0$ and polarization field $\mathbf{E}_1 = -\lambda\mathbf{P}$ due to the surface charge density $\hat{\mathbf{n}} \cdot \mathbf{P}$

Table 8.1. Depolarization factors λ of typical ellipsoids

Type of ellipsoid	Axis	λ
Sphere	Any	$\frac{4\pi}{3}$
Thin slap	Normal	4π
Thin slap	Parallel	0
Long cylinder	Along axis	0
Long circular cylinder	Normal to axis	2π

(a) Inside the sample (where r can approach 0) and
(b) Outside the sample (where r can approach ∞),

2. Imposing boundary conditions
   (a) **E** well behaved as $r \to 0$,
   (b) $\mathbf{E} \to \mathbf{E}_0$ as $r \to \infty$,
   (c) $\mathbf{D}_{\text{normal}} = (\mathbf{E} + 4\pi\mathbf{P})_{\text{normal}}$ and $\mathbf{E}_{\text{trans}}$ be continuous at the surface.

For an ellipsoid with the depolarization factors λ_1, λ_2, and λ_3 along the three principal axes.

$$\lambda_1 + \lambda_2 + \lambda_3 = 4\pi. \tag{8.18}$$

Some examples are listed in Table 8.1.

8.6 Lorentz Field

Assume that we know **E**, the macroscopic field inside the solid. Now consider an atom at position **R**. Draw a sphere of radius ℓ (named as *Lorentz sphere*) about **R** where $\ell \gg a$, the interatomic spacing (see Fig. 8.4). The contribution to the microscopic field at **R** from induced dipoles on other atoms can be divided into two parts:

1. For atoms inside the sphere we will actually sum over the contribution from their individual dipole moments $\mathbf{p}_i$.
2. For atoms outside the sphere we can treat the contribution macroscopically, treating them as part of a continuum with polarization **P**.

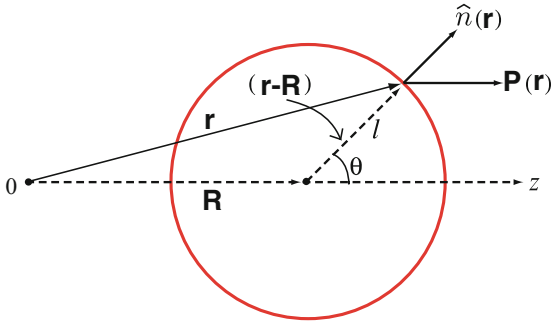


Fig. 8.4. A Lorentz sphere of radius ℓ centered at $\mathbf{R}$

The dipole moments outside the Lorentz sphere contribute a surface charge density on the surface of the Lorentz sphere, and we can write

$$\phi(\mathbf{R}) = \int_{\text{Lorentz sphere}} \mathrm{d}S \; \frac{\mathbf{P}(\mathbf{r}) \cdot \hat{\mathbf{n}}(\mathbf{r})}{|\mathbf{R} - \mathbf{r}|}. \tag{8.19}$$

The field $\mathbf{E}_2$ caused by this surface charge on the spherical cavity (Lorentz sphere) is called the *Lorentz field*:

$$\mathbf{E}_2(\mathbf{R}) = -\nabla_{\mathbf{R}}\phi(\mathbf{R}) = \int \mathrm{d}S \; \mathbf{P}(\mathbf{r}) \cdot \hat{\mathbf{n}}(\mathbf{r}) \frac{(\mathbf{R} - \mathbf{r})}{|\mathbf{R} - \mathbf{r}|^3}. \tag{8.20}$$

To evaluate this integral note, from Fig. 8.4, that

$$\begin{aligned} |\mathbf{r} - \mathbf{R}| &= \ell, \\ \mathbf{P}(\mathbf{r}) \cdot \hat{\mathbf{n}}(\mathbf{r}) &= P \cos\theta, \\ \mathrm{d}S &= 2\pi\ell^2 \sin\theta \, \mathrm{d}\theta, \text{ and} \\ \mathbf{R} - \mathbf{r} &= \ell \left(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta\right). \end{aligned}$$

Hence, we have

$$E_2(\mathbf{R}) = -\int_0^{\pi} 2\pi\ell^2 \; \mathrm{d}(\cos\theta) P \cos\theta \frac{\ell\cos\theta}{\ell^3}.$$

Only the z-component of $\mathbf{R} - \mathbf{r}$ survives the integration. We find that

$$\mathbf{E}_2(\mathbf{R}) = 2\pi\mathbf{P}\int_{-1}^{1} \mathrm{d}(\cos\theta) \; \cos^2\theta = \frac{4\pi}{3}\mathbf{P}. \tag{8.21}$$

$\mathbf{E}_2 = \frac{4\pi\mathbf{P}}{3}$ is the Lorentz field.

The final contribution $\mathbf{E}_3$ arises from the contribution of the dipoles within the Lorentz sphere (L.S.). It is given by

$$\mathbf{E}_3 = \sum_{i \in \text{L.S.}} \frac{3\left(\mathbf{p}_i \cdot \mathbf{r}_i\right)\mathbf{r}_i - r_i^2 \mathbf{p}_i}{r_i^5}. \tag{8.22}$$

This term clearly depends on the crystal structure. If $\mathbf{p}_i = \mathbf{p} = p\hat{z}$, then the field at the center of the Lorentz sphere is

$$\mathbf{E}_3 = (0, 0, E_3) = p \sum_{i \in \text{L.S.}} \frac{3z_i^2 - r_i^2}{r_i^5} \hat{z}. \tag{8.23}$$

For a crystal with cubic symmetry

$$\sum_i \frac{x_i^2}{r_i^5} = \sum_i \frac{y_i^2}{r_i^5} = \sum_i \frac{z_i^2}{r_i^5} = \frac{1}{3} \sum_i \frac{r_i^2}{r_i^5}.$$

Thus, for a cubic crystal, the two terms cancel and $E_3 = 0$. Hence, we find the local field in a cubic crystal

$$\begin{aligned} E_{\text{LF}} &= \underbrace{E_0 + E_1}_{E} + \underbrace{E_2}_{+\frac{4\pi}{3}P} + E_3 \\ &= E + \tfrac{4\pi}{3}P + 0. \end{aligned} \tag{8.24}$$

The last expression is the *Lorentz relation*. We note that, since $E_1 = -\frac{4\pi}{3}P$ for a spherical specimen, the local field at the center of a sphere of cubic crystal is simply given by

$$E_{\text{LF}}^{\text{sphere}} = E_0 - \frac{4\pi}{3}P + \frac{4\pi}{3}P + 0 = E_0.$$

8.7 Clausius–Mossotti Relation

The induced dipole moment of an atom is given, in terms of the local field, by $\mathbf{p} = \alpha \mathbf{E}_{\text{LF}}$. The polarization $\mathbf{P}$ is given, for a cubic crystal, by

$$\mathbf{P} = N\mathbf{p} = N\alpha \mathbf{E}_{\text{LF}} = N\alpha \left(\mathbf{E} + \frac{4\pi}{3} \mathbf{P} \right), \tag{8.25}$$

where N is the number of atoms per unit volume. Solving for $\mathbf{P}$ gives

$$\mathbf{P} = \frac{N\alpha}{1 - \frac{4\pi N\alpha}{3}} \mathbf{E} \equiv \chi \mathbf{E}. \tag{8.26}$$

Thus, the electrical susceptibility of the solid is

$$\chi = \frac{N\alpha}{1 - \frac{4\pi N\alpha}{3}}. \tag{8.27}$$

Since $\mathbf{D} = \varepsilon \mathbf{E}$ with $\varepsilon = 1 + 4\pi\chi$, we find that

$$\varepsilon = 1 + \frac{4\pi N\alpha}{1 - \frac{4\pi N\alpha}{3}}. \tag{8.28}$$

or

$$\varepsilon = \frac{1 + \frac{8\pi N\alpha}{3}}{1 - \frac{4\pi N\alpha}{3}}. \tag{8.29}$$

This relation between the macroscopic dielectric function ε and the atomic polarizability α is called the *Clausius–Mossotti relation*. It can also be written (solving for $\frac{4\pi N\alpha}{3}$) by

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{4\pi N\alpha}{3}. \tag{8.30}$$

8.8 Polarizability and Dielectric Functions of Some Simple Systems

The total polarizability of the atoms or ions within a unit cell can usually be separated into three parts:

1. *Electronic polarizability* α_{e}: The displacement of the electrons relative to the nucleus.
2. *Ionic polarizability* α_{i}: The displacement of an ion itself with respect to its equilibrium position.
3. *Dipolar polarizability* α_{dipole}: The orientation of any permanent dipoles by the electric field in the presence of thermal disorder.

Atoms and homonuclear diatomic molecules have no dipole moments in their ground states. Molecules like KCl, HCl, $H_2O, \ldots$ do exhibit permanent dipole moments. A typical dipole moment $p = qd$ has $q \simeq 4.8 \times 10^{-10}$ esu and $d \simeq 10^{-8}$ cm, giving $p \simeq 5 \times 10^{-18}$ stat-coulomb cm.

8.8.1 Evaluation of the Dipole Polarizability

In the absence of an electric field, the probability that a dipole $\mathbf{p}$ will be oriented within the solid angle $\mathrm{d}\Omega = \sin\theta\, \mathrm{d}\theta\, \mathrm{d}\phi$ is independent of θ and ϕ and is given by $\frac{\mathrm{d}\Omega}{4\pi}$. In the presence of a field $\mathbf{E}$, the probability is proportional to $\mathrm{d}\Omega\ \mathrm{e}^{-W/k_{\mathrm{B}}T}$, where $W = -\mathbf{p}\cdot\mathbf{E}$ is the energy of the dipole in the field $\mathbf{E}$. If we choose the z-direction parallel to $\mathbf{E}$, then the average values of p_x and p_y vanish and we have

$$\bar{p}_z = \frac{\int \mathrm{d}\Omega\ \mathrm{e}^{pE\cos\theta/k_{\mathrm{B}}T}\ p\cos\theta}{\int \mathrm{d}\Omega\ \mathrm{e}^{pE\cos\theta/k_{\mathrm{B}}T}}. \tag{8.31}$$

Let $\frac{pE}{k_{\mathrm{B}}T} = \xi$, $\cos\theta = x$ and rewrite $\bar{p}_z$ as

$$\begin{aligned}\bar{p}_z &= p\frac{\int_{-1}^{1} dx\ x\mathrm{e}^{\xi x}}{\int_{-1}^{1} dx\ \mathrm{e}^{\xi x}} = p\frac{d}{d\xi}\ln\int_{-1}^{1}\mathrm{d}x\ \mathrm{e}^{\xi x} \\ &= p\frac{\mathrm{d}}{\mathrm{d}\xi}\ln\left(\frac{2\sinh\xi}{\xi}\right) = p\left(\coth\xi - \tfrac{1}{\xi}\right).\end{aligned}$$

Thus, we can write $\bar{p}_z$ as

$$\bar{p}_z = p\mathcal{L}(\xi). \tag{8.32}$$

Here, $\mathcal{L}(\xi)$ is the *Langevin function* defined by $\mathcal{L}(\xi) = \coth\xi - \frac{1}{\xi}$. The dipole moment per unit volume is then given by

$$P = N\bar{p}_z = Np\,\mathcal{L}\left(\frac{pE}{k_{\mathrm{B}}T}\right).$$

We note that for $\xi \to \infty$, $\mathcal{L}(\xi) \to 1$, while for $\xi \to 0$, $\mathcal{L}(\xi) = \frac{\xi}{3}$. If $\xi \ll 1$, $P = \frac{Np^2E}{3k_{\mathrm{B}}T}$. At room temperature the condition is satisfied if $E \ll \frac{k_{\mathrm{B}}T}{p} \simeq 10^7$ volts/cm. The standard unit of dipole moment is the *Debye*, defined by 1 Debye $= 10^{-18}$ esu. Figure 8.5 is a sketch of the temperature dependence of an electrical polarization P.

The electronic polarizability α_{e} and the ionic polarizability α_{ion} are almost independent of temperature. Therefore, by measuring $\frac{\varepsilon-1}{\varepsilon+2} = \frac{4\pi N\alpha}{3}$ as a function of temperature one can obtain the value of p^2 from the slope (see Fig. 8.6).

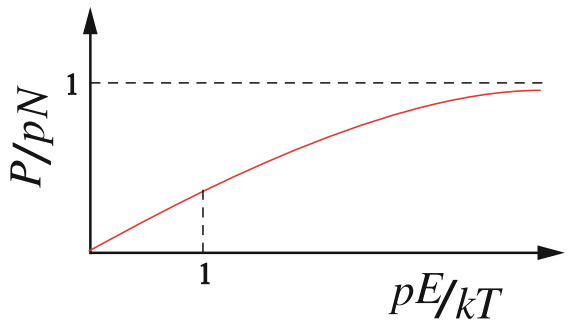


Fig. 8.5. Temperature dependence of an electrical polarization P

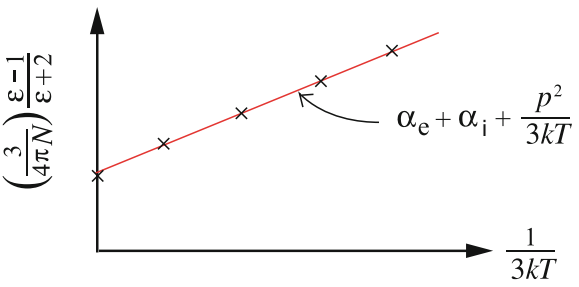


Fig. 8.6. An electrical polarizability α as a function of temperature

8.8.2 Polarizability of Bound Electrons

Assume that an electron (of charge $q = -e$) is bound harmonically to a particular location (e.g., the position of a particular ion). Its equation of motion is written by

$$m\left(\ddot{\mathbf{r}} + \frac{\dot{\mathbf{r}}}{\tau}\right) = -k\mathbf{r} - e\mathbf{E}, \tag{8.33}$$

where $-k\mathbf{r}$ is the restoring force and $\mathbf{E}$ is the perturbing electric field. Assume $\mathbf{E} \propto \mathrm{e}^{\mathrm{i}\omega t}$ and let $k = m\omega_0^2$. Then we can solve for $\mathbf{r} \propto \mathrm{e}^{\mathrm{i}\omega t}$ to obtain

$$\mathbf{r} = \frac{-e\mathbf{E}/m}{-\omega^2 + \mathrm{i}\omega/\tau + \omega_0^2} \equiv \frac{\mathbf{p}}{-e}. \tag{8.34}$$

The dipole moment of the atom will be $\mathbf{p} = -e\mathbf{r}$ and the polarization $\mathbf{P} = N\mathbf{p} = -Ne\mathbf{r} \equiv N\alpha_{\mathrm{el}}\mathbf{E}$. This gives for α_{el}

$$\alpha_{\mathrm{el}} = \frac{(e^2/m)\left[\omega_0^2 - \omega^2 - \mathrm{i}\omega/\tau\right]}{(\omega_0^2 - \omega^2)^2 + (\omega/\tau)^2}. \tag{8.35}$$

The dielectric function $\varepsilon = 1 + 4\pi N\alpha_{\mathrm{el}}$ is

$$\varepsilon(\omega) = 1 + \frac{\left(4\pi Ne^2/m\right)\left(\omega_0^2 - \omega^2 - \mathrm{i}\omega/\tau\right)}{(\omega_0^2 - \omega^2)^2 + (\omega/\tau)^2}. \tag{8.36}$$

It is clear that α_{e} has a real and an imaginary part whose frequency dependences are of the form sketched in Fig. 8.7.

8.8.3 Dielectric Function of a Metal

In a metal (e.g., Drude model) the electrons are free. This means that the restoring force vanishes (i.e., $k \to 0$) or $\omega_0 = 0$. In that case, we obtain

$$\alpha_{\mathrm{e}} = \frac{e^2/m}{-\omega^2 + \mathrm{i}\omega/\tau}, \tag{8.37}$$

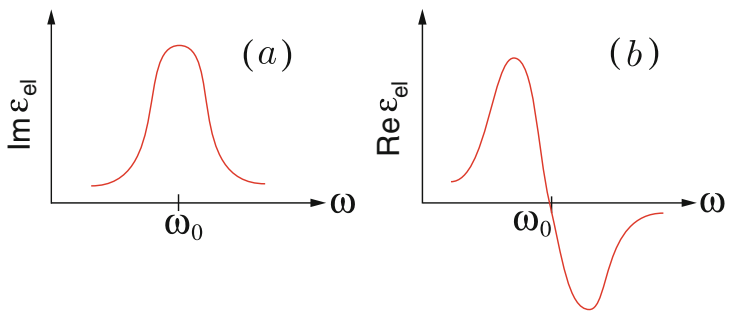


Fig. 8.7. The frequency dependence of the dielectric function ε of atoms with bound electrons. (**a**) Real part of $\varepsilon(\omega)$, (**b**) imaginary part of $\varepsilon(\omega)$

or

$$\varepsilon(\omega) = 1 - \frac{4\pi N e^2/m}{\omega^2 - \mathrm{i}\omega/\tau} = 1 - \frac{\omega_p^2}{\omega^2 - \mathrm{i}\omega/\tau}. \tag{8.38}$$

The real and imaginary parts of $\varepsilon(\omega)$ are

$$\Re\varepsilon(\omega) = 1 - \frac{\omega_p^2\tau^2}{1+\omega^2\tau^2} \tag{8.39}$$

and

$$\Im\varepsilon(\omega) = -\frac{\omega_p^2\tau^2}{\omega\tau(1+\omega^2\tau^2)}. \tag{8.40}$$

The fact that $\Re\varepsilon(\omega) < 0$ for $\omega^2 < \omega_p^2 - \frac{1}{\tau^2}$ leads to an imaginary refractive index and is responsible for the fact that metals are good reflectors.

8.8.4 Dielectric Function of a Polar Crystal

In an ionic crystal like NaCl, longitudinal optical phonons have associated with them charge displacements, which result in a macroscopic polarization field $\mathbf{P}_\mathrm{L}$. Here the subscript L stands for the lattice polarization. (See, for example, Fig. 8.8.)

The polarization field $\mathbf{P}_\mathrm{L}$ consists of two parts: (1) the displacements of the charged individual ions from their equilibrium positions and (2) the polarization of the ions themselves resulting from the displacement of the electrons relative to the nucleus under the influence of the $\mathbf{E}$ field. In determining each of these contributions to $\mathbf{P}_\mathrm{L}$, we must use the local field $\mathbf{E}_\mathrm{LF}$. We shall consider the following model of a polar crystal:

1. The material is a cubic lattice with two atoms per unit cell; the volume of the unit cell is V.
2. The charges, masses, and atomic polarizabilities of these ions are $\pm ze$, $M_\pm$, and $\alpha_\pm$
3. In addition to long range electrical forces, there is a short range restoring force that is proportional to the relative displacement of the pair of atoms in the same cell.

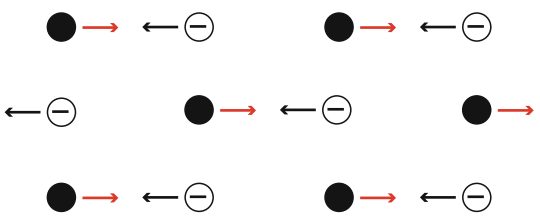


Fig. 8.8. Polarization field due to charge displacements in a polar crystal

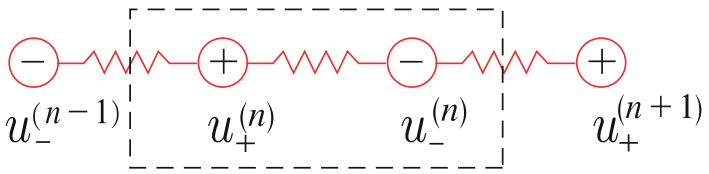


Fig. 8.9. Ionic displacements in the n^{th} unit cell and its nearest neighbor atoms

Note: Here we are considering only the $q = 0$ optical phonon, so that the ionic displacements are identical in each cell. Therefore, the restoring force can be written in such a way that: The force on $M_+^{(n)}$ depends on $u_+^{(n)} - u_-^{(n)}$ and $u_+^{(n)} - u_-^{(n-1)}$, but $u_-^{(n)} = u_-^{(n-1)}$ so the force is simply proportional to $u_+^{(n)} - u_-^{(n)}$. (See Fig. 8.9.)

We can write the equations of motion of u_+ and u_- as

$$\begin{aligned} M_+\ddot{\mathbf{u}}_+ &= -k(\mathbf{u}_+ - \mathbf{u}_-) + zeE_{\text{LF}}, \\ M_-\ddot{\mathbf{u}}_- &= k(\mathbf{u}_+ - \mathbf{u}_-) - zeE_{\text{LF}}. \end{aligned} \tag{8.41}$$

The local field $\mathbf{E}_{\text{LF}}$ is given, for cubic crystals, by (8.24):

$$\mathbf{E}_{\text{LF}} = \mathbf{E} + \frac{4\pi}{3}\mathbf{P}_{\text{L}}, \tag{8.42}$$

and

$$\mathbf{P}_{\text{L}} = \frac{ze}{V}(\mathbf{u}_+ - \mathbf{u}_-) + \frac{1}{V}(\alpha_+ + \alpha_-)\mathbf{E}_{\text{LF}}. \tag{8.43}$$

Substitute for $\mathbf{E}_{\text{LF}}$ in terms of $\mathbf{E}$ and $\mathbf{P}_{\text{L}}$, and then solve for $\mathbf{P}_{\text{L}}$. This gives us

$$\mathbf{P}_{\text{L}} = \frac{ze}{V\beta}\mathbf{r} + \frac{\alpha_+ + \alpha_-}{V\beta}\mathbf{E}. \tag{8.44}$$

Here, $\beta = 1 - \frac{4\pi}{3}\left(\frac{\alpha_+ + \alpha_-}{V}\right)$ and $\mathbf{r} = \mathbf{u}_+ - \mathbf{u}_-$. Introduce

$$\begin{aligned} \Omega_\pm^2 &\equiv \tfrac{k}{M_\pm}, \\ \Omega_{p\pm}^2 &\equiv \tfrac{4\pi e^2}{VM_\pm}, \\ \overline{M}^{-1} &\equiv M_+^{-1} + M_-^{-1}. \end{aligned} \tag{8.45}$$

The equations of motion can be rewritten

$$\begin{aligned} -\omega^2\mathbf{u}_+ &= -\Omega_+^2\mathbf{r} + \tfrac{1}{3\beta}\Omega_{p+}^2\mathbf{r} + \tfrac{ze}{M_+\beta}\mathbf{E}, \\ -\omega^2\mathbf{u}_- &= +\Omega_-^2\mathbf{r} - \tfrac{1}{3\beta}\Omega_{p-}^2\mathbf{r} - \tfrac{ze}{M_-\beta}\mathbf{E}. \end{aligned} \tag{8.46}$$

If we subtract the second equation from the first we obtain

$$\left[-\omega^2 + \left(\Omega_+^2 + \Omega_-^2\right) - \frac{\Omega_{p+}^2 + \Omega_{p-}^2}{3\beta}\right]\mathbf{r} = \frac{ze}{\beta\overline{M}}\mathbf{E}. \tag{8.47}$$

This can be rewritten

$$\mathbf{r} = -\frac{ze}{\beta\overline{M}}\left(\omega^2 + b_{11}\right)^{-1}\mathbf{E}, \tag{8.48}$$

where

$$b_{11} = -\left[\Omega_+^2 + \Omega_-^2 - \frac{\Omega_{p+}^2 + \Omega_{p-}^2}{3\beta}\right] \equiv -\omega_{\mathrm{T}}^2. \tag{8.49}$$

$-b_{11}$ is a frequency squared and it is positive since $\Omega_{p\pm}^2$ is always smaller than $\Omega_\pm$. Let us call it $+\omega_{\mathrm{T}}^2$. Since we know the expression for $\mathbf{P}_{\mathrm{L}}$ in terms of $\mathbf{r}$ and $\mathbf{E}$ we can write

$$\mathbf{P}_{\mathrm{L}} = \frac{ze}{V\beta}\left(-\frac{ze}{\beta\overline{M}}\right)\frac{\mathbf{E}}{\omega^2 - \omega_{\mathrm{T}}^2} + \frac{\alpha_+ + \alpha_-}{\beta V}\mathbf{E}. \tag{8.50}$$

Let us define

$$\begin{aligned} b_{22} &= \frac{\alpha_+ + \alpha_-}{\beta V}, \\ b_{12}^2 &= \frac{z^2e^2}{\beta^2\overline{M}V}. \end{aligned} \tag{8.51}$$

Then, we can rewrite $\mathbf{P}_{\mathrm{L}}$ as

$$\mathbf{P}_{\mathrm{L}} = \left(b_{22} - \frac{b_{12}^2}{\omega^2 - \omega_{\mathrm{T}}^2}\right)\mathbf{E} \equiv \chi\mathbf{E}. \tag{8.52}$$

Recall that the electrical susceptibility is defined by $\chi(\omega) = \frac{P_{\mathrm{L}}(\omega)}{E(\omega)}$, and the dielectric function by

$$\varepsilon(\omega) = 1 + 4\pi\chi(\omega). \tag{8.53}$$

From (8.52) for $\mathbf{P}_{\mathrm{L}}$ we find

$$\chi(\omega) = b_{22} - \frac{b_{12}^2}{\omega^2 - \omega_{\mathrm{T}}^2}. \tag{8.54}$$

The frequency ω_{T} is in the infrared ($\sim 10^{13}$/sec). If we look at frequencies much larger than ω_{T} we find

$$\chi_\infty = b_{22}. \tag{8.55}$$

For $\omega \to 0$ we find that

$$\chi_0 = b_{22} + \frac{b_{12}^2}{\omega_{\mathrm{T}}^2} = \chi_\infty + \frac{b_{12}^2}{\omega_{\mathrm{T}}^2}. \tag{8.56}$$

Therefore we can write

$$b_{12}^2 = \omega_{\mathrm{T}}^2\left(\chi_0 - \chi_\infty\right) = \frac{\omega_{\mathrm{T}}^2}{4\pi}\left(\varepsilon_0 - \varepsilon_\infty\right). \tag{8.57}$$

This, of course, is positive since ε_0 contains contributions from the displacements of the ions as well as the electronic displacements within each ion. The latter is very fast while the former is slow. The dielectric function $\varepsilon(\omega)$ can be written

$$\begin{aligned}\varepsilon(\omega) &= \varepsilon_\infty - \frac{\omega_\mathrm{T}^2}{\omega^2-\omega_\mathrm{T}^2}(\varepsilon_0-\varepsilon_\infty)\\ &= \varepsilon_\infty\left[1-\left(\frac{\varepsilon_0}{\varepsilon_\infty}-1\right)\frac{\omega_\mathrm{T}^2}{\omega^2-\omega_\mathrm{T}^2}\right].\end{aligned}$$

We define $\omega_\mathrm{L}^2 = \omega_\mathrm{T}^2\frac{\varepsilon_0}{\varepsilon_\infty} > \omega_\mathrm{T}^2$. Then, we can write

$$\varepsilon(\omega) = \varepsilon_\infty\left[\frac{\omega^2-\omega_\mathrm{L}^2}{\omega^2-\omega_\mathrm{T}^2}\right]. \tag{8.58}$$

Here, ω_L and ω_T are the TO and LO phonon frequencies, respectively. We note that $\omega_\mathrm{L} > \omega_\mathrm{T}$ since $\varepsilon_0 > \varepsilon_\infty$ in general. Values of ε_0 and ε_∞ for some polar materials are listed in Table 8.2. Instead of discussing the lattice polarization $\mathbf{P}_\mathrm{L}$, we could have discussed the lattice current density

$$\mathbf{j}_\mathrm{L} = \dot{\mathbf{P}}_\mathrm{L} = \mathrm{i}\omega\mathbf{P}_\mathrm{L} = \mathrm{i}\omega\chi(\omega)\mathbf{E}.$$

A plot of $\varepsilon(\omega)$ versus ω is shown in Fig. 8.10. At $\omega = 0$ ε has the static value ε_0, and as $\omega \to \infty$ it approaches the high-frequency value ε_∞. ε_0 is always larger than ε_∞. There is a resonance at $\omega = \omega_\mathrm{T}$.

Table 8.2. Dielectric constants ε_0 and ε_∞ for polar materials

Crystal	ε_0	ε_∞	$\varepsilon_0/\varepsilon_\infty$
LiF	8.7	1.93	4.5
NaCl	5.62	2.25	2.50
KBr	4.78	2.33	2.05
Cu_2O	8.75	4.0	2.2
PbS	17.9	2.81	6.37

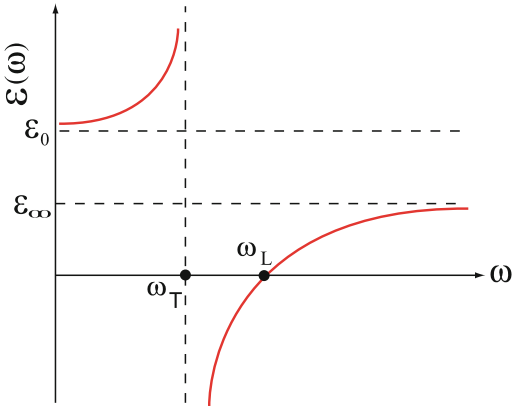


Fig. 8.10. Frequency dependence of the dielectric function $\varepsilon(\omega)$ of a polar crystal

8.9 Optical Properties

The dielectric and magnetic properties of a medium are characterized by the dielectric function $\varepsilon(\omega)$ and the magnetic permeability $\mu(\omega)$:

$$\mathbf{D} = \varepsilon\mathbf{E} \quad \text{and} \quad \mathbf{B} = \mu\mathbf{H}. \tag{8.59}$$

In terms of $\mathbf{E}$ and $\mathbf{B}$, Maxwell's equations can be written

$$\begin{aligned} \nabla \cdot \mathbf{E} &= 4\pi\rho = 4\pi(\rho_0 + \rho_\mathrm{P}) \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\tfrac{1}{c}\dot{\mathbf{B}} \\ \nabla \times \mathbf{B} &= \tfrac{1}{c}\dot{\mathbf{E}} + \tfrac{4\pi}{c}(\mathbf{j}_0 + \mathbf{j}_\mathrm{P}) + 4\pi\nabla \times \mathbf{M}. \end{aligned} \tag{8.60}$$

The last equation involves the magnetization which is normally very small. Here, we will neglect it; this is equivalent to taking $\mu = 1$ or $\mathbf{B} = \mathbf{H}$. The sources of $\mathbf{E}$ are all charges; external (ρ_0) and induced polarization (ρ_P) charge densities. The sources of $\mathbf{B}$ are the rate of change of $\mathbf{E}$ and the total current (external $\mathbf{j}_0$ and induced $\mathbf{j}_\mathrm{P}$ current densities). Recall that $\mathbf{j}_\mathrm{P} = \dot{\mathbf{P}}$ and $\nabla \cdot \mathbf{P} = -\rho_\mathrm{P}$.

> Note: Sometimes the first Maxwell equation is replaced by $\nabla \cdot \mathbf{D} = 4\pi\rho_0$. Here $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$ and as we have seen $\rho_\mathrm{P} = -\nabla \cdot \mathbf{P}$. The fourth equation is sometimes replaced by $\nabla \times \mathbf{H} = \frac{1}{c}\dot{\mathbf{D}} + \frac{4\pi}{c}\mathbf{j}_0$, which omits all polarization currents.

In this chapter, we shall ignore all magnetic effects and take $\mu(\omega) = 1$. This is an excellent approximation for most materials since the magnetic susceptibility is usually much smaller than unity. There are two extreme ways of writing the equation for $\nabla \times \mathbf{B}$:

$$\begin{aligned} \nabla \times \mathbf{B} &= \frac{\varepsilon}{c}\dot{\mathbf{E}} + \frac{4\pi}{c}\mathbf{j}_0 \text{ or} \\ \nabla \times \mathbf{B} &= \frac{1}{c}\dot{\mathbf{E}} + \frac{4\pi}{c}\left(\mathbf{j}_0 + \sigma\mathbf{E}\right) \end{aligned} \tag{8.61}$$

The first equation is just that for $\mathbf{H}$ in which we put $\mu = 1$ and $\mathbf{D} = \varepsilon\mathbf{E}$. The second equation is that for $\nabla \times \mathbf{B}$ in which we have taken $\mathbf{j}_\mathrm{P} = \sigma\mathbf{E}$ where σ is the conductivity. From this we see that $\frac{\mathrm{i}\omega}{c}\varepsilon(\omega) = \frac{\mathrm{i}\omega}{c} + \frac{4\pi}{c}\sigma(\omega)$, or

$$\varepsilon(\omega) = 1 - \frac{4\pi\mathrm{i}}{\omega}\sigma(\omega) \tag{8.62}$$

is a complex dielectric constant and simply related to the conductivity $\sigma(\omega)$. We have assumed that $\mathbf{E}$ and $\mathbf{B}$ are proportional to $\mathrm{e}^{\mathrm{i}\omega t}$.

8.9.1 Wave Equation

For the propagation of light in a material characterized by a complex dielectric function $\varepsilon(\omega)$, the external sources $\mathbf{j}_0$ and ρ_0 vanishes. Therefore, we have

$$\nabla \times \mathbf{E} = -\frac{\mathrm{i}\omega}{c}\mathbf{B}$$
$$\nabla \times \mathbf{B} = \frac{\mathrm{i}\omega\varepsilon(\omega)}{c}\mathbf{E}. \tag{8.63}$$

Later, we will consider both bulk and surface waves. We will take the normal to the surface as the z-direction and consider waves that propagate at some angle to the interface. There is no loss of generality in assuming that the wave vector $\mathbf{q} = (0, q_y, q_z)$, so that the $\mathbf{E}$ field is given by

$$\mathbf{E} = (E_x, E_y, E_z)\mathrm{e}^{\mathrm{i}(\omega t - q_y y - q_z z)}.$$

The operators ∇ and $\frac{\partial}{\partial t}$ become $-\mathrm{i}\mathbf{q}$ and $\mathrm{i}\omega$, and the two Maxwell equations for $\nabla \times \mathbf{E}$ and $\nabla \times \mathbf{B}$ can be combined to give

$$\begin{aligned}\nabla \times (\nabla \times \mathbf{E}) &= -\tfrac{\mathrm{i}\omega}{c}\nabla \times \mathbf{B} = -\tfrac{\mathrm{i}\omega}{c}\left(\tfrac{\mathrm{i}\omega\varepsilon}{c}\mathbf{E}\right) \\ &= \nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E}.\end{aligned}$$

This can be rewritten

$$\left(\frac{\omega^2}{c^2}\varepsilon(\omega) - q^2\right)\mathbf{E} + \mathbf{q}\,(\mathbf{q}\cdot\mathbf{E}) = 0. \tag{8.64}$$

This vector equation can be expressed as a matrix equation

$$\begin{pmatrix} \frac{\omega^2}{c^2}\varepsilon(\omega) - q^2 & 0 & 0 \\ 0 & \frac{\omega^2}{c^2}\varepsilon(\omega) - q_z^2 & q_y q_z \\ 0 & q_y q_z & \frac{\omega^2}{c^2}\varepsilon(\omega) - q_y^2 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = 0. \tag{8.65}$$

8.10 Bulk Modes

For an infinite homogeneous medium of dielectric function $\varepsilon(\omega)$, the nontrivial solutions are obtained by setting the determinant of the 3×3 matrix (multiplying the column vector $\mathbf{E}$ in (8.65)) equal to zero. This gives

$$\varepsilon(\omega)\left[\frac{\omega^2}{c^2}\varepsilon(\omega) - q^2\right]^2 = 0. \tag{8.66}$$

There are two *transverse modes* satisfying

$$\omega^2 = \frac{c^2 q^2}{\varepsilon(\omega)}. \tag{8.67}$$

One of these has $q_y E_y + q_z E_z \equiv \mathbf{q}\cdot\mathbf{E} = 0;\ E_x = 0$. The other has $E_x \neq 0$ and $E_y = E_z = 0$. The other mode is longitudinal and has $E_x = 0$ and $\mathbf{q} \parallel \mathbf{E}$ or $\frac{E_z}{E_y} = \frac{q_z}{q_y}$, and has the dispersion relation

$$\varepsilon(\omega) = 0. \tag{8.68}$$

First, let us look at *longitudinal modes*.

8.10.1 Longitudinal Modes

Longitudinal modes, as we have seen, are given by the zeros of the dielectric function $\varepsilon(\omega)$. For simplicity we will neglect collisions and set $\tau \to \infty$ in the various dielectric functions we have studied.

Bound Electrons

We found that (for $\tau \to \infty$)

$$\varepsilon(\omega) = 1 + \frac{4\pi N e^2/m}{\omega_0^2 - \omega^2}. \tag{8.69}$$

We have seen the quantity $4\pi N e^2/m$ before. It is just the square of the plasma frequency ω_p of a system of N electrons per unit volume. The zero of $\varepsilon(\omega)$ occurs at

$$\omega^2 = \omega_0^2 + \omega_\mathrm{p}^2 \equiv \Omega^2. \tag{8.70}$$

Free Electrons

For free electrons $\omega_0 = 0$. Therefore, the longitudinal mode (plasmon) occurs at

$$\omega^2 = \omega_\mathrm{p}^2. \tag{8.71}$$

Polar Crystal

For a polar dielectric, the dielectric function is given by

$$\varepsilon(\omega) = \varepsilon_\infty \frac{\omega^2 - \omega_\mathrm{L}^2}{\omega^2 - \omega_\mathrm{T}^2}. \tag{8.72}$$

The longitudinal mode occurs at $\omega = \omega_\mathrm{L}$, the longitudinal optical phonon frequency.

Degenerate Polar Semiconductor

For a polar semiconductor containing N free electrons per unit volume in the conduction band

$$\varepsilon(\omega) = \varepsilon_\infty \left(\frac{\omega^2 - \omega_\mathrm{L}^2}{\omega^2 - \omega_\mathrm{T}^2} \right) - \frac{\omega_\mathrm{p}^2}{\omega^2}. \tag{8.73}$$

This can be written as

$$\varepsilon(\omega) = \varepsilon_\infty \frac{(\omega^2 - \omega_+^2)(\omega^2 - \omega_-^2)}{\omega^2(\omega^2 - \omega_\mathrm{T}^2)}. \tag{8.74}$$

Here, $\omega_\pm^2$ are two solutions of the quadratic equation

$$\omega^4 - \omega^2(\omega_\mathrm{L}^2 + \tilde{\omega}_\mathrm{p}^2) + \omega_\mathrm{T}^2 \tilde{\omega}_\mathrm{p}^2 = 0, \tag{8.75}$$

where $\tilde{\omega}_\mathrm{p}^2 = \frac{\omega_\mathrm{p}^2}{\varepsilon_\infty}$ with background dielectric constant ε_∞. The modes are *coupled plasmon–LO phonon modes*. One can see where these two modes occur by plotting $\varepsilon(\omega)$ vs. ω (see Fig. 8.11).

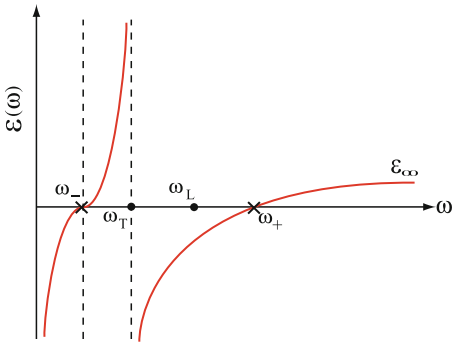


Fig. 8.11. Frequency dependence of the dielectric function $\varepsilon(\omega)$ for a degenerate polar semiconductor

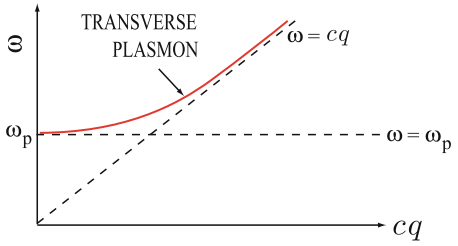


Fig. 8.12. Dispersion relations of the transverse modes in a metal

8.10.2 Transverse Modes

For transverse waves $\omega^2 = \frac{c^2q^2}{\varepsilon(\omega)}$. Again we will take the limit $\tau \to \infty$.

Dielectric

For a dielectric $\varepsilon(\omega)$ is a constant ε_0 independent of frequency. Thus, we have

$$\omega = \frac{c}{\sqrt{\varepsilon_0}} q. \tag{8.76}$$

Here, $\sqrt{\varepsilon_0}$ is called the *refractive index* n, and the velocity of light in the medium is $\frac{c}{n}$. paragraphMetal For a metal $\varepsilon = 1 - \frac{\omega_p^2}{\omega^2}$, giving

$$\omega^2 = \omega_p^2 + c^2q^2. \tag{8.77}$$

No transverse waves propagate for $\omega < \omega_p$ since $\varepsilon(\omega)$ is negative there (see Fig. 8.12).

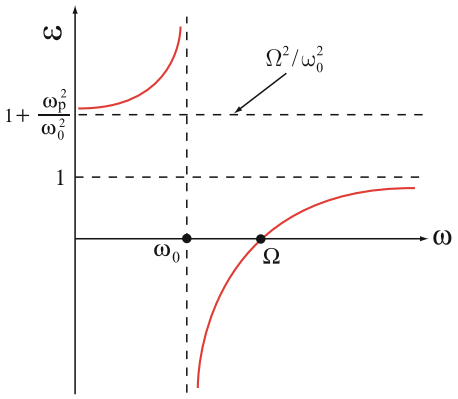


Fig. 8.13. Frequency dependence of the dielectric function $\varepsilon(\omega)$ for bound electrons

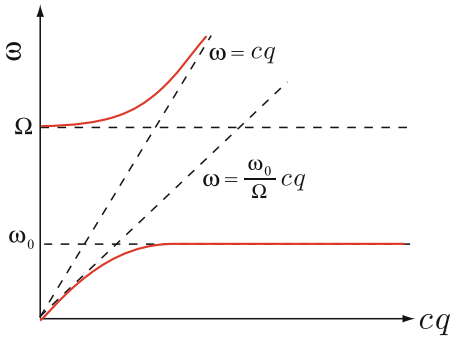


Fig. 8.14. Dispersion relation of the transverse modes for bound electrons

Bound Electrons

For bound electrons $\varepsilon = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2} = \frac{\Omega^2 - \omega^2}{\omega_0^2 - \omega^2}$ giving for the transverse mode

$$c^2 q^2 = \omega^2 \left(\frac{\Omega^2 - \omega^2}{\omega_0^2 - \omega^2} \right). \tag{8.78}$$

It is worth sketching $\varepsilon(\omega)$ versus ω (see Fig. 8.13).
The dielectric function is negative for $\omega_0 < \omega < \Omega$. The dispersion relation of the transverse mode for bound electrons given by (8.78) is sketched in Fig. 8.14. No transverse modes propagate where $\varepsilon(\omega) < 0$ since $c^2 q^2 = \omega^2 \varepsilon$ gives imaginary values of q in that region.

Polar Dielectric

In this case $\varepsilon(\omega) = \varepsilon_\infty \frac{\omega^2 - \omega_L^2}{\omega^2 - \omega_T^2}$ and again it is worth plotting $\varepsilon(\omega)$ versus ω. It is shown in Fig. 8.15. $\varepsilon(\omega)$ is negative in the region $\omega_T < \omega < \omega_L$. Plotting

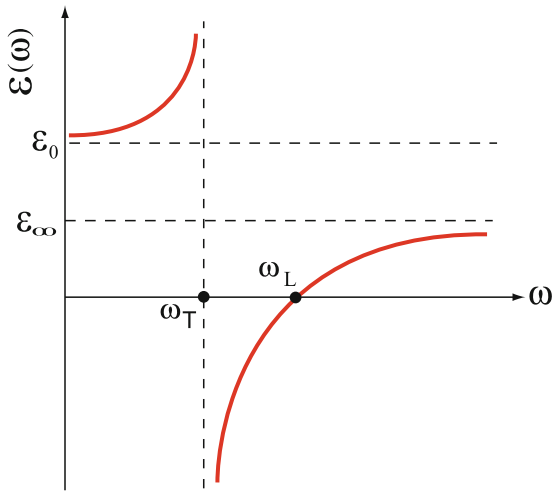


Fig. 8.15. Frequency dependence of the dielectric function $\varepsilon(\omega)$ in a polar dielectric

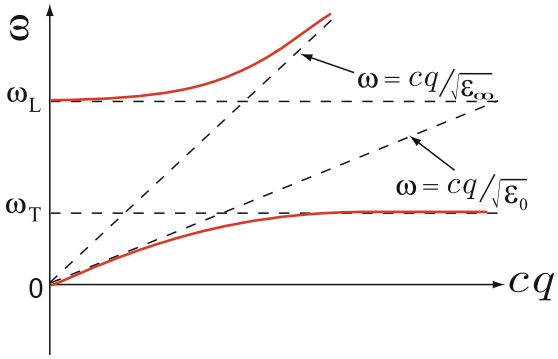


Fig. 8.16. Dispersion relation of the transverse modes in a polar dielectric

ω versus cq we have result shown in Fig. 8.16. There is a region between ω_{T} and ω_{L} where no light propagates the (*reststrahlen region*). The propagating modes are referred to as *polariton modes*. They are linear combinations of transverse phonon and electromagnetic modes.

Degenerate Polar Semiconductor

Here $\varepsilon(\omega) = \varepsilon_\infty \frac{\omega^2-\omega_{\mathrm{L}}^2}{\omega^2-\omega_{\mathrm{T}}^2} - \frac{\omega_{\mathrm{p}}^2}{\omega^2}$. We have already shown in (8.74) that this can be written in terms of ω_+ and ω_-. The equation for a transverse mode becomes

$$c^2q^2 = \varepsilon_\infty \frac{(\omega^2-\omega_+^2)(\omega^2-\omega_-^2)}{\omega^2-\omega_{\mathrm{T}}^2}. \tag{8.79}$$

In Fig. 8.11 the dielectric function $\varepsilon(\omega)$ was plotted as a function of ω in order to study the longitudinal modes. There, we found that $\varepsilon(\omega)$ was negative if

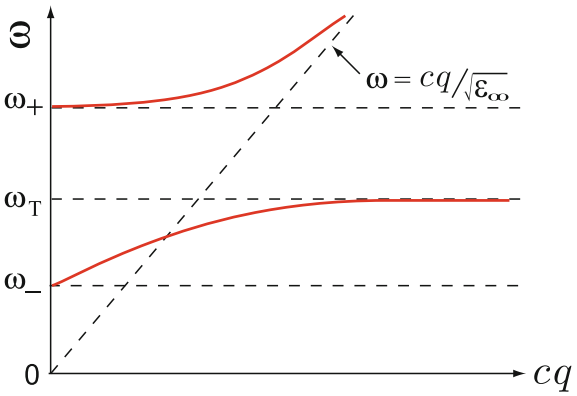


Fig. 8.17. Dispersion relation of the transverse modes in a polar semiconductor

$\omega < \omega_-$ or if $\omega_T < \omega < \omega_+$. The dispersion curve for the transverse mode is displayed in Fig. 8.17.

8.11 Reflectivity of a Solid

A vacuum–solid interface is located at $z = 0$. The solid ($z > 0$) is described by a dielectric function $\varepsilon(\omega)$ and vacuum ($z < 0$) by $\varepsilon = 1$. An incident light wave of frequency ω propagates in the z-direction as shown in Fig. 8.18.

Here we take the wave to be polarized in the $\hat{y}$-direction and $q_0 = \frac{\omega}{c}$ while $q = \sqrt{\varepsilon(\omega)}q_0$. There are three waves to consider:

1. The incident wave whose electric field is given by

$$\mathbf{E}_\mathrm{I} = \hat{y}E_\mathrm{I}\mathrm{e}^{\mathrm{i}(\omega t - q_0 z)}.$$

2. The reflected wave whose electric field is given by

$$\mathbf{E}_\mathrm{R} = \hat{y}E_\mathrm{R}\mathrm{e}^{\mathrm{i}(\omega t + q_0 z)}. \tag{8.80}$$

3. The transmitted wave whose electric field is given by

$$\mathbf{E}_\mathrm{T} = \hat{y}E_\mathrm{T}\mathrm{e}^{\mathrm{i}(\omega t - q z)}.$$

Because $\nabla \times \mathbf{E} = -\frac{1}{c}\dot{\mathbf{B}}$, $\mathbf{B} = \frac{cq_i}{\omega} \times \mathbf{E}$, where $q_i = q_0$ for the wave in vacuum and $q_i = q$ for the wave in the solid. We can easily see that

$$\begin{aligned}
\mathbf{B}_\mathrm{I} &= -\hat{x}E_\mathrm{I}\mathrm{e}^{\mathrm{i}(\omega t - q_0 z)}, \\
\mathbf{B}_\mathrm{R} &= \hat{x}E_\mathrm{R}\mathrm{e}^{\mathrm{i}(\omega t + q_0 z)}, \\
\mathbf{B}_\mathrm{T} &= -\hat{x}\sqrt{\varepsilon(\omega)}E_\mathrm{T}\mathrm{e}^{\mathrm{i}(\omega t - q z)}.
\end{aligned} \tag{8.81}$$

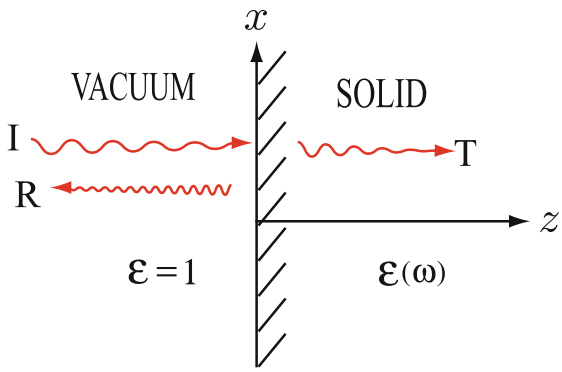


Fig. 8.18. Reflection of light wave at the interface of vacuum and solid of dielectric function $\varepsilon(\omega)$

The boundary conditions at $z = 0$ are continuity of $\mathbf{E}$ and $\mathbf{B}$. This gives

$$E_{\mathrm{I}} + E_{\mathrm{R}} = E_{\mathrm{T}}, \qquad -E_{\mathrm{I}} + E_{\mathrm{R}} = -\varepsilon^{1/2} E_{\mathrm{T}}. \tag{8.82}$$

Solving these equations for $\frac{E_{\mathrm{R}}}{E_{\mathrm{I}}}$ gives

$$\frac{E_{\mathrm{R}}}{E_{\mathrm{I}}} = \frac{1 - \sqrt{\varepsilon(\omega)}}{1 + \sqrt{\varepsilon(\omega)}}. \tag{8.83}$$

8.11.1 Optical Constants

The refractive index $n(\omega)$ and extinction coefficient $k(\omega)$ are real functions of frequency defined by

$$\varepsilon(\omega) = (n + \mathrm{i}k)^2. \tag{8.84}$$

Therefore, the *reflection coefficient*, defined by $R = |E_{\mathrm{R}}/E_{\mathrm{I}}|^2$ is given by

$$R = \frac{(1-n)^2 + k^2}{(1+n)^2 + k^2}. \tag{8.85}$$

The power absorbed is given by $P = \Re(\mathbf{j} \cdot \mathbf{E})$. But $\mathbf{j} = \sigma\mathbf{E}$ and $\sigma = \frac{\mathrm{i}\omega}{4\pi}(\varepsilon - 1)$. Therefore the power absorbed is proportional to $\frac{\omega}{4\pi}\varepsilon_2(\omega)|E|^2$. But the imaginary part of $\varepsilon(\omega)$ is just $2nk$, so that

$$\text{Power absorbed } \propto n(\omega)\, k(\omega). \tag{8.86}$$

8.11.2 Skin Effect

We have seen that for a metal $\varepsilon(\omega)$ is given by

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega - \mathrm{i}/\tau)} = 1 - \frac{\omega_p^2}{\omega^2 + 1/\tau^2} - \mathrm{i}\frac{\omega_p^2/\omega\tau}{\omega^2 + 1/\tau^2}. \tag{8.87}$$

At optical frequencies ω is usually large compared to $\frac{1}{\tau}$. Therefore, the real part of $\varepsilon(\omega)$ is large compared to the imaginary part; however, it is negative. $\varepsilon_1(\omega) \simeq -\omega_p^2/\omega^2$, since ω_p is large compared to optical frequencies for most metals. The wave vector of the transmitted wave is

$$q = \frac{\omega}{c}\varepsilon^{1/2} \simeq \frac{\omega}{c}\sqrt{\frac{-\omega_p^2}{\omega^2}} = \pm \mathrm{i}\frac{\omega_p}{c}$$

Thus, the wave

$$\mathbf{E}_\mathrm{T} = E_\mathrm{T}\hat{y}\ \mathrm{e}^{\mathrm{i}(\omega t - qz)}$$

decays with increasing z as $\mathrm{e}^{-z/\lambda}$ where $\lambda = c/\omega_p$ is called the *skin depth*. For $\omega_p \simeq 10^{16}\mathrm{sec}^{-1}$, $\lambda \approx 30\,\mathrm{nm}$. In a metal, light only penetrates this distance. This analysis assumed that $\mathbf{j}(\mathbf{r}) = \sigma\mathbf{E}(\mathbf{r})$, a local relationship between $\mathbf{j}$ and $\mathbf{E}$. If the mean free path $l = v_\mathrm{F}\tau$ is larger than λ, the skin depth, this assumption is not valid. Then one must use a more sophisticated analysis; this is referred to as the *anomalous skin effect*.

8.12 Surface Waves

In solving the equations describing the propagation of electromagnetic waves in an infinite medium, we considered the wave vector $\mathbf{q}$, which satisfied the relation

$$q^2 = \frac{\omega^2}{c^2}\varepsilon(\omega) \tag{8.88}$$

to have components q_y and q_z which were real.

At a surface ($z = 0$) separating two different dielectrics it is possible to have solutions for which q_z^2 is negative in one or both of the media. If q_z^2 is negative in both media, implying that q_z itself is imaginary, such solutions describe *surface waves*.

Let us look at the system shown in Fig. 8.19. The wave equation can be written

$$q_{zi}^2 = \frac{\omega^2}{c^2}\varepsilon_i - q_y^2, \tag{8.89}$$

where $i = \mathrm{I}$ or II. We think of ω and q_y as given and the same in each medium. Then the wave equation tells us the value of q_z^2 in each medium.

Let us assume a p-polarized wave (the s-polarization in which $\mathbf{E}$ is parallel to the surface does not usually give surface waves). We take

$$\mathbf{E} = (0, E_y, E_z)\ \mathrm{e}^{\mathrm{i}(\omega t - q_y y - q_z z)}. \tag{8.90}$$

Because there is no charge density except at the surface $z = 0$, we have $\nabla \cdot \mathbf{E} = \mathbf{q} \cdot \mathbf{E} = 0$ everywhere except at the surface. This implies that

$$q_y E_y + q_z E_z = 0, \tag{8.91}$$

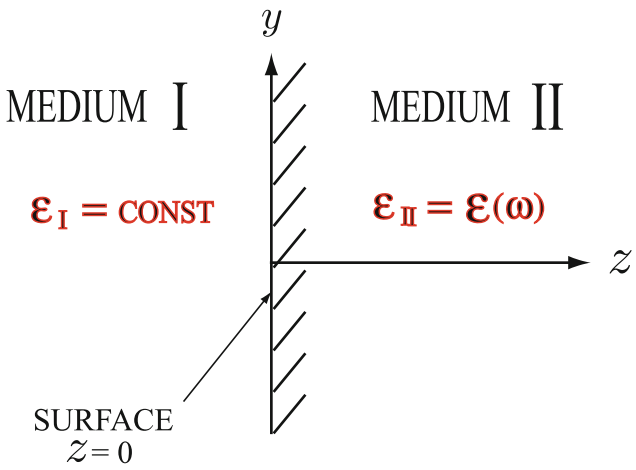


Fig. 8.19. The interface of two different media of dielectric functions ε_{I} and $\varepsilon_{\mathrm{II}}$

giving the value of E_z in each medium in terms of q_y, q_{zi}, and E_y. We take

$$\begin{aligned} E_{\mathrm{I}} &= (0, E_{y\mathrm{I}}, E_{z\mathrm{I}})\ \mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}q_y y + \alpha_{\mathrm{I}} z},\ \alpha_{\mathrm{I}}^2 = -q_{z\mathrm{I}}^2 \\ E_{\mathrm{II}} &= (0, E_{y\mathrm{II}}, E_{z\mathrm{II}})\ \mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}q_y y - \alpha_{\mathrm{II}} z},\ \alpha_{\mathrm{II}}^2 = -q_{z\mathrm{II}}^2. \end{aligned} \tag{8.92}$$

Since

$$\begin{aligned} \mathrm{e}^{-\mathrm{i}q_{z\mathrm{I}} z} &= \mathrm{e}^{\alpha_{\mathrm{I}} z}, q_{z\mathrm{I}} = \mathrm{i}\alpha_{\mathrm{I}} \\ \mathrm{e}^{-\mathrm{i}q_{z\mathrm{II}} z} &= \mathrm{e}^{-\alpha_{\mathrm{II}} z}, q_{z\mathrm{II}} = -\mathrm{i}\alpha_{\mathrm{II}}. \end{aligned}$$

Here α_{I} and α_{II} are positive and the form of $E(z)$ has been chosen to vanish as $|z| \to \infty$.

Since $E_{z_i} = -\frac{q_y}{q_{z_i}} E_y$, we obtain

$$E_{z\mathrm{I}} = -\frac{q_y}{\mathrm{i}\alpha_{\mathrm{I}}} E_{y\mathrm{I}}, \qquad E_{z\mathrm{II}} = -\frac{q_y}{-\mathrm{i}\alpha_{\mathrm{II}}} E_{y\mathrm{II}} \tag{8.93}$$

The boundary conditions are

$$\begin{aligned} &\text{(i)} \quad E_{y\mathrm{I}}(0) = E_{y\mathrm{II}}(0) = E_y(0) \\ &\text{(ii)} \quad \varepsilon_{\mathrm{I}} E_{z\mathrm{I}}(0) = \varepsilon_{\mathrm{II}} E_{z\mathrm{II}}(0) \end{aligned} \tag{8.94}$$

These conditions give us the dispersion relation of the surface wave. Substituting fields given by (8.93) into the second expression of (8.94), we have

$$\frac{\varepsilon_{\mathrm{I}}}{\alpha_{\mathrm{I}}} + \frac{\varepsilon_{\mathrm{II}}}{\alpha_{\mathrm{II}}} = 0 \qquad \text{or} \qquad \frac{\varepsilon_o}{\alpha_o} + \frac{\varepsilon(\omega)}{\alpha} = 0. \tag{8.95}$$

where

$$\alpha_{\mathrm{o}} = \sqrt{q_y^2 - \frac{\omega^2}{c^2}\varepsilon_{\mathrm{o}}} \qquad \text{and} \qquad \alpha = \sqrt{q_y^2 - \frac{\omega^2}{c^2}\varepsilon(\omega)}. \tag{8.96}$$

Non-retarded Limit

This is the case where $cq_y \gg \omega$ or c may be taken as infinite. In this limit, we have $\alpha_{\mathrm{I}} \simeq \alpha_{\mathrm{II}} \simeq q_y$ and the dispersion relation becomes

$$\varepsilon_o + \varepsilon(\omega) = 0. \tag{8.97}$$

Surface Polaritons (with Retardation Effects)

We take the dispersion relation given by (8.95) and square both sides. This gives

$$\varepsilon_o^2 \alpha^2 = \varepsilon^2(\omega)\alpha_0^2,$$

or

$$\varepsilon_o^2 \left(q_y^2 - \frac{\omega^2}{c^2}\varepsilon(\omega) \right) = \varepsilon^2(\omega) \left(q_y^2 - \frac{\omega^2}{c^2}\varepsilon_{\mathrm{o}} \right).$$

This can be simplified to

$$\left[\varepsilon_{\mathrm{o}} + \varepsilon(\omega)\right] q_y^2 = \frac{\omega^2}{c^2}\varepsilon_{\mathrm{o}}\varepsilon(\omega)$$

or

$$c^2 q_y^2 = \frac{\omega^2 \varepsilon_{\mathrm{o}} \varepsilon(\omega)}{\varepsilon_{\mathrm{o}} + \varepsilon(\omega)} \tag{8.98}$$

8.12.1 Plasmon

For a system containing n_0 free electrons of mass m

$$\varepsilon(\omega) = \varepsilon_{\mathrm{s}} - \frac{\omega_{\mathrm{p}}^2}{\omega^2}, \tag{8.99}$$

where ε_{s} is the background dielectric constant and $\omega_{\mathrm{p}}^2 = \frac{4\pi n_0 e^2}{m}$. In the non-retarded limit we find

$$\varepsilon_{\mathrm{o}} + \left(\varepsilon_{\mathrm{s}} - \frac{\omega_{\mathrm{p}}^2}{\omega^2} \right) = 0.$$

Solving for ω^2 gives the surface plasmon frequency

$$\omega_{\mathrm{SP}} = \frac{\omega_{\mathrm{p}}}{\sqrt{\varepsilon_{\mathrm{o}} + \varepsilon_{\mathrm{s}}}}. \tag{8.100}$$

Recall the bulk plasmon is given by $\varepsilon(\omega) = 0$

$$\omega_{\mathrm{BP}} = \frac{\omega_{\mathrm{p}}}{\sqrt{\varepsilon_{\mathrm{s}}}} > \omega_{\mathrm{SP}}. \tag{8.101}$$

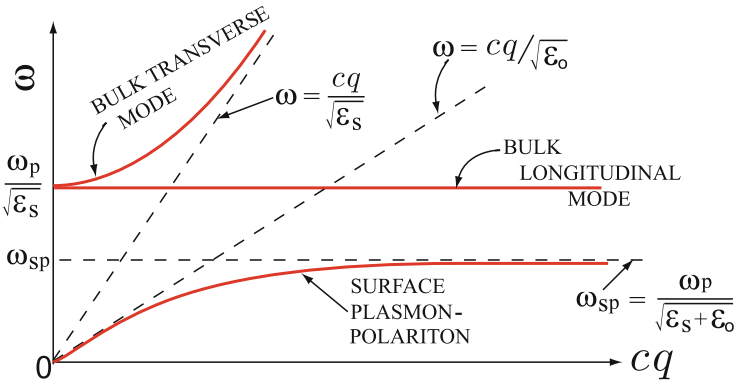


Fig. 8.20. Dispersion curves of the bulk and surface plasmon–polariton modes

So that we have $\omega_{\mathrm{SP}} < \omega_{\mathrm{BP}}$. For the surface plasmon–polariton we find

$$c^2 q_y^2 = \frac{\omega^2 \varepsilon_{\mathrm{o}} \left(\varepsilon_{\mathrm{s}} - \omega_{\mathrm{p}}^2/\omega^2\right)}{\varepsilon_{\mathrm{o}} + \varepsilon_{\mathrm{s}} - \omega_{\mathrm{p}}^2/\omega^2} \tag{8.102}$$

It is easy to see that, for very small values of cq_y, $\omega \to 0$ and we can neglect ε_{s} and $(\varepsilon_{\mathrm{o}} + \varepsilon_{\mathrm{s}})$ compared to $-\omega_{\mathrm{p}}^2/\omega^2$ on the right-hand side of (8.102).

$$\lim_{cq_y \to 0} \omega^2 \simeq \frac{c^2 q_y^2}{\varepsilon_{\mathrm{o}}}. \tag{8.103}$$

For very large cq_y, the only solution occurs when the denominator of the right hand side approaches zero.

$$\lim_{cq_y \to \infty} \omega^2 \simeq \frac{\omega_{\mathrm{p}}^2}{\varepsilon_{\mathrm{o}} + \varepsilon_{\mathrm{s}}} = \omega_{\mathrm{SP}}^2. \tag{8.104}$$

The dispersion curves of the bulk and surface plasmon–polariton modes are shown in Fig. 8.20.

8.12.2 Surface Phonon–Polariton

Here we take the dielectric function of a polar crystal

$$\varepsilon(\omega) = \varepsilon_\infty \frac{\omega^2 - \omega_{\mathrm{L}}^2}{\omega^2 - \omega_{\mathrm{T}}^2}. \tag{8.105}$$

At the interface of a dielectric ε_{I} and a polar material described by (8.105) the surface mode is given by (8.95)

$$\frac{\varepsilon_{\mathrm{I}}}{\alpha_{\mathrm{I}}} = -\frac{\varepsilon(\omega)}{\alpha}. \tag{8.106}$$

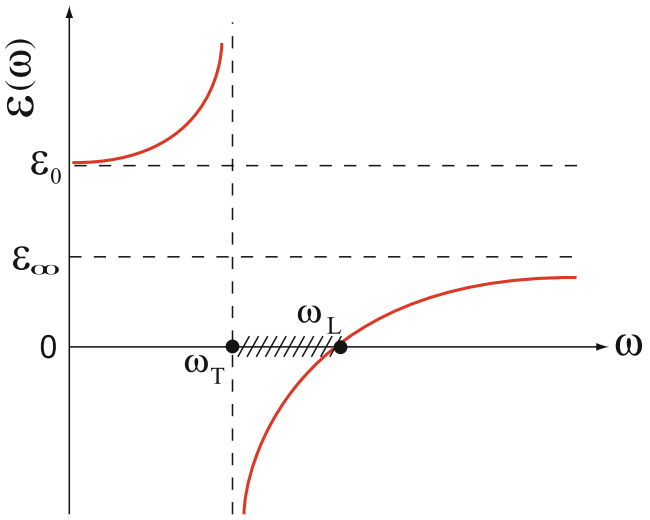


Fig. 8.21. Dielectric function $\varepsilon(\omega)$ of a polar crystal

Since ε_{I}, α_{I}, and α are all positive, this equation has a solution only in the region where $\varepsilon(\omega) < 0$. Recall that $\varepsilon(\omega)$ versus ω looks as shown in Fig. 8.21. $\varepsilon(\omega)$ is negative if $\omega_{\mathrm{T}} < \omega < \omega_{\mathrm{L}}$. The dispersion relation, (8.106), is written by

$$c^2 q_y^2 = \frac{\omega^2 \varepsilon_{\mathrm{I}} \varepsilon(\omega)}{\varepsilon_{\mathrm{I}} + \varepsilon(\omega)} = \frac{\varepsilon_{\mathrm{I}} \varepsilon_\infty \omega^2 (\omega^2 - \omega_{\mathrm{L}}^2)}{\varepsilon_{\mathrm{I}} (\omega^2 - \omega_{\mathrm{T}}^2) + \varepsilon_\infty (\omega^2 - \omega_{\mathrm{L}}^2)}. \tag{8.107}$$

The denominator can be written as

$$D \equiv (\varepsilon_{\mathrm{I}} + \varepsilon_\infty)\omega^2 - (\varepsilon_{\mathrm{I}} \omega_{\mathrm{T}}^2 + \varepsilon_\infty \omega_{\mathrm{L}}^2) = (\varepsilon_{\mathrm{I}} + \varepsilon_\infty)(\omega^2 - \omega_{\mathrm{s}}^2), \tag{8.108}$$

where the surface phonon frequency ω_{s} is given by

$$\omega_{\mathrm{s}}^2 = \frac{\varepsilon_{\mathrm{I}} \omega_{\mathrm{T}}^2 + \varepsilon_\infty \omega_{\mathrm{L}}^2}{\varepsilon_{\mathrm{I}} + \varepsilon_\infty}. \tag{8.109}$$

It is easy to show that $\omega_{\mathrm{T}}^2 < \omega_{\mathrm{s}}^2 < \omega_{\mathrm{L}}^2$. The dispersion relation can be written

$$c^2 q_y^2 = \frac{\varepsilon_{\mathrm{I}} \varepsilon_\infty \omega^2 (\omega^2 - \omega_{\mathrm{L}}^2)}{(\varepsilon_{\mathrm{I}} + \varepsilon_\infty)(\omega^2 - \omega_{\mathrm{s}}^2)}. \tag{8.110}$$

Figure 8.22 shows the right-hand side of (8.110) as a function of frequency. Since surface modes can occur only where $q_y^2 > 0$ and $\varepsilon(\omega) < 0$, we see that the surface mode is restricted to the frequency region $\omega_{\mathrm{T}} < \omega < \omega_{\mathrm{s}}$. It is not difficult to see that as $cq_y \to \infty$, the surface polariton approaches the frequency ω_{s}. It is also apparent that at $\omega = \omega_{\mathrm{T}}$, $c^2 q_y^2 = \varepsilon_{\mathrm{I}} \omega_{\mathrm{T}}^2$. This gives the dispersion curve sketched in Fig. 8.23.

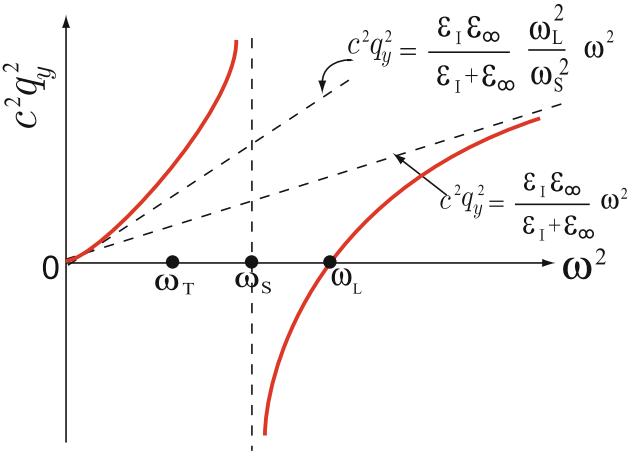


Fig. 8.22. Dispersion curves of phonon–polariton modes

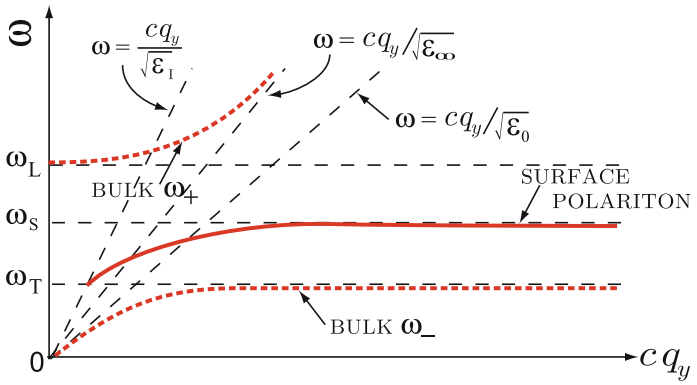


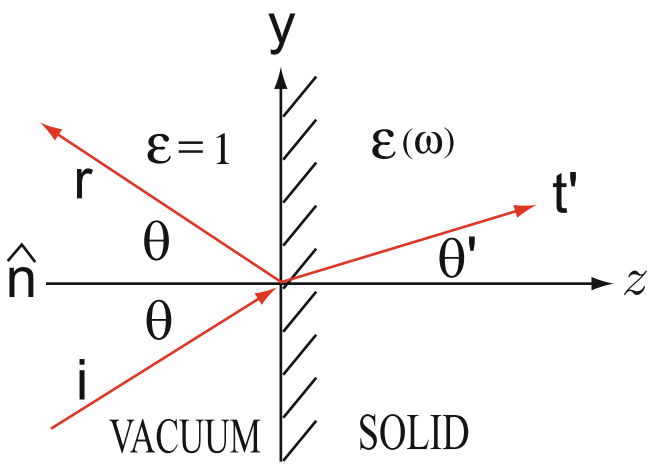
Fig. 8.23. Dispersion relation of surface phonon–polariton modes. Also shown are the bulk mode ω_+ and ω_- which can occur outside the region $\omega_T < \omega < \omega_L$

Problems

8.1. Use the eigenstates of the hydrogen atom to evaluate its atomic polarizability α.

8.2. If atomic hydrogen formed a cubic lattice, what would its static dielectric constant be?

8.3. Evaluate the reflectivity for an S-polarized and a P-polarized electromagnetic wave incident at an angle θ from vacuum on a material of dielectric function $\varepsilon(\omega)$ as illustrated in the figure below.



One can take $\mathbf{E} = (E_x, 0, 0)\mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}\mathbf{q}\cdot\mathbf{r}}$ and $\mathbf{E} = (0, E_y, E_z)\mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}\mathbf{q}\cdot\mathbf{r}}$ as the S- and P-polarized electric fields, respectively. Remember that $\mathbf{q}\cdot\mathbf{E} = 0$ and $\mathbf{q} = (0, q_y, q_z)$.

8.4. A degenerate polar semiconductor contains n_0 free electrons per unit volume in the conduction band. Its dielectric function $\varepsilon(\omega)$ is given by

$$\varepsilon(\omega) = \varepsilon_\infty \frac{\omega^2 - \omega_{\mathrm{L}}^2}{\omega^2 - \omega_{\mathrm{T}}^2} - \frac{\omega_{\mathrm{p}}^2}{\omega^2},$$

where ω_{L} and ω_{T} are the LO and TO phonon frequencies, and $\omega_{\mathrm{p}} = \sqrt{\frac{4\pi n_0 e^2}{m}}$.

1. Show that $\varepsilon(\omega)$ can be written as

$$\varepsilon(\omega) = \varepsilon_\infty \frac{(\omega^2 - \omega_-^2)(\omega^2 - \omega_+^2)}{\omega^2(\omega^2 - \omega_{\mathrm{T}}^2)}$$

and determine ω_-^2 and ω_+^2.

2. Make a sketch of $\varepsilon(\omega)$ versus ω; be sure to indicate the locations of ω_{T}, ω_{L}, ω_-, ω_+, ε_0, and ε_∞.
3. Determine the dispersion relation of the longitudinal and transverse modes, i.e. ω as a function of q.
4. In which regions of frequency are the transverse waves unable to propagate?
5. Consider a vacuum–degenerate polar semiconductor interface. Use the results obtained in the text to determine the dispersion relations of the surface modes.
6. Make a sketch of ω versus q_y (q_y is parallel to the interface) for these surface modes and for the bulk modes which have $q_z = 0$.

Summary

In this chapter, we studied dielectric properties of solids in the presence of an external electromagnetic disturbance. We first reviewed elementary electricity and magnetism, and introduced concept of local field inside a solid. Then dispersion relations of self-sustaining collective modes and reflectivity of a solid are studied for various situations. Finally, the collective modes localized near the surface of a solid are also described and dispersion relations of surface plasmon-polariton and surface phonon-polariton modes are discussed explicitly.

When an external electromagnetic disturbance is introduced into a solid, it will produce induced charge density and induced current density. These induced densities produce induced electric and magnetic fields. The *local field* $\mathbf{E}_{\mathrm{LF}}(\mathbf{r})$ at the position of an atom in a solid is given by

$$\mathbf{E}_{\mathrm{LF}} = \mathbf{E}_0 + \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3,$$

where E_0, E_1, E_2, E_3 are, respectively, the external field, *depolarization field* $(= -\lambda\mathbf{P})$, *Lorentz field* $(= \frac{4\pi\mathbf{P}}{3})$, and the field due to the dipoles within the Lorentz sphere $\left(= \sum_{i\in \mathrm{L.S.}} \frac{3(\mathbf{p}_i\cdot\mathbf{r}_i)\mathbf{r}_i - r_i^2\mathbf{p}_i}{r_i^5}\right)$. The local field at the center of a sphere of cubic crystal is simply given by

$$E_{\mathrm{LF}}^{\mathrm{sphere}} = E_0 - \frac{4\pi}{3}P + \frac{4\pi}{3}P + 0 = E_0.$$

The induced dipole moment of an atom is given by $\mathbf{p} = \alpha\mathbf{E}_{\mathrm{LF}}$. The polarization $\mathbf{P}$ is given, for a cubic crystal, by $\mathbf{P} = \frac{N\alpha}{1-\frac{4\pi N\alpha}{3}}\mathbf{E} \equiv \chi\mathbf{E}$, where N is the number of atoms per unit volume and χ is the electrical susceptibility. The electrical susceptibility and the dielectric function $(\varepsilon = 1 + 4\pi\chi)$ of the solid are

$$\chi = \frac{N\alpha}{1-\frac{4\pi N\alpha}{3}};\ \varepsilon = 1 + \frac{4\pi N\alpha}{1-\frac{4\pi N\alpha}{3}}.$$

The relation between the macroscopic dielectric function ε and the atomic polarizability α is called the *Clausius–Mossotti relation*:

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{4\pi N\alpha}{3}$$

The total polarizability of the atoms or ions within a unit cell can usually be separated into three parts:

1. *Electronic polarizability* α_{e}: The displacement of the electrons relative to the nucleus
2. *Ionic polarizability* α_{i}: The displacement of an ion itself with respect to its equilibrium position
3. *Dipolar polarizability* α_{dipole}: The orientation of any permanent dipoles by the electric field in the presence of thermal disorder.

In the presence of a field $\mathbf{E}$, the average dipole moment per unit volume is given by $\bar{p}_z = p\,\mathcal{L}\left(\frac{pE}{k_\mathrm{B}T}\right)$, where $\mathcal{L}(\xi)$ is the Langevin function. The dipolar polarizability α_{dipole} shows strong temperature dependence. The electronic polarizability α_e and the ionic polarizability α_{ion} are almost independent of temperature.

In a metal, the conduction electrons are free and the dielectric function becomes

$$\varepsilon(\omega) = 1 - \frac{4\pi N e^2/m}{\omega^2 - \mathrm{i}\omega/\tau} = 1 - \frac{\omega_p^2}{\omega^2 - \mathrm{i}\omega/\tau}.$$

In an ionic crystal, we have

$$\varepsilon(\omega) = \varepsilon_\infty \left[\frac{\omega^2 - \omega_\mathrm{L}^2}{\omega^2 - \omega_\mathrm{T}^2}\right].$$

Here, ω_L and ω_T are the TO and LO phonon frequencies, respectively. We note that $\omega_\mathrm{L} > \omega_\mathrm{T}$ since $\varepsilon_0 > \varepsilon_\infty$ in general.

For the propagation of light in a material characterized by $\varepsilon(\omega)$, the external sources $\mathbf{j}_0$ and ρ_0 vanishes. Therefore, we have

$$\nabla \times \mathbf{E} = -\frac{\mathrm{i}\omega}{c}\mathbf{B};\ \nabla \times \mathbf{B} = \frac{\mathrm{i}\omega\varepsilon(\omega)}{c}\mathbf{E}.$$

The two Maxwell equations for $\nabla \times \mathbf{E}$ and $\nabla \times \mathbf{B}$ can be combined to give a wave equation:

$$\left(\frac{\omega^2}{c^2}\varepsilon(\omega) - q^2\right)\mathbf{E} + \mathbf{q}\,(\mathbf{q}\cdot\mathbf{E}) = 0.$$

For an infinite homogeneous medium of dielectric function $\varepsilon(\omega)$, a general dispersion relation of the self-sustaining waves is written as

$$\varepsilon(\omega)\left[\frac{\omega^2}{c^2}\varepsilon(\omega) - q^2\right]^2 = 0.$$

The two *transverse modes* and one *longitudinal mode* are characterized, respectively, by

$$\omega^2 = \frac{c^2 q^2}{\varepsilon(\omega)};\ \varepsilon(\omega) = 0.$$

For the interface ($z = 0$) of two different media of dielectric functions ε_I and $\varepsilon_{\mathrm{II}}$, the boundary conditions give us the general dispersion of the surface wave:

$$\frac{\varepsilon_\mathrm{I}}{\alpha_\mathrm{I}} + \frac{\varepsilon_{\mathrm{II}}}{\alpha_{\mathrm{II}}} = 0 \qquad \text{or} \qquad \frac{\varepsilon_o}{\alpha_o} + \frac{\varepsilon(\omega)}{\alpha} = 0$$

where

$$\alpha_\mathrm{o} = \sqrt{q_y^2 - \frac{\omega^2}{c^2}\varepsilon_\mathrm{o}} \qquad \text{and} \qquad \alpha = \sqrt{q_y^2 - \frac{\omega^2}{c^2}\varepsilon(\omega)}.$$

9

Magnetism in Solids

9.1 Review of Some Electromagnetism

9.1.1 Magnetic Moment and Torque

We begin with a brief review of some elementary electromagnetism. A current distribution $\mathbf{j}(\mathbf{r})$ produces a magnetic dipole moment at the origin that is given by

$$\mathbf{m} = \frac{1}{2c}\int \mathbf{r}\times\mathbf{j}(\mathbf{r})\mathrm{d}^3r. \tag{9.1}$$

If $\mathbf{j}(\mathbf{r})$ is composed from particles of charge q_i at positions $\mathbf{r}_i$ moving with velocity $\mathbf{v}_i$, $\mathbf{j}(\mathbf{r}) = \sum_i q_i\mathbf{v}_i\delta(\mathbf{r}-\mathbf{r}_i)$, and the magnetic moment $\mathbf{m}$ is

$$\mathbf{m} = \frac{1}{2c}\sum_i q_i\mathbf{r}_i\times\mathbf{v}_i. \tag{9.2}$$

For a single particle of charge q moving in a circle of radius r_0 at velocity v_0, we have

$$m = \frac{1}{2c}qr_0v_0 \tag{9.3}$$

and $\mathbf{m}$ is perpendicular to the plane of the circle. The current $\mathbf{i}$ in the loop is given by q divided by $t = \frac{2\pi r}{v_0}$, the time to complete one circuit. Thus

$$i = \frac{qv_0}{2\pi r_0}. \tag{9.4}$$

We can use this in our expression for m to get

$$m = \frac{q}{2c}r_0v_0 = \frac{ia}{c}. \tag{9.5}$$

Here, $a = \pi r_0^2$ is the area of the loop. We can write $\mathbf{m} = \frac{i\mathbf{a}}{c}$ if we associate vector character with the area $\mathbf{a}$ of the loop.

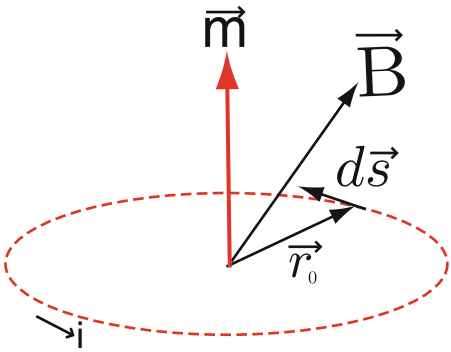


Fig. 9.1. A magnetic moment $\mathbf{m}$ due to a current loop of radius $\mathbf{r}_0$ located in a magnetic field $\mathbf{B}$

If a magnetic field were imposed on a magnetic moment, the magnetic moment would experience a torque. To show this we begin with the Lorentz force

$$\mathbf{F} = \frac{q}{c}\mathbf{v}_0 \times \mathbf{B}. \tag{9.6}$$

For a charge $\mathrm{d}q$ the force $\mathrm{d}\mathbf{F}$ is given by

$$\mathrm{d}\mathbf{F} = \frac{\mathrm{d}q}{c}\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} \times \mathbf{B} = \frac{i}{c}\mathrm{d}\mathbf{s} \times \mathbf{B}. \tag{9.7}$$

Here, $\mathrm{d}\mathbf{s}$ is an infinitesimal element of path length (see, for example, Fig. 9.1).

The torque τ is given by $\int \mathbf{r} \times \mathrm{d}\mathbf{F}$.

$$\tau = \int \mathbf{r} \times \mathrm{d}\mathbf{F} = \frac{i}{c}\int \mathbf{r} \times \mathrm{d}\mathbf{s} \times \mathbf{B}. \tag{9.8}$$

But $\int \mathbf{r} \times \mathrm{d}\mathbf{s} = \mathbf{a}$, and hence we have

$$\tau = \frac{i}{c}\mathbf{a} \times \mathbf{B} = \mathbf{m} \times \mathbf{B}. \tag{9.9}$$

9.1.2 Vector Potential of a Magnetic Dipole

If a magnetic dipole $\mathbf{m}$ is located at the origin, it produces a vector potential at position $\mathbf{r}$ given by

$$\mathbf{A}(\mathbf{r}) = \frac{\mathbf{m} \times \mathbf{r}}{r^3}. \tag{9.10}$$

Of course the magnetic field $\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}(\mathbf{r})$. If we have a magnetization (magnetic dipole moment per unit volume), then $\mathbf{A}(\mathbf{r})$ is given by

$$\mathbf{A}(\mathbf{r}) = \int \mathrm{d}^3 r' \frac{\mathbf{M}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}. \tag{9.11}$$

As we did with the electric polarization $\mathbf{P}(\mathbf{r})$, we can transform this equation into two parts.

$$\mathbf{A}(\mathbf{r}) = \int_{\mathrm{V}} \mathrm{d}^3 r' \frac{\nabla_{r'} \times \mathbf{M}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \oint_{\mathrm{S}} \mathrm{d}S' \frac{\mathbf{M}(\mathbf{r}') \times \hat{\mathbf{n}}'}{|\mathbf{r} - \mathbf{r}'|}, \tag{9.12}$$

where $\hat{\mathbf{n}}$ is a unit vector outward normal to the surface S. The volume integration is carried out over the volume V of the magnetized material. The surface integral is carried out over the surface bounding the magnetized object. Since $\mathbf{A}(\mathbf{r})$ is related to a current density by

$$\mathbf{A}(\mathbf{r}) = \frac{1}{c} \int \mathrm{d}^3 r' \frac{\mathbf{j}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \tag{9.13}$$

the vector potential produced by a magnetization is equivalent to volume distribution of current

$$\mathbf{j}_{\mathrm{M}}(\mathbf{r}) = c\nabla \times \mathbf{M}(\mathbf{r}) \tag{9.14}$$

and a surface distribution of current

$$\mathbf{j}_{\mathrm{S}}(\mathbf{r}) = c\mathbf{M}(\mathbf{r}) \times \hat{n}. \tag{9.15}$$

Recall that if $\mathbf{E} = 0$, Maxwell's equation for $\nabla \times \mathbf{B}$ is

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} (\mathbf{j}_0 + \mathbf{j}_{\mathrm{M}}) = \frac{4\pi}{c} \mathbf{j}_0 + 4\pi \nabla \times \mathbf{M}. \tag{9.16}$$

Defining $\mathbf{H} = \mathbf{B} - 4\pi\mathbf{M}$ gives

$$\nabla \times \mathbf{H} = \frac{4\pi}{c} \mathbf{j}_0 \tag{9.17}$$

which shows that $\mathbf{H}$ is that part of the field arising from the free current density $\mathbf{j}_0$. As we stated before the two Maxwell equations

$$\nabla \cdot \mathbf{E} = 4\pi \left(\rho_0 + \rho_{\mathrm{ind}}\right), \tag{9.18}$$

and

$$\nabla \times \mathbf{B} = \frac{1}{c} \dot{\mathbf{E}} + \frac{4\pi}{c} \left(\mathbf{j}_0 + \mathbf{j}_{\mathrm{ind}}\right) + 4\pi \nabla \times \mathbf{M} \tag{9.19}$$

are sometimes written in terms of $\mathbf{D}$ and $\mathbf{H}$.

$$\nabla \cdot \mathbf{D} = 4\pi\rho_0,$$

(where $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$ and $\nabla \cdot \mathbf{P} = -\rho_{\mathrm{pol}}$ with bound charge density ρ_{pol}) and

$$\nabla \times \mathbf{H} = \frac{1}{c} \dot{\mathbf{D}} + \frac{4\pi}{c} \mathbf{j}_0.$$

9.2 Magnetic Moment of an Atom

9.2.1 Orbital Magnetic Moment

Let us consider the nucleus to be fixed and evaluate the orbital contribution of the electron currents to the magnetic moment of an atom.

$$\mathbf{m} = \frac{1}{2c}\sum_i q_i \mathbf{r}_i \times \mathbf{v}_i. \tag{9.20}$$

Since $q_i = -e$ for every electron, and every electron has mass m_e, we can write

$$\mathbf{m} = -\frac{e}{2m_\mathrm{e}c}\sum_i \mathbf{r}_i \times m_\mathrm{e}\mathbf{v}_i = -\frac{e}{2m_\mathrm{e}c} \times \begin{pmatrix}\text{total angular momentum}\\ \text{of the electrons}\end{pmatrix}. \tag{9.21}$$

We know $\sum_i \mathbf{r}_i \times m_\mathrm{e}\mathbf{v}_i$ is quantized and equal to $\hbar\mathbf{L}$, where $|\mathbf{L}| = 0, 1, 2, \ldots$ and $L_z = 0, \pm 1, \pm 2, \ldots, \pm L$. Thus, we have

$$\mathbf{m} = -\frac{e\hbar}{2m_\mathrm{e}c}\mathbf{L} = -\mu_\mathrm{B}\mathbf{L}. \tag{9.22}$$

Here, $\mu_\mathrm{B} = \frac{e\hbar}{2m_\mathrm{e}c} = 0.927 \times 10^{-20}$(ergs/gauss) or 5.8×10^{-2} (meV/T) is called the **Bohr magneton**. The Bohr magneton corresponds to the magnetic moment of a $1s$ electron in H.

9.2.2 Spin Magnetic Moment

In addition to orbital angular momentum $\hbar\mathbf{L}$, each electron in an atom has an intrinsic spin angular momentum $\hbar\mathbf{s}$, giving a total spin angular momentum $\hbar\mathbf{S}$ where

$$\mathbf{S} = \sum_i \mathbf{s}_i. \tag{9.23}$$

The z-component of spin is $s_z = \pm\frac{1}{2}$, and the spin contribution to the magnetic moment is $\mp\mu_\mathrm{B}$. Thus, for each electron, there is a contribution $-2\mu_\mathrm{B}\mathbf{s}$ to the magnetic moment of an atom. If we sum over all spins, the total spin contribution to the magnetic moment is

$$\mathbf{m}_\mathrm{s} = -2\mu_\mathrm{B}\sum_i \mathbf{s}_i = -2\mu_\mathrm{B}\mathbf{S}. \tag{9.24}$$

Note that the factor of 2 appearing in this expression is not exact. It is actually given by $g = 2(1 + \frac{\alpha}{2\pi} - 2.973\frac{\alpha^2}{\pi^2} + \cdots) \simeq 2 \times 1.0011454$. However, in our discussion here we will take the g-factor as 2.

9.2.3 Total Angular Momentum and Total Magnetic Moment

The total angular momentum of an atom is given by

$$\mathbf{J} = \mathbf{L} + \mathbf{S}. \tag{9.25}$$

The total magnetic moment is given by

$$\mathbf{m} = -\mu_{\mathrm{B}} (\mathbf{L} + 2\mathbf{S}) . \tag{9.26}$$

In quantum mechanics the components of $\mathbf{J}$, $\mathbf{L}$, and $\mathbf{S}$ are operators that satisfy commutation relations. As we learned in quantum mechanics, it is possible to diagonalize J^2 and J_z simultaneously.

$$J^2 |j, j_z\rangle = j(j+1) |j, j_z\rangle ; \ j = 0, \frac{1}{2}, \frac{3}{2}, \ldots \tag{9.27}$$

$$J_z |j, j_z\rangle = j_z |j, j_z\rangle ; \ -j \leq j_z \leq j \tag{9.28}$$

Note that $j_z = 0, \pm 1, \ldots, \pm j$ or $j_z = \pm\frac{1}{2}, \pm\frac{3}{2}, \ldots, \pm j$. We can write that

$$\mathbf{m} = -\hat{g}\mu_{\mathrm{B}}\mathbf{J}. \tag{9.29}$$

This defines the operator $\hat{g}$ because we have $\mathbf{J} = \mathbf{L} + \mathbf{S}$ and

$$\hat{g}\mathbf{J} = \mathbf{L} + 2\mathbf{S}. \tag{9.30}$$

We can use these definitions to show that

$$\mathbf{J} \cdot \mathbf{J} = (\mathbf{L} + \mathbf{S}) \cdot (\mathbf{L} + \mathbf{S}) = \mathbf{L}^2 + \mathbf{S}^2 + 2\mathbf{L} \cdot \mathbf{S} \tag{9.31}$$

and

$$\hat{g}\mathbf{J} \cdot \mathbf{J} = (\mathbf{L} + \mathbf{S}) \cdot (\mathbf{L} + 2\mathbf{S}) = \mathbf{L}^2 + 2\mathbf{S}^2 + 3\mathbf{L} \cdot \mathbf{S}. \tag{9.32}$$

We can eliminate $\mathbf{L} \cdot \mathbf{S}$ and obtain

$$g_{\mathrm{L}} = \frac{3}{2} + \frac{1}{2}\frac{s(s+1) - l(l+1)}{j(j+1)} \tag{9.33}$$

This eigenvalue of $\hat{g}$ is called the **Landé g-factor**.

9.2.4 Hund's Rules

The ground state of an atom or ion with an incomplete shell is determined by Hund's rules:

1. The ground state has the maximum S consistent with the Pauli exclusion principle.
2. It has the maximum L consistent with the maximum spin multiplicity $2S + 1$ of Rule 1.
3. The J-value is given by $L - S$ when a shell is less than half filled and by $L + S$ when more than half filled.

Table 9.1. Ground state electron configurations angular momentum quantum numbers for the elements of atomic numbers $20 \leq Z \leq 29$

Z	Element	Configuration	Spectroscopic notation	S	L	J	g
20	Ca	$(3p)^6(4s)^2$	${}^1S_0$	0	0	0	–
21	Sc	$(3d)^1(4s)^2$	${}^2D_{\frac{3}{2}}$	$\frac{1}{2}$	2	$\frac{3}{2}$	$\frac{4}{5}$
22	Ti	$(3d)^2(4s)^2$	${}^3F_2$	1	3	2	$\frac{2}{3}$
23	V	$(3d)^3(4s)^2$	${}^4F_{\frac{3}{2}}$	$\frac{3}{2}$	3	$\frac{3}{2}$	$\frac{2}{5}$
24	Cr	$(3d)^5(4s)^1$	${}^7S_3$	3	0	3	2
25	Mn	$(3d)^5(4s)^2$	${}^6S_{\frac{5}{2}}$	$\frac{5}{2}$	0	$\frac{5}{2}$	2
26	Fe	$(3d)^6(4s)^2$	${}^5D_4$	2	2	4	$\frac{3}{2}$
27	Co	$(3d)^7(4s)^2$	${}^4F_{\frac{9}{2}}$	$\frac{3}{2}$	3	$\frac{9}{2}$	$\frac{4}{3}$
28	Ni	$(3d)^8(4s)^2$	${}^3F_4$	1	3	4	$\frac{5}{4}$
29	Cu	$(3d)^{10}(4s)^1$	${}^2S_{\frac{1}{2}}$	$\frac{1}{2}$	0	$\frac{1}{2}$	2

Example

Consider an ion of Fe^{2+}; it has six electrons in the $3d$ level. We can put five of them in spin up states (since d means $l = 2$ and m_l can be $-2, -1, 0, 1, 2$) and to maximize S, hence,

$$\uparrow \uparrow \uparrow \uparrow \uparrow \downarrow \text{ gives } S = 2.$$

The maximum value of L-value is given by

$$L = -2 - 1 + 0 + 1 + 2 + 2 = 2.$$

The J-value (since it is over half-filled) is

$$J = L + S = 4.$$

Therefore, we have

$$g = \frac{3}{2} + \frac{1}{2}\frac{2(3) - 2(3)}{4(5)} = \frac{3}{2}.$$

One can work out some examples listed in Table 9.1. The ground state notation is ${}^{2S+1}L_J$, where $L = 0, 1, 2, 3, 4, \ldots$ are denoted by the letters $S, P, D, F, G, \ldots$, respectively.

9.3 Paramagnetism and Diamagnetism of an Atom

In the presence of a magnetic field $\mathbf{B}$ the Hamiltonian describing the electrons in an atom can be written as

$$H = H_0 + \sum_i \frac{1}{2m}\left(\mathbf{p}_i + \frac{e}{c}\mathbf{A}(\mathbf{r}_i)\right)^2 + 2\mu_B \mathbf{B} \cdot \sum_i \mathbf{s}_i, \tag{9.34}$$

where H_0 is the nonkinetic part of the atomic Hamiltonian, $\mathbf{p}_i = -\mathrm{i}\hbar\nabla_i$, and the sum is over all electrons in an atom. For a homogeneous magnetic field $\mathbf{B}$, one can choose a vector potential of

$$\mathbf{A} = -\frac{1}{2}\mathbf{r} \times \mathbf{B}. \tag{9.35}$$

Here, we take the magnetic field $\mathbf{B}$ in the z-direction.

$$\mathbf{B} = (0, 0, B_0). \tag{9.36}$$

Then, the vector potential is given by

$$\mathbf{A} = -\frac{1}{2}B_0\left(y\hat{\mathrm{i}} - x\hat{\mathrm{j}}\right). \tag{9.37}$$

Substituting the vector potential into (9.34), we have

$$H = H_0 + \sum_i \frac{p_i^2}{2m_\mathrm{e}} + \frac{eB_0}{2m_\mathrm{e}c}\sum_i (x_i p_{iy} - y_i p_{ix}) + \frac{e^2B_0^2}{8m_\mathrm{e}c^2}\sum_i \left(x_i^2 + y_i^2\right) + 2\mu_\mathrm{B}B_0 S_z. \tag{9.38}$$

Here, we note that

$$x_i p_{iy} - y_i p_{ix} = (\mathbf{r}_i \times \mathbf{p}_i)_z = \hbar l_{\mathrm{i}z}.$$

Now, we can write the Hamiltonian as

$$H = H_0 + \sum_i \frac{p_i^2}{2m_\mathrm{e}} + \mu_\mathrm{B}(L_z + 2S_z)B_0 + \frac{e^2B_0^2}{8m_\mathrm{e}c^2}\sum_i \left(x_i^2 + y_i^2\right). \tag{9.39}$$

But $-\mu_\mathrm{B}(L_z + 2S_z)$ is simply m_z, the z-component of the magnetic moment of the atom in the absence of the applied magnetic field $\mathbf{B}$. Therefore, we have

$$H = \mathcal{H} - m_z B_0 + \frac{e^2B_0^2}{8m_\mathrm{e}c^2}\sum_i \left(x_i^2 + y_i^2\right), \tag{9.40}$$

where $\mathcal{H} = H_0 + \sum_i \frac{p_i^2}{2m_\mathrm{e}}$. In the presence of the magnetic field $\mathbf{B}_0$,

$$\begin{aligned} v_{ix} &= \frac{\partial H}{\partial p_{ix}} = \frac{1}{m}\left(p_{ix} - \frac{eB_0}{2c}y_i\right), \\ v_{iy} &= \frac{\partial H}{\partial p_{iy}} = \frac{1}{m}\left(p_{iy} + \frac{eB_0}{2c}x_i\right), \end{aligned} \tag{9.41}$$

and the magnetic moment in the presence of $\mathbf{B}_0$ is (see (9.20))

$$\boldsymbol{\mu} = \sum_i \left(-\frac{e}{2c}\mathbf{r}_i \times \mathbf{v}_i - \frac{e}{2m_\mathrm{e}c}2\hbar\mathbf{S}_i\right). \tag{9.42}$$

Using (9.41), the expression for v_{ix} and v_{iy}, one obtains

$$\mu_z = -\mu_\mathrm{B} L_z - 2\mu_\mathrm{B} S_z - \frac{e^2 B_0}{4m_\mathrm{e}c^2} \sum_i \left(x_i^2 + y_i^2\right). \tag{9.43}$$

Note that one can also obtain this result from (9.40) using the relation

$$\mu_z = -\frac{\partial H}{\partial B_0}. \tag{9.44}$$

Thus, we have μ_z, the z-component of magnetic moment of the atom in the magnetic field B_0 is given by

$$\mu_z = m_z - \frac{e^2 B_0}{4m_\mathrm{e}c^2} \sum_i \left(x_i^2 + y_i^2\right). \tag{9.45}$$

It differs from m_z, its value when $B_0 = 0$ by a term that is negative and proportional to B_0.

If the atom is in its ground state, $\overline{m_z}$, the average value of m_z is $\overline{m_z} = -\mu_\mathrm{B}(\overline{L_z + 2S_z}) = -g\mu_\mathrm{B}\overline{J_z} = -g\mu_\mathrm{B} j_z$, where $j_z = -J, -J+1, \ldots, J$. For a spherically symmetric atom, $\overline{x_i^2} = \overline{y_i^2} = \overline{z_i^2} = \frac{1}{3}\overline{r_i^2}$. Therefore, we obtain

$$\mu_z = -g\mu_\mathrm{B}\overline{J_z} - \frac{e^2 B_0}{6m_\mathrm{e}c^2} \sum_i \overline{r_i^2}. \tag{9.46}$$

The second term on the right-hand side is the origin of **diamagnetism**. If $J = 0$ (so that $\overline{J_z} = 0$), then a system containing N atoms per unit volume would produce a magnetization

$$M = -N\frac{e^2 B_0}{6m_\mathrm{e}c^2} \sum_i \overline{r_i^2}, \tag{9.47}$$

and the **diamagnetic susceptibility**

$$\chi_\mathrm{Dia} = \frac{M}{B_0} = -N\frac{e^2}{6m_\mathrm{e}c^2} \sum_i \overline{r_i^2}. \tag{9.48}$$

Here, we have assumed $\chi_\mathrm{Dia} \ll 1$ and set $\chi = \frac{M}{B}$ instead of $\frac{M}{H}$. This result was first derived by Langevin.

All substances exhibit diamagnetism. Paramagnetism occurs only in samples whose atoms possess permanent magnetic moments (i.e. $m \neq 0$ when $B_0 = 0$). All free atoms except those having complete electronic shells are paramagnetic. In solids, however, fewer substances exhibit paramagnetism because the electrons form energy bands and filled bands do not contribute to paramagnetism.

Examples of paramagnetism in solids are

1. Pauli spin paramagnetism of metals.
2. Paramagnetism due to incomplete shells.
   (a) Transition elements:

   Iron group elements with incomplete 3d shell, for example, $\left[\mathrm{Ti}^{3+}(3d^1) \sim \mathrm{Cu}^{2+}(3d^9)\right]$.

   Palladium group elements with incomplete 4d shell, for example, $\left[\mathrm{Zr}^{3+}(4d^1) \sim \mathrm{Ag}^{2+}(4d^9)\right]$.

   Platinum group elements with incomplete 5d shell, for example, $\left[\mathrm{Hf}^{3+}(5d^1) \sim \mathrm{Au}^{2+}(5d^9)\right]$.

   (b) Rare earth elements:

   Rare earth group elements (or Lanthanides) with incomplete $4f$ shell, for example, $\left[\mathrm{Ce}^{3+}(4f^1) \sim \mathrm{Yb}^{3+}(4f^{13})\right]$.

   Transuranic group elements (or Actinides) with incomplete $5f$ or 6d shells, for example, elements beyond Th.

9.4 Paramagnetism of Atoms

We have seen that the permanent magnetic dipole moment of an atom is given by

$$\mathbf{m} = -g_{\mathrm{L}}\mu_{\mathrm{B}}\mathbf{J}. \tag{9.49}$$

We will assume that the separations between atoms in the systems of interest are sufficiently large that the interactions between the atoms can be neglected. The energy of an atom in a magnetic field $\mathbf{B}$ is

$$E = -\mathbf{m}\cdot\mathbf{B} = g_{\mathrm{L}}\mu_{\mathrm{B}}Bm_J, \tag{9.50}$$

where $m_{\mathrm{J}} = -J, -J+1, \ldots, J-1, J$. The probability of finding an atom in state $|J, m_J>$ at a temperature T is

$$p(m_J) = \frac{1}{Z}\mathrm{e}^{-\beta E(m_J)}, \tag{9.51}$$

where $\beta = (k_{\mathrm{B}}T)^{-1}$ and the normalization constant Z is chosen so that $\sum_{m_J} p(m_J) = 1$. This gives

$$Z = \sum_{m_j=-J}^{J} \mathrm{e}^{-\beta g_{\mathrm{L}}\mu_{\mathrm{B}}Bm_J}. \tag{9.52}$$

Let $\beta g_{\mathrm{L}}\mu_{\mathrm{B}}B = y$. Then, $Z = \sum_{m=-J}^{J} \mathrm{e}^{-ym}$. This can be rewritten

$$\begin{aligned} Z &= \mathrm{e}^{-yJ}\left(1 + \mathrm{e}^{y} + \mathrm{e}^{2y} + \cdots + \mathrm{e}^{2Jy}\right) \\ &= \mathrm{e}^{-yJ}\left[\frac{(\mathrm{e}^{y})^{2J+1} - 1}{\mathrm{e}^{y} - 1}\right]. \end{aligned} \tag{9.53}$$

The result for Z can be rewritten

$$Z(x) = \frac{\sinh \frac{2J+1}{2J}x}{\sinh \frac{x}{2J}}, \tag{9.54}$$

where $x = yJ = \beta g_{\mathrm{L}}\mu_{\mathrm{B}}BJ$. The magnetization of a system containing N atoms per unit volume will be

$$M = -Ng_{\mathrm{L}}\mu_{\mathrm{B}}\frac{\sum_{m_J} m_J p(m_J)}{\sum_{m_J} p(m_J)} = Ng_{\mathrm{L}}\mu_{\mathrm{B}}J\frac{\partial}{\partial x}\ln Z. \tag{9.55}$$

This is usually written as

$$M = Ng_{\mathrm{L}}\mu_{\mathrm{B}}JB_J(\beta g_{\mathrm{L}}\mu_{\mathrm{B}}BJ), \tag{9.56}$$

where the function $B_J(x)$ is called the **Brillouin function**. It is not difficult to see that

$$B_J(x) = \frac{2J+1}{2J}\coth\frac{2J+1}{2J}x - \frac{1}{2J}\coth\frac{x}{2J}. \tag{9.57}$$

The argument of the Brillouin function $\frac{2J+1}{2J}\beta g_{\mathrm{L}}\mu_{\mathrm{B}}BJ$ is small compared tounity if the magnetic field B is small compared to 500 T at room temperature. Under these conditions (use $\coth z \simeq \frac{1}{z} + \frac{z}{3}$ for $z \ll 1$) one can write

$$B_J(x) \simeq \frac{x}{3}\frac{J+1}{J}, \tag{9.58}$$

and

$$M \simeq \frac{Ng_{\mathrm{L}}^2\mu_{\mathrm{B}}^2 J(J+1)}{3k_{\mathrm{B}}T}B. \tag{9.59}$$

Since $\langle \mathbf{m}\cdot\mathbf{m}\rangle = g_{\mathrm{L}}^2\mu_{\mathrm{B}}^2\langle \mathbf{J}\cdot\mathbf{J}\rangle = g_{\mathrm{L}}^2\mu_{\mathrm{B}}^2 J(J+1)$ we can write

$$\chi_{\mathrm{Para}} = \frac{M}{B} = \frac{N\langle \mathbf{m}^2\rangle}{3k_{\mathrm{B}}T} \tag{9.60}$$

for the paramagnetic susceptibility of a system of atoms of magnetic moment $\mathbf{m}$ at high temperature ($g_{\mathrm{L}}\mu_{\mathrm{B}}BJ \ll k_{\mathrm{B}}T$). This is commonly known as **Curie's law**. Notice that when J becomes very large

$$\lim_{J\to\infty} B_J(x) \Rightarrow \coth x - \frac{1}{x} = \mathcal{L}(x), \tag{9.61}$$

where $\mathcal{L}$ is the **Langevin function** that we encountered in studying electric dipole moments. Thus, the quantum mechanical result goes over to the

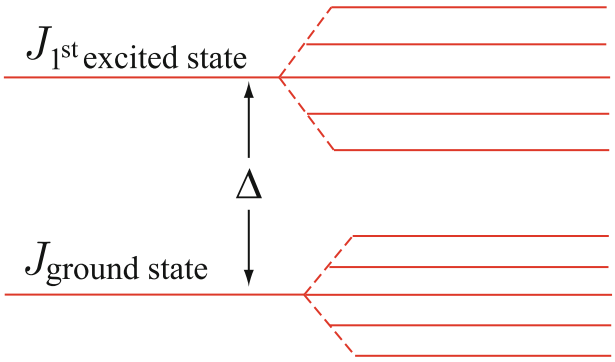


Fig. 9.2. Energy level splitting of the ground state and first excited multiplets for an atom of the total angular momentum quantum number J

classical result as $J \to \infty$, as expected. Curie's law is often written

$$M \simeq \frac{N\mu_{\mathrm{B}}^2 p^2 B}{3k_{\mathrm{B}}T}, \tag{9.62}$$

where $p = g_{\mathrm{L}}\sqrt{J(J+1)}$ is called the **effective number of Bohr magnetons**. Knowing S, L, J and g_{L} from the application of Hund's rules immediately gives us p. For example, for a Dy^{3+} ion the atomic configuration is $(4f)^9(5s)^2(5p)^6$. This results from removing two $6s$ electrons and one $4f$ electron from the neutral atom. The S-value will be $\frac{5}{2}$ (seven $4f$-electrons in $\uparrow$ and two in $\downarrow$ states), $L = 5$ (the two $\downarrow$ electrons have $m_z = 3$ and 2 to maximize L), and $J = L + S = \frac{15}{2}$, and hence $g_{\mathrm{L}} = \frac{4}{3}$ and $p = \frac{4}{3}\sqrt{\frac{15}{2}\cdot\frac{17}{2}} \simeq 10.63$.

Observed and calculated p-values agree fairly well. There are exceptions when excited state multiplets are not sufficiently high in energy (see, for example, Fig. 9.2).

Until now we have assumed $\Delta \gg k_{\mathrm{B}}T$ and $\Delta \gg g_{\mathrm{L}}\mu_{\mathrm{B}}JB$. If this is not true, higher multiplets can be important in evaluation of χ or p. Typically, for an ion with partially filled shell with nonzero value of J, $\chi_{\mathrm{Para}} \sim 10^{-2} - 10^{-3}$ at room temperature and $\chi_{\mathrm{Dia}} \sim 10^{-5}$, which is independent of temperature. Therefore, we have $\chi_{\mathrm{Para}} \sim 500\chi_{\mathrm{Dia}}$ at room temperature.

9.5 Pauli Spin Paramagnetism of Metals

If we used the classical theory of paramagnetism for a particle with magnetic moment $\mathbf{m}$, the magnetization at a temperature T (with $k_{\mathrm{B}}T \gg |\mathbf{m}\cdot\mathbf{B}|$) would be given by Curie's law

$$M = \frac{N\langle \mathbf{m}^2 \rangle B}{3k_{\mathrm{B}}T}. \tag{9.63}$$

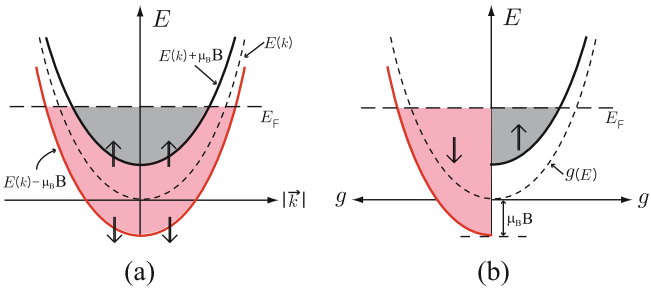


Fig. 9.3. Energy level splitting of the electron gas in the presence of the magnetic field **B**. Energy parabolas $E(k)$ (**a**) and density of states $g(E)$ (**b**) of electrons in two different spin states in the presence of Zeeman splitting

For free electrons $\mathbf{m} = -2\mu_\mathrm{B}\mathbf{s}$ and $\langle \mathbf{m}^2 \rangle = 4\mu^2 \langle \mathbf{s} \cdot \mathbf{s} \rangle = 4\mu^2 s(s+1)$. Since $s = \frac{1}{2}$ this gives $\langle \mathbf{m}^2 \rangle = 3\mu^2$ and[1]

$$\chi_\mathrm{classical} = \frac{M}{B} = \frac{n_0 \mu_\mathrm{B}^2}{k_\mathrm{B} T}. \tag{9.64}$$

As we discussed earlier, this is not what is observed experimentally. In metals the observed susceptibility is approximately independent of temperature and two orders of magnitude smaller than the value of $\chi_\mathrm{classical}$ evaluated at room temperature.

The qualitative explanation is exactly the same as that by which the Sommerfeld model explained the electronic contribution to the specific heat. At a temperature T only electrons whose energy lies within a shell of width $k_\mathrm{B}T$ about the Fermi energy are effectively free. Other electrons are inefficient because of the Pauli exclusion principle. If we replace n_0 by n_eff, where

$$n_\mathrm{eff} \simeq n_0 \frac{k_\mathrm{B} T}{\zeta}. \tag{9.65}$$

The spin susceptibility becomes $\chi_\mathrm{QM} \simeq \frac{n_0 \mu_\mathrm{B}^2}{\zeta} \simeq \frac{k_\mathrm{B} T}{\zeta} \chi_\mathrm{classical} = \frac{n_0 \mu_\mathrm{B}^2}{\zeta}$.

To obtain χ_QM more rigorously, we simply assume that in the presence of the magnetic field **B** the energy of an electron is changed by an amount

$$\delta E = \pm \mu_\mathrm{B} B = -\mathbf{m} \cdot \mathbf{B}; \quad \mathbf{m} = -g_\mathrm{L} \mu_\mathrm{B} \mathbf{S}$$

depending on whether its spin is up or down relative to the direction of **B**.

The number of particles of spin up (or down) per unit volume is

$$n_\pm = \frac{1}{2} \int_0^\infty \mathrm{d}E f_0(E) g\left(E \mp \mu_\mathrm{B} B\right), \tag{9.66}$$

where $+$ and $-$ in the subscript of $n_\pm$ correspond to the cases of spin up $(+)$ and spin down $(-)$ states, respectively. (See Fig. 9.3.) We evaluated many

[1] Here, n_0 is the number of free electrons per unit volume in a metal.

integrals over Fermi functions in Chap. 3. Remember that the total number of states per unit volume with energy less than ε is given by

$$G(\varepsilon) = \int_0^{\varepsilon} g(\varepsilon)\mathrm{d}\varepsilon = n_0 \left(\frac{k}{k_\mathrm{F}}\right)^3 = n_0 \left(\frac{\varepsilon}{\varepsilon_\mathrm{F}}\right)^{3/2}.$$

Using these results we can obtain

$$n_\pm = \frac{1}{2}\left[G(\zeta \mp \mu_\mathrm{B}B) + \frac{\pi^2}{6}(k_\mathrm{B}T)^2 g'(\zeta \mp \mu_\mathrm{B}B)\right]. \tag{9.67}$$

The magnetization M is equal to $\mu_\mathrm{B}(n_- - n_+)$. Expanding for $\zeta \gg \mu_\mathrm{B}B$ and $k_\mathrm{B}T \ll \zeta$ leads to

$$M \simeq \mu_\mathrm{B}^2 B\left[g(\zeta) + \frac{\pi^2}{6}(k_\mathrm{B}T)^2 g''(\zeta)\right]. \tag{9.68}$$

The chemical potential is determined by requiring the number of particles to be $n_0 = n_- + n_+$. This gives

$$n_0 = G(\zeta) + \frac{\pi^2}{6}(k_\mathrm{B}T)^2 g'(\zeta) + O(\mu_\mathrm{B}^2 B^2). \tag{9.69}$$

To order $\mu_\mathrm{B}^2 B^2$, we note that

$$\zeta = \zeta_0 - \frac{\pi^2}{6}(k_\mathrm{B}T)^2 \frac{g'(\zeta_0)}{g(\zeta_0)}. \tag{9.70}$$

Using $g(\zeta) = \frac{3}{2}\frac{n_0}{\zeta_0}\left(\frac{\varepsilon}{\zeta_0}\right)^{1/2}$ gives

$$\chi_\mathrm{QM} = \frac{3n_0\mu_\mathrm{B}^2}{2\zeta_0}\left[1 - \frac{\pi^2}{12}\left(\frac{k_\mathrm{B}T}{\zeta_0}\right)^2 + \cdots\right] \tag{9.71}$$

for the Pauli spin (paramagnetic) susceptibility of a metal.

9.6 Diamagnetism of Metals

According to classical mechanics there should be no diamagnetism of a free electron gas. Consider the effect of a magnetic field $\mathbf{B}$ on the motion of an electron. The force acting on the electron is

$$\mathbf{F} = -\frac{e}{c}\mathbf{v} \times \mathbf{B}. \tag{9.72}$$

This force is always perpendicular to $\mathbf{v}$, therefore $\mathbf{F} \cdot d\mathbf{l} = \mathbf{F} \cdot \mathbf{v}\mathrm{d}t = 0$. Thus, no work is done on the electrons by the field $\mathbf{B}$ and their energy is unchanged.

Further the distribution function depends only on E, T, N and will also be unchanged. Thus, there can be no induced currents and no diamagnetism.

Quantum mechanics gives a different answer. Landau was the first to derive the **diamagnetic susceptibility** of metals. We will not rederive his result in full, but simply show how the result comes about in a quantum mechanical calculation.

Let $\mathbf{A} = (0, xB, 0)$ be the vector potential of a dc magnetic field $\mathbf{B}$. The Hamiltonian for a single electron is (here we shall neglect the intrinsic magnetic moment of the electron)

$$H = \frac{1}{2m}\left[p_x^2 + \left(p_y + \frac{e}{c}Bx\right)^2 + p_z^2\right]. \tag{9.73}$$

Recall that $\mathbf{p} = -\mathrm{i}\hbar\nabla$. The Schrödinger equation is

$$-\frac{\hbar^2}{2m}\left[\frac{\partial^2}{\partial x^2} + \left(\frac{\partial}{\partial y} + i\frac{eB}{\hbar c}x\right)^2 + \frac{\partial^2}{\partial z^2}\right]\Psi = E\Psi. \tag{9.74}$$

Since the Hamiltonian is independent of y and z, let us assume a solution of the form

$$\Psi(x, y, z) = \mathrm{e}^{\mathrm{i}k_y y + \mathrm{i}k_z z}\phi(x). \tag{9.75}$$

The equation which $\phi(x)$ must satisfy is

$$\left[\frac{\partial^2}{\partial x^2} - \left(k_y + \frac{eB}{\hbar c}x\right)^2 - k_z^2 + \frac{2mE}{\hbar^2}\right]\phi(x) = 0 \tag{9.76}$$

If we let $x' = x + \frac{\hbar k_y}{m\omega_\mathrm{c}}$ this equation becomes

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x'^2} + \frac{1}{2}m\omega_\mathrm{c}^2 x'^2\right)\phi(x') = \left(E - \frac{\hbar^2 k_z^2}{2m}\right)\phi(x'). \tag{9.77}$$

This is just the equation for a simple harmonic oscillator of mass m and characteristic frequency ω_c. The energy levels are

$$E - \frac{\hbar^2 k_z^2}{2m} = \hbar\omega_\mathrm{c}\left(n + \frac{1}{2}\right);\ n = 0, 1, 2, \ldots \tag{9.78}$$

Thus, the eigenfunctions and eigenvalues for an electron in the presence of a magnetic field $\mathbf{B}$ are

$$|nk_y k_z\rangle = L^{-1}\mathrm{e}^{\mathrm{i}k_y y + \mathrm{i}k_z z}\phi_n\left(x + \frac{\hbar k_y}{m\omega_\mathrm{c}}\right). \tag{9.79}$$

$$E_n(k_y, k_z) = \frac{\hbar^2 k_z^2}{2m} + \hbar\omega_\mathrm{c}\left(n + \frac{1}{2}\right). \tag{9.80}$$

We note that the eigenvalues $E_n(k_y, k_z)$ are independent of k_y. The allowed values of k_y and k_z are determined by imposing periodic boundary conditions. If we require the particles to be in a cube of length L, then because the center of each oscillator must be in the box, the range of possible k_y values must be

$$\text{Range of values of } k_y \leq \frac{m\omega_c L}{\hbar}. \tag{9.81}$$

The total number of allowed values of k_y, for a given k_z, is

$$\frac{L}{2\pi} \times \text{ Range of values of } k_y = \frac{m\omega_c L^2}{2\pi\hbar} = \frac{BL^2}{hc/e}. \tag{9.82}$$

Thus for each value of n, k_z (and spin $\mathbf{s}$) there are $\frac{m\omega_c L^2}{2\pi\hbar}$ energy states. Consider the following schematic plot of the energy levels for a given k_z shown in Fig. 9.4. In quantum mechanics, a dc magnetic field can alter the distribution of energy levels. Thus, there can be a change in the energy of the system and this can result in a diamagnetic current. The diamagnetic susceptibility turns out to be

$$\chi_L = -\frac{n_0}{2\zeta_0}\left(\frac{e\hbar}{2m^*c}\right)^2 = -\frac{n_0\mu_B^2}{2\zeta_0}\left(\frac{m}{m^*}\right)^2. \tag{9.83}$$

Notice that we expect m^* (not m) to appear because the diamagnetism is associated with the orbital motion of the electrons. This is justifiable only if the cyclotron radius is much larger than the interatomic spacing.

$$r_c = \frac{v_F}{\omega_c} = \frac{v_F}{eB_0/m^*c}. \tag{9.84}$$

Typically, $v_F \approx 10^8\,\text{cm/s}$ and $\omega_c \approx 1.76 \times 10^7 B_0$. Thus for $B_0 \sim 10^5$ gauss, $r_c \approx 10^{-4}\,\text{cm} \sim 10^4\,\text{Å} \gg$ lattice constant. The total magnetic susceptibility of a metal is

$$\chi_{QM} = \chi_P + \chi_L = \frac{3n_0\mu_B^2}{2\zeta_0}\left[1 - \frac{1}{3}\left(\frac{m}{m^*}\right)^2\right] \tag{9.85}$$

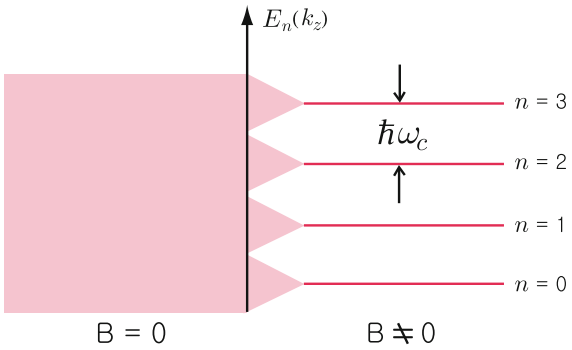


Fig. 9.4. Energy level splitting of the electron gas due to orbital quantizations in the presence of the magnetic field **B**

9.7 de Haas–van Alphen Effect

We have seen that the energy levels for an electron in a magnetic field look like, as is shown in Fig. 9.5,

$$E_n(k_y, k_z) = \frac{\hbar^2 k_z^2}{2m} + \hbar\omega_c \left(n + \frac{1}{2}\right).$$

The Fermi energy ζ is a slowly varying function of $\mathbf{B}$. As we increase $\mathbf{B}$, the distance between Landau levels increases, and at $k_z = 0$ levels pass through the Fermi energy. As the Landau level at $k_z = 0$ passes through the Fermi level, the internal energy abruptly decreases.

Let us consider a simple situation, where the Fermi energy ζ is between two orbits. Let us assume that $T = 0$ so that we have a perfectly sharp Fermi surface. As we increase the field, the states are raised in energy so that the lowest occupied state approaches ζ in energy. Of course, all the energies in the presence of the field will be higher than those in the absence of the field by an amount $\frac{1}{2}\hbar\omega_c$. As the levels approach the Fermi energy, the free energy of the electron gas approaches a maximum. As the highest occupied level passes the Fermi surface, it begins to empty, thus decreasing the free energy of the electron gas. When the Fermi level lies below the cyclotron level the energy of the electron gas is again a minimum. Thus, we can see how the free energy is a periodic function of the magnetic field. Now, since many physically observable properties of the system are derived from the free energy (such as the magnetization), we see that they, too, are periodic functions of the magnetic field. The periodic oscillation of the diamagnetic susceptibility of a metal at low temperatures is known as the **de Haas–van Alphen effect**. The de Haas–van Alphen effect arises from the periodic variation of the total energy of an electron gas as a function of a static magnetic field. The energy variation is easily observed in experiments as a periodic variation in the magnetic moment of the metal.

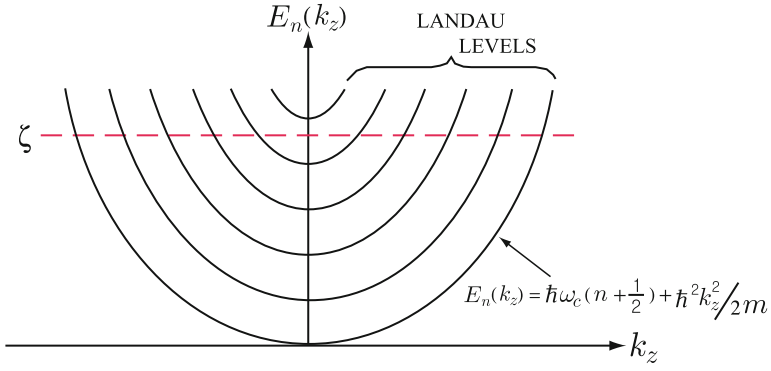


Fig. 9.5. Schematics of energy levels for an electron in a magnetic field $\mathbf{B}$

Density of States

Look at $G(E)$, the number of states per unit volume of energy less than E.

$$G(E) = \frac{1}{L^3} \sum_{\substack{nk_yk_z \\ E_{nk_z} < E}} 1 = \frac{1}{L^3} \left(\frac{L}{2\pi}\right)^2 2 \sum_{\substack{n \\ E_{nk_z} < E}} \int \mathrm{d}k_y \, \mathrm{d}k_z \; 1. \tag{9.86}$$

We added a factor 2 to take account of spin. Since $\int \mathrm{d}k_y = \frac{m\omega_c L}{\hbar}$, we have

$$G(E) = \frac{1}{L^3} \left(\frac{L}{2\pi}\right)^2 2\frac{m\omega_c L}{\hbar} \sum_{\substack{n \\ E_{nk_z} < E}} \int \mathrm{d}k_z,$$

where $E_n(k_y, k_z) = \frac{\hbar^2 k_z^2}{2m} + \hbar\omega_c(n + \frac{1}{2})$. Define $\kappa_n = \frac{\sqrt{2m}}{\hbar}\left(E - \hbar\omega_c(n + \frac{1}{2})\right)^{1/2}$. For each value of n the k_z integration goes from $-\kappa_n$ to κ_n. This gives

$$G(E) = \frac{m\omega_c}{2\pi^2\hbar} \sum_{n=0}^{n_{\mathrm{Max}}} 2\frac{\sqrt{2m}}{\hbar} \left(E - \hbar\omega_c\left(n + \frac{1}{2}\right)\right)^{1/2}. \tag{9.87}$$

The density of states $g(E)$ is $\frac{\mathrm{d}G}{\mathrm{d}E}$.

$$g(E) = \frac{m\omega_c}{2\pi^2\hbar} \frac{\sqrt{2m}}{\hbar} \sum_{n=0}^{n_{\mathrm{Max}}} \left(E - \hbar\omega_c\left(n + \frac{1}{2}\right)\right)^{-1/2}. \tag{9.88}$$

We note that the $g(E)$ has square root singularities at $E = \hbar\omega_c(n + \frac{1}{2})$ as illustrated in Fig. 9.6.

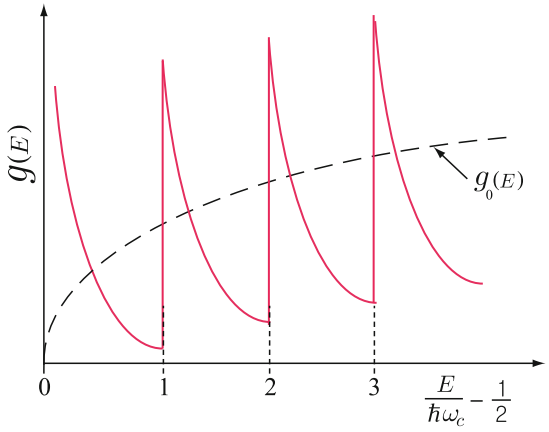


Fig. 9.6. Schematic plot of the density of states for an electron gas in a magnetic field **B**

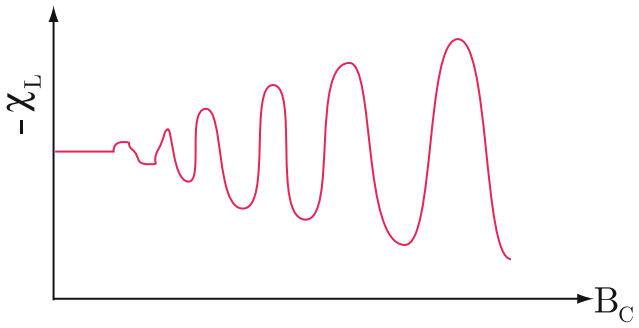


Fig. 9.7. Schematic plot of the magnetic field dependence of the diamagnetic susceptibility

In the limit as $\hbar\omega_c \to 0$, the sum on n can be replaced by an integral

$$\sum_{n=0}^{n_{\mathrm{Max}}} \longrightarrow \frac{1}{\hbar\omega_c}\int_0^E \mathrm{d}x,$$

where $x = n\hbar\omega_c$. In this case the $g(E)$ reduces to the free particle density of states for $\mathbf{B} = 0$, i.e. $g_0(E) = \frac{\sqrt{2}m^{3/2}}{\pi^2\hbar^3}E^{1/2}$.

Because $g(E)$ has square root singularities, the internal energy, the magnetic susceptibility, etc. show oscillations (see Fig. 9.7). As $\hbar\omega_c$ changes with N fixed (or N changes with $\hbar\omega_c$ fixed), the Fermi level passes through the bottoms of different Landau levels. Let $\zeta \gg \mu_B B$ or $\zeta \gg \hbar\omega_c$. Then the electron gas occupies states in many different cyclotron levels. At low temperatures all cyclotron levels are partially occupied up to a limiting energy ζ, which might lie between the threshold energies of the nth and $(n-1)$th cyclotron levels. As B-field increases, the energy and the number of states in each subband increase, and hence the threshold energy likewise increases. Since the total number of electrons is given, there is a continuous rearrangement of the electrons with increasing the field. When the threshold energy $E_n = (n+\frac{1}{2})\hbar\omega_c$ grows from a value below ζ to a value above ζ, the electrons in the nth band fall back into states in the $(n-1)$th band, decreasing the total energy. As the field increases, it rises again until the next threshold energy E_{n-1} exceeds the ζ. As a result ζ itself becomes (weakly) periodic. The separation between the individual threshold energy is $\hbar\omega_c$. Therefore, $\hbar\omega_c \gg k_B T$ is an important condition; otherwise the electron distribution in the region of ζ is so widely spread that oscillations are smoothed out. When the condition $\zeta \gg \hbar\omega_c$ is no longer fulfilled, all the electrons are in the lowest subband and the oscillations cease. This limit is called the **quantum limit**.

The oscillations in the magnetic susceptibility are observed in experiments when $\hbar\omega_c \gg k_B T$ in high purity (low scattering) samples. Oscillations in electrical conductivity are called the **Shubnikov–de Haas oscillations**. Both

the de Haas–van Alphen oscillations and Shubnikov–de Haas oscillations are useful in studying electronic properties of metals and semiconductor quantum structures.

9.8 Cooling by Adiabatic Demagnetization of a Paramagnetic Salt

The entropy of a paramagnetic salt is the sum of the entropy due to phonons and the entropy due to the magnetic moments.

$$S = S_{\mathrm{p}} + S_{\mathrm{m}}. \tag{9.89}$$

If the paramagnetic ion has angular momentum J, then the ground state in the absence of any applied magnetic field must be $2J+1$ fold degenerate. This is because m_J can have any value between $-J$ and $+J$ with equal probability. For a system containing N paramagnetic ions (noninteracting), the total degeneracy is $(2J+1)^N$, and the magnetic contribution to the entropy is

$$S_{\mathrm{m}}(B=0) = k_{\mathrm{B}} \ln(2J+1)^N = N k_{\mathrm{B}} \ln(2J+1). \tag{9.90}$$

Introduce a magnetic field B (neglect local field corrections treating the ions as noninteracting magnetic ions). Then the magnetic entropy must be given by

$$S_{\mathrm{m}}(B) = -N k_{\mathrm{B}} \sum_{m_J=-J}^{J} p(m_J) \ln p(m_J), \tag{9.91}$$

where

$$p(m_J) = Z^{-1} \mathrm{e}^{-\frac{g_{\mathrm{L}}\mu_{\mathrm{B}}B}{k_{\mathrm{B}}T} m_J}. \tag{9.92}$$

Here, we have used the relation $S(B,T) = k_{\mathrm{B}} \frac{\partial}{\partial T}(T \ln Z)$ where the normalization constant Z is defined so that $\sum_{m_J} p(m_J) = 1$, giving

$$Z = \sum_{m_J} \mathrm{e}^{-\frac{g_{\mathrm{L}}\mu_{\mathrm{B}}B}{k_{\mathrm{B}}T} m_J}. \tag{9.93}$$

Substitute the expression for $p(m_J)$ into $S_{\mathrm{m}}(B)$ to have

$$S_{\mathrm{m}}(B) = N k_{\mathrm{B}} \ln Z + N k_{\mathrm{B}} \frac{g_{\mathrm{L}}\mu_{\mathrm{B}}B}{k_{\mathrm{B}}T} \overline{m}_J. \tag{9.94}$$

We note that the magnetization is given by $M = -N g_{\mathrm{L}} \mu_{\mathrm{B}} \overline{m}_J$ so that

$$S_{\mathrm{m}}(B) = N k_{\mathrm{B}} \ln Z - \frac{MB}{T}. \tag{9.95}$$

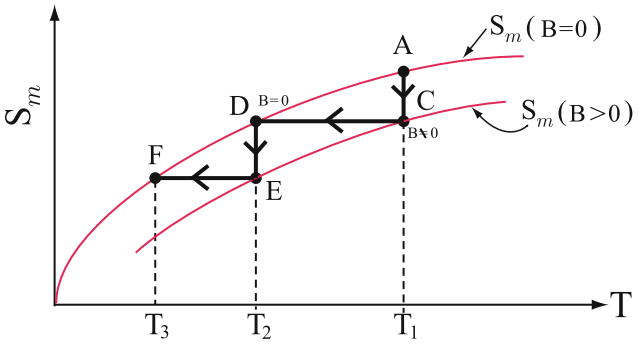


Fig. 9.8. Schematic plot of the process of cooling by adiabatic demagnetization of a paramagnetic salt

Notice that $S_{\mathrm{m}}(B)$ depends only on the product $\beta B = \frac{B}{k_{\mathrm{B}}T}$. Thus, we have

$$S_{\mathrm{m}}(B) - S_{\mathrm{m}}(0) = Nk_{\mathrm{B}} \ln \frac{Z}{2J+1} - \frac{MB}{T}. \tag{9.96}$$

It is easy to see that this quantity is always negative. This agrees with the intuitive idea that the system is more disordered in the absence of the magnetic field. The phonon contribution to the entropy is essentially independent of magnetic field.

Now, consider the following process (see Fig. 9.8):

1. Apply a magnetic field B under isothermal conditions. This takes one from point A to point C in the S_{m} versus T plane.
2. Now, isolate the salt from the heat bath and adiabatically remove the magnetic field to arrive at D.

This process has lowered the temperature from T_1 to T_2. The process can be repeated. In an ideal system $S_{\mathrm{m}}(B = 0)$ should approach zero as T approaches zero. In practice there is a lower limit in T that can be reached; it is due to the internal magnetic fields (i.e. coupling of magnetic moments to one another) in the paramagnetic salt.

9.9 Ferromagnetism

Some materials possess a spontaneous magnetic moment; that is, even in the absence of an applied magnetic field they have a magnetization M. The value of the spontaneous magnetic moment per unit volume is called the **spontaneous magnetization**, $M_{\mathrm{s}}(T)$. The temperature T_{c} above which the spontaneous magnetization vanishes is called the **Curie temperature**.

The simplest way to account for the spontaneous alignment is by postulating the existence of an internal field H_{E}, called the **Weiss field**, which causes

the magnetic moments of the atoms to line up. The value of $H_{\rm E}$ is determined from the Curie temperature $T_{\rm c}$ by the relation $g\mu_{\rm B}JH_{\rm E} \simeq k_{\rm B}T_{\rm c}$ to be

$$H_{\rm E} = \frac{k_{\rm B}T_c}{g\mu_{\rm B}J}. \tag{9.97}$$

Typically, $H_{\rm E}$ has a value of about 500 Tesla. We shall see that effective field is not of magnetic origin. If we take $\mu_{\rm B}$ divided by the volume of a unit cell, we obtain $\frac{\mu_{\rm B}}{a^3} \simeq 10^3$ gauss $\ll H_{\rm E}$. Weiss assumed that the effective field $H_{\rm E}$ was proportional to the magnetization, i.e.

$$H_{\rm E} = \lambda M. \tag{9.98}$$

For $T > T_{\rm c}$, the magnetic susceptibility obeys Curie's law, but now $H + H_{\rm E}$ would replace H

$$M = \frac{C}{T}\left(H + H_{\rm E}\right) = \frac{C}{T}\left(H + \lambda M\right)$$

Therefore, we have

$$M = \frac{C}{T - \lambda C}H = \frac{C}{T - T_{\rm c}}H \tag{9.99}$$

Since $C = \frac{Ng_{\rm L}^2\mu_{\rm B}^2 J(J+1)}{3k_{\rm B}}$, the molecular field parameter can be written

$$\lambda^{-1} = \frac{Ng_{\rm L}^2\mu_{\rm B}^2 J(J+1)}{3k_{\rm B}T_{\rm c}}. \tag{9.100}$$

For Fe, we have $\lambda \simeq 5,000$.

Problems

9.1. Consider a volume V bounded by a surface S filled with a magnetization $\mathbf{M}(\mathbf{r}')$ that depends on the position $\mathbf{r}'$. The vector potential $\mathbf{A}$ produced by a magnetization $\mathbf{M}(\mathbf{r})$ is given by

$$\mathbf{A}(\mathbf{r}) = \int \mathrm{d}^3 r' \frac{\mathbf{M}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}.$$

(a) Show that $\nabla' \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} \right) = \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}$.

(b) Use this result together with the divergence theorem to show that $\mathbf{A}(\mathbf{r})$ can be written as

$$\mathbf{A}(\mathbf{r}) = \int_{\mathrm{V}} \mathrm{d}^3 r' \frac{\nabla_{r'} \times \mathbf{M}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \oint_{\mathrm{S}} \mathrm{d}S' \frac{\mathbf{M}(\mathbf{r}') \times \hat{\mathbf{n}}'}{|\mathbf{r} - \mathbf{r}'|},$$

where $\hat{\mathbf{n}}$ is a unit vector outward normal to the surface S. The volume integration is carried out over the volume V of the magnetized material. The surface integral is carried out over the surface bounding the magnetized object.

9.2. Demonstarate for yourself that Table 9.1 is correct by placing $\uparrow$ or $\downarrow$ arrows according to Hund's rules as shown below for Cr of atomic configuration $(3d)^5(4s)^1$.

l_z	2	1	0	−1	−2
3d-shell	↑	↑	↑	↑	↑
4s-shell			↑		

Clearly $S = \frac{1}{2} \times 6 = 3$, $L = 0$, $J = L + S = 3$, and

$$g = \frac{3}{2} + \frac{1}{2} \frac{3(3+1) - 0(0+1)}{3(3+1)} = 2.$$

Therefore, the spectroscopic notation of Cr is ${}^7S_3$.

Use Hund's rules (even though they might not be appropriate for every case) to make a similar table for Y^{39}, Nb^{41}, Tc^{43}, La^{57}, Dy^{66}, W^{74}, and Am^{95}.

9.3. A system of N electrons is confined to move on the $x-y$ plane. A magnetic field $\mathbf{B} = B\hat{z}$ is perpendicular to the plane.

(a) Show that the eigenstates of an electron are written by

$$\varepsilon_{n\sigma}(k) = \hbar\omega_{\mathrm{c}} \left(n + \frac{1}{2} \right) - \mu_{\mathrm{B}} B \sigma_z,$$

$$\psi_{n\sigma}(k,x,y) = \mathrm{e}^{iky} u_n(x + \frac{\hbar k}{m\omega_\mathrm{c}})\eta_\sigma .$$

Here, $k = \frac{2\pi}{L} \times j$, where $j = -\frac{N}{2}, -\frac{N}{2}+1, \ldots, \frac{N}{2}-1$, and η_σ is a spin eigenfunction.

(b) Determine the density of states $g_\sigma(\varepsilon)$ for electrons of spin σ. Remember that each cyclotron level can accomodate $N_\mathrm{L} = \frac{BL^2}{hc/e}$ electrons of each spin.

(c) Determine $G_\sigma(\varepsilon)$, the total number of states per unit area.

(d) Describe qualitatively how the chemical potential at $T = 0$ changes as the magnetic field is increased from zero to a value larger than $(\frac{hc}{e})\frac{N_\mathrm{L}}{L^2}$.

9.4. Show that $S_\mathrm{m}(B,T) < S_\mathrm{m}(0,T)$ by showing that $\mathrm{d}S(B,T) = \partial_B S|_{T,V}\mathrm{d}B + \partial_T S|_{B,V}\mathrm{d}T$ and that $\partial_B S(B,T)|_{T,V} < 0$ for all values of $\frac{g_\mathrm{L}\mu_\mathrm{B}B}{k_\mathrm{B}T}$ if $J \neq 0$. Here, $\partial_T S|_{B,V}$ is just $\frac{c_\mathrm{v}}{T}$.

Summary

The total angular momentum and magnetic moment of an atom are given by

$$\mathbf{J} = \mathbf{L} + \mathbf{S}.; \ \mathbf{m} = -\mu_{\mathrm{B}} (\mathbf{L} + 2\mathbf{S}) = -\hat{g}\mu_{\mathrm{B}}\mathbf{J}.$$

Here, the eigenvalue of the operator $\hat{g}$ is the **Landé g-factor** written as

$$g_{\mathrm{L}} = \frac{3}{2} + \frac{1}{2}\frac{s(s+1) - l(l+1)}{j(j+1)}.$$

The ground state of an atom or ion with an incomplete shell is determined by Hund's rules:

1. The ground state has the maximum S consistent with the Pauli exclusion principle.
2. It has the maximum L consistent with the maximum spin multiplicity $2S + 1$ of Rule 1.
3. The J-value is given by $L - S$ when a shell is less than half filled and by $L + S$ when more than half filled.

In the presence of a magnetic field $\mathbf{B}$ the Hamiltonian describing the electrons in an atom is written as

$$H = H_0 + \sum_i \frac{1}{2m}\left(\mathbf{p}_i + \frac{e}{c}\mathbf{A}(\mathbf{r}_i)\right)^2 + 2\mu_{\mathrm{B}}\mathbf{B} \cdot \sum_i \mathbf{s}_i,$$

where H_0 is the nonkinetic part of the atomic Hamiltonian and the sum is over all electrons in an atom. For a homogeneous magnetic field $\mathbf{B}$ in the z-direction, we have $\mathbf{A} = -\frac{1}{2}B_0\left(y\hat{\mathrm{i}} - x\hat{\mathrm{j}}\right)$. In this gauge, the Hamiltonian becomes

$$H = \mathcal{H} - m_z B_0 + \frac{e^2 B_0^2}{8m_{\mathrm{e}}c^2}\sum_i \left(x_i^2 + y_i^2\right),$$

where $\mathcal{H} = H_0 + \sum_i \frac{p_i^2}{2m_{\mathrm{e}}}$ and $m_z = \mu_{\mathrm{B}}(L_z + 2S_z)$. In the presence of $\mathbf{B}_0$, the z-component of magnetic moment of the atom becomes

$$\mu_z = m_z - \frac{e^2 B_0}{6m_{\mathrm{e}}c^2}\sum_i \overline{r_i^2}.$$

The second term on the right-hand side is the origin of **diamagnetism**. If $J = 0$ (so that $\overline{J_z} = 0$), the (Langevin) diamagnetic susceptibility is given by

$$\chi_{\mathrm{Dia}} = \frac{M}{B_0} = -N\frac{e^2}{6m_{\mathrm{e}}c^2}\sum_i \overline{r_i^2}.$$

The energy of an atom in a magnetic field $\mathbf{B}$ is $E = g_{\mathrm{L}}\mu_{\mathrm{B}}Bm_J$, where $m_{\mathrm{J}} = -J, -J+1, \ldots, J-1, J$. The magnetization of a system containing N

atoms per unit volume is written as $M = Ng_{\rm L}\mu_{\rm B}JB_J(\beta g_{\rm L}\mu_{\rm B}BJ)$, where the function $B_J(x)$ is called the **Brillouin function**. If the magnetic field B is small compared to 500 T at room temperature, M becomes

$$M \simeq \frac{Ng_{\rm L}^2\mu_{\rm B}^2 J(J+1)}{3k_{\rm B}T}B,$$

and we obtain the **Curie's law** for the paramagnetic susceptibility:

$$\chi_{\rm Para} = \frac{M}{B} = \frac{N\left\langle \mathbf{m}^2 \right\rangle}{3k_{\rm B}T}$$

at high temperature, $(g_{\rm L}\mu_{\rm B}BJ \ll k_{\rm B}T)$.

In the presence of the magnetic field $\mathbf{B}$, the number of electrons of spin up (or down) per unit volume is

$$n_{\pm} = \frac{1}{2}\int_0^{\infty} dE f_0(E) g\left(E \mp \mu_{\rm B}B\right).$$

For $\zeta \gg \mu_{\rm B}B$ and $k_{\rm B}T \ll \zeta$, the magnetization $M(=\mu_{\rm B}(n_- - n_+))$ reduces to

$$M \simeq \mu_{\rm B}^2 B\left[g(\zeta) + \frac{\pi^2}{6}(k_{\rm B}T)^2 g''(\zeta)\right],$$

with $\zeta = \zeta_0 - \frac{\pi^2}{6}(k_{\rm B}T)^2\frac{g'(\zeta_0)}{g(\zeta_0)}$. Since $g(\zeta) = \frac{3}{2}\frac{n_0}{\zeta_0}\left(\frac{\varepsilon}{\zeta_0}\right)^{1/2}$, we obtain the (quantum mechanical) expression

$$\chi_{\rm QM} = \frac{3n_0\mu_{\rm B}^2}{2\zeta_0}\left[1 - \frac{\pi^2}{12}\left(\frac{k_{\rm B}T}{\zeta_0}\right)^2 + \cdots\right]$$

for the Pauli spin (paramagnetic) susceptibility of a metal.

In quantum mechanics, a dc magnetic field can alter the distribution of the electronic energy levels and the orbital states of an electron are described by the eigenfunctions and eigenvalues given by

$$|nk_yk_z\rangle = L^{-1}\mathrm{e}^{\mathrm{i}k_y y + \mathrm{i}k_z z}\phi_n\left(x + \frac{\hbar k_y}{m\omega_{\rm c}}\right);\ E_n(k_y,k_z) = \frac{\hbar^2k_z^2}{2m} + \hbar\omega_{\rm c}\left(n + \frac{1}{2}\right).$$

The quantum mechanical (Landau) diamagnetic susceptibility of a metal becomes

$$\chi_{\rm L} = -\frac{n_0}{2\zeta_0}\left(\frac{e\hbar}{2m^*c}\right)^2 = -\frac{n_0\mu_{\rm B}^2}{2\zeta_0}\left(\frac{m}{m^*}\right)^2.$$

Appearance of m^* (not m) indicates that the diamagnetism is associated with the orbital motion of the electrons.

In a metal, as we increase $\mathbf{B}$, the Landau level at $k_z = 0$ passes through the Fermi energy ζ and the internal energy abruptly decreases. Many physically observable properties of the system are periodic functions of the magnetic field. The periodic oscillation of the diamagnetic susceptibility of a metal at low temperatures is known as the **de Haas–van Alphen effect**. Oscillations in electrical conductivity are called the **Shubnikov–de Haas oscillations**.

Part II

Advanced Topics in Solid-State Physics

10

Magnetic Ordering and Spin Waves

10.1 Ferromagnetism

10.1.1 Heisenberg Exchange Interaction

The origin of the Weiss effective field is found in the *exchange field* between the interacting electrons on different atoms. For simplicity, assume that atoms A and B are neighbors and that each atom has one electron. Let Ψ_{A} and Ψ_{B} be the wave functions of the electron on atoms A and B, respectively. The Pauli principle requires that the wave function for the pair of electrons be antisymmetric. If we label the two indistinguishable electrons 1 and 2, this means

$$\Psi(1,2) = -\Psi(2,1). \tag{10.1}$$

The wave function for an electron has a spatial part and a spin part. Let $\eta_{i\uparrow}$ and $\eta_{i\downarrow}$ be the spin eigenfunctions for electron i in spin up and spin down states, respectively. There are two possible ways of obtaining an antisymmetric wave function for the pair (1,2).

$$\Psi_{\mathrm{I}} = \Phi_{\mathrm{S}}(r_1, r_2)\chi_{\mathrm{A}}(1,2) \tag{10.2}$$

$$\Psi_{\mathrm{II}} = \Phi_{\mathrm{A}}(r_1, r_2)\chi_{\mathrm{S}}(1,2). \tag{10.3}$$

The wave function Ψ_{I} has a symmetric space part and an antisymmetric spin part, and the wave function Ψ_{II} has an antisymmetric space part and a symmetric spin part. In (10.2) and (10.3), the space parts are

$$\Phi_{\substack{\mathrm{S}\\\mathrm{A}}}(r_1, r_2) = \frac{1}{\sqrt{2}}\left[\Psi_{\mathrm{a}}(1)\Psi_{\mathrm{b}}(2) \pm \Psi_{\mathrm{a}}(2)\Psi_{\mathrm{b}}(1)\right], \tag{10.4}$$

and χ_{A} and χ_{S} are the spin wave functions for the singlet ($s = 0$) spin state (which is antisymmetric) and for the triplet ($s = 1$) spin state (which is symmetric).

$$\chi_{\mathrm{A}}(1,2) = \frac{1}{\sqrt{2}}\left[\eta_{1\uparrow}\eta_{2\downarrow} - \eta_{1\downarrow}\eta_{2\uparrow}\right]; \qquad S_z = 0 \tag{10.5}$$

$$\chi_{\mathrm{S}}(1,2) = \begin{cases} \eta_{1\uparrow}\eta_{2\uparrow} & S_z = 1 \\ \frac{1}{\sqrt{2}}\left[\eta_{1\uparrow}\eta_{2\downarrow} + \eta_{1\downarrow}\eta_{2\uparrow}\right] & S_z = 0 \\ \eta_{1\downarrow}\eta_{2\downarrow} & S_z = -1 \end{cases} \tag{10.6}$$

If we consider the electron–electron interaction

$$V = \frac{e^2}{r_{12}}, \tag{10.7}$$

we can evaluate the expectation value of V in state Ψ_{I} or in state Ψ_{II}. Since V is independent of spin it is simple enough to see that

$$\begin{aligned} &\langle \Psi_{\mathrm{I}} | V | \Psi_{\mathrm{I}} \rangle = \langle \Phi_{\mathrm{S}} | V | \Phi_{\mathrm{S}} \rangle \\ &= \langle \Psi_{\mathrm{a}}(1)\Psi_{\mathrm{b}}(2) | V | \Psi_{\mathrm{a}}(1)\Psi_{\mathrm{b}}(2) \rangle + \langle \Psi_{\mathrm{a}}(1)\Psi_{\mathrm{b}}(2) | V | \Psi_{\mathrm{a}}(2)\Psi_{\mathrm{b}}(1) \rangle . \end{aligned} \tag{10.8}$$

When we do the same for Ψ_{II}, we obtain

$$\begin{aligned} &\langle \Psi_{\mathrm{II}} | V | \Psi_{\mathrm{II}} \rangle = \langle \Phi_{\mathrm{A}} | V | \Phi_{\mathrm{A}} \rangle \\ &= \langle \Psi_{\mathrm{a}}(1)\Psi_{\mathrm{b}}(2) | V | \Psi_{\mathrm{a}}(1)\Psi_{\mathrm{b}}(2) \rangle - \langle \Psi_{\mathrm{a}}(1)\Psi_{\mathrm{b}}(2) | V | \Psi_{\mathrm{a}}(2)\Psi_{\mathrm{b}}(1) \rangle . \end{aligned} \tag{10.9}$$

The two terms are called the *direct* and *exchange terms* and labeled V_{d} and $\mathcal{J}$, respectively. Thus, the expectation value of the Coulomb interaction between electrons is given by

$$\langle V \rangle = \begin{cases} V_{\mathrm{d}} + \mathcal{J} \text{ for the singlet state } (S = 0) \\ V_{\mathrm{d}} - \mathcal{J} \text{ for the triplet state } (S = 1). \end{cases} \tag{10.10}$$

Now, $\mathbf{S} = \hat{\mathbf{s}}_1 + \hat{\mathbf{s}}_2$ and $\mathbf{S}^2 = (\hat{\mathbf{s}}_1 + \hat{\mathbf{s}}_2)^2 = \hat{\mathbf{s}}_1^2 + \hat{\mathbf{s}}_2^2 + 2\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2$. Therefore, $\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2 = \frac{1}{2}(\hat{\mathbf{s}}_1 + \hat{\mathbf{s}}_2)^2 - \frac{1}{2}\hat{\mathbf{s}}_1^2 - \frac{1}{2}\hat{\mathbf{s}}_2^2 = \frac{1}{2}S(S+1) - \frac{3}{4}$. Here, we have used the fact that the operator $\mathbf{S}^2$ has eigenvalues $S(S+1)$ and $\hat{\mathbf{s}}_1^2$ and $\hat{\mathbf{s}}_2^2$ have eigenvalues $\frac{1}{2}(\frac{1}{2}+1) = \frac{3}{4}$. Thus, one can write

$$\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2 = \begin{cases} -\frac{3}{4} \text{ if } S = 0 \\ \frac{1}{4} \text{ if } S = 1. \end{cases}$$

Then, we write

$$\langle V \rangle = V_{\mathrm{d}} + \mathcal{J}\left(1 - \mathbf{S}^2\right) = V_{\mathrm{d}} - \frac{1}{2}\mathcal{J} - 2\mathcal{J}\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2. \tag{10.11}$$

Here, $-2\mathcal{J}\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2$ denotes the contribution to the energy from a pair of atoms (or ions) located at sites 1 and 2. For a large number of atoms we need only sum over all pairs to get

$$E = \text{constant} - \frac{1}{2}\sum_{i \neq j} 2\mathcal{J}_{ij}\hat{\mathbf{s}}_i \cdot \hat{\mathbf{s}}_j. \tag{10.12}$$

Normally, one assumes that $\mathcal{J}_{ij}$ is nonzero only for nearest neighbors and perhaps next nearest neighbors. The factor $\frac{1}{2}$ is introduced in order to avoid double counting of the interaction. The introduction of the interaction term $-2\mathcal{J}\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2$ is the source of the Weiss internal field which produces ferromagnetism. If z is the number of nearest neighbors of each atom i, then for atom i we have

$$E_{\text{ex}} = -2\mathcal{J}z\overline{S^2} = -g_{\text{L}}\mu_{\text{B}}\overline{S}H_{\text{E}}. \tag{10.13}$$

10.1.2 Spontaneous Magnetization

From our study of paramagnetism we know that

$$M = Ng_{\text{L}}\mu_{\text{B}}SB_{\text{S}}(x), \tag{10.14}$$

where $x = \frac{g_{\text{L}}\mu_{\text{B}}SB_{\text{Local}}}{k_{\text{B}}T}$. Here, B_{Local} is $B + \lambda M$, i.e. it includes the Weiss field. If we plot M vs. x, we get the behavior shown in Fig. 10.1. But for $B = 0$, $B_{\text{Local}} = \lambda M$. Therefore, $x = \frac{g_{\text{L}}\mu_{\text{B}}S\lambda M}{k_{\text{B}}T}$. If we plot this straight line x vs. M on the panel of Fig. 10.1 for different temperatures T we find the behavior shown in Fig. 10.2. Solutions (intersections) occur only at $(M = 0, x = 0)$ for $T > T_{\text{c}}$. For $T < T_{\text{c}}$ there is a solution at some nonzero value of M, i.e.

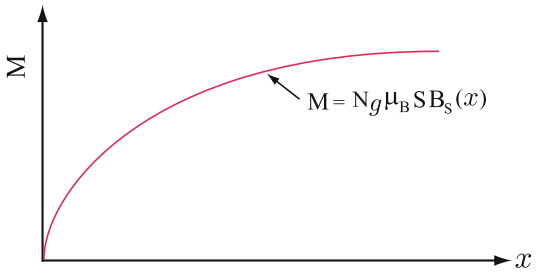


Fig. 10.1. Schematic plot of the magnetization M of a paramagnet as a function of the dimensionless parameter x defined by $x = \frac{g_{\text{L}}\mu_{\text{B}}SB_{\text{Local}}}{k_{\text{B}}T}$

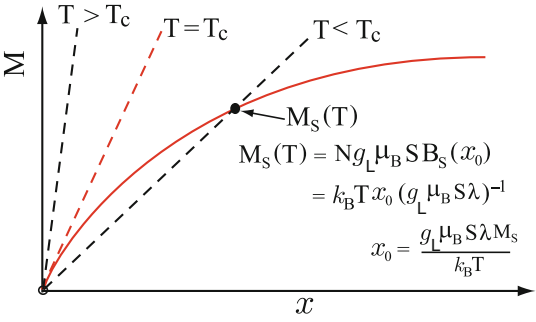


Fig. 10.2. Schematic plot of the magnetization M of a paramagnet for various different temperatures, in the absence of an external magnetic field, as a function of the dimensionless parameter x defined by $x = \frac{g_{\text{L}}\mu_{\text{B}}SB_{\text{Local}}}{k_{\text{B}}T}$

$$M_{\mathrm{S}}(T) = N g_{\mathrm{L}} \mu_{\mathrm{B}} S B_{\mathrm{S}}(x_0), x_0 = \frac{g_{\mathrm{L}} \mu_{\mathrm{B}} S \lambda M_{\mathrm{S}}}{k_{\mathrm{B}} T}.$$

The Curie temperature T_{C} is the temperature, at which the gradient of the line $M = \frac{k_{\mathrm{B}} T x}{g_{\mathrm{L}} \mu_{\mathrm{B}} S \lambda}$ and the curve $M = N g_{\mathrm{L}} \mu_{\mathrm{B}} S B_{\mathrm{S}}(x)$ are equal at the origin. Recall that, for small x, $B_{\mathrm{S}}(x) = \frac{(S+1)x}{3S} + O(x^3)$. Then the T_{C} is given by

$$T_{\mathrm{C}} = \frac{\lambda N \left[g_{\mathrm{L}} \mu_{\mathrm{B}} \sqrt{S(S+1)} \right]^2}{3 k_{\mathrm{B}}}. \tag{10.15}$$

It is not difficult to see that $M_{\mathrm{S}}(T)$ vs. T looks like Fig. 10.3.

If a finite external magnetic field B_0 is applied, then we have

$$M = \frac{k_{\mathrm{B}} T x}{g_{\mathrm{L}} \mu_{\mathrm{B}} S \lambda} - \frac{B_0}{\lambda}. \tag{10.16}$$

Plotting this straight line on the M–x plane gives the behavior shown in Fig. 10.4.

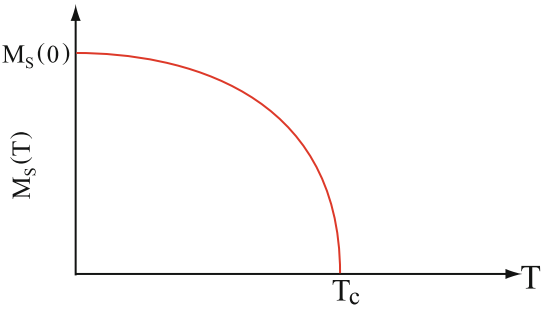


Fig. 10.3. Schematic plot of the spontaneous magnetization M_{S} as a function of temperature T

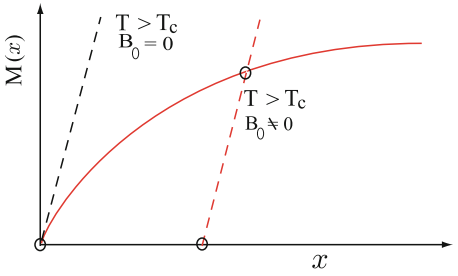


Fig. 10.4. Schematic plot of the magnetization M of a paramagnet, in the presence of an external magnetic field B_0, as a function of the dimensionless parameter x defined by $x = \frac{g_{\mathrm{L}} \mu_{\mathrm{B}} S B_{\mathrm{Local}}}{k_{\mathrm{B}} T}$

10.1.3 Domain Structure

If all the magnetic moments in a finite sample are lined up, then there will be flux emerging from the sample as shown in Fig. 10.5. There is an energy density $\frac{1}{8\pi}\mathbf{H}(\mathbf{r})\cdot\mathbf{B}(\mathbf{r})$ associated with this flux emerging from the sample, and the total *emergence energy* is given by

$$U = \frac{1}{8\pi}\int d^3r\, \mathbf{H}(\mathbf{r})\cdot\mathbf{B}(\mathbf{r}) \tag{10.17}$$

The emergence energy can be lowered by introducing a domain structures as shown in Fig. 10.6. To have more than a single domain, one must have a domain wall, and the domain wall has a positive energy per unit area.

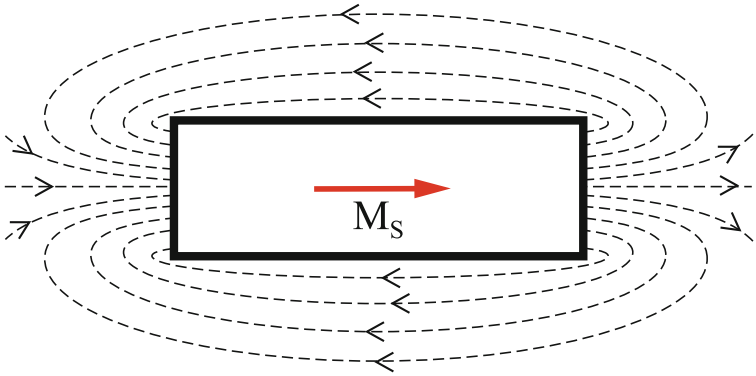


Fig. 10.5. Schematic plot of the magnetic flux around a sample with a single domain of finite spontaneous magnetization

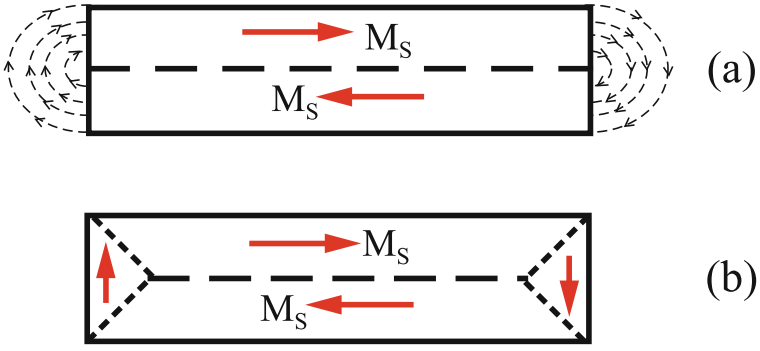


Fig. 10.6. Domain structures in a sample with finite spontaneous magnetization. (**a**) Pair of domains, (**b**) domains of closure

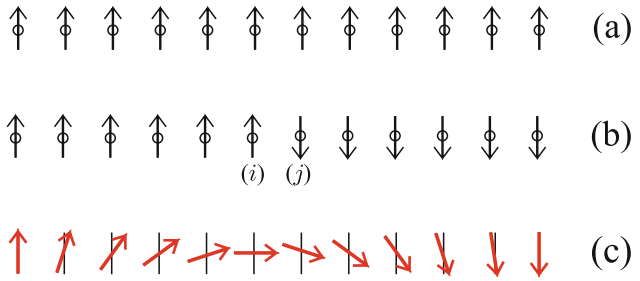


Fig. 10.7. A chain of magnetic spins interacting via Heisenberg exchange interaction. (**a**) Single domain, (**b**) a domain wall, (**c**) gradual spin flip

10.1.4 Domain Wall

Consider a chain of magnetic spins (Fig. 10.7a) interacting via Heisenberg exchange interaction

$$H_{\text{ex}} = -2\mathcal{J}\sum_{\langle i,j\rangle} \mathbf{s}_i \cdot \mathbf{s}_j,$$

where the sum is over all pairs of nearest neighbors. Compare the energy of this configuration with that having an abrupt domain wall as shown in Fig. 10.7b. Only spins (i) and (j) have a misalignment so that

$$\begin{aligned}\Delta E &= H_{\text{ex}}(i\uparrow, j\downarrow) - H_{\text{ex}}(i\uparrow, j\uparrow)\\ &= -2\mathcal{J}(\tfrac{1}{2})(-\tfrac{1}{2}) - \left[-2\mathcal{J}(\tfrac{1}{2})(\tfrac{1}{2})\right] = \mathcal{J}.\end{aligned} \tag{10.18}$$

Energetically it is more favorable to have the spin flip gradually as shown in Fig. 10.7c. If we assume the angle between each neighboring pair in the domain wall is Φ, we can write

$$(E_{\text{ex}})_{ij} = -2\mathcal{J}\mathbf{s}_i \cdot \mathbf{s}_j = -2\mathcal{J}s_i s_j \cos\Phi. \tag{10.19}$$

Now, if the spin turns through an angle Φ_0 ($\Phi_0 = \pi$ in the case shown in Fig. 10.7b) in N steps, where N is large, then $\Phi_{ij} \simeq \frac{\Phi_0}{N}$ within the domain wall, and we can approximate $\cos\Phi_{ij}$ by $\cos\Phi_{ij} \approx 1 - \frac{1}{2}\frac{\Phi_0^2}{N^2}$. Therefore, the exchange energy for a neighboring spin pair will be

$$(E_{\text{ex}})_{ij} = -2\mathcal{J}S^2\left(1 - \frac{1}{2}\frac{\Phi_0^2}{N^2}\right). \tag{10.20}$$

The increase in exchange energy due to the domain wall will be

$$E_{\text{ex}} = N\left(\mathcal{J}S^2\frac{\Phi_0^2}{N^2}\right) = \mathcal{J}S^2\frac{\Phi_0^2}{N}. \tag{10.21}$$

Clearly, the exchange energy is lower if the domain wall is very wide. In fact, if no other energies were involved, the domain wall width Na (where a is the atomic spacing) would be infinite. However, there is another energy involved, the *anisotropy energy*. Let us consider it next.

10.1.5 Anisotropy Energy

We realize that crystals are not spherically symmetric, but have finite point group symmetry. In real crystals, certain directions are *easy* to magnetize and others are *hard*. For example, Co is a hexagonal crystal. It is easy to magnetize Co along the hexagonal axis, but hard to magnetize it along any axis perpendicular to the hexagonal axis. The excess energy needed to magnetize the crystal in a direction that makes an angle θ with the hexagonal axis can be written

$$\frac{E_{\mathrm{A}}}{V} = K_1 \sin^2\theta + K_2 \sin^4\theta > 0. \tag{10.22}$$

For Fe, a cubic crystal, the $\langle 100\rangle$ directions are easy axes and the $\langle 111\rangle$ directions are hard. The anisotropy energy must reflect the cubic symmetry of the lattice. If we define $\alpha_i = \cos\theta_i$ as shown in Fig. 10.8, then an approximation to the anisotropy energy can be written

$$\frac{E_{\mathrm{A}}}{V} \approx K_1\left(\alpha_x^2\alpha_y^2 + \alpha_y^2\alpha_z^2 + \alpha_z^2\alpha_x^2\right) + K_2\alpha_x^2\alpha_y^2\alpha_z^2. \tag{10.23}$$

The constants K_1 and K_2 in (10.22) and (10.23) are called *anisotropy constants*. They are very roughly of the order of $10^5\,\mathrm{erg/cm^3}$.

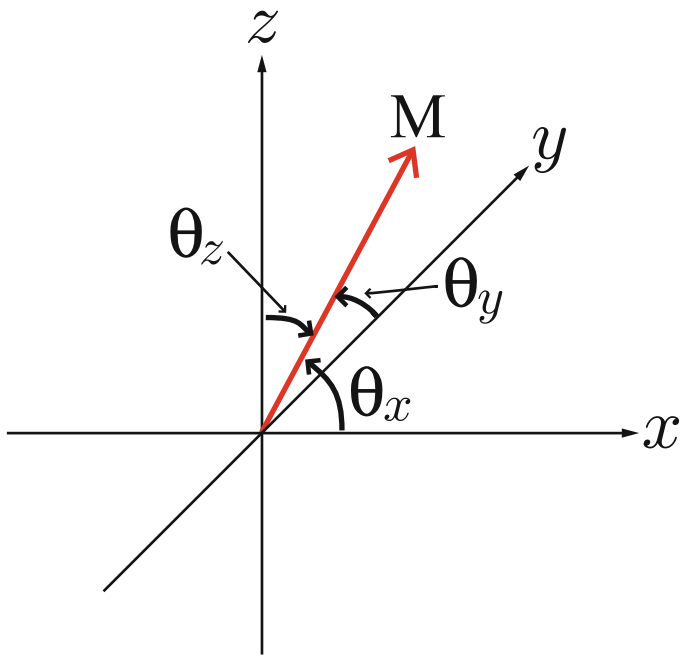


Fig. 10.8. Orientation of the magnetization with respect to the crystal axes in a cubic lattice

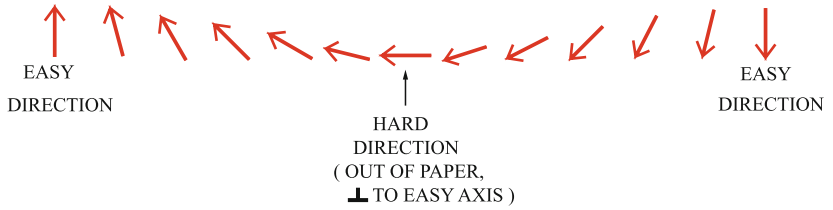


Fig. 10.9. Rotation of the magnetization in a domain wall

Clearly, if we make a domain wall, we must rotate the magnetization away from one easy direction and into another easy direction (see, for example, Fig. 10.9). To get an order of magnitude estimate of the domain wall thickness we can write the energy per unit surface area as the sum of the exchange contribution $\sigma_{\rm ex}$ and the anisotropy contribution $\sigma_{\rm A}$

$$\sigma_{\rm ex} = \frac{E_{\rm ex}}{a^2} = \frac{\mathcal{J}S^2\pi^2}{Na^2}, \tag{10.24}$$

where a is the atomic spacing. The anisotropy energy will be proportional to the anisotropy constant (energy per unit volume) times the number of spins times a.

$$\sigma_{\rm A} \approx KNa \simeq 10^2 - 10^7\,{\rm J/m^3}. \tag{10.25}$$

Thus the total energy per unit area will be

$$\sigma = \frac{\pi^2\mathcal{J}S^2}{Na^2} + KNa. \tag{10.26}$$

The σ has a minimum as a function of N, since the exchange part wants N to be very large and the anisotropy part wants it very small. At the minimum we have

$$N \simeq \left(\frac{\pi^2\mathcal{J}S^2}{Ka^3}\right)^{1/2} \approx 300. \tag{10.27}$$

The width of the domain wall is $\delta = Na \simeq \pi S(\frac{J}{Ka})^{1/2}$, and the energy per unit area of the domain wall is $\sigma \simeq 2\pi S(\frac{JK}{a})^{1/2}$.

10.2 Antiferromagnetism

For a Heisenberg ferromagnet we had an interaction Hamiltonian given by

$$\mathcal{H} = -2\mathcal{J}\sum_{\langle i,j\rangle} \mathbf{s}_i \cdot \mathbf{s}_j, \tag{10.28}$$

and the exchange constant $\mathcal{J}$ was positive. This made $\mathbf{s}_i$ and $\mathbf{s}_j$ align parallel to one another so that the energy was minimized. It is not uncommon to have

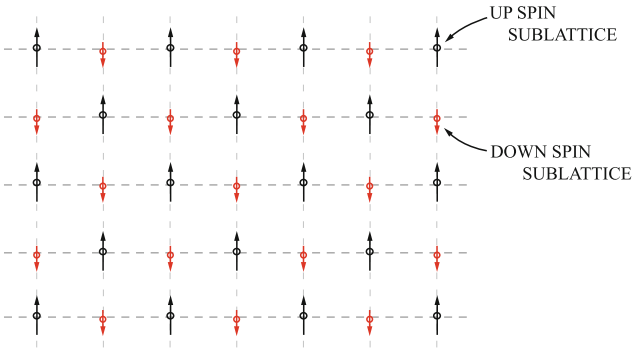


Fig. 10.10. Sublattice structure of spins in a ferrimagnet

spin systems in which $\mathcal{J}$ is negative. Then the Hamiltonian

$$\mathcal{H} = 2\,|\mathcal{J}| \sum_{\langle i,j \rangle} \mathbf{s}_i \cdot \mathbf{s}_j \tag{10.29}$$

will attempt to align the neighboring spins antiparallel. Materials with $\mathcal{J} < 0$ are called *antiferromagnets*.

For a antiferromagnet, the magnetic susceptibility increases as the temperature increases up to the transition temperature $T_{\mathrm{N}} = \frac{|\mathcal{J}|}{k_{\mathrm{B}}}$, known as the *Néel temperature*. Above T_{N}, the antiferromagnetic crystal is in the standard paramagnetic state.

10.3 Ferrimagnetism

In an antiferromagnet we can think of two different *sublattices* as shown in Fig. 10.10. If the two sublattices happened to have a different spin on each (e.g. up sublattice has $s = \frac{3}{2}$, down sublattice has $s = 1$), then instead of an antiferromagnet for $J < 0$, we have a *ferrimagnet*.

10.4 Zero-Temperature Heisenberg Ferromagnet

In the presence of an applied magnetic field B_0 oriented in the z-direction, the Hamiltonian of a Heisenberg ferromagnet can be written

$$\mathcal{H} = -\sum_{i,j} \mathcal{J}(\mathbf{R}_i - \mathbf{R}_j)\mathbf{S}_i \cdot \mathbf{S}_j - g\mu_{\mathrm{B}} B_0 \sum_i S_{iz}. \tag{10.30}$$

Here, we take the usual practice that the symbol $\mathbf{S}_i$ represents the total angular momentum of the ith ion and is parallel to the magnetic moment of the

ion, rather than opposite to the moment as was given by (9.26). The exchange integral $\mathcal{J}$ is defined as a half of the difference between the singlet and triplet energies. Let us define the operators $S^{\pm}$ by

$$S^{\pm} = S_x \pm \mathrm{i}S_y. \tag{10.31}$$

Remember that we can write $\mathbf{S}$ as

$$\mathbf{S} = \frac{1}{2}\sigma, \tag{10.32}$$

where σ_x, σ_y, σ_y are the Pauli spin matrices given by

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{10.33}$$

Let us choose units in which $\hbar = 1$. Then, S_x, S_y, and S_z satisfy the commutation relations

$$\begin{aligned} [S_x, S_y]_- &= \mathrm{i}S_z, \\ [S_y, S_z]_- &= \mathrm{i}S_x, \\ [S_z, S_x]_- &= \mathrm{i}S_y. \end{aligned} \tag{10.34}$$

We will be using the symbols S and S_z for quantum mechanical operators and for numbers associated with eigenvalues. Where confusion might arise we will write $\hat{\mathbf{S}}$ and $\hat{S}_z$ for the quantum mechanical operators. From quantum mechanics we know that $\hat{\mathbf{S}}^2$ and $\hat{S}_z$ can be diagonalized in the same representation since they commute. We usually write

$$\begin{aligned} \hat{\mathbf{S}}^2|S, S_z> &= S(S+1)|S, S_z\rangle, \\ \hat{S}_z|S, S_z\rangle &= \ S_z|S, S_z\rangle. \end{aligned} \tag{10.35}$$

Let us look at $\hat{S}^+$ operating on the state $|S, S_z\rangle$. We recall that

$$\begin{aligned} \left[\hat{\mathbf{S}}^2, \hat{S}^+\right] &= 0, \\ \left[\hat{S}_z, \hat{S}^+\right] &= \hat{S}^+. \end{aligned} \tag{10.36}$$

We can write

$$\hat{\mathbf{S}}^2\hat{S}^+|S, S_z\rangle = \hat{S}^+\hat{\mathbf{S}}^2|S, S_z\rangle + \left[\hat{\mathbf{S}}^2, \hat{S}^+\right]|S, S_z\rangle. \tag{10.37}$$

The second term vanishes because the commutator is zero, and $\hat{\mathbf{S}}^2|S, S_z\rangle = S(S+1)|S, S_z\rangle$ giving

$$\hat{\mathbf{S}}^2\hat{S}^+|S, S_z\rangle = S(S+1)\hat{S}^+|S, S_z\rangle. \tag{10.38}$$

Perform the same operation for $\hat{S}_z$ operating on $\hat{S}^+|S, S_z\rangle$ to have

$$\hat{S}_z\hat{S}^+|S, S_z\rangle = \hat{S}^+\hat{S}_z|S, S_z\rangle + \left[\hat{S}_z, \hat{S}^+\right]|S, S_z\rangle. \tag{10.39}$$

Use the fact that $\left[\hat{S}_z, \hat{S}^+\right] = \hat{S}^+$ and $\hat{S}_z|S, S_z\rangle = S_z|S, S_z\rangle$. This gives

$$\hat{S}_z\hat{S}^+|S, S_z\rangle = (S_z + 1)\hat{S}^+|S, S_z\rangle. \tag{10.40}$$

This means that $\hat{S}^+|S, S_z\rangle$ is proportional to $|S, S_z + 1\rangle$. To determine the normalization constant we write

$$\hat{S}^+|S, S_z\rangle = N|S, S_z + 1\rangle,$$

and note that

$$\{N|S, S_z + 1\rangle\}^\dagger = N^* < S, S_z + 1|$$

and

$$\{\hat{S}_z|S, S_z\rangle^\dagger = \langle S, S_z|\hat{S}_z$$

Thus, we have

$$|N|^2\langle S, S_z + 1|S, S_z + 1\rangle = \langle S, S_z|\hat{S}^-\hat{S}^+|S, S_z\rangle$$
$$= \langle S, S_z|(\hat{S}_x^2 + \hat{S}_y^2 - \hat{S}_z)|S, S_z\rangle = \langle S, S_z|(\hat{\mathbf{S}}^2 - \hat{S}_z^2 - \hat{S}_z)|S, S_z\rangle$$

giving for N

$$N = \left[S(S + 1) - S_z^2 - S_z\right]^{1/2}.$$

We can then show that

$$\begin{aligned}\hat{S}^+|S, S_z\rangle &= \sqrt{(S - S_z)(S + 1 + S_z)}|S, S_z + 1\rangle \\ \hat{S}^-|S, S_z\rangle &= \sqrt{(S + S_z)(S + 1 - S_z)}|S, S_z - 1\rangle\end{aligned} \tag{10.41}$$

Now, note that

$$S_{ix}S_{jx} + S_{iy}S_{jy} = \frac{1}{2}\left(S_i^+S_j^- + S_i^-S_j^+\right). \tag{10.42}$$

These are all operators, but we omit the $\hat{}$ over the S. The Heisenberg Hamiltonian, (10.30), becomes

$$\mathcal{H} = -\sum_{i,j}\mathcal{J}_{ij}S_{iz}S_{jz} - \frac{1}{2}\sum_{i,j}\mathcal{J}_{ij}\left(S_i^+S_j^- + S_i^-S_j^+\right) - g\mu_{\mathrm{B}}B_0\sum_i S_{iz}. \tag{10.43}$$

It is rather clear that the ground state will be obtained when all the spins are aligned parallel to one another and to the magnetic field $\mathbf{B}_0$. Let us define this state as $|\mathrm{GS}>$ or $|0>$. We can write

$$|0> = \prod_i |S, S\rangle_i. \tag{10.44}$$

Here, $|S, S\rangle_i$ is the state of the ith spin in which $\hat{S}_{iz}$ has the eigenvalue $S_z = S$, its maximum value. It is clear that $\hat{S}_i^+$ operating on $|0\rangle$ gives zero for every position i in the crystal. Therefore, $\mathcal{H}$ operating on $|0\rangle$ gives

$$\mathcal{H}|0\rangle = -\left(\sum_{i,j} \mathcal{J}_{ij}\hat{S}_{iz}\hat{S}_{jz} + g\mu_{\mathrm{B}}B_0\sum_i \hat{S}_{iz}\right)|0> . \tag{10.45}$$

Equation (10.45) shows that the state, in which all the spins are parallel and aligned along $\mathbf{B}_0 = (0, 0, B_0)$, so that S_z takes its maximum value S, has the lowest energy. The ground state energy becomes

$$E_0 = -S^2\sum_{i,j} \mathcal{J}_{ij} - Ng\mu_{\mathrm{B}}B_0S. \tag{10.46}$$

If $\mathcal{J}_{ij} = \mathcal{J}$ for nearest neighbor pairs and zero otherwise, then $\sum_{i,j} 1 = Nz$, where z is the number of nearest neighbors. Then E_0 reduces to

$$\begin{aligned} E_0 &= -S^2Nz\mathcal{J} - Ng\mu_{\mathrm{B}}B_0S \\ &= -g\mu_{\mathrm{B}}NS\left[B_0 + z\frac{\mathcal{J}S}{g^2\mu_{\mathrm{B}}^2}\right]. \end{aligned} \tag{10.47}$$

10.5 Zero-Temperature Heisenberg Antiferromagnet

If $\mathcal{J}$ is replaced by $-\mathcal{J}$ so that the exchange interaction tends to align neighboring spins in opposite directions, the ground state of the system is not quite simple. In fact, it has been solved exactly only for the special case of spin $\frac{1}{2}$ atoms in a one-dimensional chain by Hans Bethe. Let us set the applied magnetic field $B_0 = 0$. Then the Hamiltonian is given by

$$\mathcal{H} = \sum_{i,j} \mathcal{J}_{ij}\mathbf{S}_i \cdot \mathbf{S}_j. \tag{10.48}$$

If we assume that each sublattice acts as the ground state of the ferromagnet, but has S_z oriented in opposite directions on sublattices A and B, we would write a trial wave function

$$\Phi_{\mathrm{Trial}} = \prod_{\substack{\mathrm{i}\,\in\,\mathrm{A} \\ j\,\in\,\mathrm{B}}} |S, S\rangle_i|S, -S\rangle_j. \tag{10.49}$$

Remember that the Hamiltonian is

$$\mathcal{H} = \sum_{i,j} \mathcal{J}_{ij}\left(S_{iz}S_{jz} + \frac{1}{2}S_i^+S_j^- + \frac{1}{2}S_i^-S_j^+\right). \tag{10.50}$$

The $S_{iz}S_{jz}$ term would take its lowest possible value with this wave function, but unfortunately $S_i^-S_j^+$ operating on Φ_{Trial} would give a new wave function in which sublattice A has one atom with S_z having the value $S - 1$ and sublattice B has one with $S_z = -S + 1$. Thus, Φ_{Trial} is not an eigenfunction of $\mathcal{H}$.

10.6 Spin Waves in Ferromagnet

The Heisenberg Hamiltonian for a system (with unit volume) consisting of N spins with the nearest neighbor interaction can be written

$$\mathcal{H} = -2\mathcal{J}\sum_{\langle i,j\rangle}\hat{\mathbf{S}}_i\cdot\hat{\mathbf{S}}_j - g\mu_{\mathrm{B}}B_0\sum_i S_{iz}, \tag{10.51}$$

where the symbol $\langle i,j\rangle$ below the $\sum$ implies a sum over all distinct pairs of nearest neighbors. The constants of the motion are $\hat{\mathbf{S}}^2 = \sum_i \hat{\mathbf{S}}_i \cdot \sum_j \hat{\mathbf{S}}_j$ and $\hat{S}_z = \sum_j \hat{S}_{jz}$, where $\hat{\mathbf{S}} = \sum_j \hat{\mathbf{S}}_j$. The eigenvalues of $\hat{\mathbf{S}}^2$ and $\hat{S}_z$ are given by

$$\begin{aligned}\hat{\mathbf{S}}^2|0\rangle &= NS(NS+1)|0\rangle\\ \hat{S}_z|0\rangle &= NS|0\rangle.\end{aligned} \tag{10.52}$$

The ground state satisfies the equation

$$\mathcal{H}|0\rangle = -\left(g\mu_{\mathrm{B}}B_0NS + \mathcal{J}NzS^2\right)|0\rangle \tag{10.53}$$

10.6.1 Holstein–Primakoff Transformation

If we write $\hat{\mathbf{S}}_i\cdot\hat{\mathbf{S}}_j$ in terms of x, y, and z components of the spin operators, the Heisenberg Hamiltonian becomes

$$\mathcal{H} = -2\mathcal{J}\sum_{\langle i,j\rangle}\left(\hat{S}_{ix}\hat{S}_{jx} + \hat{S}_{iy}\hat{S}_{jy} + \hat{S}_{iz}\hat{S}_{jz}\right) - g\mu_{\mathrm{B}}B_0\sum_i \hat{S}_{iz}. \tag{10.54}$$

We can write

$$\hat{S}_{ix}\hat{S}_{jx} + \hat{S}_{iy}\hat{S}_{jy} = \frac{1}{2}\hat{S}_i^+\hat{S}_j^- + \frac{1}{2}\hat{S}_i^-\hat{S}_j^+. \tag{10.55}$$

Now, the Hamiltonian is rewritten

$$\mathcal{H} = -2\mathcal{J}\sum_{\langle i,j\rangle}\left(\frac{1}{2}\hat{S}_i^+\hat{S}_j^- + \frac{1}{2}\hat{S}_i^-\hat{S}_j^+ + \hat{S}_{iz}\hat{S}_{jz}\right) - g\mu_{\mathrm{B}}B_0\sum_i \hat{S}_{iz}. \tag{10.56}$$

The spin state of each atom is characterized by the value of S_z, which can take on any value between $-S$ and S separated by a step of unity. Because we are interested in low lying states, we will consider excited states in which the value of S_{iz} does not differ too much from its ground state value S. It is convenient to introduce an operator $\hat{n}_j$ defined by

$$\hat{n}_j = S_j - \hat{S}_{jz} = S - \hat{S}_{jz}. \tag{10.57}$$

$\hat{n}_j$ is called the *spin deviation operator*; it takes on the eigenvalues 0, 1, 2, $\cdots$, 2S telling us how much the value of S_z on site j differs from its ground

state value S. We now define $a_j^\dagger$ and its Hermitian conjugate a_j by

$$\hat{n}_j = a_j^\dagger a_j. \tag{10.58}$$

$a_j^\dagger$ and a_j are creation and annihilation operators for the j^{th} atom. We will require a_j and $a_j^\dagger$ to satisfy the commutation relation $[a_j, a_j^\dagger] = 1$, since a spin deviation looks like a boson. Notice that $a_j^\dagger$, which creates one spin deviation on site j, acts like the lowering operator S_j^- while a_j acts, by destroying one spin deviation on site j, like S_j^+. Therefore, we expect $a_j^\dagger$ to be proportional to S_j^- and a_j to be proportional to S_j^+. One can determine the coefficient by noting that

$$[\hat{S}^+, \hat{S}^-] = 2\hat{S}_z = 2(S - \hat{n}). \tag{10.59}$$

If we introduce the Holstein–Primakoff transformation to boson creation and annihilation operators $a_j^\dagger$ and a_j

$$\hat{S}_j^+ = (2S_j - \hat{n}_j)^{1/2} a_j \quad \text{and} \quad \hat{S}_j^- = a_j^\dagger (2S_j - \hat{n}_j)^{1/2} \tag{10.60}$$

and substitute into the expression for the commutator of $\hat{S}^+$ with $\hat{S}^-$, we obtain

$$[\hat{S}^+, \hat{S}^-] = 2(S - \hat{n}) \tag{10.61}$$

if $[a, a^\dagger] = 1$. The proof of (10.61) is given below. We want to show that $[\hat{S}^+, \hat{S}^-] = 2(S - \hat{n})$ if $[a, a^\dagger] = 1$. We start by defining $\hat{G} = (2S - \hat{n})^{1/2}$. Then, we can write

$$\begin{aligned}
[\hat{S}^+, \hat{S}^-] &= [\hat{G}a, a^\dagger \hat{G}] = \hat{G}[a, a^\dagger \hat{G}] + [\hat{G}, a^\dagger \hat{G}]a \\
&= \hat{G}a^\dagger [a, \hat{G}] + \hat{G}^2 + [\hat{G}, a^\dagger]\hat{G}a \\
&= \hat{G}^2 + \hat{n}\hat{G}^2 - a^\dagger \hat{G}^2 a.
\end{aligned}$$

But, we note that

$$\begin{aligned}
-a^\dagger \hat{G}^2 a &= -a^\dagger (2S - \hat{n})a = -a^\dagger \{[2S - \hat{n}, a] + a(2S - \hat{n})\} \\
&= -a^\dagger \left\{ -[\hat{n}, a] + a\hat{G}^2 \right\} = -a^\dagger \left\{ -[a^\dagger a, a] + a\hat{G}^2 \right\} \\
&= -a^\dagger \left\{ -[a^\dagger, a]a + a\hat{G}^2 \right\} = -\hat{n} - \hat{n}\hat{G}^2.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
[\hat{S}^+, \hat{S}^-] &= \hat{G}^2 + \hat{n}\hat{G}^2 - a^\dagger \hat{G}^2 a \\
&= \hat{G}^2 + \hat{n}\hat{G}^2 - \hat{n} - \hat{n}\hat{G}^2 \\
&= \hat{G}^2 - \hat{n} = 2(S - \hat{n}) = 2\hat{S}_z.
\end{aligned}$$

To obtain this result we had to require $[a^\dagger, a] = -1$. If we substitute (10.59) and (10.60) into the Hamiltonian, (10.56), we obtain

$$\mathcal{H} = -2\mathcal{J}S\sum_{\langle i,j\rangle}\left\{\sqrt{1-\frac{\hat{n}_i}{2S}}a_i a_j^\dagger\sqrt{1-\frac{\hat{n}_j}{2S}} + a_i^\dagger\sqrt{1-\frac{\hat{n}_i}{2S}}\sqrt{1-\frac{\hat{n}_j}{2S}}a_j + S\left(1-\frac{\hat{n}_i}{S}\right)\left(1-\frac{\hat{n}_j}{S}\right)\right\} - g\mu_{\mathrm{B}}B_0 S\sum_i\left(1-\frac{\hat{n}_i}{S}\right). \tag{10.62}$$

So far, we have made no approximation other than those inherent in the Heisenberg model. Now, we will make the approximation that $\langle\hat{n}_i\rangle \ll 2S$ for all states of interest. Therefore, in an expansion of the operator $\sqrt{1-\frac{\hat{n}_i}{2S}}$ we will keep only terms up to those linear in $\hat{n}_i$, i.e.

$$\sqrt{1-\frac{\hat{n}_i}{2S}} \simeq 1-\frac{\hat{n}_i}{4S}+\cdots. \tag{10.63}$$

We make this substitution into the Heisenberg Hamiltonian and write $\mathcal{H}$ as

$$\mathcal{H} = E_0 + \mathcal{H}_0 + \mathcal{H}_1. \tag{10.64}$$

Here, E_0 is the ground state energy that we obtained by assuming that the ground state wave function was $|0\rangle = \prod_i |S, S_z = S\rangle_i$.

$$\begin{aligned} E_0 &= -2S^2\textstyle\sum_{\langle i,j\rangle}\mathcal{J}_{ij} - Ng\mu_{\mathrm{B}}B_0 \\ &= -z\mathcal{J}NS^2 - g\mu_{\mathrm{B}}B_0 NS. \end{aligned} \tag{10.65}$$

$\mathcal{H}_0$ is the part of the Hamiltonian that is quadratic in the spin deviation creation and annihilation operators.

$$\mathcal{H}_0 = (g\mu_{\mathrm{B}}B_0 + 2z\mathcal{J}S)\sum_i\hat{n}_i - 2\mathcal{J}S\sum_{\langle i,j\rangle}\left(a_i a_j^\dagger + a_i^\dagger a_j\right). \tag{10.66}$$

$\mathcal{H}_1$ includes all higher terms. To fourth order in $a^\dagger$'s and a's the expression for $\mathcal{H}_1$ is given explicitly by

$$\begin{aligned} \mathcal{H}_1 = -2\mathcal{J}\textstyle\sum_{\langle i,j\rangle}\Big(&\hat{n}_i\hat{n}_j - \tfrac{1}{4}\hat{n}_i a_i a_j^\dagger - \tfrac{1}{4}a_i a_j^\dagger\hat{n}_j - \tfrac{1}{4}\hat{n}_j a_i^\dagger a_j - \tfrac{1}{4}a_i^\dagger a_j\hat{n}_i\Big) \\ &+ \text{higher order terms.} \end{aligned} \tag{10.67}$$

Let us concentrate on $\mathcal{H}_0$. It is apparent that $a_i^\dagger a_j$ transfers a spin deviation from the j^{th} atom to the ith atom. Thus, a state with a spin deviation on the jth atom is not an eigenstate of $\mathcal{H}$. This problem is similar to that which we encountered in studying lattice vibrations. By this, we mean that spin deviations on neighboring sites are coupled together in the same way that atomic displacements of neighboring atoms are coupled in lattice dynamics. As we did in studying phonons, we will introduce new variables that we call *magnon*

or *spin wave* variables defined as follows:

$$b_{\mathbf{k}} = N^{-1/2} \sum_j \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}_j}\, a_j \quad \text{and} \quad b_{\mathbf{k}}^\dagger = N^{-1/2} \sum_j \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}_j}\, a_j^\dagger. \tag{10.68}$$

As usual the inverse can be written

$$a_j = N^{-1/2} \sum_{\mathbf{k}} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}_j}\, b_{\mathbf{k}} \quad \text{and} \quad a_j^\dagger = N^{-1/2} \sum_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}_j}\, b_{\mathbf{k}}^\dagger. \tag{10.69}$$

It is straightforward (but left as an exercise) to show, because $[a_j, a_{j'}] = [a_j^\dagger, a_{j'}^\dagger] = 0$ and $[a_j, a_{j'}^\dagger] = \delta_{jj'}$, that

$$[b_{\mathbf{k}}, b_{\mathbf{k}'}] = \left[b_{\mathbf{k}}^\dagger, b_{\mathbf{k}'}^\dagger\right] = 0 \quad \text{and} \quad \left[b_{\mathbf{k}}, b_{\mathbf{k}'}^\dagger\right] = \delta_{\mathbf{k}\mathbf{k}'}. \tag{10.70}$$

Substitute into $\mathcal{H}_0$ the expression for spin deviation operators in terms of the magnon operators; this gives

$$\begin{aligned}\mathcal{H}_0 = {} & (g\mu_{\mathrm{B}}B_0 + 2z\mathcal{J}S) \sum_j N^{-1} \sum_{\mathbf{k}\mathbf{k}'} \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_j} b_{\mathbf{k}}^\dagger b_{\mathbf{k}'} \\ & - 2\mathcal{J}SN^{-1} \sum_{\langle j,l\rangle} \sum_{\mathbf{k}\mathbf{k}'} \left(\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}_l - \mathrm{i}\mathbf{k}'\cdot\mathbf{x}_j} b_{\mathbf{k}'} b_{\mathbf{k}}^\dagger + \mathrm{e}^{\mathrm{i}\mathbf{k}'\cdot\mathbf{x}_j - \mathrm{i}\mathbf{k}\cdot\mathbf{x}_l} b_{\mathbf{k}'}^\dagger b_{\mathbf{k}}\right)\end{aligned} \tag{10.71}$$

We introduce $\boldsymbol{\delta}$, one of the nearest neighbor vectors connecting neighboring sites and write $\mathbf{x}_l = \mathbf{x}_j + \boldsymbol{\delta}$ in the summation $\sum_{\langle j,l\rangle}$, so that it becomes $\frac{1}{2}\sum_j \sum_\delta = \frac{1}{2} zN$. We also make use of the fact that

$$\sum_j \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_j} = \delta_{\mathbf{k}\mathbf{k}'} N. \tag{10.72}$$

Then, $\mathcal{H}_0$ can be expressed as

$$\mathcal{H}_0 = (g\mu_{\mathrm{B}}B_0 + 2z\mathcal{J}S) \sum_{\mathbf{k}} b_{\mathbf{k}}^\dagger b_{\mathbf{k}} - \mathcal{J}S \sum_{\mathbf{k}} \sum_{\boldsymbol{\delta}} \left(\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\boldsymbol{\delta}} b_{\mathbf{k}} b_{\mathbf{k}}^\dagger + \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\boldsymbol{\delta}} b_{\mathbf{k}}^\dagger b_{\mathbf{k}}\right). \tag{10.73}$$

We now define

$$\gamma_{\mathbf{k}} = z^{-1} \sum_{\boldsymbol{\delta}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\boldsymbol{\delta}}. \tag{10.74}$$

If there is a center of symmetry about each atom then $\gamma_{-\mathbf{k}} = \gamma_{\mathbf{k}}$. Further, since $\sum_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}} = 0$ unless $\mathbf{R} = 0$, it is apparent that $\sum_{\mathbf{k}} \gamma_{\mathbf{k}} = 0$. Using these results in our expression for $\mathcal{H}_0$ gives

$$\mathcal{H}_0 = \sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}} b_{\mathbf{k}}^\dagger b_{\mathbf{k}}, \tag{10.75}$$

where

$$\hbar\omega_{\mathbf{k}} = 2z\mathcal{J}S(1 - \gamma_{\mathbf{k}}) + g\mu_{\mathrm{B}}B_0. \tag{10.76}$$

Thus, if we neglect $\mathcal{H}_1$, we have for the Hamiltonian of a state containing magnons

$$\mathcal{H} = -\left(g\mu_{\mathrm{B}}B_0 NS + z\mathcal{J}NS^2\right) + \sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}} b_{\mathbf{k}}^{\dagger} b_{\mathbf{k}}. \tag{10.77}$$

This tells us that the elementary excitations are waves (remember $b_{\mathbf{k}}^{\dagger} = N^{-1/2}\sum_j \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}_j} a_j^{\dagger}$ is a linear combination of spin deviations shared equally in amplitude by all sites) of energy $\hbar\omega_{\mathbf{k}}$. Provided that we stay at low enough temperature so that $\langle\hat{n}_j\rangle \ll S$, this approximation is rather good. At higher temperatures, where many spin waves are excited, the higher terms (spin wave–spin wave interactions) become important.

10.6.2 Dispersion Relation for Magnons

For long wave lengths $|\mathbf{k}\cdot\boldsymbol{\delta}| \ll 1$. In this region, we can expand $\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\boldsymbol{\delta}}$ in powers of $\mathbf{k}$ to get

$$\gamma_{\mathbf{k}} = z^{-1}\sum_{\boldsymbol{\delta}} \left(1 + \mathrm{i}\mathbf{k}\cdot\boldsymbol{\delta} - \frac{(\mathbf{k}\cdot\boldsymbol{\delta})^2}{2} + \cdots\right). \tag{10.78}$$

Using $\sum_{\boldsymbol{\delta}} 1 = z$, and $\sum_{\boldsymbol{\delta}} \boldsymbol{\delta} = 0$ gives

$$\gamma_{\mathbf{k}} \approx 1 - \frac{1}{2z}\sum_{\boldsymbol{\delta}} (\mathbf{k}\cdot\boldsymbol{\delta})^2. \tag{10.79}$$

Thus, $z(1-\gamma_{\mathbf{k}}) \simeq \frac{1}{2}\sum_{\boldsymbol{\delta}} (\mathbf{k}\cdot\boldsymbol{\delta})^2$ and in this limit we have

$$\hbar\omega_{\mathbf{k}} = g\mu_{\mathrm{B}}B_0 + \mathcal{J}S\sum_{\boldsymbol{\delta}} (\mathbf{k}\cdot\boldsymbol{\delta})^2. \tag{10.80}$$

For a simple cubic lattice $|\boldsymbol{\delta}| = a$ and $\sum_{\boldsymbol{\delta}} (\mathbf{k}\cdot\delta)^2 = 2k^2a^2$ giving

$$\hbar\omega_{\mathbf{k}} = g\mu_{\mathrm{B}}B_0 + 2\mathcal{J}Sa^2k^2. \tag{10.81}$$

In a simple cubic lattice, the magnon energy is of the same form as the energy of a free particle in a constant potential $\varepsilon = V_0 + \frac{\hbar^2k^2}{2m^*}$ where $V_0 = g\mu_{\mathrm{B}}B_0$ and $\frac{1}{m^*} = \frac{4\mathcal{J}Sa^2}{\hbar^2}$.

The dispersion relation we have been considering is appropriate for a Bravais lattice. In reciprocal space the $\mathbf{k}$ values will, as is usual in crystalline materials, be restricted to the first Brillouin zone. For a lattice with more than one spin per unit cell, *optical magnons* as well as *acoustic magnons* are found, as is shown in Fig. 10.11.

10.6.3 Magnon–Magnon Interactions

The terms in $\mathcal{H}_1$ that we have omitted involve more than two spin deviation creation and annihilation operators. These terms are responsible for magnon–magnon scattering just as cubic and quartic anharmonic terms are responsible

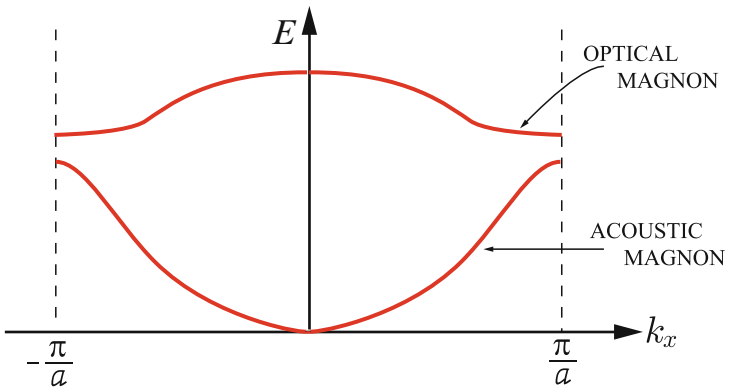


Fig. 10.11. Magnon dispersion curves

for phonon–phonon scattering. Freeman J. Dyson studied the leading terms associated with magnon–magnon scattering.[1] Rigorous treatment of magnon–magnon scattering is mathematically difficult.

10.6.4 Magnon Heat Capacity

If the external magnetic field is zero and if magnon–magnon interactions are neglected, then we can write the magnon frequency as $\omega_{\mathbf{k}} = Dk^2$ for small values of k. Here $D = 2\mathcal{J}Sa^2$. The internal energy per unit volume associated with these excitations is given by (we put $\hbar = 1$ for convenience)

$$U = \frac{1}{V}\sum_{\mathbf{k}} \omega_{\mathbf{k}} \langle n_k \rangle \tag{10.82}$$

where $\langle n_k \rangle = \frac{1}{\mathrm{e}^{\omega_{\mathbf{k}}/\Theta}-1}$ is the Bose–Einstein distribution function since magnons are Bosons. Converting the sum to an integral over $\mathbf{k}$ gives

$$U = \frac{1}{(2\pi)^3}\int_{\mathrm{BZ}} \mathrm{d}^3k \frac{Dk^2}{\mathrm{e}^{Dk^2/\Theta}-1}. \tag{10.83}$$

Let $Dk^2 = \Theta x^2$; then U becomes

$$U = \frac{D}{2\pi^2}\left(\frac{\Theta}{D}\right)^{5/2}\int \mathrm{d}x \frac{x^4}{\mathrm{e}^{x^2}-1}. \tag{10.84}$$

Here, we have used $\mathrm{d}^3k = 4\pi k^2 \mathrm{d}k$. Let $x^2 = y$ and set the upper limit at $y_{\mathrm{M}} = \left(\frac{Dk_{\mathrm{M}}}{\Theta}\right)^2$. Then, we find

$$U = \frac{\Theta^{5/2}}{4\pi^2 D^{3/2}}\int_0^{y_{\mathrm{M}}} \mathrm{d}y \frac{y^{3/2}}{\mathrm{e}^y - 1}. \tag{10.85}$$

[1] F.J. Dyson, Phys. Rev. **102**, 1217 (1956).

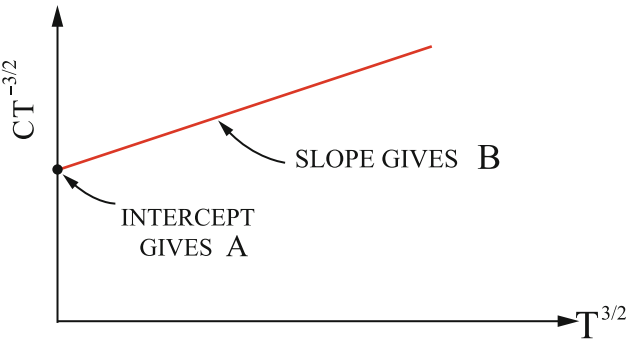


Fig. 10.12. Specific heat of an insulating ferromagnet

For very low temperatures $\Theta \ll \omega_{\mathrm{M}}$ and no serious error is made by replacing y_{M} by ∞. Then, the integral becomes

$$\int_0^\infty \mathrm{d}y \frac{y^{3/2}}{\mathrm{e}^y - 1} = \Gamma\left(\frac{5}{2}\right) \zeta\left(\frac{5}{2}, 1\right). \tag{10.86}$$

Here, $\Gamma(x)$ and $\zeta(a, b)$ are the Γ function and Riemann zeta function, respectively: $\Gamma(\frac{5}{2}) = \frac{3}{2} \cdot \frac{1}{2}\Gamma(\frac{1}{2}) = \frac{3}{4}\sqrt{\pi}$ and $\zeta(\frac{5}{2}, 1) \approx 1.341$. Thus for U we obtain

$$U \simeq \frac{0.45}{\pi^2} \frac{\Theta^{5/2}}{D^{3/2}} \tag{10.87}$$

and for the specific heat due to magnons

$$C = \frac{\partial U}{\partial T} = 0.113 k_{\mathrm{B}} \left(\frac{\Theta}{D}\right)^{3/2} \tag{10.88}$$

For an insulating ferromagnet the specific heat contains contributions due to phonons and due to magnons. At low temperatures we have

$$C = AT^{3/2} + BT^3 \tag{10.89}$$

Plotting $CT^{-3/2}$ as a function of $T^{3/2}$ at low temperature should give a straight line. (See, for example, Fig. 10.12.) For the ideal Heisenberg ferromagnet YIG (yttrium iron garnet) D has a value approximately $0.8\,\mathrm{erg} \cdot \mathrm{cm}^2$ implying an effective mass $m^* \simeq 6\, m_{\mathrm{e}}$.

10.6.5 Magnetization

The thermal average of the magnetization at a temperature T is referred to as the *spontaneous magnetization* at temperature T. It is given by

$$M_{\mathrm{s}} = \frac{g\mu_{\mathrm{B}}}{V} \left(NS - \left\langle \sum_{\mathbf{k}} b_k^\dagger b_k \right\rangle \right). \tag{10.90}$$

The first term is just the zero temperature value where $S_z = NS$ and $g\mu_\mathrm{B} = 2\mu$. The second term results from the presence of spin deviations $\hat{n}_j$. Remember that

$$\begin{aligned}\textstyle\sum_j \hat{n}_j &= \textstyle\sum_j a_j^\dagger a_j = \sum_j \frac{1}{N} \sum_{\mathbf{k}\mathbf{k}'} \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_j} b_\mathbf{k}^\dagger b_{\mathbf{k}'} \\ &= \textstyle\sum_{\mathbf{k}\mathbf{k}'} b_\mathbf{k}^\dagger b_{\mathbf{k}'} \delta_{\mathbf{k}\mathbf{k}'} = \sum_\mathbf{k} b_\mathbf{k}^\dagger b_\mathbf{k}.\end{aligned} \tag{10.91}$$

We can define

$$\Delta M = M_\mathrm{s}(0) - M_\mathrm{s}(T) = \frac{2\mu}{V} \sum_\mathbf{k} \langle n_k \rangle \,,$$

where $\langle n_k \rangle = \frac{1}{\mathrm{e}^{Dk^2/\Theta}-1}$. Replacing the sum over the wave number k by an integral in the usual way gives

$$\begin{aligned}\Delta M &= \frac{2\mu}{(2\pi)^3} 4\pi \int \frac{\mathrm{d}k\, k^2}{\mathrm{e}^{Dk^2/\Theta} - 1} \\ &= \frac{\mu}{2\pi^2} \left(\frac{\Theta}{D}\right)^{3/2} \int_0^{y_\mathrm{M}} \frac{\mathrm{d}y\, y^{1/2}}{\mathrm{e}^y - 1}.\end{aligned} \tag{10.92}$$

Again if $\Theta \ll \hbar\omega_\mathrm{M}$, y_M can be replaced by ∞. Then the definite integral has the value $\Gamma(\frac{3}{2})\zeta(\frac{3}{2}, 1)$, and we obtain for ΔM

$$\Delta M = 0.117\mu \left(\frac{\Theta}{D}\right)^{3/2} = 0.117\mu \left(\frac{\Theta}{2a^2 S\mathcal{J}}\right)^{3/2}. \tag{10.93}$$

For $M_\mathrm{s}(T)$ we can write

$$M_\mathrm{s}(T) = \frac{N}{V} 2\mu S - 0.117 \frac{\mu}{a^3} \left(\frac{\Theta}{2S\mathcal{J}}\right)^{3/2}. \tag{10.94}$$

For simple cubic, bcc, and fcc lattices, $\frac{N}{V}$ has the values $1/a^3$, $2/a^3$, and $4/a^3$, respectively. Thus, we can write

$$M_\mathrm{s}(T) \simeq \frac{2\mu S}{a^3} \left[\alpha - 0.02 \frac{\Theta^{3/2}}{S^{5/2}\mathcal{J}^{3/2}}\right], \tag{10.95}$$

where $\alpha = 1, 2, 4$ for simple cubic, bcc, and fcc lattices, respectively. The $T^{3/2}$ dependence of the magnetization is a well-known result associated with the presence of noninteracting spin waves. Higher order terms in $\frac{\Theta}{\mathcal{J}}$ are obtained if the full expression for $\gamma_\mathbf{k}$ is used instead of just the long wave length expansion (correct up to k^2 term) and the k-integral is performed over the first Brillouin zone and not integrated to infinity. The first nonideal magnon term, resulting from magnon–magnon interactions, is a term of order $\left(\frac{\Theta}{\mathcal{J}}\right)^4$. F.J. Dyson obtained this term correctly in a classic paper in the mid 1950s.[2]

[2] F.J. Dyson, Phys. Rev **102**, 1230 (1956).

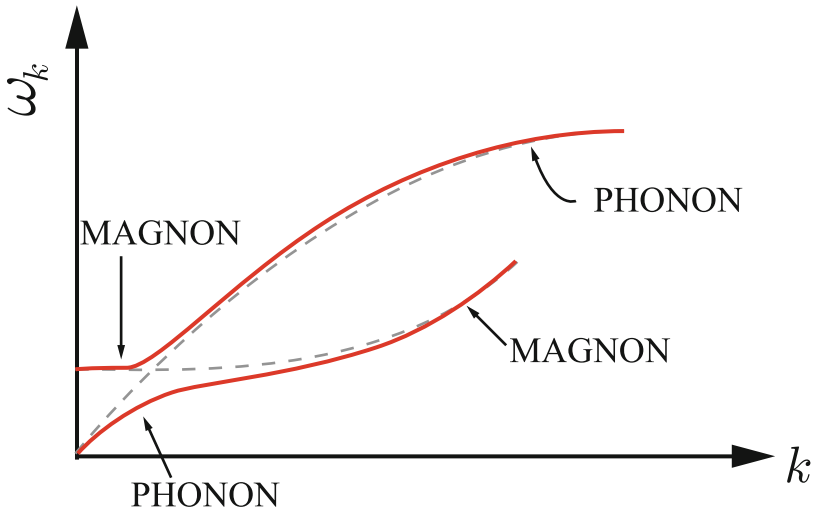


Fig. 10.13. Coupling of magnons–phonons

10.6.6 Experiments Revealing Magnons

Among the many experiments which demonstrate the existence of magnons, a few important ones are as follows:

1. The existence of side bands in ferromagnetic resonances. The uniform precession mode in a ferromagnetic resonance experiment excites a $k = 0$ spin wave. In a ferromagnetic film, it is possible to couple to modes with wave length λ satisfying $\frac{1}{2}\lambda = \frac{d}{n}$ where d is the thickness of the film. This gives resonances at magnon wave numbers $k_n = \frac{n\pi}{d}$.
2. The existence of inelastic neutron scattering peaks associated with magnons.
3. The coupling of magnons to phonons in ferromagnetic crystals (see Fig. 10.13).

10.6.7 Stability

We started with a Heisenberg Hamiltonian $\mathcal{H} = -\mathcal{J}\sum_{\langle i,j\rangle} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j$. In the ferromagnetic ground state the spins are aligned. However, the direction of the resulting magnetization is arbitrary (since $\mathcal{H}$ has complete rotational symmetry) so that the ground state is degenerate. If one selects a certain direction for $\mathbf{M}$ as the starting point of magnon theory, the system is found to be unstable. Infinitesimal amount of thermal energy excites a very large number of spin waves (remembering that when $B_0 = 0$ the $k = 0$ spin waves have zero energy). The difficulty of having an unstable ground state with $\mathbf{M}$ in a particular direction is removed by removing the degeneracy caused by spherical symmetry of the Hamiltonian. This is accomplished by either

1. Applying a field $\mathbf{B}_0$ in a particular direction or
2. Introducing an effective anisotropy field B_{A}.

For $\Theta \ll \mu_{\mathrm{B}}|\mathbf{B}|$ where $\mathbf{B}$ is either $\mathbf{B}_0$ or $\mathbf{B}_{\mathrm{A}}$, only small deviations from the ground state occur. The anisotropy field is a mathematical convenience which accounts for anisotropic interaction in real crystals. It is not so important in ferromagnets, but it is very important in antiferromagnets

10.7 Spin Waves in Antiferromagnets

The Heisenberg Hamiltonian of an antiferromagnet has $\mathcal{J} > 0$ so that

$$\mathcal{H} = +2\mathcal{J} \sum_{\langle i,j \rangle} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j - g\mu_{\mathrm{B}} \mathbf{B}_0 \cdot \sum_i \hat{\mathbf{S}}_i, \tag{10.96}$$

where the sum is over all possible distinct nearest neighbor pairs. The state in which all N spins on sublattice 1 are $\uparrow$ and all N spins on sublattice 2 are $\downarrow$ is a highly degenerate state because the direction for $\uparrow$ (or $\downarrow$) is completely arbitrary. This degeneracy is not removed by introducing an external field $\mathbf{B}_0$. For $|\mathbf{B}_0|$ not too large, the spins align themselves antiferromagnetically in the plane perpendicular to $\mathbf{B}_0$. However, the direction of a given sublattice magnetization is still arbitrary in that plane.

Lack of stability can be overcome by introducing an anisotropy field B_{A} with the following propeties:

1. $\mathbf{B}_{\mathrm{A}}$ is in the $+z$-direction at sites in sublattice 1.
2. $\mathbf{B}_{\mathrm{A}}$ is in the $-z$-direction at sites in sublattice 2.
3. $\mu_{\mathrm{B}} B_{\mathrm{A}}$ is not too small (compared to $\frac{1}{N}\mathcal{J}$).

Then, the Heisenberg Hamiltonian for an antiferromagnet in the presence of an applied field $\mathbf{B}_0 = B_0 \hat{z}$ and an anisotropy field $\mathbf{B}_{\mathrm{A}}$ can be written

$$\mathcal{H} = +\mathcal{J} \sum_{\langle i,j \rangle} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j - g\mu_{\mathrm{B}}(B_{\mathrm{A}} + B_0) \sum_{l \in \mathrm{a}} \hat{S}^{\mathrm{a}}_{lz} + g\mu_{\mathrm{B}}(B_{\mathrm{A}} - B_0) \sum_{p \in \mathrm{b}} \hat{S}^{\mathrm{b}}_{pz}. \tag{10.97}$$

The superscripts a and b refer to the two sublattices. In the limit where $B_{\mathrm{A}} \to \infty$ while $\mathcal{J} \to 0$ and $B_0 \to 0$, the ground state will have

$$\begin{aligned} S^{\mathrm{a}}_{lz} &= S \quad \text{for all } l \in \mathrm{a} \\ S^{\mathrm{b}}_{pz} &= -S \quad \text{for all } p \in \mathrm{b} \end{aligned} \tag{10.98}$$

This state is not true ground state of the system when B_{A} and $\mathcal{J}$ are both finite. The spin wave theory of an antiferromagnet can be carried out in analogy with the treatment for the ferromagnet. We introduce spin deviations from the "$B_{\mathrm{A}} \to \infty$ ground state" by writing

$$\begin{aligned} S^{\mathrm{a}}_{lz} &= S - a^\dagger_l a_l \quad \text{for all } l \in \mathrm{a} \\ S^{\mathrm{b}}_{pz} &= -(S - b^\dagger_p b_p) \quad \text{for all } p \in \mathrm{b}, \end{aligned} \tag{10.99}$$

where the spin deviation operators satisfy commutation relations $\left[a_l, a_l^\dagger\right] = 1$ and $\left[b_p, b_p^\dagger\right] = 1$. Once again it is easy to show that

$$\begin{array}{l} \hat{S}_l^{\mathrm{a}+} = (2S - \hat{n}_l)^{1/2} a_l; \quad \hat{S}_l^{\mathrm{a}-} = a_l^\dagger (2S - \hat{n}_l)^{1/2} \\ \hat{S}_p^{\mathrm{b}+} = b_p^\dagger (2S - \hat{m}_p)^{1/2}; \; \hat{S}_p^{\mathrm{b}-} = (2S - \hat{m}_p)^{1/2} b_p. \end{array} \tag{10.100}$$

Here, $\hat{n}_l = a_l^\dagger a_l$ and $\hat{m}_p = b_p^\dagger b_p$. In spin wave theory, we assume $\langle \hat{n}_l \rangle \ll 2S$ and $\langle \hat{m}_p \rangle \ll 2S$ and expand the square roots keeping only linear terms in $\hat{n}_l$ and $\hat{m}_p$. The Hamiltonian can then be written

$$\mathcal{H} = E_0 + \mathcal{H}_0 + \mathcal{H}_1. \tag{10.101}$$

Here, E_0 is the ground state energy given by

$$E_0 = -2Nz\mathcal{J}S^2 - 2g\mu_{\mathrm{B}} B_{\mathrm{A}} NS, \tag{10.102}$$

and $\mathcal{H}_0$ is the part of the Hamiltonian that is quadratic in the spin deviation creation and annihilation operators

$$\begin{array}{l} \mathcal{H}_0 = 2\mathcal{J}S \sum_{\langle l,p \rangle}{}' \left(a_l b_p + a_l^\dagger b_p^\dagger + \hat{n}_l + \hat{m}_p \right) \\ \qquad + g\mu_{\mathrm{B}} \left(B_{\mathrm{A}} + B_0 \right) \sum_{l \in \mathrm{a}} \hat{n}_l + g\mu_{\mathrm{B}} \left(B_{\mathrm{A}} - B_0 \right) \sum_{p \in \mathrm{b}} \hat{m}_p. \end{array} \tag{10.103}$$

The sum of products of a's and b's is over nearest neighbor pairs. $\mathcal{H}_1$ is a sum of an infinite number of terms each containing at least four a or b operators or their Hermitian conjugates. We can again introduce spin wave variables

$$\begin{array}{l} c_{\mathbf{k}} = N^{-1/2} \sum_l \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}_l} \, a_l, \quad c_{\mathbf{k}}^\dagger = N^{-1/2} \sum_l \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}_l} \, a_l^\dagger, \\ d_{\mathbf{k}} = N^{-1/2} \sum_p \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}_p} \, b_p, \; d_{\mathbf{k}}^\dagger = N^{-1/2} \sum_p \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}_p} \, b_p^\dagger. \end{array} \tag{10.104}$$

In terms of the spin wave variables we can rewrite $\mathcal{H}_0$ as

$$\begin{array}{l} \mathcal{H}_0 = 2z\mathcal{J}S \sum_{\mathbf{k}} \left(\gamma_{\mathbf{k}} c_{\mathbf{k}}^\dagger d_{\mathbf{k}}^\dagger + \gamma_{\mathbf{k}} c_{\mathbf{k}} d_{\mathbf{k}} + c_{\mathbf{k}}^\dagger c_{\mathbf{k}} + d_{\mathbf{k}}^\dagger d_{\mathbf{k}} \right) \\ \qquad + g\mu_{\mathrm{B}} \left(B_{\mathrm{A}} + B_0 \right) \sum_{\mathbf{k}} c_{\mathbf{k}}^\dagger c_{\mathbf{k}} + g\mu_{\mathrm{B}} \left(B_{\mathrm{A}} - B_0 \right) \sum_{\mathbf{k}} d_{\mathbf{k}}^\dagger d_{\mathbf{k}}. \end{array} \tag{10.105}$$

Here, we have introduced

$$\gamma_{\mathbf{k}} = z^{-1} \sum_{\boldsymbol{\delta}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\boldsymbol{\delta}} = \gamma_{-\mathbf{k}}$$

once again. We are going to forget all about $\mathcal{H}_1$, and consider for the moment that the entire Hamiltonian is given by $\mathcal{H}_0 + E_0$. $\mathcal{H}_0$ is still not in a trivial form. We can easily put it into normal form as follows:

1. Define new operators $\alpha_{\mathbf{k}}$ and $\beta_{\mathbf{k}}$

$$\alpha_{\mathbf{k}} = u_{\mathbf{k}} c_{\mathbf{k}} - v_{\mathbf{k}} d_{\mathbf{k}}^\dagger \;\; ; \;\; \beta_{\mathbf{k}} = u_{\mathbf{k}} d_{\mathbf{k}} - v_{\mathbf{k}} c_{\mathbf{k}}^\dagger, \tag{10.106}$$

where $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ are real and satisfy $u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2 = 1$.

2. Solve these equations (and their Hermitian conjugates) for the c's and d's in terms of α and β. We can write

$$c_{\mathbf{k}} = u_{\mathbf{k}}\alpha_{\mathbf{k}} + v_{\mathbf{k}}\beta^{\dagger}_{\mathbf{k}} \;\; ; \;\; c^{\dagger}_{\mathbf{k}} = u_{\mathbf{k}}\alpha^{\dagger}_{\mathbf{k}} + v_{\mathbf{k}}\beta_{\mathbf{k}} \tag{10.107}$$

and

$$d_{\mathbf{k}} = v_{\mathbf{k}}\alpha^{\dagger}_{\mathbf{k}} + u_{\mathbf{k}}\beta_{\mathbf{k}} \;\; ; \;\; d^{\dagger}_{\mathbf{k}} = v_{\mathbf{k}}\alpha_{\mathbf{k}} + u_{\mathbf{k}}\beta^{\dagger}_{\mathbf{k}}. \tag{10.108}$$

3. Substitute (10.107) and (10.108) in $\mathcal{H}_0$ to have

$$\begin{aligned}
\mathcal{H}_0 = 2zS\mathcal{J}\textstyle\sum_{\mathbf{k}} \Big\{ & \gamma_{\mathbf{k}} \Big[u_{\mathbf{k}}v_{\mathbf{k}}(\alpha^{\dagger}_{\mathbf{k}}\alpha_{\mathbf{k}} + \beta_{\mathbf{k}}\beta^{\dagger}_{\mathbf{k}} + \alpha_{\mathbf{k}}\alpha^{\dagger}_{\mathbf{k}} + \beta^{\dagger}_{\mathbf{k}}\beta_{\mathbf{k}}) \\
& + u^2_{\mathbf{k}}(\alpha^{\dagger}_{\mathbf{k}}\beta^{\dagger}_{\mathbf{k}} + \alpha_{\mathbf{k}}\beta_{\mathbf{k}}) + v^2_{\mathbf{k}}(\beta^{\dagger}_{\mathbf{k}}\alpha^{\dagger}_{\mathbf{k}} + \beta_{\mathbf{k}}\alpha_{\mathbf{k}}) \Big] \\
& + u^2_{\mathbf{k}}\alpha^{\dagger}_{\mathbf{k}}\alpha_{\mathbf{k}} + v^2_{\mathbf{k}}\beta_{\mathbf{k}}\beta^{\dagger}_{\mathbf{k}} + u_{\mathbf{k}}v_{\mathbf{k}}(\alpha^{\dagger}_{\mathbf{k}}\beta^{\dagger}_{\mathbf{k}} + \beta_{\mathbf{k}}\alpha_{\mathbf{k}}) \\
& + v^2_{\mathbf{k}}\alpha_{\mathbf{k}}\alpha^{\dagger}_{\mathbf{k}} + u^2_{\mathbf{k}}\beta_{\mathbf{k}}\beta^{\dagger}_{\mathbf{k}} + u_{\mathbf{k}}v_{\mathbf{k}}(\alpha_{\mathbf{k}}\beta_{\mathbf{k}} + \beta^{\dagger}_{\mathbf{k}}\alpha^{\dagger}_{\mathbf{k}}) \Big\} \\
& + g\mu_{\mathrm{B}}(B_{\mathrm{A}} + B_0)\textstyle\sum_{\mathbf{k}} \Big[u^2_{\mathbf{k}}\alpha^{\dagger}_{\mathbf{k}}\alpha_{\mathbf{k}} + v^2_{\mathbf{k}}\beta_{\mathbf{k}}\beta^{\dagger}_{\mathbf{k}} + u_{\mathbf{k}}v_{\mathbf{k}}(\alpha^{\dagger}_{\mathbf{k}}\beta^{\dagger}_{\mathbf{k}} + \beta_{\mathbf{k}}\alpha_{\mathbf{k}}) \Big] \\
& + g\mu_{\mathrm{B}}(B_{\mathrm{A}} - B_0)\textstyle\sum_{\mathbf{k}} \Big[v^2_{\mathbf{k}}\alpha_{\mathbf{k}}\alpha^{\dagger}_{\mathbf{k}} + u^2_{\mathbf{k}}\beta^{\dagger}_{\mathbf{k}}\beta_{\mathbf{k}} + u_{\mathbf{k}}v_{\mathbf{k}}(\alpha_{\mathbf{k}}\beta_{\mathbf{k}} + \beta^{\dagger}_{\mathbf{k}}\alpha^{\dagger}_{\mathbf{k}}) \Big].
\end{aligned} \tag{10.109}$$

We can regroup these terms as follows:

$$\begin{aligned}
\mathcal{H}_0 = {} & 2\textstyle\sum_{\mathbf{k}} \left[2zS\mathcal{J}\left(\gamma_{\mathbf{k}}u_{\mathbf{k}}v_{\mathbf{k}} + v^2_{\mathbf{k}}\right) + g\mu_{\mathrm{B}}B_{\mathrm{A}}v^2_{\mathbf{k}} \right] \\
& + \textstyle\sum_{\mathbf{k}} \left[2zS\mathcal{J}\left(2\gamma_{\mathbf{k}}u_{\mathbf{k}}v_{\mathbf{k}} + u^2_{\mathbf{k}} + v^2_{\mathbf{k}}\right) + g\mu_{\mathrm{B}}B_{\mathrm{A}}(u^2_{\mathbf{k}} + v^2_{\mathbf{k}}) + g\mu_{\mathrm{B}}B_0 \right] \alpha^{\dagger}_{\mathbf{k}}\alpha_{\mathbf{k}} \\
& + \textstyle\sum_{\mathbf{k}} \left[2zS\mathcal{J}\left(2\gamma_{\mathbf{k}}u_{\mathbf{k}}v_{\mathbf{k}} + u^2_{\mathbf{k}} + v^2_{\mathbf{k}}\right) + g\mu_{\mathrm{B}}B_{\mathrm{A}}(u^2_{\mathbf{k}} + v^2_{\mathbf{k}}) - g\mu_{\mathrm{B}}B_0 \right] \beta^{\dagger}_{\mathbf{k}}\beta_{\mathbf{k}} \\
& + \textstyle\sum_{\mathbf{k}} \left\{ 2zS\mathcal{J}\left[\gamma_{\mathbf{k}}(u^2_{\mathbf{k}} + v^2_{\mathbf{k}}) + 2u_{\mathbf{k}}v_{\mathbf{k}}\right] + 2g\mu_{\mathrm{B}}B_{\mathrm{A}}u_{\mathbf{k}}v_{\mathbf{k}} \right\} \left(\alpha^{\dagger}_{\mathbf{k}}\beta^{\dagger}_{\mathbf{k}} + \alpha_{\mathbf{k}}\beta_{\mathbf{k}}\right).
\end{aligned} \tag{10.110}$$

4. We put the Hamiltonian in diagonal form by requiring the coefficient of the last term to vanish. We define ω_{e} and ω_{A} by

$$\omega_{\mathrm{e}} = 2\mathcal{J}zS \;\; \text{and} \;\; \omega_{\mathrm{A}} = g\mu_{\mathrm{B}}B_{\mathrm{A}}. \tag{10.111}$$

We must solve

$$\omega_{\mathrm{e}}\left[\gamma_{\mathbf{k}}(u^2_{\mathbf{k}} + v^2_{\mathbf{k}}) + 2u_{\mathbf{k}}v_{\mathbf{k}}\right] + 2\omega_{\mathrm{A}}u_{\mathbf{k}}v_{\mathbf{k}} = 0, \tag{10.112}$$

remembering that $u^2_{\mathbf{k}} = 1 + v^2_{\mathbf{k}}$. Then, (10.112) reduces to

$$\frac{1 + 2v^2_{\mathbf{k}}}{2v_{\mathbf{k}}\sqrt{1 + v^2_{\mathbf{k}}}} = -\frac{1}{\gamma_{\mathbf{k}}}\left(\frac{\omega_{\mathrm{A}}}{\omega_{\mathrm{e}}} + 1\right).$$

Solving for $v^2_{\mathbf{k}}$ gives

$$v^2_{\mathbf{k}} = -\frac{1}{2} + \frac{1}{2}\frac{\omega_{\mathrm{A}} + \omega_{\mathrm{e}}}{\sqrt{(\omega_{\mathrm{A}} + \omega_{\mathrm{e}})^2 - \gamma^2_{\mathbf{k}}\omega^2_{\mathrm{e}}}}. \tag{10.113}$$

Thus, we have

$$u^2_{\mathbf{k}} = \frac{1}{2} + \frac{1}{2}\frac{\omega_{\mathrm{A}} + \omega_{\mathrm{e}}}{\sqrt{(\omega_{\mathrm{A}} + \omega_{\mathrm{e}})^2 - \gamma^2_{\mathbf{k}}\omega^2_{\mathrm{e}}}}, \tag{10.114}$$

and since

$$u_{\mathbf{k}}v_{\mathbf{k}} = -\frac{1}{2}\frac{\gamma_{\mathbf{k}}\omega_{\mathrm{e}}}{\omega_{\mathrm{A}}+\omega_{\mathrm{e}}}(u_{\mathbf{k}}^2+v_{\mathbf{k}}^2),$$

we have

$$u_{\mathbf{k}}v_{\mathbf{k}} = -\frac{1}{2}\frac{\gamma_{\mathbf{k}}\omega_{\mathrm{e}}}{\sqrt{(\omega_{\mathrm{A}}+\omega_{\mathrm{e}})^2-\gamma_{\mathbf{k}}^2\omega_{\mathrm{e}}^2}}. \tag{10.115}$$

Now, let us write the Hamiltonian in a diagonal form

$$\mathcal{H}_0 = C + \sum_{\mathbf{k}}\left[(\omega_{\mathbf{k}}+g\mu_{\mathrm{B}}B_0)\alpha_{\mathbf{k}}^{\dagger}\alpha_{\mathbf{k}} + (\omega_{\mathbf{k}}-g\mu_{\mathrm{B}}B_0)\beta_{\mathbf{k}}^{\dagger}\beta_{\mathbf{k}}\right] \tag{10.116}$$

where

$$\begin{aligned}\omega_{\mathbf{k}} &= 2zS\mathcal{J}\left(2\gamma_{\mathbf{k}}u_{\mathbf{k}}v_{\mathbf{k}}+u_{\mathbf{k}}^2+v_{\mathbf{k}}^2\right)+g\mu_{\mathrm{B}}B_{\mathrm{A}}(u_{\mathbf{k}}^2+v_{\mathbf{k}}^2)\\ &= \sqrt{(\omega_{\mathrm{A}}+\omega_{\mathrm{e}})^2-\gamma_{\mathbf{k}}^2\omega_{\mathrm{e}}^2}.\end{aligned} \tag{10.117}$$

The constant C is given by

$$\begin{aligned}C &= 2\textstyle\sum_{\mathbf{k}}\left[2zS\mathcal{J}\left(\gamma_{\mathbf{k}}u_{\mathbf{k}}v_{\mathbf{k}}+v_{\mathbf{k}}^2\right)+g\mu_{\mathrm{B}}B_{\mathrm{A}}v_{\mathbf{k}}^2\right]\\ &= \textstyle\sum_{\mathbf{k}}\left[\omega_{\mathbf{k}}-(\omega_{\mathrm{A}}+\omega_{\mathrm{e}})\right].\end{aligned} \tag{10.118}$$

Thus, to this order of approximation we have

$$\begin{aligned}\mathcal{H} = &-2NzJS^2-2g\mu_{\mathrm{B}}B_{\mathrm{A}}NS+\textstyle\sum_{\mathbf{k}}\left[\omega_{\mathbf{k}}-(\omega_{\mathrm{A}}+\omega_{\mathrm{e}})\right]\\ &+\textstyle\sum_k(\omega_{\mathbf{k}}+\omega_{\mathrm{B}})\alpha_{\mathbf{k}}^{\dagger}\alpha_{\mathbf{k}}+\sum_k(\omega_{\mathbf{k}}-\omega_{\mathrm{B}})\beta_{\mathbf{k}}^{\dagger}\beta_{\mathbf{k}},\end{aligned} \tag{10.119}$$

where

$$\omega_{\mathrm{B}} = g\mu_{\mathrm{B}}B_0. \tag{10.120}$$

10.7.1 Ground State Energy

In the ground state

$$\left\langle 0\left|\alpha_{\mathbf{k}}^{\dagger}\alpha_{\mathbf{k}}\right|0\right\rangle = \left\langle 0\left|\beta_{\mathbf{k}}^{\dagger}\beta_{\mathbf{k}}\right|0\right\rangle = 0.$$

Thus the ground state energy is given by

$$E_{\mathrm{GS}} = -2NzJS^2-2g\mu_{\mathrm{B}}B_{\mathrm{A}}NS+\sum_{\mathbf{k}}\left[\omega_{\mathbf{k}}-(\omega_{\mathrm{A}}+\omega_{\mathrm{e}})\right]. \tag{10.121}$$

Let us consider the case $B_0 = B_{\mathrm{A}} = 0$; thus $\omega_{\mathrm{A}} \to 0$ and $\omega_{\mathbf{k}} \to \omega_{\mathrm{e}}(1-\gamma_{\mathbf{k}}^2)^{1/2}$. But ω_{e} is simply $2\mathcal{J}zS$. Hence for $B_0 = B_{\mathrm{A}} = 0$ the ground state energy is given by

$$E_{\mathrm{GS}} = -2NzJS^2 - N\omega_{\mathrm{e}} + \omega_{\mathrm{e}}\sum_{\mathbf{k}}(1-\gamma_{\mathbf{k}}^2)^{1/2}. \tag{10.122}$$

By using $\omega_{\rm e} = 2\mathcal{J}zS$, this can be rewritten by

$$E_{\rm GS} = -2Nz\mathcal{J}S\left[S + 1 - N^{-1}\sum_{\mathbf{k}}\sqrt{1-\gamma_{\mathbf{k}}^2}\right]. \tag{10.123}$$

Let us define $\beta = z\left(1 - N^{-1}\sum_{\mathbf{k}}\sqrt{1-\gamma_{\mathbf{k}}^2}\right)$; then $E_{\rm GS}$ can be written as

$$E_{\rm GS} = -2Nz\mathcal{J}S(S + z^{-1}\beta). \tag{10.124}$$

For a simple cubic lattice $\beta \simeq 0.58$. For other crystal structures β has slightly different values.

10.7.2 Zero Point Sublattice Magnetization

For very large anisotropy field $B_{\rm A}$, the magnetization of sublattice a is $g\mu_{\rm B}NS$ while that of sublattice b is equal in magnitude and opposite in direction. When $B_{\rm A} \to 0$, the resulting antiferromagnetic state will have a sublattice magnetization that differs from the value of $B_{\rm A} \to \infty$. Then magnetization is given by

$$M(T) = \frac{g\mu_{\rm B}}{V}\left\langle 0|\hat{\mathcal{S}}_z|0\right\rangle, \tag{10.125}$$

where the total spin operator $\hat{\mathcal{S}}_z$ is given, for sublattice a, by

$$\hat{\mathcal{S}}_z = \sum_{l\in{\rm a}} S_{lz}^{\rm a} = NS - \sum_l a_l^\dagger a_l. \tag{10.126}$$

But $\sum_l a_l^\dagger a_l = \sum_{\mathbf{k}} c_{\mathbf{k}}^\dagger c_{\mathbf{k}}$, and the $c_{\mathbf{k}}$ and $c_{\mathbf{k}}^\dagger$ can be written in terms of the operators $\alpha_{\mathbf{k}}$, $\alpha_{\mathbf{k}}^\dagger$, $\beta_{\mathbf{k}}$, and $\beta_{\mathbf{k}}^\dagger$ to get

$$\hat{\mathcal{S}}_z = NS - \sum_{\mathbf{k}}\left(u_{\mathbf{k}}\alpha_{\mathbf{k}}^\dagger + v_{\mathbf{k}}\beta_{\mathbf{k}}\right)\left(u_{\mathbf{k}}\alpha_{\mathbf{k}} + v_{\mathbf{k}}\beta_{\mathbf{k}}^\dagger\right). \tag{10.127}$$

Multiplying out the product appearing in the sum we can write

$$\Delta\hat{\mathcal{S}} \equiv NS - \hat{\mathcal{S}}_z = \sum_{\mathbf{k}}\left\{v_{\mathbf{k}}^2 + u_{\mathbf{k}}^2\alpha_{\mathbf{k}}^\dagger\alpha_{\mathbf{k}} + v_{\mathbf{k}}^2\beta_{\mathbf{k}}^\dagger\beta_{\mathbf{k}} + u_{\mathbf{k}}v_{\mathbf{k}}\left(\alpha_{\mathbf{k}}^\dagger\beta_{\mathbf{k}}^\dagger + \alpha_{\mathbf{k}}\beta_{\mathbf{k}}\right)\right\}. \tag{10.128}$$

At zero temperature the ground state $|0>$ contains no excitations so that $\alpha_{\mathbf{k}}|0\rangle = \beta_{\mathbf{k}}|0\rangle = 0$. Thus, at $T = 0$, $\Delta\mathcal{S}(T) = \left\langle 0|\Delta\hat{\mathcal{S}}|0\right\rangle$ has a value $\Delta\mathcal{S}_0$ given by

$$\Delta\mathcal{S}_0 = \sum_{\mathbf{k}} v_{\mathbf{k}}^2 = -\frac{1}{2}\sum_{\mathbf{k}}\left[1 - \frac{\omega_{\rm A} + \omega_{\rm e}}{\sqrt{(\omega_{\rm A} + \omega_{\rm e})^2 - \gamma_{\mathbf{k}}^2\omega_{\rm e}^2}}\right]. \tag{10.129}$$

Let us put $\omega_{\mathrm{A}} = 0$ corresponding to $B_{\mathrm{A}} \to 0$. This gives

$$\Delta\mathcal{S}_0 = -\frac{N}{2} + \frac{1}{2}\sum_{\mathbf{k}} \frac{1}{\sqrt{1-\gamma_{\mathbf{k}}^2}}. \tag{10.130}$$

If we define $\beta' = zN^{-1}\sum_{\mathbf{k}} \left(1 - \frac{1}{\sqrt{1-\gamma_{\mathbf{k}}^2}}\right)$, then we have

$$\Delta\mathcal{S}_0 = -\frac{1}{2}\frac{\beta' N}{z}. \tag{10.131}$$

For a simple cubic lattice $z = 6$ and β' has the value 0.94 giving for $\Delta\mathcal{S}_0$ the value $-0.078N$.

10.7.3 Finite Temperature Sublattice Magnetization

At a finite temperature it is apparent from Eqs.(10.128) and (10.129) that

$$\Delta\mathcal{S}(T) = \Delta\mathcal{S}_0 + \sum_{\mathbf{k}} \left[u_{\mathbf{k}}^2 \left\langle \alpha_{\mathbf{k}}^\dagger \alpha_{\mathbf{k}} \right\rangle + v_{\mathbf{k}}^2 \left\langle \beta_{\mathbf{k}}^\dagger \beta_{\mathbf{k}} \right\rangle\right]. \tag{10.132}$$

But the excitations described by the creation operators $\alpha_{\mathbf{k}}^\dagger$ and $\beta_{\mathbf{k}}^\dagger$ have energies $\omega_{\mathbf{k}} \pm \omega_{\mathrm{B}}$ (the sign $-$ goes with $\beta_{\mathbf{k}}^\dagger$), so that

$$\left\langle \alpha_{\mathbf{k}}^\dagger \alpha_{\mathbf{k}} \right\rangle = \frac{1}{\mathrm{e}^{\beta(\omega_{\mathbf{k}}+\omega_{\mathrm{B}})} - 1} \quad \text{and} \quad \left\langle \beta_{\mathbf{k}}^\dagger \beta_{\mathbf{k}} \right\rangle = \frac{1}{\mathrm{e}^{\beta(\omega_{\mathbf{k}}-\omega_{\mathrm{B}})} - 1}. \tag{10.133}$$

In these equations $\beta = \frac{1}{k_{\mathrm{B}}T}$, $\omega_{\mathrm{B}} = 2\mu_{\mathrm{B}}B_0$, and $\omega_{\mathbf{k}} = \sqrt{(\omega_{\mathrm{A}} + \omega_{\mathrm{e}})^2 - \gamma_{\mathbf{k}}^2\omega_{\mathrm{e}}^2}$. At low temperature only very low frequency or small wave number modes will be excited. Remember that

$$\gamma_{\mathbf{k}} = z^{-1}\sum_{\boldsymbol{\delta}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\boldsymbol{\delta}}$$

where $\boldsymbol{\delta}$ indicates the nearest neighbors of the atom at the origin. To order k^2 for a simple cubic lattice

$$\gamma_{\mathbf{k}} = z^{-1}\sum_{\boldsymbol{\delta}} \left(1 - \frac{(\mathbf{k}\cdot\boldsymbol{\delta})^2}{2}\right) = 1 - \frac{k^2a^2}{z}.$$

Thus, the excitation energies $\varepsilon_k \equiv \omega_{\mathbf{k}} \pm \omega_{\mathrm{B}}$ are approximated, in the long wave length limit ($k^2a^2 \ll \frac{z\omega_{\mathrm{A}}}{\omega_{\mathrm{e}}} \ll 1$), by

$$\begin{aligned} \varepsilon_k &\simeq [\omega_{\mathrm{A}}(\omega_{\mathrm{A}} + 2\omega_{\mathrm{e}})]^{1/2} \sqrt{1 + \frac{\omega_{\mathrm{e}}^2}{\omega_{\mathrm{A}}(2\omega_{\mathrm{e}} + \omega_{\mathrm{A}})}\frac{k^2a^2}{z}} \pm \omega_{\mathrm{B}} \\ &\approx [\omega_{\mathrm{A}}(\omega_{\mathrm{A}} + 2\omega_{\mathrm{e}})]^{1/2} + \frac{1}{2z}\frac{\omega_{\mathrm{e}}^2}{\sqrt{\omega_{\mathrm{A}}(\omega_{\mathrm{A}} + 2\omega_{\mathrm{e}})}} k^2a^2 \pm \omega_{\mathrm{B}}. \end{aligned} \tag{10.134}$$

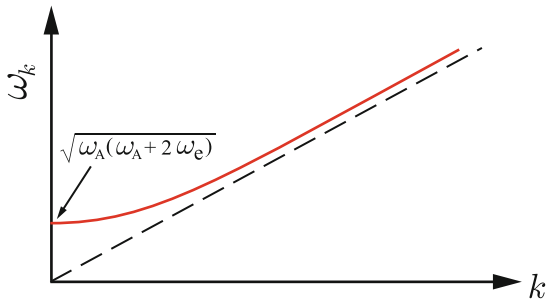


Fig. 10.14. Antiferromagnetic spin wave excitation energies in the long wave length limit

Thus, the *uniform mode* of antiferromagnetic resonance is given, in the presence of an applied field, by

$$\varepsilon_{k=0} = \sqrt{\omega_A(\omega_A + 2\omega_e)} \pm \omega_B. \tag{10.135}$$

In the long wave length limit, but in the region of $1 \gg k^2a^2 \gg z\frac{\omega_A}{\omega_e}\left(2 + \frac{\omega_A}{\omega_e}\right)$, we expect the behavior given by

$$\varepsilon_k \simeq \frac{\omega_e ka}{\sqrt{z}} \pm \omega_B \approx 2\sqrt{z}\mathcal{J}Sak \pm \omega_B. \tag{10.136}$$

Figure 10.14 shows the excitation energies $\omega_{\mathbf{k}}$ as a function of wave number k in the long wave length limit.

Let us make an approximation like the Debye approximation of lattice dynamics in the absence of an applied field. Replace the first Brillouin zone by a sphere of radius k_M, where

$$\frac{1}{(2\pi)^3}\frac{4}{3}\pi k_M^3 = \frac{N}{V}$$

to have

$$\varepsilon_k \simeq \frac{\Theta_N}{k_M}k. \tag{10.137}$$

Here Θ_N is the value of ε_k at $k = k_M$. With a use of this approximation for ε_k of both the + and - (or $\alpha_{\mathbf{k}}$ and $\beta_{\mathbf{k}}$) modes one can evaluate the spin fluctuation

$$\Delta\mathcal{S}(T) = \Delta\mathcal{S}_0 + \sum_{\mathbf{k}} \frac{u_{\mathbf{k}}^2 + v_{\mathbf{k}}^2}{e^{\varepsilon_k/\Theta} - 1}. \tag{10.138}$$

Using our expressions for $v_{\mathbf{k}}^2$ (and $u_{\mathbf{k}}^2 = 1 + v_{\mathbf{k}}^2$), replacing the k-summation by an integral, and evaluation for $\Theta \ll \Theta_N$ we have

$$\Delta\mathcal{S}(T) = \Delta\mathcal{S}_0 + \frac{\sqrt{3}}{12}\left(6\pi^2\right)^{2/3} N \left(\frac{\Theta}{\Theta_N}\right)^2. \tag{10.139}$$

10.7.4 Heat Capacity due to Antiferromagnetic Magnons

For $\Theta < \omega_0 \left(\equiv \sqrt{\omega_{\mathrm{A}}(\omega_{\mathrm{A}} + 2\omega_{\mathrm{e}})} \right)$, the heat capacity will vary with temperature as $\mathrm{e}^{-\mathrm{const}/T}$, since the probability of exciting a magnon will be exponentially small. For somewhat higher temperatures (but not too high since we are assuming small $|k|$) where modes with $\omega_{\mathbf{k}} \simeq \frac{\Theta_{\mathrm{N}}}{k_{\mathrm{M}}} k$ are excited, the specific heat is very much like the low temperature Debye specific heat (the temperature region in question is defined by $\omega_0 \ll \Theta \ll \Theta_{\mathrm{N}}$). The internal energy will be given by

$$U = 2 \sum_{\mathbf{k}} \frac{\omega_{\mathbf{k}}}{\mathrm{e}^{\omega_{\mathbf{k}}/\Theta} - 1}. \tag{10.140}$$

Here, we have two antiferromagnetic magnons for every value of $\mathbf{k}$, instead of three as for phonons, and the factor of 2 results from counting two types of spin excitations, $\alpha_{\mathbf{k}}^{\dagger}$ and $\beta_{\mathbf{k}}^{\dagger}$ type modes. Replacing the sum by an integral and replacing the upper limit k_{M} by infinity, as in the low temperature Debye specific heat, gives

$$U = N \frac{(k_{\mathrm{M}} a)^3}{15} \frac{\Theta^4}{\Theta_{\mathrm{N}}^3} \pi^2 = N \frac{2\pi^4}{5} \frac{\Theta^4}{\Theta_{\mathrm{N}}^3}. \tag{10.141}$$

For the specific heat per particle one obtains

$$C = \frac{8\pi^4}{5} \left(\frac{\Theta}{\Theta_{\mathrm{N}}} \right)^3. \tag{10.142}$$

10.8 Exchange Interactions

Here, we briefly describe various kinds of exchange interactions which are the underlying sources of the long range magnetic ordering.

1. *Direct exchange* is the kind of exchange we discussed when we investigate the simple Heisenberg exchange interaction. The magnetic ions interact through the direct Coulomb interaction among the electrons on the two ions as a result of their wave function overlap.
2. *Superexchange* is the underlying mechanism of a number of ionic solids, such as MnO and MnF_2, showing magnetic ground states. Even in the absence of direct overlap between the electrons on different magnetic ions sharing a nonmagnetic ion (one with closed electronic shells and located in between the magnetic ions), the two magnetic ions can have exchange interaction mediated by the nonmagnetic ion. (See, for example, Fig. 10.15.)
3. *Indirect exchange* is the magnetic interaction between magnetic moments localized in a metal (such as rare earth metals) through the mediation of conduction electrons in the metal. It is a metallic analogue of

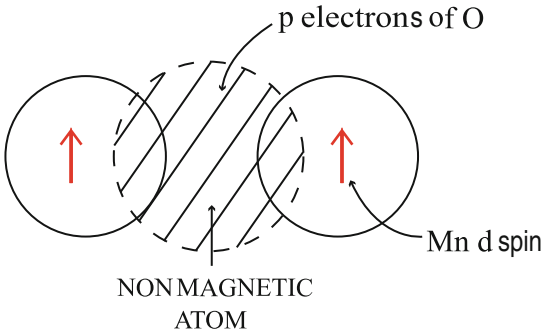


Fig. 10.15. Schematic illustration of superexchange coupling in a magnetic oxide. Two Mn ions (each having unpaired electron in a d orbital) are separated by an oxygen ion having two p electrons

superexchange in ionic insulators and is also called as the *Ruderman-Kittel-Kasuya-Yosida (RKKY) interaction*. For example, the unpaired f electrons in the rare earths are magnetic and they can be coupled to f electrons in a neighboring rare earth ion through the exchange interaction via nonmagnetic conduction electrons.

4. *Double exchange* coupling is the ferromagnetic superexchange in an extended system. The double exchange explains the ferromagnetic coupling between magnetic ions of mixed valency. For example, $La_{1-x}Sr_xMnO_3 (0 \leq x \leq 0.175)$ shows ferromagnetic metallic behavior below room temperature. In this material, a fraction x of the Mn ions are Mn^{4+} and $1-x$ are Mn^{3+}, because La exists as La^{3+} and Sr exists as Sr^{2+}.
5. *Itinerant ferromagnetism* occurs in solids (such as Fe, Co, Ni, $\cdots$) containing the magnetic moments associated with the delocalized electrons, known as *itinerant electrons*, wandering through the sample.

10.9 Itinerant Ferromagnetism

Most of our discussion up to now has simply assumed a Heisenberg $\mathcal{J}_{ij}\mathbf{S}_i \cdot \mathbf{S}_j$ type interaction of localized spins. The atomic configurations of some of the atoms in the $3d$ transition metal series are Sc $(3d)^1(4s)^2$, Ti $(3d)^2(4s)^2$, V $(3d)^3(4s)^2$, Cr $(3d)^5(4s)^1$, Mn $(3d)^5(4s)^2$, Fe $(3d)^6(4s)^2$, Co $(3d)^7(4s)^2$, Ni $(3d)^8(4s)^2$, Cu $(3d)^{10}(4s)^1$. If we simply calculate the band structure of these materials, completely ignoring the possibility of magnetic order, we find that the density of states of the solid has a large and relatively narrow set of peaks associated with the $3d$ bands, and a broad but low peak associated with the $4s$ bands as is sketched in Fig. 10.16. The position of the Fermi level determines whether the d bands are partially filled or completely filled. For transition metals with partially filled d bands, the electrons participating in the magnetic states are itinerant.

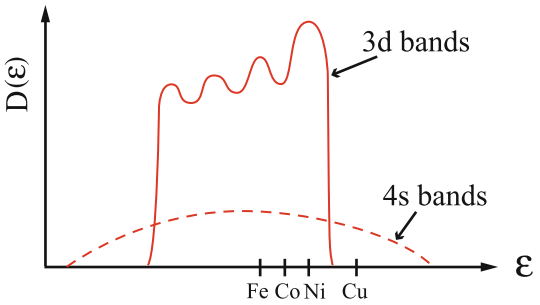


Fig. 10.16. Schematic illustration of the density of states of the transition metals

10.9.1 Stoner Model

In order to account for itinerant ferromagnetism, Stoner introduced a very simple model with the following properties:

1. The Bloch bands obtained in a band structure calculation are maintained.
2. By adding an exchange energy to the Bloch bands a spin splitting, described by an internal mean field, can be obtained.
3. States with spin antiparallel $(-)$ to the internal field are lowered in energy relative to those with parallel $(+)$ spin.

We can write for spin up $(+)$ and spin down $(-)$ electrons

$$E_-(k) \simeq \frac{\hbar k^2}{2m^*} \quad \text{and} \quad E_+(k) \simeq \frac{\hbar k^2}{2m^*} + \Delta, \tag{10.143}$$

where Δ is the spin splitting. The spin split Bloch bands and Fermi surfaces for spin up and spin down electrons are illustrated in Fig. 10.17 in the presence of spin splitting Δ.

10.9.2 Stoner Excitations

A single particle excitation in which an electron with wave vector $\mathbf{k}$ and spin down $(-)$ is excited to an empty state with wave vector $\mathbf{k}+\mathbf{q}$ and spin up $(+)$ has energy

$$\begin{aligned} E &= E_+(\mathbf{k}+\mathbf{q}) - E_-(\mathbf{k}) \\ &= \frac{\hbar^2(\mathbf{k}+\mathbf{q})^2}{2m^*} + \Delta - \frac{\hbar k^2}{2m^*} \\ &= \frac{\hbar^2}{m^*}\mathbf{q}\cdot\left(\mathbf{k}+\frac{\mathbf{q}}{2}\right) + \Delta. \end{aligned} \tag{10.144}$$

These Stoner single particle excitations define the single particle continuum shown in Fig. 10.18. The single particle continuum of possible values of $|\mathbf{k}|$ for different values of $|\mathbf{q}|$ are hatched. Clearly when $q = 0$, the excitations all have energy Δ. These are single particle excitations. In addition Stoner found

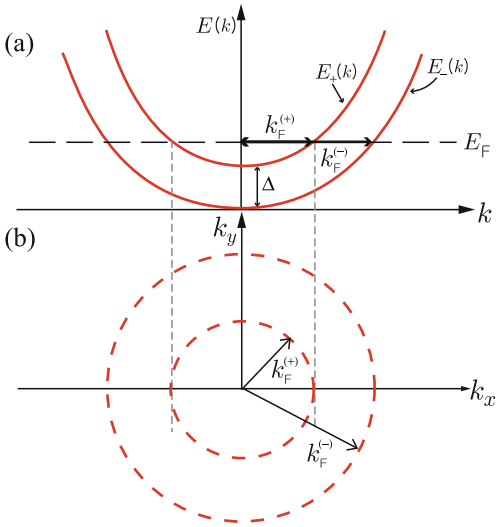


Fig. 10.17. Schematic illustration of the spin split Bloch bands in the Stoner model. (**a**) Energy dispersion of the Bloch bands in the presence of spin splitting Δ (**b**) The Fermi surfaces for spin up and spin down electrons

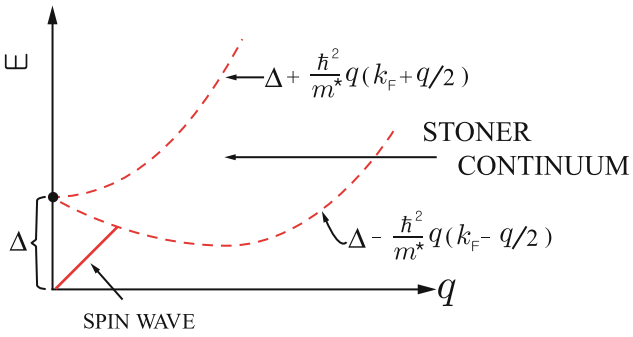


Fig. 10.18. Schematic illustration of the energy dispersion of the Stoner excitations and spin wave modes. The hatched area shows the single particle continuum of possible values of $|\mathbf{k}|$ for different values of $|\mathbf{q}|$

spin waves of an itinerant ferromagnet that started at the origin ($E = 0$ at $q = 0$) and intersected the single particle continuum at q_c, a finite value of q. The spin wave excitation is also indicated in Fig. 10.18.

10.10 Phase Transition

Near T_c, the ferromagnet is close to a phase transition. Many observable properties should display interesting behavior as a function of $T - T_c$ (see, for example, Fig. 10.19).

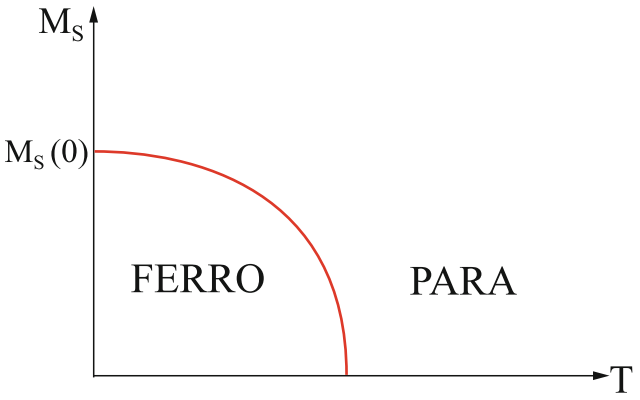


Fig. 10.19. Schematic illustration of the temperature dependence of the spontaneous magnetization

Here, we list only a few of the interesting examples.

1. *Magnetization*: As T increases toward T_c the spontaneous magnetization must vanish as
$$M(T) \approx (T_c - T)^{\beta} \text{ with } \beta > 0.$$
2. *Susceptibility*: As T decreases toward T_c in the paramagnetic state, the magnetic susceptibility $\chi(T)$ must diverge as
$$\chi(T) \approx (T - T_c)^{-\gamma} \text{ with } \gamma > 0.$$
3. *Specific heat*: As T decreases toward T_c in the paramagnetic state, the specific heat has a characteristic singularity given by
$$C(T) \approx (T - T_c)^{-\alpha} \text{ with } \alpha > 0.$$

In the *mean field theory*, where the interactions are replaced by their values in the presence of a self-consistently determined average magnetization, we find $\beta = \frac{1}{2}$ and $\gamma = 1$ for all dimensions. The mean field values do not agree with experiments or with several exactly solvable theoretical models for T very close to T_c. For example,

1. $\beta = \frac{1}{8}$ in the 2 dimensional Ising model.
2. $\beta \simeq \frac{1}{3}$ in the 3 dimensional Heisenberg model.
3. $\gamma \simeq 1.25$ for most 3 dimensional phase transitions instead of the mean field predictions of $\gamma = 1$.

In the early 1970s K. G. Wilson developed the *renormalization group theory* of phase transitions to describe the behavior of systems in the region $T \simeq T_c$.

Problems

10.1. Show that spin operators satisfy $[\hat{S}^2, \hat{S}^+] = 0$ and $[\hat{S}_z, \hat{S}^+] = \hat{S}^+$. Evaluate the commutator $[S^+, S^-]$ and $[S^\pm, S_z]$, and show that $S^\pm$ act as raising and lowering operators.

10.2. If $b_{\mathbf{k}} = N^{-1/2} \sum_j \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}_j} a_j$ and $b_{\mathbf{k}}^\dagger = N^{-1/2} \sum_j \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}_j} a_j^\dagger$ are spin wave operators in terms of spin deviation operators, show that $[a_j, a_{j'}] = [a_j^\dagger, a_{j'}^\dagger]$=0 and $[a_j, a_{j'}^\dagger] = \delta_{jj'}$ implies $[b_{\mathbf{k}}, b_{\mathbf{k}'}] = \left[b_{\mathbf{k}}^\dagger, b_{\mathbf{k}'}^\dagger\right] = 0$ and $\left[b_{\mathbf{k}}, b_{\mathbf{k}'}^\dagger\right] = \delta_{\mathbf{k}\mathbf{k}'}$.

10.3. In the text the Heisenberg Hamiltonian was written as

$$\mathcal{H} = -2\mathcal{J}S \sum_{\langle i,j\rangle} \left\{ \sqrt{1 - \frac{\hat{n}_i}{2S}}\, a_i a_j^\dagger \sqrt{1 - \frac{\hat{n}_j}{2S}} + a_i^\dagger \sqrt{1 - \frac{\hat{n}_i}{2S}} \sqrt{1 - \frac{\hat{n}_j}{2S}}\, a_j \right.$$
$$\left. + S\left(1 - \frac{\hat{n}_i}{S}\right)\left(1 - \frac{\hat{n}_j}{S}\right) \right\} - g\mu_\mathrm{B} B_0 S \sum_i \left(1 - \frac{\hat{n}_i}{S}\right),$$

where $\hat{n}_j = a_j^\dagger a_j$ and $a_j^\dagger$ (a_j) creates (annihilates) a spin deviation on site j. Expand the square roots for small $\hat{n}$ and show that the results for $\mathcal{H}_0$ and $\mathcal{H}_1$ agree with the expressions shown in (10.66) and (10.67), respectively.

10.4. Evaluate $\omega_{\mathbf{k}}$, the spin wave frequency, for arbitray $\mathbf{k}$ within the first Brillouin zone of a simple cubic lattice. Expand the result for small k and compare it with the result given by (10.81).

10.5. An antiferromagnet can be described by $H = \sum_{\langle i,j\rangle} \mathcal{J}_{ij} \mathbf{S}_i \cdot \mathbf{S}_j$, where $\mathcal{J}_{ij} > 0$. Show that the ground state energy E_0 of the Heisenberg antiferromagnet must satisfy

$$-S(S+1) \sum_{i,j} \mathcal{J}_{ij} \le E_0 \le -S^2 \sum_{i,j} \mathcal{J}_{ij}.$$

Hint: for the upper bound one can use the trial wave function

$$\Phi_\mathrm{Trial} = \prod_{\substack{i \in \mathrm{A} \\ j \in \mathrm{B}}} |S, S\rangle_i |S, -S\rangle_j,$$

where $|\, S, \pm S\rangle_k$ is the state with $S_z = \pm S$ on site k.

10.6. Prove that operators α_k's and β_k's defined in terms of spin wave operators

$$\alpha_{\mathbf{k}} = u_{\mathbf{k}} c_{\mathbf{k}} - v_{\mathbf{k}} d_{\mathbf{k}}^\dagger \quad \text{and} \quad \beta_{\mathbf{k}} = u_{\mathbf{k}} d_{\mathbf{k}} - v_{\mathbf{k}} c_{\mathbf{k}}^\dagger$$

satisfy the standard commutation rules. Here, $u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2 = 1$. (See (10.106))

10.7. Discuss spin wave excitations in a two-dimensional square lattice.

Summary

In this chapter, we studied magnetic ordering and spin wave excitations of magnetic solids. We first reviewed Heisenberg exchange interactions of atoms and then discussed spontaneous magnetization and domain wall properties of ferromagnets. The zero-temperature properties of Heisenberg ferromagnets and antiferromagnets are described. Spin wave excitations and magnon heat capacities of ferromagnets and antiferromagnets are also discussed. Finally, Stoner model is introduced as an illustration of itinerant ferromagnetism.

The Heisenberg interaction Hamiltonian is given by

$$\mathcal{H} = -2\mathcal{J} \sum_{\langle i,j \rangle} \mathbf{s}_i \cdot \mathbf{s}_j,$$

where the sum is over all pairs of nearest neighbors. The exchange constant $\mathcal{J}$ is positive (negative) for ferromagnets (antiferromagnets). For a chain of magnetic spins, it is more favorable energetically to have the spin flip gradually. If the spin turns through an angle Φ_0 in N steps, where N is large, the increase in exchange energy due to the domain wall is $E_{\text{ex}} = \mathcal{J}S^2 \frac{\Phi_0^2}{N}$. The exchange energy is lower if the domain wall is very wide.

In the presence of an applied magnetic field B_0 oriented in the z-direction, the Hamiltonian of a Heisenberg ferromagnet becomes

$$\mathcal{H} = -\sum_{i,j} \mathcal{J}_{ij} S_{iz} S_{jz} - \frac{1}{2} \sum_{i,j} \mathcal{J}_{ij} \left(S_i^+ S_j^- + S_i^- S_j^+ \right) - g\mu_{\text{B}} B_0 \sum_i S_{iz}.$$

In the ground state all the spins are aligned parallel to one another and to the magnetic field $\mathbf{B}_0$: $|0\rangle = \prod_i |S, S\rangle_i$. The ground state energy becomes

$$E_0 = -S^2 \sum_{i,j} \mathcal{J}_{ij} - N g \mu_{\text{B}} B_0 S.$$

For Heisenberg antiferromagnets, $\mathcal{J}$ is replaced by $-\mathcal{J}$ but a trial wave function $\Phi_{\text{Trial}} = \prod_{\substack{i \in \text{A} \\ j \in \text{B}}} |S, S\rangle_i |S, -S\rangle_j$ is not an eigenfunction of $\mathcal{H}$.

Low lying excitations of ferromagnet can be studied by introducing *spin deviation operator* $\hat{n}_j$ defined by

$$\hat{n}_j = S_j - \hat{S}_{jz} = S - \hat{S}_{jz} \equiv a_j^\dagger a_j.$$

With a use of the Holstein–Primakoff transformation to operators $a_j^\dagger$ and a_j

$$\hat{S}_j^+ = (2S_j - \hat{n}_j)^{1/2} a_j \quad \text{and} \quad \hat{S}_j^- = a_j^\dagger (2S_j - \hat{n}_j)^{1/2},$$

the Heisenberg Hamiltonian can be written, in the limit of $\langle \hat{n}_i \rangle \ll 2S$, as

$$\mathcal{H} = E_0 + \mathcal{H}_0 + \mathcal{H}_1.$$

Here, E_0, $\mathcal{H}_0$, and $\mathcal{H}_1$ are given, respectively, by

$$\begin{aligned} E_0 &= -z\mathcal{J}NS^2 - g\mu_{\mathrm{B}}B_0NS, \\ \mathcal{H}_0 &= (g\mu_{\mathrm{B}}B_0 + 2z\mathcal{J}S)\textstyle\sum_i \hat{n}_i - 2\mathcal{J}S\sum_{<i,j>}\left(a_ia_j^\dagger + a_i^\dagger a_j\right), \\ \mathcal{H}_1 &= -2\mathcal{J}\textstyle\sum_{<i,j>}\left(\hat{n}_i\hat{n}_j - \frac{1}{4}\hat{n}_ia_ia_j^\dagger - \frac{1}{4}a_ia_j^\dagger\hat{n}_j - \frac{1}{4}\hat{n}_ja_i^\dagger a_j - \frac{1}{4}a_i^\dagger a_j\hat{n}_i\right) \\ &\quad + \text{higher order terms.} \end{aligned}$$

Introducing *spin wave variables* defined by

$$b_{\mathbf{k}} = N^{-1/2}\sum_j \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}_j}\,a_j \quad \text{and} \quad b_{\mathbf{k}}^\dagger = N^{-1/2}\sum_j \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}_j}\,a_j^\dagger,$$

$\mathcal{H}_0$ becomes $\mathcal{H}_0 = \sum_{\mathbf{k}}\hbar\omega_{\mathbf{k}}b_{\mathbf{k}}^\dagger b_{\mathbf{k}}$, where $\hbar\omega_{\mathbf{k}} = 2z\mathcal{J}S(1-\gamma_{\mathbf{k}}) + g\mu_{\mathrm{B}}B_0$. Thus, if we neglect $\mathcal{H}_1$, we have for the Hamiltonian of a state containing magnons

$$\mathcal{H} = -\left(g\mu_{\mathrm{B}}B_0NS + z\mathcal{J}NS^2\right) + \sum_{\mathbf{k}}\hbar\omega_{\mathbf{k}}b_{\mathbf{k}}^\dagger b_{\mathbf{k}}.$$

We note that, at low enough temperature, the elementary excitations are waves of energy $\hbar\omega_{\mathbf{k}}$.

At low temperature, the internal energy and magnon specific heat are given by

$$U \simeq \frac{0.45}{\pi^2}\frac{\Theta^{5/2}}{D^{3/2}} \text{ and } C = \frac{\partial U}{\partial T} = 0.113k_{\mathrm{B}}\left(\frac{\Theta}{D}\right)^{3/2}.$$

The *spontaneous magnetization* at temperature T is given by

$$M_{\mathrm{s}} = \frac{g\mu_{\mathrm{B}}}{V}\left(NS - \left\langle\sum_{\mathbf{k}} b_k^\dagger b_k\right\rangle\right).$$

At low temperature, $M_{\mathrm{s}}(T)$ becomes $M_{\mathrm{s}}(T) = \frac{N}{V}2\mu S - 0.117\frac{\mu}{a^3}\left(\frac{\Theta}{2S\mathcal{J}}\right)^{3/2}$.

In the presence of an applied field $\mathbf{B}_0 = B_0\hat{z}$ and an anisotropy field $\mathbf{B}_{\mathrm{A}}$, the Heisenberg Hamiltonian of an antiferromagnet can be written

$$\mathcal{H} = +\mathcal{J}\sum_{<i,j>}\hat{\mathbf{S}}_i\cdot\hat{\mathbf{S}}_j - g\mu_{\mathrm{B}}(B_{\mathrm{A}}+B_0)\sum_{l\in\mathrm{a}}\hat{S}_{lz}^{\mathrm{a}} + g\mu_{\mathrm{B}}(B_{\mathrm{A}}-B_0)\sum_{p\in\mathrm{b}}\hat{S}_{pz}^{\mathrm{b}}.$$

In the absence of magnon–magnon interaction, the ground-state energy is given by

$$E_{\mathrm{GS}} = -2Nz\mathcal{J}S^2 - 2g\mu_{\mathrm{B}}B_{\mathrm{A}}NS + \sum_{\mathbf{k}}\left[\omega_{\mathbf{k}} - (\omega_{\mathrm{A}}+\omega_{\mathrm{e}})\right].$$

The internal energy due to antiferromagnetic magnons is given by

$$U = 2\sum_{\mathbf{k}}\frac{\omega_{\mathbf{k}}}{\mathrm{e}^{\omega_{\mathbf{k}}/\Theta}-1}.$$

The low temperature specific heat per particle becomes

$$C = \frac{8\pi^4}{5}\left(\frac{\Theta}{\Theta_{\mathrm{N}}}\right)^3.$$

11

Many Body Interactions – Introduction

11.1 Second Quantization

The Hamiltonian of a many particle system is usually of the form

$$H = \sum_i H_0(i) + \frac{1}{2}\sum_{i\neq j} V_{ij}. \tag{11.1}$$

Here, $H_0(i)$ is the single particle Hamiltonian describing the ith particle, and V_{ij} is the interaction between the ith and jth particles. Suppose we know the single particle eigenfunctions and eigenvalues

$$H_0|k\rangle = \varepsilon_k|k\rangle.$$

We can construct a basis set for the many particle wave functions by taking products of single particle wave functions. We actually did this for bosons when we discussed phonon modes of a crystalline lattice. We wrote

$$|n_1, n_2, \ldots, n_k, \ldots\rangle = (n_1!n_2!\cdots n_k!)^{-1/2}\left(a_1^\dagger\right)^{n_1}\left(a_2^\dagger\right)^{n_2}\cdots\left(a_k^\dagger\right)^{n_k}\cdots|0\rangle. \tag{11.2}$$

This represents a state in which the mode 1 contains n_1 excitations, ..., the mode k contains n_k excitations. Another way of saying it is that there are n_1 phonons of wave vector k_1, n_2 phonons of wave vector $k_2, \ldots$. The creation and annihilation operators $a^\dagger$ and a satisfy

$$\left[a_k, a_{k'}^\dagger\right]_- = \delta_{kk'};\ \ [a_k, a_{k'}]_- = \left[a_k^\dagger, a_{k'}^\dagger\right]_- = 0.$$

The commutation relations assure the symmetry of the state vector under interchange of a pair of particles since

$$a_k^\dagger a_{k'}^\dagger = a_{k'}^\dagger a_k^\dagger.$$

The single particle part is given by

$$\sum_i H_0(i) = \sum_k \varepsilon_k n_k, \tag{11.3}$$

where $\varepsilon_k = \langle k|H_0|k\rangle$ and $n_k = a_k^\dagger a_k$.

For Fermions, the single particle states can be singly occupied or empty. This means that n_k can take only two possible values, 0 or 1. It is convenient to introduce operators $c_k^\dagger$ and its Hermitian conjugate c_k and to require them to satisfy anticommutation relations

$$\begin{aligned} \left[c_k, c_{k'}^\dagger\right]_+ &\equiv c_k c_{k'}^\dagger + c_{k'}^\dagger c_k = \delta_{kk'}, \\ [c_k, c_{k'}]_+ &= \left[c_k^\dagger, c_{k'}^\dagger\right]_+ = 0. \end{aligned} \tag{11.4}$$

These relations assure occupancy of 0 or 1 since $\left(c_k^\dagger\right)^2 = 0$ and $(c_k)^2 = 0$:

$$\begin{aligned} \left[c_k^\dagger, c_k^\dagger\right]_+ &= 2c_k^\dagger c_k^\dagger = 0 \\ [c_k, c_k]_+ &= 2c_k c_k = 0 \end{aligned}$$

from the anticommutation relations given by (11.4). It is convenient to order the possible values of the quantum number k (e.g. the smallest k's first). Then, an eigenfunction can be written

$$|0_1, 1_2, 0_3, 0_4, 1_5, 1_6, \ldots, 1_k, \ldots\rangle = \cdots c_k^\dagger \cdots c_6^\dagger c_5^\dagger c_2^\dagger |0_1, 0_2, \ldots 0_k, \ldots, 0_n, \ldots\rangle.$$

The order is important, because interchanging $c_6^\dagger$ and $c_5^\dagger$ leads to

$$|0_1, 1_2, 0_3, 0_4, 1_6, 1_5, \ldots, 1_k, \ldots\rangle = -|0_1, 1_2, 0_3, 0_4, 1_5, 1_6, \ldots, 1_k, \ldots\rangle.$$

The kinetic (or single particle) energy part is given by

$$\sum_{\substack{k \\ \text{occupied}}} \langle k|H_0|k\rangle c_k^\dagger c_k = \sum_k \varepsilon_k c_k^\dagger c_k = \sum_k \varepsilon_k n_k. \tag{11.5}$$

The more difficult question is "How do we represent the interaction term in the second quantization or *occupation number representation*?".

In the coordinate representation the many particle product functions must be either symmetric for Bosons or antisymmetric for Fermions. Let us write out the case for Fermions

$$\Phi = \frac{1}{\sqrt{N!}} \sum_P (-)^P P \{\phi_\alpha(1)\phi_\beta(2) \cdots \phi_\omega(N)\} \tag{11.6}$$

Here, $\sum_P$ means sum over all permutations and $(-)^P$ is -1 for odd permutations and $+1$ for even permutations. For example, for a three particle state

the wave function $\Phi_{\alpha\beta\gamma}(1,2,3)$ can be written

$$\begin{aligned}\Phi_{\alpha\beta\gamma} = \tfrac{1}{\sqrt{3!}}\,[\,&\phi_\alpha(1)\phi_\beta(2)\phi_\gamma(3) - \phi_\alpha(1)\phi_\beta(3)\phi_\gamma(2) + \phi_\alpha(2)\phi_\beta(3)\phi_\gamma(1)\\ &- \phi_\alpha(2)\phi_\beta(1)\phi_\gamma(3) + \phi_\alpha(3)\phi_\beta(1)\phi_\gamma(2) - \phi_\alpha(3)\phi_\beta(2)\phi_\gamma(1)]\,.\end{aligned} \tag{11.7}$$

Such antisymmetrized product functions are often written as Slater determinants

$$\Phi = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_\alpha(1) & \phi_\alpha(2) & \cdots & \phi_\alpha(N) \\ \phi_\beta(1) & \phi_\beta(2) & \cdots & \phi_\beta(N) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_\omega(1) & \phi_\omega(2) & \cdots & \phi_\omega(N) \end{vmatrix}. \tag{11.8}$$

Look at V_{12} operating on a two particle wave function $\Phi_{\alpha\beta}(1,2)$. We assume that $V_{12} = V(|\mathbf{r_1} - \mathbf{r_2}|) = V(r_{12}) = V_{21}$. Then

$$V_{12}\Phi_{\alpha\beta}(1,2) = \frac{1}{\sqrt{2}}V_{12}\,[\phi_\alpha(1)\phi_\beta(2) - \phi_\beta(1)\phi_\alpha(2)]\,.$$

The matrix element $\langle\Phi_{\gamma\delta}|V_{12}|\Phi_{\alpha\beta}\rangle$ becomes

$$\begin{aligned}\langle\Phi_{\gamma\delta}|V_{12}|\Phi_{\alpha\beta}\rangle = &\frac{1}{2}\langle\gamma\delta|V_{12}|\alpha\beta\rangle + \frac{1}{2}\langle\delta\gamma|V_{12}|\beta\alpha\rangle\\ &-\frac{1}{2}\langle\gamma\delta|V_{12}|\beta\alpha\rangle - \frac{1}{2}\langle\delta\gamma|V_{12}|\alpha\beta\rangle.\end{aligned} \tag{11.9}$$

Since $\langle\gamma\delta|V_{12}|\alpha\beta\rangle = \int \mathrm{d}^3r_1\mathrm{d}^3r_2\,\phi^*_\gamma(1)\phi^*_\delta(2)V(r_{12})\phi_\alpha(1)\phi_\beta(2)$, we can see that it must be equal to $\langle\delta\gamma|V_{12}|\beta\alpha\rangle$ by simple interchange of the dummy variables r_1 and r_2. Thus, we find, for two-particle wave function, that

$$\langle\Phi_{\gamma\delta}|V_{12}|\Phi_{\alpha\beta}\rangle = \langle\gamma\delta|V_{12}|\alpha\beta\rangle - \langle\gamma\delta|V_{12}|\beta\alpha\rangle. \tag{11.10}$$

Just as we found in discussing the Heisenberg exchange interaction, we find that the antisymmetry leads to a direct term and an exchange term. Had we been considering Bosons instead of Fermions, a plus sign would have appeared in $\Phi_{\alpha\beta}(1,2)$ and in the expression for the matrix element.

Exactly the same result can be obtained by writing

$$V_{12} = \sum_{\lambda\lambda'\mu\mu'}\langle\lambda'\mu'|V_{12}|\lambda\mu\rangle c^\dagger_{\lambda'}c^\dagger_{\mu'}c_\mu c_\lambda, \tag{11.11}$$

and

$$|\Phi_{\alpha\beta}\rangle = \frac{1}{\sqrt{2}}c^\dagger_\beta c^\dagger_\alpha|0\rangle, \tag{11.12}$$

where $|0\rangle$ is the *vacuum state*, which contains no particles. It is clear that

$$V_{12}|\Phi_{\alpha\beta}\rangle = \frac{1}{\sqrt{2}}\sum_{\lambda\lambda'\mu\mu'}\langle\lambda'\mu'|V_{12}|\lambda\mu\rangle c^\dagger_{\lambda'}c^\dagger_{\mu'}c_\mu c_\lambda c^\dagger_\beta c^\dagger_\alpha|0\rangle$$

will vanish unless (1) $\lambda = \beta$ and $\mu = \alpha$ or (2) $\lambda = \alpha$ and $\mu = \beta$. From this, we see that

$$V_{12}|\Phi_{\alpha\beta}\rangle = \frac{1}{\sqrt{2}} \sum_{\lambda'\mu'} \left[\langle\lambda'\mu'|V_{12}|\beta\alpha\rangle - \langle\lambda'\mu'|V_{12}|\alpha\beta\rangle\right] c^\dagger_{\lambda'} c^\dagger_{\mu'}|0\rangle.$$

Taking the scalar product with $\langle\Phi_{\gamma\delta}| = \frac{1}{\sqrt{2}}\langle 0|c_\gamma c_\delta$ gives

$$\langle\Phi_{\gamma\delta}|V_{12}|\Phi_{\alpha\beta}\rangle = \frac{1}{2} \sum_{\lambda'\mu'} \left[\langle\lambda'\mu'|V_{12}|\beta\alpha\rangle - \langle\lambda'\mu'|V_{12}|\alpha\beta\rangle\right] \langle 0|c_\gamma c_\delta c^\dagger_{\lambda'} c^\dagger_{\mu'}|0\rangle. \tag{11.13}$$

The matrix element $\langle 0|c_\gamma c_\delta c^\dagger_{\lambda'} c^\dagger_{\mu'}|0\rangle$ will vanish unless (1) $\delta = \lambda'$ and $\gamma = \mu'$ or (2) $\gamma = \lambda'$ and $\delta = \mu'$. The final result can be seen to be

$$\langle\Phi_{\gamma\delta}|V_{12}|\Phi_{\alpha\beta}\rangle = \langle\gamma\delta|V_{12}|\alpha\beta\rangle - \langle\gamma\delta|V_{12}|\beta\alpha\rangle. \tag{11.14}$$

If we consider the operator $\frac{1}{2}\sum_{i\neq j} V_{ij}$ we need only note that we can consider a particular pair i, j first. Then, when V_{ij} operates on a many particle wave function

$$\frac{1}{\sqrt{N!}} \sum_P (-)^P P\left\{\phi_\alpha(1)\phi_\beta(2)\cdots\phi_\omega(N)\right\} = c^\dagger_\alpha c^\dagger_\beta \cdots c^\dagger_\omega|0\rangle \tag{11.15}$$

only particles i and j can change their single particle states. All the rest of the particles must remain in the same single particle states.

The final result is that the Hamiltonian of a many particle system with two body interactions can be written

$$H = \sum_{kk'} \langle k'|H_0|k\rangle c^\dagger_{k'} c_k + \frac{1}{2} \sum_{kk'll'} \langle k'l'|V|kl\rangle c^\dagger_{k'} c^\dagger_{l'} c_l c_k. \tag{11.16}$$

The operators c_k and $c^\dagger_{k'}$ satisfy either commutation relations for Bosons

$$\left[c_k, c^\dagger_{k'}\right]_- = \delta_{kk'}, \quad \text{and} \quad [c_k, c_{k'}]_- = \left[c^\dagger_k, c^\dagger_{k'}\right]_- = 0. \tag{11.17}$$

or anticommutation relations for Fermions

$$\left[c_k, c^\dagger_{k'}\right]_+ = \delta_{kk'}, \quad \text{and} \quad [c_k, c_{k'}]_+ = \left[c^\dagger_k, c^\dagger_{k'}\right]_+ = 0. \tag{11.18}$$

11.2 Hartree–Fock Approximation

Now, we are all familiar with the second quantized notation for a system of interacting particles. We can write

$$H = \sum_i \varepsilon_i c^\dagger_i c_i + \frac{1}{2} \sum_{ijkl} \langle ij|V|kl\rangle c^\dagger_i c^\dagger_j c_l c_k. \tag{11.19}$$

Here, $c_i^\dagger$ creates a particle in the state ϕ_i, and

$$\langle ij|V|kl\rangle = \int d\mathbf{x}d\mathbf{x}'\, \phi_i^*(\mathbf{x})\phi_j^*(\mathbf{x}')V(\mathbf{x},\mathbf{x}')\phi_{\mathbf{k}}(\mathbf{x})\phi_l(\mathbf{x}'). \tag{11.20}$$

Remember that

$$\langle ij|V|kl\rangle = \langle ji|V|lk\rangle \tag{11.21}$$

if V is a symmetric function of $\mathbf{x}$ and $\mathbf{x}'$. In this notation, $H_0 = \sum_i \varepsilon_i c_i^\dagger c_i$ is the Hamiltonian for a noninteracting system. It is simply the sum of the product of the energy ε_i of the state ϕ_i and the number operator $n_i = c_i^\dagger c_i$. The *Hartree–Fock approximation* is obtained by replacing the product of the four operators $c_i^\dagger c_j^\dagger c_l c_k$ by a c-number (actually a ground state expectation value of a $c^\dagger c$ product) multiplying a $c^\dagger c$; that is

$$\begin{aligned} c_i^\dagger c_j^\dagger c_l c_k \approx{}& c_i^\dagger \langle c_j^\dagger c_l\rangle c_k + c_j^\dagger c_l \langle c_i^\dagger c_k\rangle \\ & - c_i^\dagger c_l \langle c_j^\dagger c_k\rangle - c_j^\dagger c_k \langle c_i^\dagger c_l\rangle. \end{aligned} \tag{11.22}$$

By $\langle\hat{\Omega}\rangle$ we mean the expectation value of $\hat{\Omega}$ in the Hartree–Fock ground state, which we are trying to determine. Because this is a diagonal matrix element, we see that

$$\langle c_j^\dagger c_l\rangle = \delta_{jl}\,\overline{n}_j. \tag{11.23}$$

Furthermore, momentum conservation requires

$$\langle ij|V|jk\rangle = \langle ij|V|ji\rangle\delta_{ik},\ \text{etc.}$$

Then, one obtains for the Hartree–Fock Hamiltonian

$$H = \sum_i E_i c_i^\dagger c_i, \tag{11.24}$$

where

$$E_i = \varepsilon_i + \sum_j \overline{n_j}\left[\langle ij|V|ij\rangle - \langle ij|V|ji\rangle\right]. \tag{11.25}$$

One can think of E_i as the eigenvalue of a one particle Schrödinger equation

$$\begin{aligned} H_{\mathrm{HF}}\phi_i(\mathbf{x}) \equiv{}& \left\{\frac{p^2}{2m} + \int \mathrm{d}^3x' V(\mathbf{x},\mathbf{x}') \textstyle\sum_j \overline{n}_j \phi_j^*(\mathbf{x}')\phi_j(\mathbf{x}')\right\}\phi_i(\mathbf{x}) \\ & - \int \mathrm{d}^3x' V(\mathbf{x},\mathbf{x}') \textstyle\sum_j \overline{n}_j \phi_j^*(\mathbf{x}')\phi_i(\mathbf{x}')\phi_j(\mathbf{x}) = E_i\phi_i. \end{aligned} \tag{11.26}$$

Do not think the Hartree–Fock approximation is trivial. One must assume a ground state configuration to compute $\langle c_j^\dagger c_l\rangle$. One then solves the "one particle" problem and hopes that the solution is such that the ground state of the N particle system, determined by filling the N lowest energy single particle states just solved for, is identical to the ground state assumed in computing $\langle c_j^\dagger c_l\rangle$. If it is not, the problem has not been solved.

11.2.1 Ferromagnetism of a degenerate electron gas in Hartree–Fock Approximation

One can easily verify that plane wave eigenfunctions

$$\phi_{\mathbf{k}s}(\mathbf{x}) = \Omega^{-1/2} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} \eta_s,$$

with single particle energy

$$\varepsilon_{\mathbf{k}s} = \frac{\hbar^2 k^2}{2m}$$

form a set of solutions of the single particle Hartree–Fock Hamiltonian.

If the ground state is assumed to be the *paramagnetic state*, in which the N lowest energy levels are occupied (each $\mathbf{k}$ state is occupied by one electron of spin $\uparrow$ and one of spin $\downarrow$) then one obtains

$$E_{\mathbf{k}s} = \varepsilon_{\mathbf{k}s} + E_{\mathrm{X}s}(\mathbf{k}) \tag{11.27}$$

where

$$E_{\mathrm{X}s}(\mathbf{k}) = -\sum_{\mathbf{k}'} n_{\mathbf{k}'} \langle \mathbf{k}\mathbf{k}' | V | \mathbf{k}'\mathbf{k} \rangle. \tag{11.28}$$

Here, we assumed that the nuclei are fixed in a given configuration and pictured as a fixed source of a static potential. The matrix element $\langle \mathbf{k}\mathbf{k}' | V | \mathbf{k}'\mathbf{k} \rangle = \frac{4\pi e^2}{|\mathbf{k}-\mathbf{k}'|^2}$, and the sum over $\mathbf{k}'$ can be performed to obtain

$$E_{\mathbf{k}s} = \frac{\hbar^2 k^2}{2m} - \frac{e^2 k_{\mathrm{F}}}{2\pi} \left[2 + \frac{k_{\mathrm{F}}^2 - k^2}{k k_{\mathrm{F}}} \ln \left(\frac{k_{\mathrm{F}} + k}{k_{\mathrm{F}} - k} \right) \right]. \tag{11.29}$$

The total energy of the paramagnetic state is

$$E_{\mathrm{P}} = \sum_{\mathbf{k}s} n_{\mathbf{k}s} \left[\varepsilon_{\mathbf{k}s} + \frac{1}{2} E_{\mathrm{X}s}(\mathbf{k}) \right]. \tag{11.30}$$

The $\frac{1}{2}$ in front of $E_{\mathrm{X}s}$ prevents double counting. This sum gives

$$E_{\mathrm{P}} = N \left[\frac{3}{5} \frac{\hbar^2 k_{\mathrm{F}}^2}{2m} - \frac{3}{4\pi} e^2 k_{\mathrm{F}} \right] \simeq N \left[\frac{2.21}{r_{\mathrm{s}}^2} - \frac{0.916}{r_{\mathrm{s}}} \right] \text{ Ryd.} \tag{11.31}$$

One can easily see that $E_{\mathbf{k}s}$ is a monotonically increasing function of k, so that the assumption about the ground state, viz. that all $\mathbf{k}$ state for which $k < k_{\mathrm{F}}$ are occupied, is in agreement with the procedure of filling the N lowest energy eigenstates of the single particle Hartree–Fock Hamiltonian.

Instead of assuming the paramagnetic ground state, we could assume that only states of spin $\uparrow$ are occupied, and that they are singly occupied for all $k < 2^{1/3} k_{\mathrm{F}}$. Then one finds that

$$E_{k\uparrow} = \frac{\hbar^2 k^2}{2m} - \frac{2^{1/3} e^2 k_{\mathrm{F}}}{2\pi}\left[2 + \frac{2^{2/3}k_{\mathrm{F}}^2 - k^2}{2^{1/3}k_{\mathrm{F}}k}\ln\left(\frac{2^{1/3}k_{\mathrm{F}} + k}{2^{1/3}k_{\mathrm{F}} - k}\right)\right]$$
$$E_{k\downarrow} = \frac{\hbar^2 k^2}{2m}. \tag{11.32}$$

This state is a solution to the Hartree–Fock problem only if

$$E_{k\uparrow}|_{k=2^{1/3}k_{\mathrm{F}}} < 0, \tag{11.33}$$

otherwise some of the spin down states would be occupied in the ground state. This condition is satisfied if

$$\frac{1}{a_0 k_{\mathrm{F}}} > \frac{\pi}{2^{2/3}} \simeq \frac{3.142}{1.588} = 1.98 \tag{11.34}$$

It is convenient to introduce the parameter r_{s} defined by

$$\frac{4\pi}{3}\left(a_0 r_{\mathrm{s}}\right)^3 = \frac{V}{N} = \frac{3\pi^2}{k_{\mathrm{F}}^3}.$$

Then, we have

$$\left(\frac{4}{9\pi}\right)^{1/3} r_{\mathrm{s}} = \left(a_0 k_{\mathrm{F}}\right)^{-1},$$

or

$$r_{\mathrm{s}} = \left(\frac{9\pi}{4}\right)^{1/3} a_0^{-1} k_{\mathrm{F}}^{-1} \simeq \frac{1.92}{a_0 k_{\mathrm{F}}}. \tag{11.35}$$

If we sum over k to get the energy of the ferromagnetic state

$$E_{\mathrm{F}} = \sum E_{k\uparrow} = N\left[2^{2/3}\frac{3}{5}\frac{\hbar^2 k_{\mathrm{F}}^2}{2m} - 2^{1/3}\frac{3}{4\pi}e^2 k_{\mathrm{F}}\right]. \tag{11.36}$$

Comparing E_{F} with E_{P}, we see that

$$E_{\mathrm{F}} < E_{\mathrm{P}} \quad \text{if} \quad a_0 k_{\mathrm{F}} < \frac{5}{2\pi}\frac{1}{2^{1/3}+1},$$

which corresponds to

$$r_{\mathrm{s}} > 5.45, \tag{11.37}$$

though the Hartree–Fock solution exists if $r_{\mathrm{s}} \geq 3.8$. The present, Hartree–Fock, treatment neglects *correlation effects* and cannot be expected to describe accurately the behavior of metals. The present treatment does however point up the fact that the exchange energy prefers parallel spin orientation, but the cost in kinetic energy is high for a ferromagnetic spin arrangement. Actually Cs has $r_{\mathrm{s}} \simeq 5.6$ and does not show ferromagnetic behavior; this is not too surprising.

11.3 Spin Density Waves

We have seen that the exchange energy favors parallel spin alignment, but that the cost in kinetic energy is high. Overhauser[1] proposed a solution of the Hartree–Fock problem in which the spins are locally parallel, but the spin polarization rotates as one moves through the crystal. This type of state enhances the (negative) exchange energy but does not cost as much in kinetic energy.

For example, an Overhauser *spiral spin density wave* could exist with a net fractional spin polarization given by

$$\mathbf{P}_\perp(\mathbf{r}) = P_{\perp 0}(\hat{\mathbf{x}}\cos Qz + \hat{\mathbf{y}}\sin Qz). \tag{11.38}$$

Overhauser showed that such a spin polarization $\mathbf{P}_\perp(\mathbf{r})$ can result from taking basis functions of the form

$$|\phi_{\mathbf{k}}\rangle = a_k|\mathbf{k}\uparrow\rangle + b_k|\mathbf{k}+\mathbf{Q}\downarrow\rangle.$$

In order that $\langle\phi_{\mathbf{k}}|\phi_{\mathbf{k}}\rangle = 1$, it is necessary that $a_k^2 + b_k^2 = 1$. This condition assures that there is no fluctuation in the charge density associated with the wave. Thus, without loss of generality we can take $a_k = \cos\theta_k$ and $b_k = \sin\theta_k$ and write

$$|\phi_{\mathbf{k}}\rangle = \cos\theta_k|\mathbf{k}\uparrow\rangle + \sin\theta_k|\mathbf{k}+\mathbf{Q}\downarrow\rangle. \tag{11.39}$$

The fractional spin polarization at a point $\mathbf{r} = \mathbf{r}_0$ is given by

$$\mathbf{P}(\mathbf{r}_0) = \frac{\Omega}{N}\sum_{k\ \text{occupied}}\langle\phi_{\mathbf{k}}|\underline{\sigma}\delta(\mathbf{r}-\mathbf{r}_0)|\phi_{\mathbf{k}}\rangle. \tag{11.40}$$

Here, $\underline{\sigma} = \sigma_x\hat{x} + \sigma_y\hat{y} + \sigma_z\hat{z}$, where σ_x, σ_y, σ_z are Pauli spin matrices, so that

$$\underline{\sigma} = \begin{pmatrix} \hat{\mathbf{z}} & \hat{\mathbf{x}} - \mathrm{i}\hat{\mathbf{y}} \\ \hat{\mathbf{x}} + \mathrm{i}\hat{\mathbf{y}} & -\hat{\mathbf{z}} \end{pmatrix}. \tag{11.41}$$

We can write

$$|\mathbf{k}\uparrow\rangle = |\mathbf{k}\rangle|\uparrow\rangle = \Omega^{-1/2}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\begin{pmatrix}1\\0\end{pmatrix}$$

and

$$|\mathbf{k}+\mathbf{Q}\downarrow\rangle = |\mathbf{k}+\mathbf{Q}\rangle|\downarrow\rangle = \Omega^{-1/2}\mathrm{e}^{\mathrm{i}(\mathbf{k}+\mathbf{Q})\cdot\mathbf{r}}\begin{pmatrix}0\\1\end{pmatrix}.$$

Then

$$\begin{aligned}
&\langle\uparrow|\underline{\sigma}|\uparrow\rangle = \hat{\mathbf{z}}\\
&\langle\uparrow|\underline{\sigma}|\downarrow\rangle = \hat{\mathbf{x}} - \mathrm{i}\hat{\mathbf{y}}\\
&\langle\downarrow|\underline{\sigma}|\uparrow\rangle = \hat{\mathbf{x}} + \mathrm{i}\hat{\mathbf{y}}\\
&\langle\downarrow|\underline{\sigma}|\downarrow\rangle = -\hat{\mathbf{z}}.
\end{aligned} \tag{11.42}$$

[1] A.W. Overhauser, Phys. Rev. **128**, 1437 (1962).

Evaluating $\langle\phi_{\mathbf{k}}|\underline{\sigma}\,\delta(\mathbf{r}-\mathbf{r}_0)|\phi_{\mathbf{k}}\rangle$ gives

$$\begin{aligned}&\langle\phi_{\mathbf{k}}|\underline{\sigma}\delta(\mathbf{r}-\mathbf{r}_0)|\phi_{\mathbf{k}}\rangle\\ &=\tfrac{1}{\Omega}\left\{\begin{array}{l}\cos^2\theta_k\langle\uparrow|\underline{\sigma}|\uparrow\rangle+\sin^2\theta_k\langle\downarrow|\underline{\sigma}|\downarrow\rangle\\ +\cos\theta_k\sin\theta_k\left[\mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_0}\ \langle\uparrow|\underline{\sigma}|\downarrow\rangle+\mathrm{e}^{-\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_0}\langle\downarrow|\underline{\sigma}|\uparrow\rangle\right]\end{array}\right\}.\end{aligned}\tag{11.43}$$

Gathering together the terms allows us to express $\mathbf{P}(\mathbf{r}_0)$ as

$$\mathbf{P}(\mathbf{r}_0)=P_{\parallel}\hat{\mathbf{z}}+P_{\perp}\left(\hat{\mathbf{x}}\cos\mathbf{Q}\cdot\mathbf{r}_0+\hat{\mathbf{y}}\sin\mathbf{Q}\cdot\mathbf{r}_0\right),\tag{11.44}$$

where

$$P_{\parallel}=\frac{1}{8\pi^3 n}\int_{\text{occupied}}\cos 2\theta_k\,\mathrm{d}^3k,\tag{11.45}$$

and

$$P_{\perp}=\frac{1}{8\pi^3 n}\int_{\text{occupied}}\sin 2\theta_k\,\mathrm{d}^3k.\tag{11.46}$$

Here, $n=\frac{N}{\Omega}$ and the integral is over all occupied states $|\phi_{\mathbf{k}}\rangle$. We will not worry about $P_{\parallel}$ because ultimately we will consider a linear combination of two spiral spin density waves (called a *linear spin density wave*) for which the $P_{\parallel}$'s cancel.

It is worth noting that the density at point $\mathbf{r}_0$ is given by

$$\begin{aligned}n(\mathbf{r}_0)&=\textstyle\sum_{\mathbf{k}}\langle\phi_{\mathbf{k}}|\underline{\mathbf{1}}\delta(\mathbf{r}-\mathbf{r}_0)|\phi_{\mathbf{k}}\rangle\\ &=\tfrac{1}{\Omega}\textstyle\sum_{\mathbf{k}}\left(\cos^2\theta_k+\sin^2\theta_k\right)=\frac{N}{\Omega}.\end{aligned}\tag{11.47}$$

When the unit matrix $\underline{\mathbf{1}}$ is replaced by $\underline{\sigma}$, it is reasonable to expect the spin density.

One can form a wave function orthogonal to $|\phi_{\mathbf{k}}\rangle$:

$$|\psi_{\mathbf{k}}\rangle=-\sin\theta_k|\mathbf{k}\uparrow\rangle+\cos\theta_k|\mathbf{k}+\mathbf{Q}\downarrow\rangle.\tag{11.48}$$

So far, we have ignored these states (i.e. assumed they were unoccupied). We shall see that this turns out to be correct for the Hartree–Fock spin density wave ground state.

Recall that the Hartree–Fock wave functions $\phi_{\mathbf{k}}(\mathbf{x})$ satisfy (11.26)

$$\begin{aligned}H_{\mathrm{HF}}\phi_{\mathbf{k}}(\mathbf{x})\equiv&\left\{\tfrac{p^2}{2m}+\textstyle\int\mathrm{d}\mathbf{x}'V(\mathbf{x},\mathbf{x}')\sum_{\mathbf{q}}\overline{n}_{\mathbf{q}}\phi^*_{\mathbf{q}}(\mathbf{x}')\phi_{\mathbf{q}}(\mathbf{x}')\right\}\phi_{\mathbf{k}}(\mathbf{x})\\ &-\textstyle\int\mathrm{d}\mathbf{x}'V(\mathbf{x},\mathbf{x}')\sum_{\mathbf{q}}\overline{n}_{\mathbf{q}}\phi^*_{\mathbf{q}}(\mathbf{x}')\phi_{\mathbf{k}}(\mathbf{x}')\phi_{\mathbf{q}}(\mathbf{x})=E_{\mathbf{k}}\phi_{\mathbf{k}}.\end{aligned}\tag{11.49}$$

We can write H_{HF} as

$$H_{\mathrm{HF}}=\frac{p^2}{2m}+U+A,\tag{11.50}$$

where

$$U(\mathbf{x})=\int d\mathbf{x}'V(\mathbf{x},\mathbf{x}')\sum_{q\ \text{occupied}}\phi^*_{\mathbf{q}}(\mathbf{x}')\phi_{\mathbf{q}}(\mathbf{x}')\tag{11.51}$$

and

$$A\psi(\mathbf{x}) = -\int d\mathbf{x}' V(\mathbf{x},\mathbf{x}') \sum_{q \text{ occupied}} \phi^*_{\mathbf{q}}(\mathbf{x}')\psi(\mathbf{x}')\phi_{\mathbf{q}}(\mathbf{x}). \tag{11.52}$$

$V(\mathbf{x},\mathbf{x}')$ can be written as

$$V(\mathbf{x},\mathbf{x}') = \sum_{\mathbf{q}\neq 0} V_q \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')}. \tag{11.53}$$

Now, consider the matrix elements of A (with the Hartree–Fock ground state assumed to be made up of the lowest energy $\phi_{\mathbf{k}}$ states) between plane wave states.

$$\langle \boldsymbol{\ell}\underline{\sigma}|A|\boldsymbol{\ell}'\underline{\sigma}'\rangle = -\sum_{\mathbf{k}}{}' \sum_{q\neq 0} V_q \langle\phi_{\mathbf{k}}|\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{x}'}|\boldsymbol{\ell}'\underline{\sigma}'\rangle\langle\boldsymbol{\ell}\underline{\sigma}|\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{x}}|\phi_{\mathbf{k}}\rangle, \tag{11.54}$$

where $\sum'_{\mathbf{k}}$ means sum over all occupied states $|\phi_{\mathbf{k}}\rangle$. Now use the expressions

$$\begin{aligned} |\phi_{\mathbf{k}}\rangle &= \cos\theta_k|\mathbf{k}\uparrow\rangle + \sin\theta_k|\mathbf{k}+\mathbf{Q}\downarrow\rangle \\ \langle\phi_{\mathbf{k}}| &= \langle\mathbf{k}\uparrow|\cos\theta_k + \langle\mathbf{k}+\mathbf{Q}\downarrow|\sin\theta_k \end{aligned} \tag{11.55}$$

to obtain

$$\begin{aligned} &\langle\boldsymbol{\ell}\underline{\sigma}|A|\boldsymbol{\ell}'\underline{\sigma}'\rangle \\ &= -\textstyle\sum'_{\mathbf{k}}\sum_{q\neq 0} V_q \left\{\langle\mathbf{k}\uparrow|\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{x}'}|\boldsymbol{\ell}'\sigma'\rangle\cos\theta_k + \langle\mathbf{k}+\mathbf{Q}\downarrow|\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{x}}|\boldsymbol{\ell}'\sigma'\rangle\sin\theta_k\right\} \\ &\quad\times\left\{\langle\boldsymbol{\ell}\sigma|\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{x}'}|\mathbf{k}\uparrow\rangle\cos\theta_k + \langle\boldsymbol{\ell}\sigma|\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{x}}|\mathbf{k}+\mathbf{Q}\downarrow\rangle\sin\theta_k\right\}. \end{aligned} \tag{11.56}$$

Because $\mathrm{e}^{\pm\mathrm{i}\mathbf{q}\cdot\mathbf{x}}$ is spin independent, we can use $\langle\sigma|\uparrow\rangle = \delta_{\sigma\uparrow}$, $\langle\sigma|\downarrow\rangle = \delta_{\sigma\downarrow}$, *etc.* to obtain

$$\begin{aligned} &\langle\boldsymbol{\ell}\underline{\sigma}|A|\boldsymbol{\ell}'\underline{\sigma}'\rangle \\ &= -\textstyle\sum'_{\mathbf{k}}\sum_{q\neq 0} V_q\ (\langle\mathbf{k}+\mathbf{q}|\boldsymbol{\ell}'\rangle\delta_{\sigma'\uparrow}\cos\theta_k + \langle\mathbf{k}+\mathbf{Q}+\mathbf{q}|\boldsymbol{\ell}'\rangle\delta_{\sigma'\downarrow}\sin\theta_k) \\ &\qquad\times(\langle\boldsymbol{\ell}|\mathbf{k}+\mathbf{q}\rangle\delta_{\sigma\uparrow}\cos\theta_k + \langle\boldsymbol{\ell}|\mathbf{k}+\mathbf{Q}+\mathbf{q}\rangle\delta_{\sigma\downarrow}\sin\theta_k). \end{aligned} \tag{11.57}$$

For $\sigma = \uparrow$ and $\sigma' = \downarrow$ we find

$$\langle\boldsymbol{\ell}\uparrow|A|\boldsymbol{\ell}'\downarrow\rangle = -\sum_{\mathbf{k}}{}' \sum_{q\neq 0} V_q\, \delta_{\mathbf{k}+\mathbf{Q}+\mathbf{q},\boldsymbol{\ell}'} \sin\theta_k\, \delta_{\boldsymbol{\ell},\mathbf{k}+\mathbf{q}} \cos\theta_k, \tag{11.58}$$

which can be rewritten

$$\langle\boldsymbol{\ell}\uparrow|A|\boldsymbol{\ell}'\downarrow\rangle = -\sum_{\mathbf{k}}{}' V_{\ell-k} \sin\theta_k\cos\theta_k\, \delta_{\boldsymbol{\ell}',\boldsymbol{\ell}+\mathbf{Q}}. \tag{11.59}$$

Thus, the Hartree–Fock exchange term A has off diagonal elements mixing the simple plane wave states $|\boldsymbol{\ell}\uparrow\rangle$ and $|\boldsymbol{\ell}+\mathbf{Q}\downarrow\rangle$. It is straightforward to see that

$$\langle \boldsymbol{\ell} \downarrow |A| \boldsymbol{\ell}' \uparrow \rangle = -\sum_{\mathbf{k}}{}' V_{\ell'-k} \sin\theta_k \cos\theta_k \, \delta_{\boldsymbol{\ell}', \boldsymbol{\ell}-\mathbf{Q}}, \tag{11.60}$$

so that A also couples $|\boldsymbol{\ell} \downarrow\rangle$ to $|\boldsymbol{\ell} - \mathbf{Q} \uparrow\rangle$. The spin diagonal terms are

$$\langle \boldsymbol{\ell} \uparrow |A| \boldsymbol{\ell} \uparrow \rangle = -\sum_{\mathbf{k}}{}' V_{\ell-k} \cos^2\theta_k \tag{11.61}$$

and

$$\langle \boldsymbol{\ell} + \mathbf{Q} \downarrow |A| \boldsymbol{\ell} + \mathbf{Q} \downarrow \rangle = -\sum_{\mathbf{k}}{}' V_{\ell-k} \sin^2\theta_k. \tag{11.62}$$

Then, we need to solve the problem given by

$$\left(\frac{p^2}{2m} + A_{\mathrm{D}} + A_{\mathrm{OD}} - E_{\mathbf{k}} \right) \Psi_{\mathbf{k}} = 0, \tag{11.63}$$

where

$$A_{\mathrm{D}} = -\begin{pmatrix} \sum_{\mathbf{k}'}{}' V_{k-k'} \cos^2\theta_{k'} & 0 \\ 0 & \sum_{\mathbf{k}'}{}' V_{k-k'} \sin^2\theta_{k'} \end{pmatrix} \tag{11.64}$$

and

$$A_{\mathrm{OD}} = -g_k \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{11.65}$$

We can simply take $|\Psi_{\mathbf{k}}\rangle = \cos\theta_k |\mathbf{k} \uparrow\rangle + \sin\theta_k |\mathbf{k} + \mathbf{Q} \downarrow\rangle$ and observe that (11.63) becomes

$$\begin{pmatrix} \varepsilon_{\mathbf{k}\uparrow} - E_{\mathbf{k}} & -g_k \\ -g_k & \varepsilon_{\mathbf{k}+\mathbf{Q}\downarrow} - E_{\mathbf{k}} \end{pmatrix} \begin{pmatrix} \cos\theta_k \\ \sin\theta_k \end{pmatrix} = 0. \tag{11.66}$$

In this matrix equation, g_k denotes the amplitude of the off-diagonal contribution of the exchange term A

$$g_k = \langle \mathbf{k} \uparrow |A| \mathbf{k} + \mathbf{Q} \downarrow \rangle = \sum_{\mathbf{k}'}{}' V_{\mathbf{k}-\mathbf{k}'} \sin\theta_{k'} \cos\theta_{k'}, \tag{11.67}$$

and $\varepsilon_{\mathbf{k}\uparrow}$ and $\varepsilon_{\mathbf{k}+\mathbf{Q}\downarrow}$ are the free electron energies plus the diagonal parts (A_{D}) of the one electron exchange energy

$$\begin{aligned} \varepsilon_{\mathbf{k}\uparrow} &= \frac{\hbar^2 k^2}{2m} - \sum_{\mathbf{k}'}{}' V_{|\mathbf{k}-\mathbf{k}'|} \cos^2\theta_{k'} \\ \varepsilon_{\mathbf{k}+\mathbf{Q}\downarrow} &= \frac{\hbar^2 (\mathbf{k}+\mathbf{Q})^2}{2m} - \sum_{\mathbf{k}'}{}' V_{|\mathbf{k}-\mathbf{k}'|} \sin^2\theta_{k'}. \end{aligned} \tag{11.68}$$

The eigenvalues $E_{\mathbf{k}}$ are determined from (11.66) by setting the determinant of the 2×2 matrix equal to zero. This gives

$$E_{\mathbf{k}\pm} = \frac{1}{2}\left(\varepsilon_{\mathbf{k}\uparrow} + \varepsilon_{\mathbf{k}+\mathbf{Q}\downarrow}\right) \pm \left[\frac{1}{4}\left(\varepsilon_{\mathbf{k}\uparrow} - \varepsilon_{\mathbf{k}+\mathbf{Q}\downarrow}\right)^2 + g_k^2\right]^{1/2}. \tag{11.69}$$

The eigenfunctions corresponding to $E_{\mathbf{k}\pm}$ are given by (11.48) and (11.55), respectively. The values of $\cos\theta_k$ are determined from (11.66) using the eigenvalues $E_{\mathbf{k}-}$ given above. This gives

$$\cos\theta_k = \frac{g_k}{\left[g_k^2 + (\varepsilon_{\mathbf{k}\uparrow} - E_{\mathbf{k}-})^2\right]^{1/2}}. \tag{11.70}$$

We note that the square modulus of the eigenfunction is a constant, and thus a charge density wave does not accompany a spin density wave.

Solution of the Integral Equation

We have to solve the integral equation (11.67), which is rewritten as

$$g_l = \int V_{l-k} \cos\theta_k \sin\theta_k \frac{\mathrm{d}^3 k}{(2\pi)^3}. \tag{11.71}$$

Here, $\cos\theta_k$ is given by (11.70), and the ground state eigenvalue $E_{\mathbf{k}}$ is itself a function of θ_k and hence of g_k. This equation is extremely complicated, and no solution is known for the general case. To obtain some feeling for what is happening we study the simple case where $V_{\boldsymbol{\ell}-\mathbf{k}}$ is constant instead of being given by $\frac{4\pi e^2}{|\boldsymbol{\ell}-\mathbf{k}|^2}$. We take $V_{\boldsymbol{\ell}-\mathbf{k}} = \gamma$; this corresponds to replacing the Coulomb interaction by a δ-function interaction. Obviously g_k will be independent of k in this case, and the integral equation becomes

$$g = \gamma \int \frac{d^3 k}{(2\pi)^3} \frac{g\left(\varepsilon_{\mathbf{k}\uparrow} - E_{\mathbf{k}}\right)}{g^2 + (\varepsilon_{\mathbf{k}\uparrow} - E_{\mathbf{k}})^2} \tag{11.72}$$

where

$$\varepsilon_{\mathbf{k}\uparrow} - E_{\mathbf{k}} = \frac{1}{2}(\varepsilon_{\mathbf{k}\uparrow} - \varepsilon_{\mathbf{k}+\mathbf{Q}\downarrow}) + \left[\frac{1}{4}(\varepsilon_{\mathbf{k}\uparrow} - \varepsilon_{\mathbf{k}+\mathbf{Q}\downarrow})^2 + g^2\right]^{1/2}. \tag{11.73}$$

By direct substitution we have

$$\frac{g\left(\varepsilon_{\mathbf{k}\uparrow} - E_{\mathbf{k}}\right)}{g^2 + (\varepsilon_{\mathbf{k}\uparrow} - E_{\mathbf{k}-})^2} = \frac{g}{2\left[\frac{1}{4}(\varepsilon_{\mathbf{k}\uparrow} - E_{\mathbf{k}})^2 + g^2\right]^{1/2}}, \tag{11.74}$$

and the integral equation becomes

$$g = \gamma \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \frac{g}{2\left[\frac{1}{4}(\varepsilon_{\mathbf{k}\uparrow} - E_{\mathbf{k}})^2 + g^2\right]^{1/2}}. \tag{11.75}$$

We replace $\varepsilon_{\mathbf{k}\uparrow} - E_{\mathbf{k}}$ in (11.75) by

$$\begin{aligned} \varepsilon_{\mathbf{k}\uparrow} - E_{\mathbf{k}} &\approx -\frac{\hbar^2}{m} Q(k_z + \frac{Q}{2}) \\ &= 2\left(\frac{\partial \varepsilon}{\partial k_z}\right)_{k_z = -\frac{Q}{2}} \left(k_z + \frac{Q}{2}\right). \end{aligned} \tag{11.76}$$

Here, we note that we left out a term $-\gamma \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \cos^2 \theta_k$. This is the same term which appeared in $P_\parallel$, and it had better vanish when we evaluate it using the solution to the integral equation for θ_k. Now let us introduce

$$\mu = \left(-\frac{\partial \varepsilon}{\partial k_z}\right)_{k_z = -Q/2}. \tag{11.77}$$

Then, we have

$$1 = \frac{\gamma}{(2\pi)^3} \int \frac{\mathrm{d}^3 k}{2\sqrt{g^2 + \mu^2 \left(k_z + \frac{Q}{2}\right)^2}}. \tag{11.78}$$

Take the region of integration to be a circular cylinder of radius $\kappa_\perp$ and of length $\kappa_\parallel$, centered at $k_z = -\frac{Q}{2}$ as shown in Fig. 11.1. Then (11.78) becomes

$$1 = \frac{\gamma}{(2\pi)^3} \int_{-\kappa_\parallel/2}^{\kappa_\parallel/2} \frac{\pi \kappa_\perp^2 \, d(k_z + Q/2)}{2\sqrt{g^2 + \mu^2 \left(k_z + \frac{Q}{2}\right)^2}} = \frac{\gamma \kappa_\perp^2}{16\pi^2 \mu} 2 \sinh^{-1}\left(\frac{\kappa_\parallel \mu}{2g}\right). \tag{11.79}$$

Thus, we obtain

$$g = \frac{\kappa_\parallel \mu}{2 \sinh\left(\frac{8\pi^2 \mu}{\gamma \kappa_\perp^2}\right)} \tag{11.80}$$

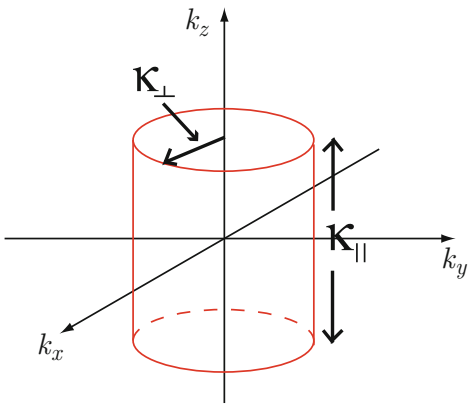


Fig. 11.1. Region of integration for (11.78). The cylinder axis is parallel to the z axis

One can now return to the equation for the ground state energy W

$$W_{\mathrm{GS}} = \frac{\hbar^2}{2m}\int \frac{\mathrm{d}^3k}{(2\pi)^3}\left[k^2\cos^2\theta_k + (\mathbf{k}+\mathbf{Q})^2\sin^2\theta_k\right] - \frac{1}{2}\int \frac{\mathrm{d}^3k\,\mathrm{d}^3k'}{(2\pi)^6}V_{k-k'}\cos^2\left(\theta_k - \theta'_k\right) \tag{11.81}$$

and replace $V_{k-k'}$ by the constant γ and substitute

1. $g = 0$ for the trivial solution corresponding to the usual paramagnetic state and
2. $g = \frac{\kappa_\parallel \mu}{2\sinh\left(\frac{8\pi^2\mu}{\gamma\kappa_\perp^2}\right)}$ for the spin density wave state

If we again take the occupied region in k space to be a cylinder of radius $\kappa_\perp$ and length $\kappa_\parallel$ centered at $k_z = -\frac{Q}{2}$, we obtain the deformation energy of the spin density wave state

$$W_{\mathrm{SDW}} - W_{\mathrm{P}} = -\frac{\mu\kappa_\parallel^2\kappa_\perp^2}{32\pi^2}\left[\coth\left(\frac{8\pi^2\mu}{\gamma\kappa_\perp^2}\right) - 1\right] < 0. \tag{11.82}$$

This quantity is negative definite, so that the spin density wave state always has the lower energy than the paramagnetic state, i.e.

$$W_{\mathrm{P}} > W_{\mathrm{SDW}}.$$

Note that in evaluation of W_{P} as well as of W_{SDW}, the occupied region of k space was taken to be a cylinder of radius $\kappa_\perp$ and length $\kappa_\parallel$ centered at $k_z = -\frac{Q}{2}$. The result does not prove that the spin density wave has lower energy than the actual paramagnetic ground state (which will be a sphere in k space instead of a cylinder, and hence have a smaller kinetic energy than the cylinder. Overhauser gave a general (but somewhat difficult) proof that a spin density wave state always exists which has lower energy than the paramagnetic state in the Hartree–Fock approximation.

The proof involves taking the wave vector of the spin density wave $\mathbf{Q}$ to be slightly smaller than $2k_{\mathrm{F}}$. Then, the spin up states at k_z are coupled by the exchange interaction to the spin down states at $k_z + Q$ as shown in Fig. 11.2. The gap (at $|k_z| = Q/2$) of the strongly coupled states causes a repopulation of k-space as indicated schematically (for the states that were spin $\downarrow$ without the spin density wave coupling) in Fig. 11.3. The flattened areas occur, of course, at the energy gap centered at $k_z = -\frac{Q}{2}$. The *repopulation energy* depends on $\kappa_\perp$, which is given by $\kappa_\perp = \sqrt{k_{\mathrm{F}}^2 - Q^2/4}$ and is much smaller than k_{F}. The dependence of the energy on the value of $\kappa_\perp$ can be used to demonstrate that the kinetic energy increase due to repopulation is small for $\kappa_\perp \ll k_{\mathrm{F}}$ and that in the Hartree–Fock approximation a spin density wave state always exists with energy lower than that of the paramagnetic state.

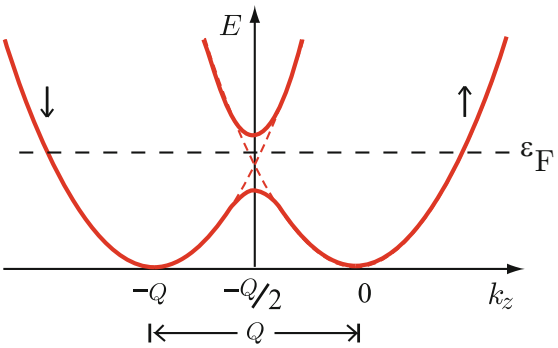


Fig. 11.2. Energies of ↑ and ↓ spin electrons as a function of k_z. The spin ↑ and spin ↓ minima have been separated by Q the wave number of the spin density wave. The thin solid lines omit the spin density wave exchange energy. The thick solid lines include it and give rise to the spin density wave gap. Near the gap, the eigenstates are linear combinations of $|k_z \uparrow\rangle$ and $|k_z + Q \downarrow\rangle$

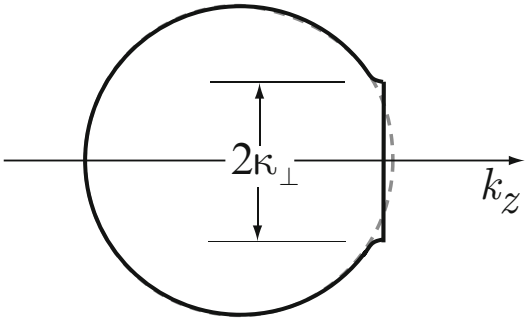


Fig. 11.3. Schematics of repopulation in k-space from originally occupied $k \downarrow$ states (inside sphere of radius k_F denoted by dashed circle) to inside of solid curve terminated plane $k_z = Q/2$ at which spin density wave gas occurs. The size of $\kappa_\perp$ is exaggerated for sake of clarity

For a spiral spin density wave the flat surface at $|k_z| = Q/2$ occur at opposite sides of the Fermi surface for spin ↑ and spin ↓ electrons. Near these positions in k-space, one cannot really speak of spin ↑ and spin ↓ electrons since the eigenstates are linear combinations of the two spins with comparable amplitudes. Far away from these regions (on the opposite sides of the Fermi surface) the spin states are essentially unmixed. The total energy can be lowered by introducing a left-handed spiral spin density wave in addition to the right-handed one. The resulting spiral spin density waves form a *linear spin density wave* which has flat surfaces separated by the wave vector $\mathbf{Q}$ of the spin density wave at both sides of the Fermi surface for each of the spins.

11.3.1 Comparison with Reality

It is not at all clear what correlation effects will do to the balance which gave $W_{\mathrm{P}} > W_{\mathrm{SDW}}$. So far no one has performed correlation calculations using the spin density wave state as a starting point. Experiment seems to show that at low temperatures the ground state of some metals, for example chromium, is a spin density wave state. Shortly after introducing spin density wave states, Overhauser.[2] introduced the idea of charge density wave states In a charge density wave state the spin magnetization vanishes everywhere, but the electron charge density has an oscillating position dependence. For a spin density wave distortion, exchange favors the distortion but correlation does not. For a charge density wave distortion, both exchange and correlation favor the distortion. However, the electrostatic (Hartree) energy associated with the charge density wave is large and unfavorable unless some other charge distortion cancels it. For soft metals like Na, K, and Pb, Overhauser claims the ground state is a charge density wave state. Some other people believe it is not. There is absolutely no doubt (from experiment) that the layered compounds like TaS_2 (and many others) have charge density wave ground states. There are many experimental results for Na, K, and Pb that do not fit the nearly free electron paramagnetic ground state, which Overhauser can explain with the charge density wave model. At the moment, the question is not completely settled. In the charge density wave materials, the large electrostatic energy (due to the Hartree field produced by the electronic charge density distortion) must be compensated by an equal and opposite distortion associated with the lattice.

11.4 Correlation Effects–Divergence of Perturbation Theory

Correlation effects are those electron–electron interaction effects which come beyond the exchange term. Picturesquely we can represent the exchange term as shown in Fig. 11.4. The diagrams corresponding to the next order in perturbation theory are the second-order terms shown in Fig. 11.5 for (a) *direct* and (b) *exchange* interactions, respectively The second-order perturbation to

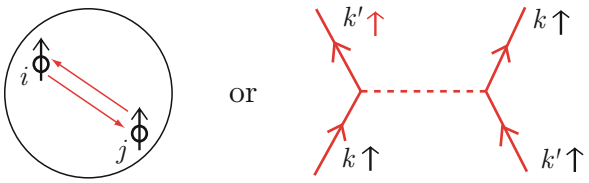


Fig. 11.4. Diagrammatic representation of the exchange interaction in the lowest order

[2] A.W. Overhauser, Phys. Rev. **167**, 691 (1968).

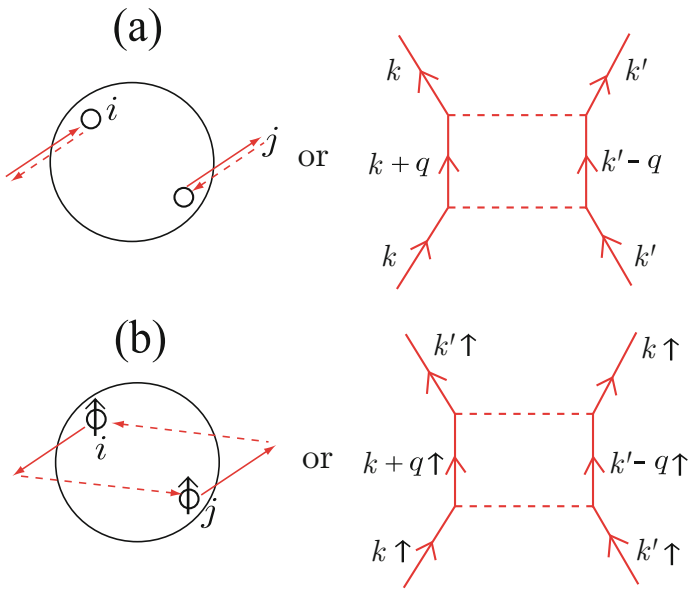


Fig. 11.5. Diagrammatic representation of the (**a**) direct and (**b**) exchange interactions in the second-order perturbation calculation

the energy is

$$E_2 = \sum_m \frac{\langle \Phi_0 | H' | \Phi_m \rangle \langle \Phi_m | H' | \Phi_0 \rangle}{E_0 - E_m} \tag{11.83}$$

Look at one term H_{ij} of $H' = \frac{1}{2}\sum_{i \neq j} V_{ij}$.

$$E_{2\mathrm{D}}(\mathbf{k}_i, \mathbf{k}_j) = \sum_{q \neq 0} \frac{\left| \left\langle \mathrm{e}^{\mathrm{i}\mathbf{k}_i \cdot \mathbf{x}} \mathrm{e}^{\mathrm{i}\mathbf{k}_j \cdot \mathbf{x}'} \left| \sum_{q_1} \frac{4\pi e^2}{q_1^2} \mathrm{e}^{\mathrm{i}\mathbf{q}_1 \cdot (\mathbf{x} - \mathbf{x}')} \right| \mathrm{e}^{\mathrm{i}(\mathbf{k}_i + \mathbf{q}) \cdot \mathbf{x}} \mathrm{e}^{\mathrm{i}(\mathbf{k}_j - \mathbf{q}) \cdot \mathbf{x}'} \right\rangle \right|^2}{E_{\mathbf{k}_i} + E_{\mathbf{k}_j} - \left(E_{\mathbf{k}_i + \mathbf{q}} + E_{\mathbf{k}_j - \mathbf{q}} \right)}, \tag{11.84}$$

where the subscript D denotes the contribution of the direct interaction. Equation (11.84) can be reduced to

$$E_{2\mathrm{D}}(k_i, k_j) = -m(4\pi e^2)^2 \sum_{q \neq 0} \frac{1}{q^4} \frac{1}{q^2 + \mathbf{q} \cdot (\mathbf{k}_i - \mathbf{k}_j)}, \tag{11.85}$$

where we have set $\hbar = 1$. Thus, we have

$$E_{2\mathrm{D}}^{\mathrm{Total}} = \frac{1}{2} \sum_{\substack{\mathbf{k}_i \neq \mathbf{k}_j; k_i, k_j < k_\mathrm{F} \\ |\mathbf{k}_i + \mathbf{q}|, |\mathbf{k}_j - \mathbf{q}| > k_\mathrm{F}}} E_{2\mathrm{D}}(k_i, k_j). \tag{11.86}$$

Summing over spins and converting sums to integrals we have[3]

[3] M. Gell-Mann, K.A. Brueckner, Phys. Rev. **106**, 364 (1957).

$$E_{\mathrm{2D}}^{\mathrm{Total}} = -\frac{e^4 m}{16\pi^7}\int \frac{\mathrm{d}^3 q}{q^4}\int_{\substack{k_i < k_{\mathrm{F}} \\ |\mathbf{k}_i+\mathbf{q}| > k_{\mathrm{F}}}} \mathrm{d}^3 k_i \int_{\substack{k_j < k_{\mathrm{F}} \\ |\mathbf{k}_j+\mathbf{q}| > k_{\mathrm{F}}}} \mathrm{d}^3 k_j \frac{1}{q^2 + \mathbf{q}\cdot(\mathbf{k}_i+\mathbf{k}_j)}. \tag{11.87}$$

It is not difficult to see that $E_{\mathrm{2D}}^{\mathrm{Total}}$ diverges because of the presence of the factor q^{-4}. In a similar way, one can show that

$$E_{\mathrm{2X}}^{\mathrm{Total}} = \frac{e^4 m}{32\pi^7}\int \frac{\mathrm{d}^3 q}{q^2}\int_{\substack{k_i < k_{\mathrm{F}} \\ |\mathbf{k}_i+\mathbf{q}| > k_{\mathrm{F}}}} \mathrm{d}^3 k_i \int_{\substack{k_j < k_{\mathrm{F}} \\ |\mathbf{k}_j+\mathbf{q}| > k_{\mathrm{F}}}} \mathrm{d}^3 k_j \frac{1}{q^2+\mathbf{q}\cdot(\mathbf{k}_i+\mathbf{k}_j)} \times \frac{1}{(\mathbf{q}+\mathbf{k}_i+\mathbf{k}_j)^2}. \tag{11.88}$$

This number is finite and has been evaluated numerically (a very complicated numerical job) with the result

$$E_{\mathrm{2X}}^{\mathrm{Total}} = N\cdot\frac{e^2}{2a_0}\times 0.046 \pm 0.002. \tag{11.89}$$

All terms beyond second-order diverge because of the factor $(\frac{1}{q^2})^m$ coming from the matrix elements of the Coulomb interaction.

The divergence results from the long range of the Coulomb interaction. Gell-Mann and Brueckner overcame the divergence difficulty by formally summing the divergent perturbation expansion to infinite order.

What they were actually accomplishing was, essentially, taking account of screening. For the present, we will concentrate on understanding something about screening in an electron gas. Later, we will discuss the diagrammatic type expansions and the ground state energy.

11.5 Linear Response Theory

We will investigate the self-consistent (Hartree) field set up by some external disturbance in an electron gas. In order to accomplish this it is very useful to introduce the single particle density matrix.

11.5.1 Density Matrix

Suppose that we have a statistical ensemble of N systems labelled $k = 1, 2, 3, \ldots, N$. Let the normalized wave function for the kth system in the ensemble be given by Ψ_k. Expand Ψ_k in terms of a complete orthonormal set of basis functions ϕ_n

$$\Psi_k = \sum_n c_{nk}\phi_n;\ \sum_n \mid c_{nk} \mid^2 = 1. \tag{11.90}$$

The expectation value of some quantum mechanical operator $\hat{A}$ in the k^{th} system of the ensemble is

$$A_k = \int \mathrm{d}\tau \, \Psi_k^* \hat{A} \Psi_k. \tag{11.91}$$

The statistical average (over all systems in the ensemble) is given by

$$\begin{aligned} \langle A \rangle &= \tfrac{1}{N} \textstyle\sum_{k=1}^{N} A_k, \\ &= \tfrac{1}{N} \textstyle\sum_{k=1}^{N} \int \mathrm{d}^3\tau \, \Psi_k^* \hat{A} \Psi_k. \end{aligned} \tag{11.92}$$

Substitute for Ψ_k in terms of the basis functions ϕ_n. This gives

$$\langle A \rangle = \frac{1}{N} \sum_{k=1}^{N} \sum_{m,n} c_{mk}^* c_{nk} \langle \phi_m | \hat{A} | \phi_n \rangle. \tag{11.93}$$

But $\langle \phi_m | \hat{A} | \phi_n \rangle = A_{mn}$, the (m, n) matrix element of $\hat{A}$ in the representation $\{\phi_n\}$. Now define a density matrix $\hat{\rho}$ as follows

$$\rho_{nm} = \frac{1}{N} \sum_{k=1}^{N} c_{mk}^* c_{nk}. \tag{11.94}$$

Then, $\langle A \rangle$ can be written

$$\langle A \rangle = \sum_{m,n} \rho_{nm} A_{mn} = \sum_{n} \left(\hat{\rho} \hat{A} \right)_{nn} = \mathrm{Tr} \left(\hat{\rho} \hat{A} \right). \tag{11.95}$$

This states that the ensemble average of a quantum mechanical operator $\hat{A}$ is simply the trace of the product of the density operator (or *density matrix*) and the operator $\hat{A}$.

11.5.2 Properties of Density Matrix

From the definition (11.94), it is clear that $\rho_{nm}^* = \rho_{mn}$. Because the unit operator **1** must have an ensemble average of unity, we have that

$$\mathrm{Tr}\hat{\rho} = 1. \tag{11.96}$$

Because $\mathrm{Tr}\hat{\rho} = \sum \rho_{\mathrm{nn}} = 1$, it is clear that $0 \leq \rho_{nn} \leq 1$ for every n. ρ_{nn} is simply the probability that the state ϕ_n is realized in the ensemble.

11.5.3 Change of Representation

Let S be a unitary matrix that transforms the orthonormal basis set $\{\phi_n\}$ into a new orthonormal basis set $\{\tilde{\phi}_n\}$. If we write

$$\tilde{\phi}_l = \sum_n \phi_n S_{nl}, \tag{11.97}$$

then we have

$$\phi_m = \sum_l S^*_{ml} \tilde{\phi}_l. \tag{11.98}$$

It can be proved by remembering that, because S is unitary, $S^{-1} = S^\dagger = \tilde{S}^*$ or $S^*_{ml} = (S^{-1})_{lm}$. Now, write the wave function for the k^{th} system in the ensemble, in terms of new basis functions, as

$$\Psi_k = \sum_l \tilde{c}_{lk} \tilde{\phi}_l. \tag{11.99}$$

Remember that

$$\Psi_k = \sum_n c_{nk} \phi_n. \tag{11.100}$$

By substituting (11.98) in (11.100), we find

$$\Psi_k = \sum_n c_{nk} \sum_l S^*_{nl} \tilde{\phi}_l = \sum_l \left(\sum_n c_{nk} S^*_{nl} \right) \tilde{\phi}_l. \tag{11.101}$$

By comparing (11.101) with (11.99), we find

$$\tilde{c}_{lk} = \sum_n c_{nk} S^*_{nl}. \tag{11.102}$$

The expression for the density matrix in the new representation is

$$\tilde{\rho}_{lp} = \frac{1}{N} \sum_{k=1}^{N} \tilde{c}^*_{pk} \tilde{c}_{lk}. \tag{11.103}$$

Now, use (11.102) and its complex conjugate in (11.103) to obtain

$$\begin{aligned} \tilde{\rho}_{lp} &= \tfrac{1}{N} \textstyle\sum_{k=1}^{N} \sum_m c^*_{mk} S_{mp} \sum_n c_{nk} S^*_{nl} \\ &= \textstyle\sum_{mn} S_{mp} \rho_{nm} S^*_{nl}, \end{aligned} \tag{11.104}$$

since $\rho_{nm} = \frac{1}{N} \sum_{k=1}^{N} c^*_{mk} c_{nk}$. But (11.104) can be rewritten

$$\tilde{\rho}_{lp} = \sum_{mn} \left(S^{-1} \right)_{ln} \rho_{nm} S_{mp}$$

or

$$\tilde{\rho} = S^{-1} \hat{\rho} S = S^\dagger \hat{\rho} S. \tag{11.105}$$

The average (over the ensemble) of an operator $\hat{A}$ is given in the new representation by

$$\begin{aligned}\langle \tilde{A} \rangle &= \mathrm{Tr}\left(\tilde{\rho}\tilde{A}\right) = \mathrm{Tr}\left(S^{-1}\rho S S^{-1} A S\right) \\ &= \mathrm{Tr}\left(S^{-1}\rho A S\right).\end{aligned}$$

But the trace of a product of two matrices is independent of the order, i.e. $\mathrm{Tr}AB = \mathrm{Tr}BA$. Therefore, we have

$$\langle \tilde{A} \rangle = \mathrm{Tr}\rho\hat{A} = \langle A \rangle. \tag{11.106}$$

This means that the ensemble average of a quantum mechanical operator $\hat{A}$ is independent of the representation chosen for the density matrix.

11.5.4 Equation of Motion of Density Matrix

The Schrödinger equation for the k^{th} system in the ensemble can be written

$$\mathrm{i}\hbar\dot{\Psi}_k = \hat{H}\Psi_k. \tag{11.107}$$

Expressing Ψ_k in terms of the basis functions ϕ_m gives

$$\mathrm{i}\hbar\sum_m \dot{c}_{mk}\phi_m = \hat{H}\sum_m c_{mk}\phi_m. \tag{11.108}$$

Taking the scalar product with ϕ_n gives

$$\mathrm{i}\hbar\dot{c}_{nk} = \sum_l \langle n|\hat{H}|l\rangle c_{lk} = \sum_l H_{nl}c_{lk}. \tag{11.109}$$

The complex conjugate of (11.107) can be written

$$-\mathrm{i}\hbar\dot{\Psi}_k^* = \hat{H}^\dagger\Psi_k^*. \tag{11.110}$$

Expressing Ψ_k^* in terms of the basis functions ϕ_l gives

$$-\mathrm{i}\hbar\sum_l \dot{c}_{lk}^*\phi_l^* = \sum_l \hat{H}^\dagger c_{lk}^*\phi_l^*. \tag{11.111}$$

Now, multiply by ϕ_n and integrate using

$$\int \mathrm{d}^3\tau\phi_l^*\phi_n = \delta_{ln}$$

and

$$\int \mathrm{d}^3\tau\phi_l^*\hat{H}^\dagger\phi_n = H^\dagger{}_{ln} = H_{ln},$$

since the Hamiltonian is a Hermitian operator. This gives

$$\mathrm{i}\hbar\dot{c}_{nk}^* = -\sum_l c_{lk}^* H_{ln}. \tag{11.112}$$

Now, look at the time rate of change of ρ_{mn}.

$$\mathrm{i}\hbar\dot{\rho}_{mn} = \frac{1}{N}\sum_{k=1}^{N}\mathrm{i}\hbar\left[\dot{c}^*_{nk}c_{mk} + c^*_{nk}\dot{c}_{mk}\right]. \tag{11.113}$$

Now, use (11.109) and (11.112) for $\dot{c}_{nk}$ and $\dot{c}^*_{nk}$ to have

$$\begin{aligned}\mathrm{i}\hbar\dot{\rho}_{mn} &= \tfrac{1}{N}\textstyle\sum_{k=1}^{N}\left[-\sum_l H_{ln}c^*_{lk}c_{mk} + \sum_l c^*_{nk}H_{ml}c_{lk}\right],\\ &= \textstyle\sum_l\left[-H_{ln}\rho_{ml} + \rho_{ln}H_{ml}\right].\end{aligned} \tag{11.114}$$

We can reorder the terms as follows:

$$\begin{aligned}\mathrm{i}\hbar\dot{\rho}_{mn} &= \textstyle\sum_l\left[H_{ml}\rho_{ln} - \rho_{ml}H_{ln}\right],\\ &= \left(H\rho - \rho H\right)_{mn}.\end{aligned} \tag{11.115}$$

This is the equation of motion of the density matrix

$$\mathrm{i}\hbar\dot{\rho} = [H,\rho]_-. \tag{11.116}$$

11.5.5 Single Particle Density Matrix of a Fermi Gas

Suppose that the single particle Hamiltonian H_0 has eigenvalues ε_n and eigenfunctions $|n\rangle$.

$$H_0|n\rangle = \varepsilon_n|n\rangle. \tag{11.117}$$

Corresponding to H_0, there is a single particle density matrix ρ_0 which is time independent and represents the equilibrium distribution of particles among the single particle states at temperature T. Because $\dot{\rho}_0 = 0$, H_0 and ρ_0 must commute. Thus, ρ_0 can be diagonalized by the eigenfunctions of H_0. We can write

$$\rho_0|n\rangle = f_0(\varepsilon_n)|n\rangle. \tag{11.118}$$

For $f_0(\varepsilon_n) = \left[\exp(\frac{\varepsilon_n-\zeta}{\Theta}) + 1\right]^{-1}$, ρ_0 is the single particle density matrix for the grand ensemble with $\Theta = k_{\mathrm{B}}T$ and the chemical potential ζ.

11.5.6 Linear Response Theory

We consider a degenerate electron gas and ask what happens when some external disturbance is introduced. For example, we might think of adding an external charge density (like a proton) to the electron gas. The electrons will respond to the disturbance, and ultimately set up a self-consistent field. We want to know what the self-consistent field is, how the external charge density is screened etc.[4]

[4] See, for eample, M.P. Greene, H.J. Lee, J.J. Quinn, S. Rodriguez, Phys. Rev. **177**, 1019 (1969).

Let the Hamiltonian in the absence of the self-consistent field be simply given by $\mathcal{H}_0 = \frac{p^2}{2m}$

$$\mathcal{H}_0|k\rangle = \varepsilon_k|k\rangle. \tag{11.119}$$

H_0 is time independent, thus the equilibrium density matrix ρ_0 must be independent of time. This means

$$[\mathcal{H}_0, \rho_0] = 0, \tag{11.120}$$

and therefore

$$\rho_0|k\rangle = f_0(\varepsilon_k)|k\rangle, \tag{11.121}$$

where

$$f_0(\varepsilon_k) = \frac{1}{\mathrm{e}^{\frac{\varepsilon_k - \zeta}{\Theta}} + 1} \tag{11.122}$$

is the Fermi–Dirac distribution function. Now, let us introduce some external disturbance. It will set up a self-consistent electromagnetic fields $\big[\mathbf{E}(\mathbf{r}, t), \mathbf{B}(\mathbf{r}, t)\big]$. We can describe these fields in terms of a scalar potential ϕ and a vector potential $\mathbf{A}$

$$\mathbf{B} = \nabla \times \mathbf{A}, \tag{11.123}$$

and

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \nabla\phi. \tag{11.124}$$

The Hamiltonian in the presence of the self-consistent field is written as

$$\mathcal{H} = \frac{1}{2m}\left(\mathbf{p} + \frac{e}{c}\mathbf{A}\right)^2 - e\phi. \tag{11.125}$$

This can be written as $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$, where up to terms linear in the self-consistent field

$$\mathcal{H}_1 = \frac{e}{2c}\left(\mathbf{v}_0 \cdot \mathbf{A} + \mathbf{A} \cdot \mathbf{v}_0\right) - e\phi. \tag{11.126}$$

Here, $\mathbf{v}_0 = \frac{\mathbf{p}}{m}$. Now, let $\rho = \rho_0 + \rho_1$, where ρ_1 is a small deviation from ρ_0 caused by the self-consistent field. The equation of motion of ρ is

$$\frac{\partial \rho}{\partial t} + \frac{i}{\hbar}\left[H, \rho\right]_- = 0. \tag{11.127}$$

Linearizing with respect to the self-consistent field gives

$$\frac{\partial \rho_1}{\partial t} + \frac{\mathrm{i}}{\hbar}\left[H_0, \rho_1\right]_- + \frac{\mathrm{i}}{\hbar}\left[H_1, \rho_0\right]_- = 0. \tag{11.128}$$

We shall investigate the situation in which $\mathbf{A}$, ϕ, ρ_1 are of the form $\mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}\mathbf{q}\cdot\mathbf{r}}$. Taking matrix elements gives

$$\langle k|\rho_1|k'\rangle = \frac{f_0(\varepsilon_{k'}) - f_0(\varepsilon_k)}{\varepsilon_{k'} - \varepsilon_k - \hbar\omega}\langle k|\mathcal{H}_1|k'\rangle. \tag{11.129}$$

Let us consider a most general component of $\mathbf{A}(\mathbf{r},t)$ and $\phi(\mathbf{r},t)$

$$\begin{aligned}\mathbf{A}(\mathbf{r},t) &= \mathbf{A}(\mathbf{q},\omega)\mathrm{e}^{\mathrm{i}\omega t-\mathrm{i}\mathbf{q}\cdot\mathbf{r}},\\ \phi(\mathbf{r},t) &= \phi(\mathbf{q},\omega)\mathrm{e}^{\mathrm{i}\omega t-\mathrm{i}\mathbf{q}\cdot\mathbf{r}}.\end{aligned} \tag{11.130}$$

Thus, we have

$$\mathcal{H}_1 = \left[\frac{e}{c}\mathbf{A}(\mathbf{q},\omega)\cdot\frac{1}{2}\left(\mathbf{v}_0\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{r}}+\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{r}}\mathbf{v}_0\right)-e\phi(\mathbf{q},\omega)\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{r}}\right]\mathrm{e}^{\mathrm{i}\omega t}. \tag{11.131}$$

Define the operator $\mathbf{V}_q$ by

$$\mathbf{V}_q = \frac{1}{2}\mathbf{v}_0\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}+\frac{1}{2}\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}\mathbf{v}_0. \tag{11.132}$$

Then, the matrix element of $\mathcal{H}_1$ can be written

$$\langle k|\mathcal{H}_1|k'\rangle = \frac{e}{c}\mathbf{A}(\mathbf{q},\omega)\cdot\langle k|\mathbf{V}_{-q}|k'\rangle - e\phi(\mathbf{q},\omega)\langle k|\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{r}}|k'\rangle. \tag{11.133}$$

We want to know the charge and current densities at a position $\mathbf{r}_0$ at time t. We can write

$$\begin{aligned}\mathbf{j}(\mathbf{r}_0,t) &= \mathrm{Tr}\left[-e\left(\tfrac{1}{2}\mathbf{v}\delta(\mathbf{r}-\mathbf{r}_0)+\tfrac{1}{2}\delta(\mathbf{r}-\mathbf{r}_0)\mathbf{v}\right)\hat{\rho}\right],\\ n(\mathbf{r}_0,t) &= \mathrm{Tr}\left[-e\delta(\mathbf{r}-\mathbf{r}_0)\hat{\rho}\right].\end{aligned} \tag{11.134}$$

Here, $-e\left[\frac{1}{2}\mathbf{v}\delta(\mathbf{r}-\mathbf{r}_0)+\frac{1}{2}\delta(\mathbf{r}-\mathbf{r}_0)\mathbf{v}\right]$ is the operator for the current density at position $\mathbf{r}_0$, while $-e\delta(\mathbf{r}-\mathbf{r}_0)$ is the charge density operator. The velocity operator $\mathbf{v}=\frac{1}{m}(\mathbf{p}+\frac{e}{c}\mathbf{A})=\mathbf{v}_0+\frac{e}{mc}\mathbf{A}$ is the velocity operator in the presence of the self-consistent field. Because $\mathbf{v}_0=\frac{\mathbf{p}}{m}$ contains the differential operator $-\frac{\mathrm{i}\hbar}{m}\nabla$, it is important to express operator like $\mathbf{v}_0\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}$ and $\mathbf{v}_0\delta(\mathbf{r}-\mathbf{r}_0)$ in the symmetric form to make them Hermitian operators.

It is easy to see that

$$\begin{aligned}\mathbf{j}(\mathbf{r}_0,t) = &-\frac{e^2}{mc}\textstyle\sum_k\langle k|\mathbf{A}(\mathbf{r},t)\delta(\mathbf{r}-\mathbf{r}_0)\hat{\rho}_0|k\rangle\\ &-e\textstyle\sum_k\langle k|\left[\frac{1}{2}\mathbf{v}_0\delta(\mathbf{r}-\mathbf{r}_0)+\frac{1}{2}\delta(\mathbf{r}-\mathbf{r}_0)\mathbf{v}_0\right]\hat{\rho}_1|k\rangle.\end{aligned} \tag{11.135}$$

$\delta(\mathbf{r}-\mathbf{r}_0)$ can be written as

$$\delta(\mathbf{r}-\mathbf{r}_0) = \Omega^{-1}\sum_{\mathbf{q}}\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}_0)}. \tag{11.136}$$

It is clear that $\langle k|\mathbf{A}(\mathbf{r},t)\delta(\mathbf{r}-\mathbf{r}_0)\rho_0|k\rangle = \Omega^{-1}\mathbf{A}(\mathbf{r}_0,t)f_0(\varepsilon_k)$. Here, of course, Ω is the volume of the system. For $\mathbf{j}(\mathbf{r}_0,t)$ we obtain

$$\mathbf{j}(\mathbf{r}_0,t) = -\frac{e^2n_0}{mc}\mathbf{A}(\mathbf{r}_0,t)-\frac{e}{\Omega}\sum_{\mathbf{k},\mathbf{k}',\mathbf{q}}\langle k'|\mathbf{V}_q|k\rangle\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{r}_0}\langle k|\hat{\rho}_1|k'\rangle. \tag{11.137}$$

But we know the matrix element $\langle k|\rho_1|k'\rangle$ from (11.129). Taking the Fourier transform of $\mathbf{j}(\mathbf{r}_0,t)$ gives

$$\begin{aligned}\mathbf{j}(\mathbf{q},\omega)= &-\frac{e^2n_0}{mc}\mathbf{A}(\mathbf{q},\omega)-\frac{e^2}{\Omega c}\sum_{k,k'}\frac{f_0(\varepsilon_{k'})-f_0(\varepsilon_k)}{\varepsilon_{k'}-\varepsilon_k-\hbar\omega}\langle k'|\mathbf{V}_q|k\rangle\langle k|\mathbf{V}_q|k'\rangle\cdot\mathbf{A}(\mathbf{q},\omega)\\ &+\frac{e^2}{\Omega}\sum_{\mathbf{k},\mathbf{k}'}\frac{f_0(\varepsilon_{k'})-f_0(\varepsilon_k)}{\varepsilon_{k'}-\varepsilon_k-\hbar\omega}\langle k'|\mathbf{V}_q|k\rangle\langle k'|\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}|k\rangle.\end{aligned} \tag{11.138}$$

This equation can be written as

$$\mathbf{j}(\mathbf{q},\omega)=-\frac{\omega_p^2}{4\pi c}\left[(\mathbf{1}+\underline{\mathbf{I}})\cdot\mathbf{A}(\mathbf{q},\omega)+\mathbf{K}\phi(\mathbf{q},\omega)\right]. \tag{11.139}$$

Here, $\omega_p^2=\frac{4\pi n_0e^2}{m}$ is the plasma frequency of the electron gas whose density is $n_0=\frac{N}{\Omega}$, and $\mathbf{1}$ is the unit tensor. The tensor $\underline{\mathbf{I}}(\mathbf{q},\omega)$ and the vector $\mathbf{K}(\mathbf{q},\omega)$ are given by

$$\begin{aligned}\underline{\mathbf{I}}(\mathbf{q},\omega) &= \frac{m}{N}\sum_{\mathbf{k},\mathbf{k}'}\frac{f_0(\varepsilon_{k'})-f_0(\varepsilon_k)}{\varepsilon_{k'}-\varepsilon_k-\hbar\omega}\langle k'|\mathbf{V}_q|k\rangle\langle k'|\mathbf{V}_q|k\rangle^*,\\ \mathbf{K}(\mathbf{q},\omega) &= \frac{mc}{N}\sum_{\mathbf{k},\mathbf{k}'}\frac{f_0(\varepsilon_{k'})-f_0(\varepsilon_k)}{\varepsilon_{k'}-\varepsilon_k-\hbar\omega}\langle k'|\mathbf{V}_q|k\rangle\langle k'|\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}|k\rangle.\end{aligned} \tag{11.140}$$

For the plane wave wave functions $|k\rangle=\Omega^{-1/2}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}$ the matrix elements are easily evaluated

$$\begin{aligned}\langle k'|\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}|k\rangle &= \delta_{k',k+q},\\ \langle k'|\mathbf{V}_q|k\rangle &= \frac{\hbar}{m}\left(\mathbf{k}+\tfrac{\mathbf{q}}{2}\right)\delta_{k',k+q}.\end{aligned} \tag{11.141}$$

11.5.7 Gauge Invariance

The transformations

$$\begin{aligned}\mathbf{A}' &= \mathbf{A}+\nabla\chi=\mathbf{A}-\mathrm{i}\mathbf{q}\chi\\ \phi' &= \phi-\tfrac{1}{c}\dot{\chi}=\phi-\mathrm{i}\tfrac{\omega}{c}\chi\end{aligned} \tag{11.142}$$

leave the fields $\mathbf{E}$ and $\mathbf{B}$ unchanged. Therefore, such a change of gauge must leave $\mathbf{j}$ unchanged. Substitution into the expression for $\mathbf{j}$ gives the condition

$$(\mathbf{1}+\underline{\mathbf{I}})\cdot(-\mathrm{i}\mathbf{q})+\mathbf{K}(-\mathrm{i}\frac{\omega}{c})=0,\ \text{or}\ \mathbf{q}+\underline{\mathbf{I}}\cdot\mathbf{q}+\frac{\omega}{c}\mathbf{K}=0. \tag{11.143}$$

Clearly no generality is lost by choosing the z-axis parallel to $\mathbf{q}$. Then, because the summand is an odd function of k_x or k_y we have

$$I_{xz}=I_{zx}=I_{yz}=I_{zy}=I_{xy}=I_{yx}=K_x=K_y=0. \tag{11.144}$$

Thus, two of the three components of the relation

$$\mathbf{q} + \underline{\mathbf{I}} \cdot \mathbf{q} + \frac{\omega}{c}\mathbf{K} = 0$$

hold automatically. It remains to be proven that

$$q + I_{zz}q + \frac{\omega}{c}K_z = 0. \tag{11.145}$$

We demonstrate this by writing I_{zz} and K_z in the following form

$$I_{zz}=\frac{\hbar^2}{mN}\left\{\sum_k -\frac{f_0(\varepsilon_k)}{\varepsilon_{k+q}-\varepsilon_k-\hbar\omega}\left(k_z+\frac{q}{2}\right)^2+\sum_k \frac{f_0(\varepsilon_{k+q})}{\varepsilon_{k+q}-\varepsilon_k-\hbar\omega}\left(k_z+\frac{q}{2}\right)^2\right\} \tag{11.146}$$

In the second term, let $k+q=\tilde{k}$ so that $k=\tilde{k}-q$; then let the dummy variable $\tilde{k}$ equal $-k$ to have

$$\frac{f_0(\varepsilon_{k+q})}{\varepsilon_{k+q}-\varepsilon_k-\hbar\omega}\left(k_z+\frac{q}{2}\right)^2 \to \frac{f_0(\varepsilon_k)}{\varepsilon_k-\varepsilon_{k+q}-\hbar\omega}\left(-k_z-\frac{q}{2}\right)^2.$$

With this replacement qI_{zz} can be written

$$qI_{zz} = -\frac{\hbar^2}{mN}\sum_k f_0(\varepsilon_k)\left(k_z+\frac{q}{2}\right)^2\left[\frac{q}{\varepsilon_{k+q}-\varepsilon_k-\hbar\omega}+\frac{q}{\varepsilon_{k+q}-\varepsilon_k+\hbar\omega}\right]. \tag{11.147}$$

Do exactly the same for K_z to get

$$\frac{\omega}{c}K_z = \frac{1}{N}\sum_k f_0(\varepsilon_k)\left(k_z+\frac{q}{2}\right)\left[\frac{\hbar\omega}{\varepsilon_{k+q}-\varepsilon_k-\hbar\omega}-\frac{\hbar\omega}{\varepsilon_{k+q}-\varepsilon_k+\hbar\omega}\right]. \tag{11.148}$$

Adding qI_{zz} to $\frac{\omega}{c}K_z$ gives

$$\begin{aligned} qI_{zz}+\frac{\omega}{c}K_z = &-\frac{1}{N}\sum_k f_0(\varepsilon_k)\left(k_z+\frac{q}{2}\right) \\ &\left[\frac{\frac{\hbar^2}{m}q(k_z+q/2)-\hbar\omega}{\varepsilon_{k+q}-\varepsilon_k-\hbar\omega}+\frac{\frac{\hbar^2}{m}q(k_z+q/2)+\hbar\omega}{\varepsilon_{k+q}-\varepsilon_k+\hbar\omega}\right]. \end{aligned} \tag{11.149}$$

But $\varepsilon_{k+q}-\varepsilon_k=\frac{\hbar^2}{m}q(k_z+q/2)$, therefore the term in square brackets is equal to 2, and hence we have

$$qI_{zz}+\frac{\omega}{c}K_z = -\frac{1}{N}\sum_k f_0(\varepsilon_k)\left(k_z+\frac{q}{2}\right)\times 2. \tag{11.150}$$

The first term $\sum_k f_0(\varepsilon_k)k_z = 0$ since it is an odd function of k_z. The second term is $-\frac{q}{N}\sum_k f_0(\varepsilon_k) = -q$. This gives $qI_{zz}+\frac{\omega}{c}K_z = -q$, meaning that

(11.145) is satisfied and our result is gauge invariant. Because we have established gauge invariance, we may now choose any gauge. Let us take $\phi = 0$; then we have

$$\mathbf{E}(\mathbf{q},\omega) = -\frac{\mathrm{i}\omega}{c}\mathbf{A}(\mathbf{q},\omega) \tag{11.151}$$

for the fields having time dependence of $\mathrm{e}^{\mathrm{i}\omega t}$. Substitute this for $\mathbf{A}$ and obtain

$$\mathbf{j}(\mathbf{q},\omega) = -\frac{n_0 e^2}{mc}\frac{\mathrm{i}}{\omega}\left[\mathbf{1} + \underline{\mathbf{I}}(\mathbf{q},\omega)\right]\cdot\mathbf{E}(\mathbf{q},\omega). \tag{11.152}$$

We can write this equation as $\mathbf{j}(\mathbf{q},\omega) = \underline{\sigma}(\mathbf{q},\omega)\cdot\mathbf{E}(\mathbf{q},\omega)$, where $\underline{\sigma}$, the conductivity tensor is given by

$$\underline{\sigma}(\mathbf{q},\omega) = \frac{\omega_p^2}{4\pi\mathrm{i}\omega}\left[\mathbf{1} + \underline{\mathbf{I}}(\mathbf{q},\omega)\right]. \tag{11.153}$$

Recall that

$$\underline{\mathbf{I}}(\mathbf{q},\omega) = \frac{m}{N}\sum_{\mathbf{k},\mathbf{k}'}\frac{f_0(\varepsilon_{k'}) - f_0(\varepsilon_k)}{\varepsilon_{k'} - \varepsilon_k - \hbar\omega}\langle k'|\mathbf{V}_q|k\rangle\langle k'|\mathbf{V}_q|k\rangle^*. \tag{11.154}$$

The gauge invariant result[5]

$$\mathbf{j}(\mathbf{q},\omega) = \underline{\sigma}(\mathbf{q},\omega)\cdot\mathbf{E}(\mathbf{q},\omega) \tag{11.155}$$

corresponds to a nonlocal relationship between current density and electric field

$$\mathbf{j}(\mathbf{r},t) = \int \mathrm{d}^3 r' \underline{\sigma}(\mathbf{r}-\mathbf{r}',t)\cdot\mathbf{E}(\mathbf{r}',t). \tag{11.156}$$

This can be seen by simply writing

$$\begin{aligned}\mathbf{j}(\mathbf{q}) &= \textstyle\int \mathrm{d}^3 r \mathbf{j}(\mathbf{r})\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}},\\ \underline{\sigma}(\mathbf{q}) &= \textstyle\int \mathrm{d}^3(\mathbf{r}-\mathbf{r}')\underline{\sigma}(\mathbf{r}-\mathbf{r}')\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')},\\ \mathbf{E}(\mathbf{q}) &= \textstyle\int \mathrm{d}^3 r' \mathbf{E}(\mathbf{r}')\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}'},\end{aligned} \tag{11.157}$$

and substituting into (11.155). Ohm's law $\mathbf{j}(\mathbf{r}) = \underline{\sigma}(\mathbf{r})\cdot\mathbf{E}(\mathbf{r})$, which is the local relation between $\mathbf{j}(\mathbf{r})$ and $\mathbf{E}(\mathbf{r})$, occurs when $\underline{\sigma}(\mathbf{q})$ is independent of $\mathbf{q}$ or, in other words, when

$$\underline{\sigma}(\mathbf{r}-\mathbf{r}') = \underline{\sigma}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}').$$

Evaluation of $\underline{\mathbf{I}}(\mathbf{q},\omega)$

We can see by symmetry that $I_{xx} = I_{yy}$ and I_{zz} are the only non-vanishing components of $\underline{\mathbf{I}}$. The integration over $\mathbf{k}$ can be performed to obtain explicit expressions for I_{xx} and I_{zz}. We demonstrate this for I_{zz}

[5] See, for eample, M.P. Greene, H.J. Lee, J.J. Quinn, S. Rodriguez, Phys. Rev. **177**, 1019 (1969) for three-dimensional case and K.S. Yi, J.J. Quinn, Phys. Rev. B **27**, 1184 (1983) for quasi two-dimensional case.

$$I_{zz}(q,\omega) = \frac{m}{N}\sum_{\mathbf{k}} \frac{f_0(\varepsilon_{k+q}) - f_0(\varepsilon_k)}{\varepsilon_{k+q} - \varepsilon_k - \hbar\omega} \frac{\hbar^2}{m^2}\left(k_z + \frac{q}{2}\right)^2. \tag{11.158}$$

We can actually return to (11.147) and convert the sum over $\mathbf{k}$ to an integral to obtain

$$I_{zz}(q,\omega) = -\frac{\hbar^2}{mN}\left(\frac{L}{2\pi}\right)^3 2\int \mathrm{d}^3k f_0(\varepsilon_k)\left[\frac{\left(k_z+\frac{q}{2}\right)^2}{\frac{\hbar^2}{m}q\left(k_z+\frac{q}{2}\right)-\hbar\omega} + \frac{\left(k_z+\frac{q}{2}\right)^2}{\frac{\hbar^2}{m}q\left(k_z+\frac{q}{2}\right)+\hbar\omega}\right]. \tag{11.159}$$

For zero temperature, $f_0(\varepsilon_k) = 1$ if $k < k_\mathrm{F}$ and zero otherwise. This gives

$$I_{zz}(q,\omega) = -\frac{1}{4\pi^2 n_0 q}\int_{-k_\mathrm{F}}^{k_\mathrm{F}} \mathrm{d}k_z (k_\mathrm{F}^2 - k_z^2)\left(k_z+\frac{q}{2}\right)^2\left[\frac{1}{k_z+\frac{q}{2}-\frac{m\omega}{\hbar q}} + \frac{1}{k_z+\frac{q}{2}+\frac{m\omega}{\hbar q}}\right]. \tag{11.160}$$

It is convenient to introduce dimensionless variables z, x, and u defined by

$$z = \frac{q}{2k_\mathrm{F}},\ x = \frac{k_z}{k_\mathrm{F}},\ \text{and}\ u = \frac{\omega}{qv_F}. \tag{11.161}$$

Then, I_{zz} can be written

$$I_{zz}(z,u) = -\frac{3}{8z}\int_{-1}^{1} \mathrm{d}x(1-x^2)(x+z)^2\left[\frac{1}{x+z-u} + \frac{1}{x+z+u}\right]. \tag{11.162}$$

If we define $\mathcal{I}_n$ by

$$\mathcal{I}_n = \int_{-1}^{1} \mathrm{d}x\, x^n\left[\frac{1}{x+z-u} + \frac{1}{x+z+u}\right], \tag{11.163}$$

then I_{zz} can be written

$$I_{zz}(z,u) = -\frac{3}{8z}\left[-\mathcal{I}_4 - 2z\mathcal{I}_3 + (1-z^2)\mathcal{I}_2 + 2z\mathcal{I}_1 + z^2\mathcal{I}_0\right]. \tag{11.164}$$

From standard integral tables one can find

$$\int \mathrm{d}x\frac{x^n}{x+a} = \frac{1}{n}x^n - \frac{a}{n-1}x^{n-1} + \frac{a^2}{n-2}x^{n-2} - \cdots + (-a)^n \ln(x+a). \tag{11.165}$$

Using this result to evaluate $\mathcal{I}_n$ and substituting the results into (11.164) we find

$$I_{zz}(z,u) = -\left(1+\frac{3}{2}u^2\right) - \frac{3u^2}{8z}\left\{\left[1-(z-u)^2\right]\ln\left(\frac{z-u+1}{z-u-1}\right) + \left[1-(z+u)^2\right]\ln\left(\frac{z+u+1}{z+u-1}\right)\right\}. \tag{11.166}$$

In exactly the same way, one can evaluate $I_{xx}(= I_{yy})$ to obtain

$$I_{xx}(z,u) = \frac{3}{8}\left(z^2 + 3u^2 - \frac{5}{3}\right) - \frac{3}{32z}\left\{\left[1-(z-u)^2\right]\ln\left(\frac{z-u+1}{z-u-1}\right)\right. \\ \left. + \left[1-(z+u)^2\right]\ln\left(\frac{z+u+1}{z+u-1}\right)\right\}. \tag{11.167}$$

11.6 Lindhard Dielectric Function

In general the electromagnetic properties of a material can be described by two tensors $\underline{\varepsilon}(\mathbf{q},\omega)$ and $\underline{\mu}(\mathbf{q},\omega)$, where

$$\mathbf{D}(\mathbf{q},\omega) = \underline{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{E}(\mathbf{q},\omega) \text{ and } \mathbf{H}(\mathbf{q},\omega) = \underline{\mu}^{-1}(\mathbf{q},\omega)\cdot\mathbf{B}(\mathbf{q},\omega). \tag{11.168}$$

For a degenerate electron gas in the absence of a dc magnetic field $\underline{\varepsilon}(\mathbf{q},\omega)$ and $\underline{\mu}(\mathbf{q},\omega)$ will be scalars. In his now classic paper "On the properties of a gas of charged particles", Jens Lindhard[6] used, instead of $\underline{\varepsilon}(\mathbf{q},\omega)$ and $\underline{\mu}(\mathbf{q},\omega)$, the *longitudinal* and *transverse* dielectric functions defined by

$$\varepsilon^{(l)} = \varepsilon \text{ and } \varepsilon^{(\mathrm{tr})} = \varepsilon^{(l)} + \frac{c^2q^2}{\omega^2}\left(1-\mu^{-1}\right). \tag{11.169}$$

Lindhard found this notation to be convenient because he always worked in the particular gauge in which $\mathbf{q}\cdot\mathbf{A} = 0$. In this gauge the Maxwell equation for $\nabla\times\mathbf{B} = \frac{1}{c}\dot{\mathbf{E}} + \frac{4\pi}{c}(\mathbf{j}_{\mathrm{ind}} + \mathbf{j}_0)$ can be written, for the fields of the form $\mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}\mathbf{q}\cdot\mathbf{r}}$,

$$-\mathrm{i}\mathbf{q}\times(-\mathrm{i}\mathbf{q}\times\mathbf{A}) = \frac{\mathrm{i}\omega}{c}\mathbf{E} + \frac{4\pi}{c}\underline{\sigma}\cdot\mathbf{E} + \frac{4\pi}{c}\mathbf{j}_0. \tag{11.170}$$

But defining

$$\underline{\varepsilon} = 1 - \frac{4\pi\mathrm{i}}{\omega}\underline{\sigma},$$

and using $\mathbf{E} = \mathrm{i}\mathbf{q}\phi - \frac{\mathrm{i}\omega}{c}\mathbf{A}$ allows us to rewrite (11.170) as

$$q^2\left(1 - \frac{\omega^2}{c^2q^2}\varepsilon^{(\mathrm{tr})}\right)\mathbf{A} = -\frac{\omega}{c}\varepsilon^{(l)}\mathbf{q}\phi + \frac{4\pi}{c}\mathbf{j}_0. \tag{11.171}$$

Here, we have made use of the fact that $\underline{\varepsilon}\cdot\mathbf{q}$ involves only $\varepsilon^{(l)}$, while $\underline{\varepsilon}\cdot\mathbf{A}$ involves only $\varepsilon^{(\mathrm{tr})}$ since $\mathbf{q}\cdot\mathbf{A} = 0$. If we compare (11.171) with the similar equation obtained from $\nabla\times\mathbf{H} = \frac{1}{c}\dot{\mathbf{D}} + \frac{4\pi}{c}\mathbf{j}_0$ when $\mathbf{H}$ is set equal to $\mu^{-1}\mathbf{B}$ and $\mathbf{D} = \varepsilon\mathbf{E}$, viz.

[6] J. Lindhard, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **28**, 8 (1954); ibid., **27**, 15 (1953).

$$q^2\left(\mu^{-1}-\frac{\omega^2}{c^2q^2}\varepsilon\right)\mathbf{A}=-\frac{\omega}{c}\varepsilon\mathbf{q}\phi+\frac{4\pi}{c}\mathbf{j}_0, \tag{11.172}$$

we see that

$$\varepsilon=\varepsilon^{(l)} \text{ and } \mu^{-1}-\frac{\omega^2}{c^2q^2}\varepsilon^{(l)}=1-\frac{\omega^2}{c^2q^2}\varepsilon^{(\mathrm{tr})}. \tag{11.173}$$

This last equation is simply rewritten

$$\varepsilon^{(\mathrm{tr})}=\varepsilon^{(l)}+\frac{c^2q^2}{\omega^2}\left(1-\mu^{-1}\right). \tag{11.174}$$

We have chosen $\mathbf{q}$ to be in the z-direction, hence

$$\varepsilon^{(l)}=1-\frac{4\pi\mathrm{i}}{\omega}\sigma_{zz} \text{ and } \varepsilon^{(\mathrm{tr})}=1-\frac{4\pi\mathrm{i}}{\omega}\sigma_{xx}. \tag{11.175}$$

Thus, we have

$$\begin{aligned}\varepsilon^{(l)}(q,\omega) &= 1-\tfrac{\omega_p^2}{\omega^2}\left[1+I_{zz}(q,\omega)\right]\\ \varepsilon^{(\mathrm{tr})}(q,\omega) &= 1-\tfrac{\omega_p^2}{\omega^2}\left[1+I_{xx}(q,\omega)\right].\end{aligned} \tag{11.176}$$

11.6.1 Longitudinal Dielectric Constant

It is quite apparent from the expression for I_{zz} that $\varepsilon^{(\mathrm{l})}$ has an imaginary part, because for certain values of z and u, the arguments appearing in the logarithmic functions in I_{zz} are negative. Recall that

$$\ln(x+\mathrm{i}y)=\frac{1}{2}\ln(x^2+y^2)+\mathrm{i}\arctan\frac{y}{x}. \tag{11.177}$$

One can write $\varepsilon^{(l)}=\varepsilon_1^{(l)}+\mathrm{i}\varepsilon_2^{(l)}$. It is not difficult to show that

$$\varepsilon_2^{(l)}=3u^2\frac{\omega_p^2}{\omega^2}\times\begin{cases}\frac{\pi}{2}u & \text{for } z+u<1\\ \frac{\pi}{8z}\left[1-(z-u)^2\right] & \text{for } |z-u|<1<z+u\\ 0 & \text{for } |z-u|>1\end{cases} \tag{11.178}$$

The correct sign of $\varepsilon_2^{(l)}$ can be obtained by giving ω a small positive imaginary part (then $\mathrm{e}^{\mathrm{i}\omega t}\to 0$ as $t\to\infty$) which allows one to go to zero after evaluation of $\varepsilon_2^{(l)}$. The meaning of $\varepsilon_2^{(l)}$ is not difficult to understand. Suppose that an effective electric field of the form

$$\mathbf{E}=\mathbf{E}_0\mathrm{e}^{-\mathrm{i}\omega t+\mathrm{i}\mathbf{q}\cdot\mathbf{r}}+\text{c.c.} \tag{11.179}$$

perturbs the electron gas. We can write $\mathbf{E}=-\nabla\phi$ and then $\phi_0=\frac{\mathrm{i}E_0}{q}$. The perturbation acting on the electrons is $H'=-e\phi$. The power (dissipated in the system of unit volume) involving absorption or emission processes of

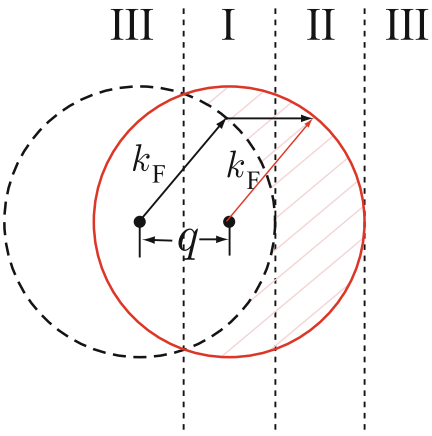


Fig. 11.6. Region of the integration indicated in (11.184)

energy $\hbar\omega$ is given by $\mathcal{P}(\mathbf{q},\omega) = \hbar\omega W(\mathbf{q},\omega)$. Here, $W(\mathbf{q},\omega)$ is the transition rate per unit volume, which is given by the standard Fermi golden rule. Then, we can write the absorption power by

$$\mathcal{P}(\mathbf{q},\omega) = \frac{2\pi}{\hbar}\frac{1}{\Omega}\sum_{\substack{k<k_F\\ k'>k_F}} \left|\langle k'| - e\phi_0 e^{i\mathbf{q}\cdot\mathbf{r}}|k\rangle\right|^2 \hbar\omega\,\delta(\varepsilon_{k'} - \varepsilon_k - \hbar\omega). \quad (11.180)$$

This results in

$$\mathcal{P}(\mathbf{q},\omega) = \frac{2\pi}{\hbar}\frac{1}{\Omega}\sum_{\substack{k<k_F\\ |\mathbf{k}+\mathbf{q}|>k_F}} e^2\left|\phi_0\right|^2 \hbar\omega\,\delta(\varepsilon_{k+q} - \varepsilon_k - \hbar\omega). \quad (11.181)$$

Now, convert the sums to integrals to obtain

$$\mathcal{P}(\mathbf{q},\omega) = \frac{2\pi}{\hbar}\frac{e^2}{\Omega}\left(\frac{E_0}{q}\right)^2 \hbar\omega\, 2\left(\frac{L}{2\pi}\right)^3 \int' \mathrm{d}^3k\,\delta(\varepsilon_{k+q} - \varepsilon_k - \hbar\omega). \quad (11.182)$$

The prime in the integral denotes the conditions $k < k_F$ and $|\mathbf{k}+\mathbf{q}| > k_F$ (see Fig. 11.6). Now, write $\int \mathrm{d}^3k = \int \mathrm{d}k_z\,\mathrm{d}^2k_\perp$. Thus

$$\mathcal{P}(\mathbf{q},\omega) = \frac{e^2\omega E_0^2}{2\pi^2 q^2}\int_{\substack{k<k_F\\ |\mathbf{k}+\mathbf{q}|>k_F}} \mathrm{d}k_z\,\mathrm{d}^2k_\perp\,\delta\left(\frac{\hbar^2 q}{m}\left(k_z + \frac{q}{2}\right) - \hbar\omega\right). \quad (11.183)$$

Integrating over k_z and using $\delta(ax) = \frac{1}{a}\delta(x)$ gives $k_z = \frac{m\omega}{\hbar q} - \frac{q}{2}$ so that

$$\mathcal{P}(\mathbf{q},\omega) = \frac{me^2\omega E_0^2}{2\pi\hbar^2 q^3}\int'_{k_z=\frac{m\omega}{\hbar q}-\frac{q}{2}} \mathrm{d}^2k_\perp. \quad (11.184)$$

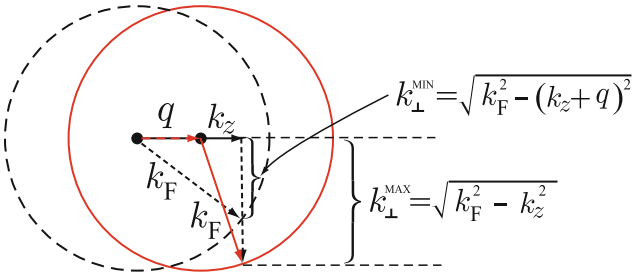


Fig. 11.7. Range of the values of $k_\perp$ appearing in (11.184)

The solid sphere in Fig. 11.6 represents $|\mathbf{k}| = k_\mathrm{F}$. The dashed sphere has $|\mathbf{k}-\mathbf{q}| = k_\mathrm{F}$. Only electrons in the hatched region can be excited to unoccupied states by adding wave vector $\mathbf{q}$ to the initial value of $\mathbf{k}$. We divide the hatched region into part I and part II. In region I, $-\frac{q}{2} < k_z < k_\mathrm{F} - q$, where $k_z = \frac{m\omega}{\hbar q} - \frac{q}{2}$. Thus

$$-\frac{q}{2} < \frac{m\omega}{\hbar q} - \frac{q}{2} < k_\mathrm{F} - q. \tag{11.185}$$

Divide (11.185) by k_F to obtain

$$-z < u - z < 1 - 2z,$$

where $z = \frac{q}{2k_\mathrm{F}}$, $x = \frac{k_z}{k_\mathrm{F}}$, and $u = \frac{\omega}{qv_F}$. Now, add $2z$ to each term to have

$$z < u + z < 1 \text{ or } u + z < 1. \tag{11.186}$$

In this region, the values of $k_\perp$ must be located between the following limits (see Fig. 11.7):

$$k_\mathrm{F}^2 - \left(\frac{m\omega}{\hbar q} + \frac{q}{2}\right)^2 < k_\perp^2 < k_\mathrm{F}^2 - \left(\frac{m\omega}{\hbar q} - \frac{q}{2}\right)^2.$$

Therefore, we have

$$\int d^2k_\perp = \left[k_\mathrm{F}^2 - \left(\frac{m\omega}{\hbar q} - \frac{q}{2}\right)^2\right] - \left[k_\mathrm{F}^2 - \left(\frac{m\omega}{\hbar q} + \frac{q}{2}\right)^2\right] = \frac{2m\omega}{\hbar}. \tag{11.187}$$

Substituting into (11.184) gives

$$\mathcal{P}(\mathbf{q},\omega) = \frac{\omega}{2\pi} \mid E_0 \mid^2 \frac{me^2}{\hbar^2 q^3} \frac{2m\omega}{\hbar}. \tag{11.188}$$

Here, we recall that energy dissipated per unit time in the system of volume Ω is also given by $\mathcal{E} = \int_\Omega \mathbf{j}\cdot\mathbf{E}\,\mathrm{d}^3r = 2\sigma_1(\mathbf{q},\omega)|E_0|^2\Omega$ and we have that $\varepsilon(q,\omega) = 1 + \frac{4\pi\mathrm{i}}{\omega}\sigma(q,\omega)$ following the form $\mathrm{e}^{-\mathrm{i}\omega t}$ for the time dependence of

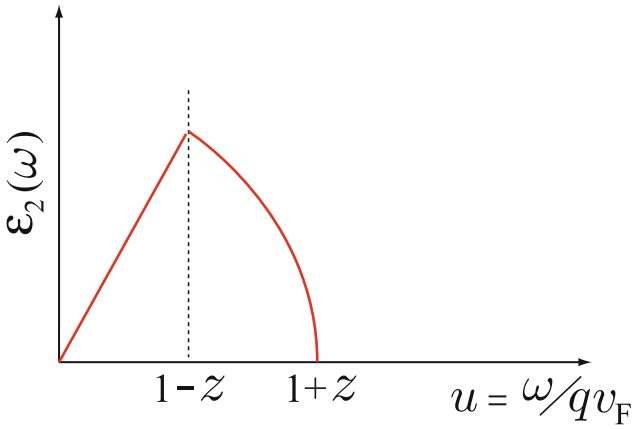


Fig. 11.8. Frequency dependence of $\varepsilon_2^{(l)}(\omega)$ the imaginary part of the dielectric function

the fields. The power dissipation per unit volume is then written

$$\mathcal{P}(\mathbf{q},\omega) = \frac{\omega}{2\pi}\varepsilon_2(q,\omega) \mid E_0 \mid^2 . \tag{11.189}$$

We note that $\varepsilon_2(\mathbf{q},\omega)$, the imaginary part of the dielectric function determines the energy dissipation in the matter due to a field $\mathbf{E}$ of wave vector $\mathbf{q}$ and frequency ω. By comparing (11.188) and (11.189), we see that, for region I,

$$\varepsilon_2^{(l)}(q,\omega) = \frac{3\omega_\mathrm{p}^2}{q^2 v_\mathrm{F}^2}\frac{\pi}{2}u \quad \text{if} \quad u+z<1. \tag{11.190}$$

In region II, $k_\mathrm{F} - q < k_z < k_\mathrm{F}$. But $k_z = \frac{m\omega}{\hbar q} - \frac{q}{2} = k_\mathrm{F}(u-z)$. Combining these and dividing by k_F we have $1-2z < u-z < 1$. Because $z<1$ in region II, the conditions can be expressed as $\mid z-u \mid < 1 < z+u$. In this case $0 < k_\perp^2 < k_\mathrm{F}^2 - k_z^2$, and, of course, $k_z = k_\mathrm{F}(u-z)$. Carrying out the algebra gives for region II

$$\varepsilon_2^{(l)}(q,\omega) = \frac{3\omega_\mathrm{p}^2}{q^2 v_\mathrm{F}^2}\frac{\pi}{8z}\left[1-(z-u)^2\right] \text{ if } \mid z-u \mid < 1 < z+u. \tag{11.191}$$

For region III, it is easy to see that $\varepsilon_2^{(l)}(\omega) = 0$. Figure 11.8 shows the frequency dependence of $\varepsilon_2^{(l)}(\omega)$. Thus, we see that the imaginary part of the dielectric function $\varepsilon_2^{(l)}(q,\omega)$ is proportional to the rate of energy dissipation due to an electric field of the form $E_0 \mathrm{e}^{-\mathrm{i}\omega t + \mathrm{i}\mathbf{q}\cdot\mathbf{r}} + \text{c.c.}$.

11.6.2 Kramers–Kronig Relation

Let $\mathbf{E}(\mathbf{x},t)$ be an electric field acting on some polarizable material. The polarization field $\mathbf{P}(\mathbf{x},t)$ will, in general, be related to $\mathbf{E}$ by an integral relationship of the form

$$\mathbf{P}(\mathbf{x},t) = \int \mathrm{d}^3x'\,\mathrm{d}t'\,\chi(\mathbf{x}-\mathbf{x}',t-t')\mathbf{E}(\mathbf{x}',t'). \tag{11.192}$$

Causality requires that $\chi(\mathbf{x}-\mathbf{x}',t-t')=0$ for all $t<t'$. That is, the polarizable material can not respond to the field until it is turned on. A well-known theorem from the theory of complex variables tells us that the Fourier transform of $\chi(t-t')$ is analytic in the upper half plane since $\mathrm{e}^{i(\omega_1+\mathrm{i}\omega_2)t}$ becomes $\mathrm{e}^{\mathrm{i}\omega_1 t}\mathrm{e}^{-\omega_2 t}$.

Theorem 1. *Given a function $f(z)$ such that $f(z)=0$ for all $z<0$, then the Fourier transform of $f(z)$ is analytic in the upper half plane.*

Take the Fourier transform of the equation for $\mathcal{P}(\mathbf{x},t)$

$$\mathbf{P}(\mathbf{q},\omega) = \int \mathrm{d}^3x\,\mathrm{d}t\,\mathbf{P}(\mathbf{x},t)\mathrm{e}^{\mathrm{i}\omega t-\mathrm{i}\mathbf{q}\cdot\mathbf{x}}. \tag{11.193}$$

Then

$$\begin{aligned}\mathbf{P}(\mathbf{q},\omega) &= \textstyle\int \mathrm{d}^3x\,\mathrm{d}t\int \mathrm{d}^3x'\,\mathrm{d}t'\,\chi(\mathbf{x}-\mathbf{x}',t-t')\mathbf{E}(\mathbf{x}',t')\mathrm{e}^{\mathrm{i}\omega t-\mathrm{i}\mathbf{q}\cdot\mathbf{x}}\\ &= \textstyle\int \mathrm{d}(\mathbf{x}-\mathbf{x}')\,\mathrm{d}(t-t')\,\chi(\mathbf{x}-\mathbf{x}',t-t')\mathrm{e}^{\mathrm{i}\omega(t-t')-\mathrm{i}\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')}\\ &\quad\times\textstyle\int \mathrm{d}^3x'\,\mathrm{d}t'\,\mathbf{E}(\mathbf{x}',t')\mathrm{e}^{\mathrm{i}\omega t'-\mathrm{i}\mathbf{q}\cdot\mathbf{x}'}.\end{aligned}$$

Therefore, we have

$$\mathbf{P}(\mathbf{q},\omega) = \chi(\mathbf{q},\omega)\mathbf{E}(\mathbf{q},\omega). \tag{11.194}$$

Here, $\chi(\mathbf{q},\omega)$ is the electrical polarizability (see (8.14)). The dielectric constant $\varepsilon(\mathbf{q},\omega)$ is related to the polarizability χ by

$$\varepsilon(\mathbf{q},\omega) = 1+4\pi\chi(\mathbf{q},\omega) \tag{11.195}$$

The theorem quoted above tells us that $\chi(\omega)$ is analytic in the upper half ω-plane. From here on we shall be interested only in the frequency dependence of $\chi(\mathbf{q},\omega)$, so for brevity we shall omit the $\mathbf{q}$ in $\chi(\mathbf{q},\omega)$. *Cauchy's theorem* states that

$$\chi(\omega) = \frac{1}{2\pi i}\int_C \frac{\chi(\omega')}{\omega'-\omega}\,\mathrm{d}\omega', \tag{11.196}$$

where the contour C must enclose the point ω and must lie completely in the region of analyticity of the complex function $\chi(\omega')$. We choose the contour lying in the upper half plane as indicated in Fig. 11.9.

As $|\omega|\to\infty$, $\chi(\omega)\to 0$ since the medium can not follow an infinitely rapidly oscillating disturbance. This allows us to discard the integral over the semicircle when its radius approaches infinity. Thus, we have

$$\chi(\omega) = \frac{1}{2\pi\mathrm{i}}\int_{-\infty}^{\infty}\frac{\chi(\omega')}{\omega'-\omega}\,\mathrm{d}\omega'. \tag{11.197}$$

We are interested in real frequencies ω, so we allow ω to approach the real axis. In doing so we must be careful to make sure that ω is enclosed by the

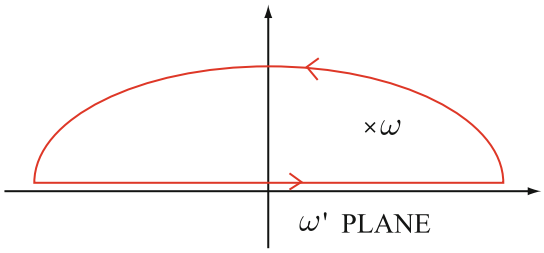


Fig. 11.9. The contour C appearing in (11.196)

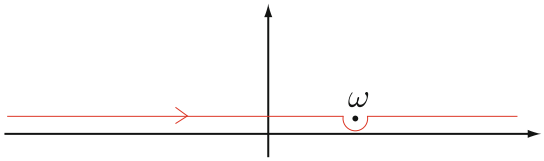


Fig. 11.10. Relevant contour C when ω approaches the real axis

original contour. One can satisfy all the conditions by deforming the contour as shown in Fig. 11.10.

Then, we have

$$\chi(\omega) = \frac{1}{2\pi \mathrm{i}} \mathcal{P} \int_{-\infty}^{\infty} \frac{\chi(\omega')}{\omega' - \omega} \, \mathrm{d}\omega' + \frac{1}{2\pi \mathrm{i}} \int_{\substack{\text{small} \\ \text{semicircle}}} \frac{\chi(\omega')}{\omega' - \omega} \, \mathrm{d}\omega', \tag{11.198}$$

where $\mathcal{P}$ denotes the principal part of the integral. We integrate the second term in (11.198) by setting $\omega' - \omega = \rho \mathrm{e}^{\mathrm{i}\phi}$ and letting $\rho \to 0$

$$\begin{aligned} \int_{\substack{\text{small} \\ \text{semicircle}}} \frac{\chi(\omega')}{\omega' - \omega} \, \mathrm{d}\omega' &= \lim_{\rho \to 0} \int_{-\pi/2}^{\pi/2} \frac{\chi(\omega + \rho \mathrm{e}^{\mathrm{i}\phi}) \rho \mathrm{e}^{\mathrm{i}\phi} \mathrm{i} \mathrm{d}\phi}{\rho \mathrm{e}^{\mathrm{i}\phi}} \\ &= \mathrm{i}\pi \chi(\omega). \end{aligned}$$

Thus, we have

$$\chi(\omega) = \frac{1}{2\pi \mathrm{i}} \mathcal{P} \int_{-\infty}^{\infty} \frac{\chi(\omega')}{\omega' - \omega} \, \mathrm{d}\omega' + \frac{1}{2\pi \mathrm{i}} \mathrm{i}\pi \chi(\omega)$$

or

$$\chi(\omega) = \frac{1}{\pi \mathrm{i}} \mathcal{P} \int_{-\infty}^{\infty} \frac{\chi(\omega')}{\omega' - \omega} \, \mathrm{d}\omega'. \tag{11.199}$$

This is the *Kramers–Kronig relation*. By writing $\chi(\omega) = \chi_1(\omega) + \mathrm{i}\chi_2(\omega)$, we can use the Kramers–Kronig relation to obtain

$$\begin{aligned} \chi_1(\omega) &= \tfrac{1}{\pi} \mathcal{P} \textstyle\int_{-\infty}^{\infty} \frac{\chi_2(\omega')}{\omega' - \omega} \, \mathrm{d}\omega' \\ \chi_2(\omega) &= -\tfrac{1}{\pi} \mathcal{P} \textstyle\int_{-\infty}^{\infty} \frac{\chi_1(\omega')}{\omega' - \omega} \, \mathrm{d}\omega', \end{aligned} \tag{11.200}$$

or in terms of $\underline{\varepsilon}$

$$\begin{aligned}\varepsilon_1(\omega) &= 1 + \tfrac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} \tfrac{\varepsilon_2(\omega')}{\omega'-\omega}\,\mathrm{d}\omega' \\ \varepsilon_2(\omega) &= -\tfrac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} \tfrac{\varepsilon_1(\omega')-1}{\omega'-\omega}\,\mathrm{d}\omega',\end{aligned} \tag{11.201}$$

where $\varepsilon_2 = 4\pi\chi_2$. Here, we note that the reality requirement on the fields $\mathbf{E}$ and $\mathbf{P}$ imposes the conditions $\chi_1(\omega) = \chi_1(-\omega)$ and $\chi_2(\omega) = -\chi_2(-\omega)$. This allows us to write

$$\varepsilon_1(\omega) = 1 + \frac{2}{\pi}\mathcal{P}\int_0^{\infty} \frac{\omega'\varepsilon_2(\omega')}{\omega'^2 - \omega^2}\,\mathrm{d}\omega'. \tag{11.202}$$

11.7 Effect of Collisions

In actual experiments, the conductivity of a metal (normal metal) is not infinite at zero frequency because the electrons collide with lattice imperfections (phonons, defects, impurities). Experimenters find it convenient to account for collisions by use of a phenomenological relaxation time τ. When collisions are included, the equation of motion of the density matrix becomes

$$\frac{\partial\rho}{\partial t} + \frac{i}{\hbar}\,[H,\rho]_- = \left(\frac{\partial\rho}{\partial t}\right)_{\mathrm{c}}. \tag{11.203}$$

The assumption of a relaxation time is equivalent to saying that

$$\left(\frac{\partial\rho}{\partial t}\right)_{\mathrm{c}} = -\frac{\rho - \tilde{\rho}_0}{\tau}. \tag{11.204}$$

Here, $\tilde{\rho}_0$ is a *local equilibrium density matrix*. We shall see that $\tilde{\rho}_0$ must be chosen with care or the treatment will be incorrect.[7] There are two requirements that $\tilde{\rho}_0$ must satisfy

1. $\tilde{\rho}_0$ must transform properly under change of gauge
2. Because collisions cannot alter the density at any point in space, $\tilde{\rho}_0$ must be chosen such that ρ and $\tilde{\rho}_0$ correspond to the same density at every point $\mathbf{r}_0$

In a Hamiltonian description of dynamics, the electromagnetic potentials appear in the Hamiltonian. But, the potentials are not unique. It turns out that the correct choice for $\tilde{\rho}_0$ which satisfies gauge invariance and conserve particle number in collisions is

$$\tilde{\rho}_0(H,\eta) = \frac{1}{\mathrm{e}^{\frac{H-\eta}{\Theta}} + 1}. \tag{11.205}$$

Here, H is the full Hamiltonian including the self-consistent potentials $(\mathbf{A}', \phi)$ and η is the *local chemical potential*. We can determine η by requiring that

[7] See, for eample, M.P. Greene, H.J. Lee, J.J. Quinn, S. Rodriguez, Phys. Rev. **177**, 1019 (1969).

$$\text{Tr } \{[\rho - \tilde{\rho}_0(H,\eta)]\,\delta(\mathbf{r}-\mathbf{r}_0)\} = 0. \tag{11.206}$$

This condition implies that the local equilibrium distribution function $\tilde{\rho}_0$ toward which the nonequilibrium distribution function ρ is relaxing has exactly the same density at every position $\mathbf{r}_0$ as the nonequilibrium distribution function does at $\mathbf{r}_0$. Of course, the local chemical potential is $\eta(\mathbf{r},t) = \zeta_0 + \zeta_1(\mathbf{r},t)$, and the value of ζ_1 is obtained by solving (11.206).

To understand this, think of the gauge in which the scalar potential $\phi(\mathbf{r})$ vanishes. In this gauge the entire Hamiltonian is kinetic. Therefore, the Hamiltonian, including the self-consistent field can be written as

$$H_{\mathrm{K}} = \frac{1}{2m}\left(\mathbf{p} + \frac{e}{c}\mathbf{A}\right)^2 = \frac{1}{2}mv^2.$$

For any gauge transformation $\mathbf{A}' = \mathbf{A} + \nabla\chi$ and $\phi' = \phi - \frac{1}{c}\dot{\chi}$, we can define $H'_{\mathrm{K}} = H' - \frac{e}{c}\dot{\chi}$. Here H' is the sum of H'_{K} and $-e\phi'$ with

$$H'_{\mathrm{K}} = \mathrm{e}^{-\frac{ie\chi}{\hbar c}} H_{\mathrm{K}} \mathrm{e}^{\frac{ie\chi}{\hbar c}}. \tag{11.207}$$

By choosing $\tilde{\rho}_0$ to depend on H_{K} we guarantee that

$$\tilde{\rho}'_0 = \mathrm{e}^{-\frac{ie\chi}{\hbar c}} \tilde{\rho}_0 \mathrm{e}^{\frac{ie\chi}{\hbar c}}$$

transforms exactly as ρ itself transforms. There are two extreme cases:

1. $H = H_{\mathrm{K}} - e\phi$, $\eta =$ constant $= \zeta_0$. All of the density variation comes from the scalar potential $\phi(\boldsymbol{r})$. (see Fig. 11.11a)
2. $H = H_{\mathrm{K}}$, $\eta = \zeta_0 + e\phi$. All of the density variation comes from the variation $\zeta_1(\boldsymbol{r})$ in the chemical potential $\eta(\boldsymbol{r})$. (see Fig. 11.11b)

Neither H nor η is gauge invariant, but their difference $H - \eta$ is. This is the quantity that appears in $\tilde{\rho}_0$.

If we let $\eta(\mathbf{r},t) = \zeta_0 + \zeta_1(\mathbf{r},t)$ where ζ_0 is the actual overall equilibrium chemical potential and $\zeta_1(\mathbf{r},t)$ is the local deviation of η from ζ_0, then we can write

$$\tilde{\rho}_0(H,\eta) = \rho_0(H_0,\zeta_0) + \rho_2. \tag{11.208}$$

The equation of motion of the density matrix is

$$\frac{\partial\rho}{\partial t} + \frac{i}{\hbar}\,[H,\rho]_- = -\frac{\rho - \tilde{\rho}_0}{\tau} \tag{11.209}$$

where $\tilde{\rho}_0 = \rho_0(H_0,\zeta_0) + \rho_2$. We can write $\rho = \tilde{\rho}_0 + \rho_1$, where ρ_1 is the deviation from the local thermal equilibrium value $\tilde{\rho}_0$. Then (11.209) becomes

$$\mathrm{i}\omega(\rho_1 + \rho_2) + \frac{\mathrm{i}}{\hbar}\,[H_0, \rho_1 + \rho_2]_- + \frac{\mathrm{i}}{\hbar}\,[H_1, \rho_0]_- = -\frac{\rho_1}{\tau}. \tag{11.210}$$

Take matrix elements as before and solve for $\langle k|\rho_1|k'\rangle$; this gives

$$\begin{aligned}\langle k|\rho_1|k'\rangle = {} & \left[\frac{\mathrm{i}\hbar/\tau}{\varepsilon_{k'} - \varepsilon_k - \hbar\omega + \mathrm{i}\hbar/\tau} - 1\right]\langle k|\rho_2|k'\rangle \\ & + \frac{f_0(\varepsilon_{k'}) - f_0(\varepsilon_k)}{\varepsilon_{k'} - \varepsilon_k - \hbar\omega + \mathrm{i}\hbar/\tau}\langle k|H_1|k'\rangle.\end{aligned} \tag{11.211}$$

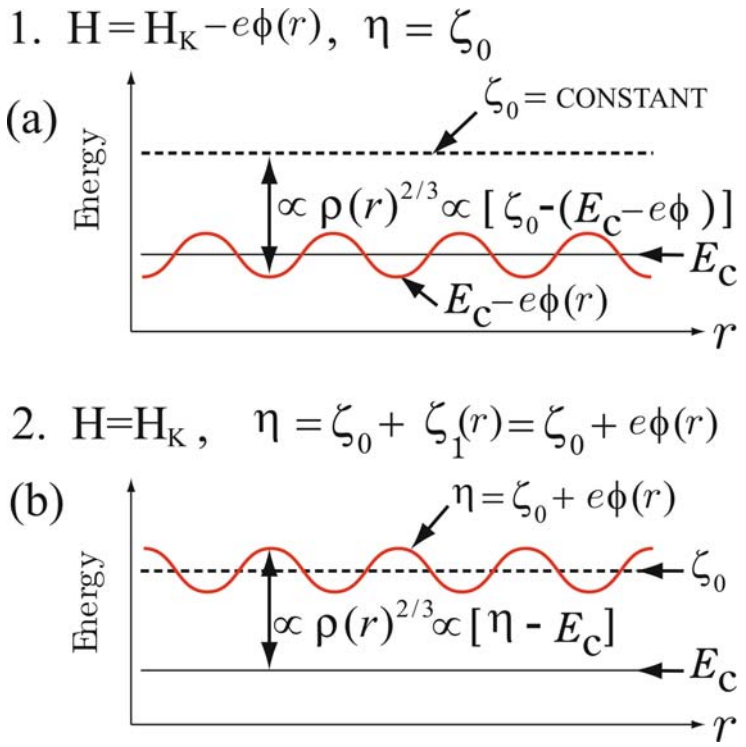


Fig. 11.11. Local chemical potential η and local density matrix $\rho(\mathbf{r})$. (**a**) η = constant$(=\zeta_0)$, (**b**) $\eta(\mathbf{r}) = \zeta_0 + e\phi(\mathbf{r})$. Here E_c denotes the bottom of the conduction band

Using the result of Problem in this chapter for $\langle k|\rho_2|k'\rangle$ in this equation gives

$$\langle k|\rho_1|k'\rangle = \left[-1 + \frac{\mathrm{i}\hbar/\tau}{\varepsilon_{k'} - \varepsilon_k - \hbar\omega + \mathrm{i}\hbar/\tau}\right] \frac{f_0(\varepsilon_{k'}) - f_0(\varepsilon_k)}{\varepsilon_{k'} - \varepsilon_k} \langle k|H_1 - \zeta_1|k'\rangle + \frac{f_0(\varepsilon_{k'}) - f_0(\varepsilon_k)}{\varepsilon_{k'} - \varepsilon_k - \hbar\omega + \mathrm{i}\hbar/\tau} \langle k|H_1|k'\rangle. \tag{11.212}$$

The parameter ζ_1 appearing in (11.212) is determined by requiring that

$$\mathrm{Tr}\,\{\rho_1 \delta(\mathbf{r} - \mathbf{r}_0)\} = 0. \tag{11.213}$$

The final result (after a lot of calculation) is

$$\mathbf{j}(\mathbf{q},\omega) = \frac{\omega_\mathrm{p}^2}{4\pi\mathrm{i}\omega} \left\{ \underline{\mathbf{1}} + \underline{\mathbf{I}} - \frac{\mathrm{i}\omega\tau}{1 + \mathrm{i}\omega\tau} \frac{(\mathbf{K}_1 - \mathbf{K}_2)(\mathbf{K}_1' - \mathbf{K}_2')}{L_1 + \mathrm{i}\omega\tau L_2} \right\} \cdot \mathbf{E}, \tag{11.214}$$

where we used the notations

$$\underline{\mathbf{I}} = \frac{\mathrm{i}\omega\tau \underline{\mathbf{I}}_1 + \underline{\mathbf{I}}_2}{1 + \mathrm{i}\omega\tau}, \tag{11.215}$$

$$\mathbf{K} = \frac{\mathrm{i}\omega\tau\mathbf{K}_1 + \mathbf{K}_2}{1 + \mathrm{i}\omega\tau}, \tag{11.216}$$

$$L_i = \frac{mc^2}{N} \sum_{kk'} \Lambda^{(i)}_{k'k} \mid \langle k'|\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}|k\rangle \mid^2, \tag{11.217}$$

$$\mathbf{K}_i = \frac{mc}{N} \sum_{kk'} \Lambda^{(i)}_{k'k} \langle k'|\mathbf{V}_\mathbf{q}|k\rangle\langle k'|\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}|k\rangle^*, \tag{11.218}$$

$$\mathbf{K}'_i = \frac{mc}{N} \sum_{kk'} \Lambda^{(i)}_{k'k} \langle k'|\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}|k\rangle\langle k'|\mathbf{V}_\mathbf{q}|k\rangle^*, \tag{11.219}$$

and

$$\underline{\mathbf{I}}_i = \frac{m}{N} \sum_{kk'} \Lambda^{(i)}_{k'k} \langle k'|\mathbf{V}_\mathbf{q}|k\rangle\langle k'|\mathbf{V}_\mathbf{q}|k\rangle^*. \tag{11.220}$$

The subscript (or superscript) i takes on the values 1 and 2, and

$$\Lambda^{(1)}_{k'k} = \frac{f_0(\varepsilon_{k'}) - f_0(\varepsilon_k)}{\varepsilon_{k'} - \varepsilon_k - \hbar\omega + \mathrm{i}\hbar/\tau} \tag{11.221}$$

and

$$\Lambda^{(2)}_{k'k} = \frac{f_0(\varepsilon_{k'}) - f_0(\varepsilon_k)}{\varepsilon_{k'} - \varepsilon_k}. \tag{11.222}$$

In the limit as $\tau \to \infty$, $\underline{\mathbf{I}} \to \underline{\mathbf{I}}_1$ and $\mathbf{K} \to \mathbf{K}_1$, hence

$$\mathbf{j}(\mathbf{q},\omega) \to \frac{\omega_\mathrm{p}^2}{4\pi\mathrm{i}\omega} \left[\underline{\mathbf{1}} + \underline{\mathbf{I}}_1\right] \cdot \mathbf{E}, \tag{11.223}$$

exactly as we had before. For $\omega\tau$ finite, there are corrections to this collisionless result that depend on $\frac{1}{\omega\tau}$.

11.8 Screening

Our original objective in considering linear response theory was to learn more about screening since we found that the long range of the Coulomb interaction was responsible for the divergence of perturbation theory beyond the first order exchange. Later on, when we mention Green's functions and the electron self energy, we will discuss some further details on *dynamic screening*, but for now, let us look at *static screening* effects.

If we set $\omega \to 0$ in (11.166), we can write

$$1 + I_{zz}(z,u) = -\frac{3}{2}u^2 \left[1 + \frac{1}{2}\left(\frac{1}{z} - z\right) \ln\left(\frac{1+z}{1-z}\right)\right]. \tag{11.224}$$

Here, $z = q/2k_\mathrm{F}$ and $u = \omega/qv_\mathrm{F}$. Substituting this result into $\varepsilon^{(l)} = 1 - \frac{\omega_\mathrm{p}^2}{\omega^2}(1 + I_{zz})$ gives

$$\varepsilon^{(l)}(q,0) = 1 + \frac{3\omega_{\rm p}^2}{q^2 v_{\rm F}^2} F(z), \tag{11.225}$$

where

$$F(z) = \frac{1}{2} + \frac{1}{4}\left(\frac{1}{z} - z\right) \ln\left(\frac{1+z}{1-z}\right). \tag{11.226}$$

Since $\ln\left(\frac{1+z}{1-z}\right) \simeq 2z(1 + \frac{z^2}{3} + \cdots)$, $F(z) \underset{z\to 0}{\longrightarrow} 1 - \frac{z^2}{3}$. For $z \gg 1$, $F(z) \simeq \frac{1}{3z^2}$ (see Fig. 11.12). For very long wave lengths, we have

$$\varepsilon^{(l)}(q) \simeq \left(1 - \frac{1}{3\pi a_0 k_{\rm F}}\right) + \frac{k_{\rm s}^2}{q^2}, \tag{11.227}$$

where $k_{\rm s} = \sqrt{\frac{4k_{\rm F}}{\pi a_0}}$ is called the *Thomas–Fermi screening wave number*. At high density $3\pi a_0 k_{\rm F} \gg 1$ so the constant term is usually approximated by unity. $\varepsilon_{\rm TF}(q) = 1 + \frac{k_{\rm s}^2}{q^2}$ is called the *Thomas–Fermi dielectric constant*. . One can certainly see that screening eliminates the divergence in perturbation theory that resulted from the $\phi_0(q) = \frac{4\pi e}{q^2}$ potential. We would write for the self-consistent screened potential by

$$\phi(q) = \frac{\phi_0(q)}{\varepsilon_{\rm TF}(q)} = \frac{4\pi e}{q^2 + k_{\rm s}^2}. \tag{11.228}$$

This potential does not diverge as $q \to 0$.

11.8.1 Friedel Oscillations

If $F(z)$ were identically equal to unity, then a point charge would give rise to the screened potential given by (11.228), which is the Yukawa potential $\phi(r) = \frac{e}{r}\mathrm{e}^{-k_{\rm s}r}$ in coordinate space. However, at $z = 1$ (or $q = 2k_{\rm F}$) $F(z)$

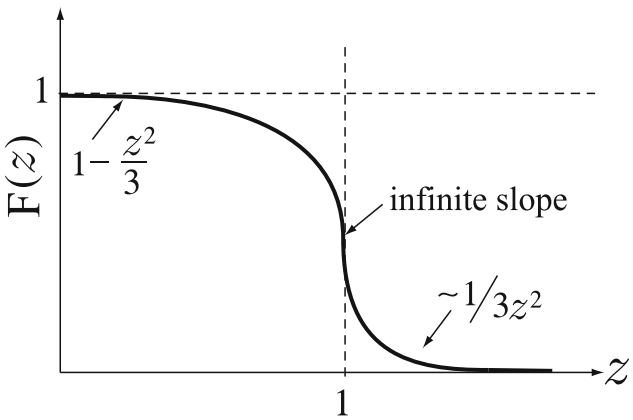


Fig. 11.12. Function $F(z)$ appearing in (11.225)

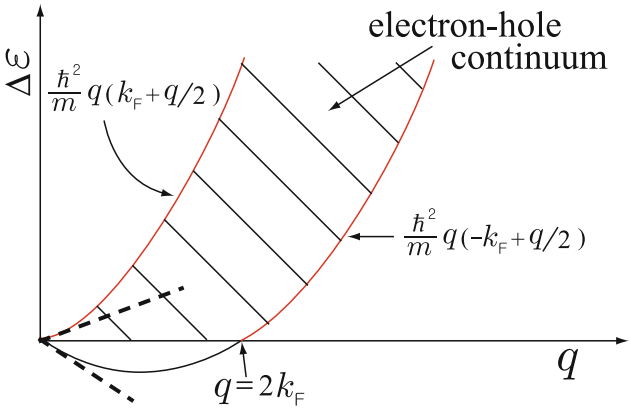


Fig. 11.13. Electron–hole pair excitation energies as a function of wave number q

drops very abruptly. In fact, $\frac{\mathrm{d}F}{\mathrm{d}z}$ has infinite slope at $z=1$ (see, for example, Fig. 11.12). The ability of the electron gas to screen disturbances of wave vector $\mathbf{q}$ drops abruptly at $q=2k_\mathrm{F}$. This is the result of the fact that pair excitations of zero energy can be created if $q<2k_\mathrm{F}$, but every pair excitation must have finite energy if $q>2k_\mathrm{F}$. This is apparent from a plot of $\Delta\varepsilon=\varepsilon_{k+q}-\varepsilon_k=\frac{\hbar^2}{m}q(k_z+\frac{q}{2})$ versus q for $k_z=\pm k_\mathrm{F}$ (see Fig. 11.13). The hatched area is called the *electron–hole continuum*. If $F(z)$ were replaced by unity, the self-consistent potential $\phi(q)$ would be written as

$$\phi(q)=\frac{4\pi e}{q^2}\frac{1}{\varepsilon^{(l)}(q)}\simeq\frac{4\pi e}{q^2+k_\mathrm{s}^2}.$$

The Fourier transform $\phi(\mathbf{r})$ is given by

$$\phi(\mathbf{r})=\int\frac{\mathrm{d}^3q}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}\phi(q) \tag{11.229}$$

and one can show that this is equal to $\phi(\mathbf{r})=\frac{e}{r}\mathrm{e}^{-k_\mathrm{s}r}$, a Yukawa potential. Because $F(z)$ is not equal to unity, but decreases rapidly around $z=1$, the potential $\phi(r)$ and the induced electron density $n_1(\mathbf{r})$ are different from the results of the simple Thomas–Fermi model. In the equation for $\phi(\mathbf{r})$ we must replace k_s^2 by $k_\mathrm{s}^2F(z)$ so that

$$\phi(\mathbf{r})=\int\frac{\mathrm{d}^3q}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}\frac{4\pi e}{q^2+k_\mathrm{s}^2F(q/2k_\mathrm{F})}. \tag{11.230}$$

The induced electron density is given by

$$n_1(q)=\frac{k_\mathrm{s}^2}{4\pi}F(z)\phi(q). \tag{11.231}$$

This can be obtained from $\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0$, where $\mathbf{j} = \underline{\sigma} \cdot \mathbf{E} = (\varepsilon^{(l)} - 1)\frac{\mathrm{i}\omega}{4\pi}(+\mathrm{i}\mathbf{q}\phi)$ and $\varepsilon^{(l)} = 1 + \frac{k_s^2}{q^2}F(z)$. After a little algebra one can show that the Fourier transform of $n_1(q)$ is given by

$$n_1(r) = \frac{12 n_0}{\pi a_0 k_\mathrm{F}} \int_0^\infty \frac{\sin 2k_\mathrm{F} r z}{2k_\mathrm{F} r z} \frac{F(z)}{1 + F(z)/(\pi a_0 k_\mathrm{F} z^2)} dz. \tag{11.232}$$

This can be written in a simpler form using

$$\frac{F(z)}{1 + F(z)/(\pi a_0 k_\mathrm{F} z^2)} = F(z) - \frac{F^2(z)}{\pi a_0 k_\mathrm{F} z^2 + F(z)}. \tag{11.233}$$

Then, $n_1(r)$ becomes

$$n_1(r) = \frac{12 n_0}{\pi a_0 k_\mathrm{F}} \left[\int_0^\infty \frac{\sin 2k_\mathrm{F} r z}{2k_\mathrm{F} r z} F(z)\, \mathrm{d}z - \int_0^\infty \frac{\sin 2k_\mathrm{F} r z}{2k_\mathrm{F} r z} \frac{F^2(z)}{\pi a_0 k_\mathrm{F} z^2 + F(z)}\, \mathrm{d}z \right]. \tag{11.234}$$

In the high density limit $\pi a_0 k_\mathrm{F} \gg 1$. Therefore in the region where $F(z)$ deviates appreciably from unity, i.e. for $z \geq 1$, $\pi a_0 k_\mathrm{F} z^2 \gg F(z)$, and we make a small error by replacing $F(z)$ in the second term of (11.234) by unity. This high-density approximation gives

$$n_1(r) = \frac{12 n_0}{\pi a_0 k_\mathrm{F}} \left[\int_0^\infty \frac{\sin 2k_\mathrm{F} r z}{2k_\mathrm{F} r z} F(z)\, \mathrm{d}z - \int_0^\infty \frac{\sin 2k_\mathrm{F} r z}{2k_\mathrm{F} r z} \frac{\mathrm{d}z}{\pi a_0 k_\mathrm{F} z^2 + 1} \right]. \tag{11.235}$$

The first integral can be evaluated exactly in terms of known functions

$$\begin{aligned} &\int_0^\infty \frac{\sin 2k_\mathrm{F} r z}{2k_\mathrm{F} r z} F(z)\, \mathrm{d}z \\ &= \frac{\pi}{2} \left\{ \frac{1}{2k_\mathrm{F} r} - \frac{1}{4k_\mathrm{F} r} \left[\frac{\sin 2k_\mathrm{F} r}{2k_\mathrm{F} r} + \cos 2k_\mathrm{F} r \right] + \frac{1}{2} \left[\frac{\pi}{2} - \mathrm{Si}(2k_\mathrm{F} r) \right] \right\} \equiv \frac{\pi}{2} f(2k_\mathrm{F} r), \end{aligned} \tag{11.236}$$

where $\mathrm{Si}(x) = \int_0^x \frac{\sin t}{t}\, \mathrm{d}t$ is the *sine integral function*. For very large values of x, the function $f(x)$ in (11.236) behaves

$$f(x) \simeq \frac{1}{x} + \frac{\cos x}{x^3} + O\left(\text{ higher orders of } \frac{1}{x} \right). \tag{11.237}$$

The second integral in (11.235) becomes

$$\int_0^\infty \frac{\sin 2k_\mathrm{F} r z}{2k_\mathrm{F} r z} \frac{\mathrm{d}z}{\pi a_0 k_\mathrm{F} z^2 + 1} = \frac{\pi}{4k_\mathrm{F} r} \left(1 - \mathrm{e}^{-\frac{r}{\sqrt{\pi a_0 k_\mathrm{F}}}} \right). \tag{11.238}$$

Therefore, for high density limit ($\pi a_0 k_\mathrm{F} \gg 1$) and large distances from the point charge impurity, the induced electron density is given by

$$n_1(r) = \frac{6n_0}{a_0 k_\mathrm{F}} \frac{\cos 2k_\mathrm{F} r}{(2k_\mathrm{F} r)^3}. \tag{11.239}$$

The oscillating behavior of the induced electron density at a wave vector $q = 2k_\mathrm{F}$ is known as a *Friedel oscillation*. Notice that the electron density induced by the presence of the point charge impurity falls off in amplitude as $\frac{1}{r^3}$. For a Yukawa potential ($\phi = \frac{e}{r}\mathrm{e}^{-k_\mathrm{s} r}$), the fall in the induced electron density is exponential.

11.8.2 Kohn Effect

When we discussed the Sommerfeld model we found a result for s_l the velocity of a longitudinal sound wave that could be written

$$\omega^2 = s_l^2 q^2 \simeq \left(\sqrt{\frac{zm}{3M}} v_\mathrm{F}\right)^2 q^2. \tag{11.240}$$

In other words the longitudinal sound velocity was given by

$$s_l = \sqrt{\frac{zm}{3M}} v_\mathrm{F}, \tag{11.241}$$

where z is the valence (charge on the positive ions), M is the ionic mass, and v_F the Fermi velocity.

This result can easily be obtained by saying that the positive ions have a *bare plasma frequency*

$$\Omega_\mathrm{p} = \sqrt{\frac{4\pi N_\mathrm{I} (ze)^2}{M}}, \tag{11.242}$$

where N_I is the number of ions per unit volume. However, the electrons will screen the charge fluctuations in the ion density, so that the actual frequency of a longitudinal sound wave of wave vector q will be

$$\omega = \frac{\Omega_\mathrm{p}}{\sqrt{\varepsilon(q,\omega)}}, \tag{11.243}$$

where $\varepsilon(q,\omega)$ is the dielectric function of the electron gas. Because the acoustic frequency is much smaller than the electron plasma frequency and $\omega \simeq s_l q \ll q v_\mathrm{F}$, we can approximate $\varepsilon(q,\omega)$ by $\varepsilon(q,0)$ in the first approximation

$$\omega^2 = \frac{\Omega_\mathrm{p}^2}{1 + \frac{k_\mathrm{s}^2}{q^2} F(z)} \simeq \frac{q^2 \Omega_\mathrm{p}^2}{q^2 + k_\mathrm{s}^2 F(z)}. \tag{11.244}$$

Let us assume $k_\mathrm{s}^2 \gg q^2$. If we take $F(z) \simeq 1$, we obtain

$$\omega^2 \simeq \frac{q^2}{k_\mathrm{s}^2} \Omega_\mathrm{p}^2. \tag{11.245}$$

But recall that $\Omega_{\rm p}^2 = 4\pi \left(\frac{n_0}{z}\right) \frac{(ze)^2}{M}$ and $k_{\rm s}^2 = \frac{4k_{\rm F}}{\pi a_0}$. Substituting into (11.245) gives the result given by (11.240). However, q need not be small compared to $k_{\rm F}$, even though $\omega \simeq s_l q$ will still be small compared to $qv_{\rm F}$ and $\omega_{\rm p}$. Then, we must keep $F(z)$ and write

$$\omega^2 \simeq \frac{s_l^2 q^2}{F(z) + \frac{\pi a_0}{4k_{\rm F}} q^2}. \tag{11.246}$$

Because $F(z)$ has an infinite first derivative at $q = 2k_{\rm F}$ (or $z = 1$), the phonon dispersion relation will show a small anomaly at $q = 2k_{\rm F}$ that is called the *Kohn anomaly*[8].

[8] W. Kohn, Phys. Rev. Lett. **2**, 393 (1959).

Problems

11.1. Let us consider the paramagnetic state of a degenerate electron gas, in which $\overline{n}_{\mathbf{k}\sigma} = 1$ for $\varepsilon_{\mathbf{k}\sigma} < \varepsilon_{\mathrm{F}}$ and zero otherwise.

(a) Show that the exchange contribution to the energy of wave vector $\mathbf{k}$ and spin σ is

$$\Sigma_{\mathrm{X}\sigma}(\mathbf{k}) = -\frac{1}{\Omega}\sum_{\mathbf{k}'}\overline{n}_{\mathbf{k}'\sigma}\frac{4\pi e^2}{|\mathbf{k}^2 - \mathbf{k}'|^2}.$$

(b) Convert the sum over $\mathbf{k}'$ to an integral and perform the integral to obtain

$$\Sigma_{\mathrm{X}\sigma}(\mathbf{k}) = -\frac{e^2 k_{\mathrm{F}}}{\pi}\left[1 + \frac{1-x^2}{2x}\ln\left|\frac{1+x}{1-x}\right|\right],$$

where $x = k/k_{\mathrm{F}}$.

(c) Plot $\Sigma_{\mathrm{X}\sigma}(k)$ as a function of $\frac{k}{k_{\mathrm{F}}}$.

(d) Show that the total energy (kinetic plus exchange) for the N particle paramagnetic state in the Hartree–Fock approximation is

$$\begin{aligned} E_{\mathrm{P}} &= \textstyle\sum_{\mathbf{k}\sigma}\overline{n}_{\mathbf{k}\sigma}\left[\dfrac{\hbar^2 k^2}{2m} + \Sigma_{\mathrm{X}\sigma}(\mathbf{k})\right] \\ &= N\left(\frac{3}{5}\frac{\hbar^2 k_{\mathrm{F}}^2}{2m} - \frac{3e^2 k_{\mathrm{F}}}{4\pi}\right). \end{aligned}$$

11.2. Consider the ferromagnetic state of a degenerate electron gas, in which $\overline{n}_{\mathbf{k}\uparrow} = 1$ for $k < k_{\mathrm{F}\uparrow}$ and $\overline{n}_{\mathbf{k}\downarrow} = 0$ for all k.

(a) Determine the Hartree–Fock energy $\varepsilon_{\mathbf{k}\sigma} = \frac{\hbar^2 k^2}{2m} + \Sigma_{\mathrm{X}\sigma}(\mathbf{k})$.

(b) Determine the value of k_{F} (Fermi wave vector of the nonmagnetic state) for which the ferromagnetic state is a valid Hartree–Fock solution.

(c) Determine the value of k_{F} for which $E_{\mathrm{F}} = \sum_{\mathrm{k}}\varepsilon_{\mathbf{k}\uparrow}$ has lower energy than E_{P} obtained in Problem 11.1.

11.3. Evaluate $I_{xx}(q,\omega)$ in the same way as we evaluated $I_{zz}(q,\omega)$, which is given by (11.166).

11.4. The longitudinal dielectric function is written as

$$\varepsilon^{(l)}(q,\omega) = 1 - \frac{\omega_p^2}{\omega^2}[1 + I_{zz}(q,\omega)].$$

Use $\ln(x+\mathrm{i}y) = \frac{1}{2}\ln(x^2+y^2) + \mathrm{i}\arctan\frac{y}{x}$ to evaluate $\varepsilon_2^{(l)}(z,u)$, the imaginary part of $\varepsilon^{(l)}(q,\omega)$, where $z = q/2k_{\mathrm{F}}$ and $u = \omega/qv_{\mathrm{F}}$.

11.5. Let us consider the static dielectric function written as

$$\varepsilon^{(l)}(q,0) = 1 + \frac{3\omega_{\mathrm{p}}^2}{q^2 v_{\mathrm{F}}^2} F(z),$$

where $z = q/2k_{\mathrm{F}}$ and $F(z) = \frac{1}{2} + \frac{1}{4}\left(\frac{1}{z} - z\right)\ln\left(\frac{1+z}{1-z}\right)$.

(a) Expand $F(z)$ in power of z for $z \ll 1$. Repeat it in power of $1/z$ for $z \gg 1$.
(b) Determine the expressions of the static dielectric function $\varepsilon^{(l)}(q,0)$ in the corresponding limits.

11.6. In the absence of a dc magnetic field, we see that $|\nu\rangle = |k_x, k_y, k_z\rangle \equiv |\mathbf{k}\rangle$, the free electron states.

(a) Show that $\langle \mathbf{k}'|\mathbf{V}_q|\mathbf{k}\rangle = \frac{\hbar}{m}(\mathbf{k} + \frac{\mathbf{q}}{2})\delta_{\mathbf{k}',\mathbf{k}+\mathbf{q}}$, where $\mathbf{V}_q = \mathbf{v}_0 \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}} + \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}\mathbf{v}_0$.
(b) Derive the Lindhard form of the conductivity tensor given by

$$\underline{\sigma}(\mathbf{q},\omega) = \frac{\omega_{\mathrm{p}}^2}{4\pi\mathrm{i}\omega}\left[\underline{1} + \frac{m}{N}\sum_{\mathbf{k}} \frac{f_0(E_{\mathbf{k}+\mathbf{q}}) - f_0(E_{\mathbf{k}})}{E_{\mathbf{k}+\mathbf{q}} - E_{\mathbf{k}} - \hbar\omega}\left(\frac{\hbar}{m}\right)^2 \left(\mathbf{k} + \frac{\mathbf{q}}{2}\right)\left(\mathbf{k} + \frac{\mathbf{q}}{2}\right)\right].$$

(c) Show that the Lindhard form of the dielectric tensor is written by

$$\underline{\varepsilon}(\mathbf{q},\omega) = \left(1 - \frac{\omega_{\mathrm{p}}^2}{\omega^2}\right)\underline{1} - \frac{m\omega_{\mathrm{p}}^2}{N\omega^2}\sum_{\mathbf{k}} \frac{f_0(E_{\mathbf{k}+\mathbf{q}}) - f_0(E_{\mathbf{k}})}{E_{\mathbf{k}+\mathbf{q}} - E_{\mathbf{k}} - \hbar\omega}\left(\frac{\hbar}{m}\right)^2 \left(\mathbf{k} + \frac{\mathbf{q}}{2}\right)\left(\mathbf{k} + \frac{\mathbf{q}}{2}\right).$$

11.7. Suppose that a system has a strong and sharp absorption line at a frequency ω_{A} and that $\varepsilon_2(\omega)$ can be approximated by

$$\varepsilon_2(\omega) = A\delta(\omega - \omega_{\mathrm{A}}) \quad \text{for } \omega > 0.$$

(a) Evaluate $\varepsilon_1(\omega)$ by using the Kronig–Kramers relation.
(b) Sketch $\varepsilon_1(\omega)$ as a function of ω.

11.8. The equation of motion of a charge $(-e)$ of mass m harmonically bound to a lattice point R_n is given by

$$m(\ddot{\mathbf{x}} + \gamma\dot{\mathbf{x}} + \omega_0^2\mathbf{x}) = -e\mathbf{E}\mathrm{e}^{\mathrm{i}\omega t}.$$

Here, $\mathbf{x} = \mathbf{r} - \mathbf{R}$, ω_0 is the oscillator frequency, and the electric field $\mathbf{E} = E\hat{x}$.

(a) Solve the equation of motion for $\mathbf{x}(t) = \mathbf{X}(\omega)\mathrm{e}^{\mathrm{i}\omega t}$.
(b) Let us consider the polarization $P(\omega) = -en_0X(\omega)$, where n_0 is the number of oscillators per unit volume. Write $P(\omega) = \alpha(\omega)E(\omega)$ and determine $\alpha(\omega)$.
(c) Plot $\alpha_1(\omega)$ and $\alpha_2(\omega)$ vs. ω, where $\alpha = \alpha_1 + \alpha_2$.
(d) Show that $\alpha(\omega)$ satisfies the Kronig–Kramers relation.

11.9. Take $H = \frac{1}{2m}\left(\mathbf{p} + \frac{e}{c}\mathbf{A}\right)^2 - e\phi$ and $H' = \frac{1}{2m}\left(\mathbf{p} + \frac{e}{c}\mathbf{A}'\right)^2 - e\phi'$ where $\mathbf{A}' = \mathbf{A} + \nabla\chi$ and $\phi' = \phi - \frac{1}{c}\dot{\chi}$.

(a) Show that $H' - \frac{e}{c}\dot{\chi} = \mathrm{e}^{-\frac{\mathrm{i}e\chi}{\hbar c}} H \mathrm{e}^{\frac{\mathrm{i}e\chi}{\hbar c}}$.
(b) Show that $\rho' = \mathrm{e}^{-\frac{\mathrm{i}e\chi}{\hbar c}} \rho \mathrm{e}^{\frac{\mathrm{i}e\chi}{\hbar c}}$ satisfies the same equation of motion, viz.
(c) $\frac{\partial \rho'}{\partial t} + \frac{\mathrm{i}}{\hbar}\left[H', \rho'\right]_- = 0$ as ρ does.

11.10. Let us take $\tilde{\rho}_0(H, \eta) = \left[\exp(\frac{H-\eta}{\Theta}) + 1\right]^{-1}$ as the local thermal equilibrium distribution function (or local equilibrium density matrix). Here $\eta(\mathbf{r}, t) = \zeta + \zeta_1(\mathbf{r}, t)$ is the local value of the chemical potential at position $\mathbf{r}$ and time t, while ζ is the overall equilibrium chemical potential. Remember that the total Hamiltonian H is written as $H = H_0 + H_1$. Write $\tilde{\rho}_0(H, \eta) = \rho_0(H_0, \zeta) + \rho_2$ and show that, to terms linear in the self-consistent field,

$$\langle \mathbf{k}|\rho_2|\mathbf{k}'\rangle = \frac{f_0(\varepsilon_{\mathbf{k}'}) - f_0(\varepsilon_{\mathbf{k}})}{\varepsilon_{\mathbf{k}'} - \varepsilon_{\mathbf{k}}} \langle \mathbf{k}|H_1 - \zeta_1|\mathbf{k}'\rangle.$$

11.11. Longitudinal sound waves in a simple metal like Na or K can be represented by the relation $\omega^2 = \frac{\Omega_\mathrm{p}^2}{\varepsilon^{(l)}(q,\omega)}$, where $\varepsilon^{(l)}(q, \omega)$ is the Lindhard dielectric function. We know that, for finite ω, $\varepsilon^{(l)}(q, \omega)$ can be written as $\varepsilon^{(l)}(q, \omega) = \varepsilon_1(q, \omega) + \mathrm{i}\varepsilon_2(q, \omega)$. This gives rise to $\omega = \omega_1 + \mathrm{i}\omega_2$, and ω_2 is proportional to the attenuation of the sound wave via excitation of conduction electrons. Estimate $\omega_2(q)$ for the case $\omega_1^2 \simeq \frac{q^2\Omega_\mathrm{p}^2}{k_\mathrm{s}^2} \gg \omega_2^2$.

Summary

In this chapter, we briefly introduced method of second quantization and Hartree–Fock approximation to describe the ferromagnetism of a degenerate electron gas and spin density wave states in solids. Equation of motion method is considered for density matrix to describe gauge invariant theory of linear responses in the presence of the most general electromagnetic disturbance. Behavior of Lindhard dielectric functions and static screening effects are examined in detail. Oscillatory behavior of the induced electron density in the presence of point charge impurity and an anomaly in the phonon dispersion relation are also discussed.

In the second quantization or occupation number representation, the Hamiltonian of a many particle system with two body interactions can be written as

$$H = \sum_k \varepsilon_k c_k^\dagger c_k + \frac{1}{2} \sum_{kk'll'} \langle k'l'|V|kl\rangle c_{k'}^\dagger c_{l'}^\dagger c_l c_k,$$

where c_k and $c_{k'}^\dagger$ satisfy commutation (anticommutation) relation for Bosons (Fermions).

The Hartree–Fock Hamiltonian is given by $H = \sum_i E_i c_i^\dagger c_i$, where

$$E_i = \varepsilon_i + \sum_j \overline{n_j} \left[\langle ij|V|ij\rangle - \langle ij|V|ji\rangle\right].$$

The Hartree–Fock ground state energy of a degenerate electron gas in the paramagnetic phase is given by $E_{\mathbf{k}s} = \frac{\hbar^2 k^2}{2m} - \frac{e^2 k_\mathrm{F}}{2\pi}\left[2 + \frac{k_\mathrm{F}^2 - k^2}{k k_\mathrm{F}} \ln\left(\frac{k_\mathrm{F}+k}{k_\mathrm{F}-k}\right)\right]$. The total energy of the paramagnetic state is

$$E_\mathrm{P} = N\left[\frac{3}{5}\frac{\hbar^2 k_\mathrm{F}^2}{2m} - \frac{3}{4\pi} e^2 k_\mathrm{F}\right].$$

If only states of spin $\uparrow$ are occupied, we have

$$E_{k\uparrow} = \frac{\hbar^2 k^2}{2m} - \frac{2^{1/3} e^2 k_\mathrm{F}}{2\pi}\left[2 + \frac{2^{2/3} k_\mathrm{F}^2 - k^2}{2^{1/3} k_\mathrm{F} k} \ln\left(\frac{2^{1/3} k_\mathrm{F} + k}{2^{1/3} k_\mathrm{F} - k}\right)\right];\ E_{k\downarrow} = \frac{\hbar^2 k^2}{2m}.$$

The total energy in the ferromagnetic phase is

$$E_\mathrm{F} = \sum E_{k\uparrow} = N\left[2^{2/3}\frac{3}{5}\frac{\hbar^2 k_\mathrm{F}^2}{2m} - 2^{1/3}\frac{3}{4\pi} e^2 k_\mathrm{F}\right].$$

The exchange energy prefers parallel spin orientation, but the cost in kinetic energy is high for a ferromagnetic spin arrangement. In a spin density wave state, the (negative) exchange energy is enhanced with no costing as much in kinetic energy. The Hartree-Fock ground state of a spiral spin density wave can be written as $|\phi_\mathbf{k}\rangle = \cos\theta_k |\mathbf{k}\uparrow\rangle + \sin\theta_k |\mathbf{k}+\mathbf{Q}\downarrow\rangle$.

In the presence of the self-consistent (Hartree) field $\{\phi, \mathbf{A}\}$, the Hamiltonian is written as $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$, where $\mathcal{H}_0$ is the Hamiltonian in the absence of the self-consistent field and $\mathcal{H}_1 = \frac{e}{2c}(\mathbf{v}_0 \cdot \mathbf{A} + \mathbf{A} \cdot \mathbf{v}_0) - e\phi$, up to terms linear in $\{\phi, \mathbf{A}\}$. Here, $\mathbf{v}_0 = \frac{\mathbf{p}}{m}$ and the equation of motion of ρ is $\frac{\partial \rho}{\partial t} + \frac{\mathrm{i}}{\hbar}[H, \rho]_- = 0$.

The current and charge densities at $(\mathbf{r}_0, t)$ are given, respectively, by

$$\mathbf{j}(\mathbf{r}_0, t) = \mathrm{Tr}\left[-e\left(\frac{1}{2}\mathbf{v}\delta(\mathbf{r} - \mathbf{r}_0) + \frac{1}{2}\delta(\mathbf{r} - \mathbf{r}_0)\mathbf{v}\right)\hat{\rho}\right];\ n(\mathbf{r}_0, t) = \mathrm{Tr}\left[-e\delta(\mathbf{r} - \mathbf{r}_0)\hat{\rho}\right].$$

Here $-e\left[\frac{1}{2}\mathbf{v}\delta(\mathbf{r} - \mathbf{r}_0) + \frac{1}{2}\delta(\mathbf{r} - \mathbf{r}_0)\mathbf{v}\right]$ is the operator for the current density at position $\mathbf{r}_0$, while $-e\delta(\mathbf{r} - \mathbf{r}_0)$ is the charge density operator. Fourier transform of $\mathbf{j}(\mathbf{r}_0, t)$ gives

$$\mathbf{j}(\mathbf{q}, \omega) = \underline{\sigma}(\mathbf{q}, \omega) \cdot \mathbf{E}(\mathbf{q}, \omega)$$

where the conductivity tensor is given by $\underline{\sigma}(\mathbf{q}, \omega) = \frac{\omega_p^2}{4\pi \mathrm{i}\omega}\left[\mathbf{1} + \underline{\mathbf{I}}(\mathbf{q}, \omega)\right]$. Here

$$\underline{\mathbf{I}}(\mathbf{q}, \omega) = \frac{m}{N}\sum_{\mathbf{k},\mathbf{k}'} \frac{f_0(\varepsilon_{k'}) - f_0(\varepsilon_k)}{\varepsilon_{k'} - \varepsilon_k - \hbar\omega}\langle k'|\mathbf{V}_q|k\rangle\langle k'|\mathbf{V}_q|k\rangle^*$$

and the operator $\mathbf{V}_q$ is defined by $\mathbf{V}_q = \frac{1}{2}\mathbf{v}_0 \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}} + \frac{1}{2}\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}\mathbf{v}_0$.

The *longitudinal* and *transverse* dielectric functions are written as

$$\varepsilon^{(l)}(q, \omega) = 1 - \frac{\omega_p^2}{\omega^2}\left[1 + I_{zz}(q, \omega)\right];\ \varepsilon^{(\mathrm{tr})}(q, \omega) = 1 - \frac{\omega_p^2}{\omega^2}\left[1 + I_{xx}(q, \omega)\right].$$

Real part (ε_1) and imaginary part (ε_2) of the dielectric function satisfy the relation

$$\varepsilon_1(\omega) = 1 + \frac{2}{\pi}\mathcal{P}\int_0^\infty \frac{\omega'\varepsilon_2(\omega')}{\omega'^2 - \omega^2}\, d\omega'.$$

The power dissipation per unit volume is then written $\mathcal{P}(\mathbf{q}, \omega) = \frac{\omega}{2\pi}\varepsilon_2(q, \omega) \mid E_0 \mid^2$.

Due to collisions of electrons with lattice imperfections, the conductivity of a normal metal is not infinite at zero frequency. In the presence of collisions, the equation of motion of the density matrix becomes, in a relaxation time approximation,

$$\frac{\partial \rho}{\partial t} + \frac{\mathrm{i}}{\hbar}[H, \rho]_- = -\frac{\rho - \tilde{\rho}_0}{\tau}.$$

Here, $\tilde{\rho}_0$ is a *local equilibrium density matrix*. Including the effect of collisions, the induced current density becomes

$$\mathbf{j}(\mathbf{q}, \omega) = \frac{\omega_\mathrm{p}^2}{4\pi\mathrm{i}\omega}\left\{\underline{\mathbf{1}} + \underline{\mathbf{I}} - \frac{\mathrm{i}\omega\tau}{1 + \mathrm{i}\omega\tau}\frac{(\mathbf{K}_1 - \mathbf{K}_2)(\mathbf{K}_1' - \mathbf{K}_2')}{L_1 + \mathrm{i}\omega\tau L_2}\right\} \cdot \mathbf{E}.$$

In the static limit, the dielectric function reduces to

$$\varepsilon^{(l)}(q, 0) = 1 + \frac{3\omega_\mathrm{p}^2}{q^2 v_\mathrm{F}^2}F(z),$$

where $F(z) = \frac{1}{2} + \frac{1}{4}\left(\frac{1}{z} - z\right)\ln\left(\frac{1+z}{1-z}\right)$ and $z = q/2k_{\mathrm{F}}$. The self-consistent screened potential is written as

$$\phi(q) = \frac{4\pi e}{q^2 + k_{\mathrm{s}}^2 F(q/2k_{\mathrm{F}})}.$$

where $k_{\mathrm{s}} = \sqrt{\frac{4k_{\mathrm{F}}}{\pi a_0}}$. For high density limit ($\pi a_0 k_{\mathrm{F}} \gg 1$) and large distances from the point charge impurity, the induced electron density is given by

$$n_1(r) = \frac{6n_0}{a_0 k_{\mathrm{F}}} \frac{\cos 2k_{\mathrm{F}} r}{(2k_{\mathrm{F}} r)^3}.$$

Electronic screening of the charge fluctuations in the ion density modifies the dispersion relation of phonons, for example,

$$\omega^2 \simeq \frac{s_l^2 q^2}{F(z) + \frac{\pi a_0}{4k_{\mathrm{F}}} q^2}$$

showing a small anomaly at $q = 2k_{\mathrm{F}}$.

12

Many Body Interactions: Green's Function Method

12.1 Formulation

Let us assume that there is a complete orthogonal set of single particle states $\phi_i(\xi)$, where $\xi = \mathbf{r}, \sigma$. By this we mean that

$$\langle \phi_i \mid \phi_j \rangle = \delta_{ij} \quad \text{and} \quad \sum_i \mid \phi_i \rangle \langle \phi_i \mid = \underline{\mathbf{1}}. \tag{12.1}$$

We can define particle field operators ψ and $\psi^\dagger$ by

$$\psi(\xi) = \sum_i \phi_i(\xi) a_i \quad \text{and} \quad \psi^\dagger(\xi) = \sum_i \phi_i^*(\xi) a_i^\dagger, \tag{12.2}$$

where a_i ($a_i^\dagger$) is an annihilation (creation) operator for a particle in state i. From the commutation relations (or anticommutation relations) satisfied by a_i and $a_j^\dagger$, we can easily show that

$$\begin{aligned} [\psi(\xi), \psi(\xi')] &= \left[\psi^\dagger(\xi), \psi^\dagger(\xi')\right] = 0, \\ \left[\psi(\xi), \psi^\dagger(\xi')\right] &= \delta(\xi - \xi'). \end{aligned} \tag{12.3}$$

The Hamiltonian of a many particle system can be written. (Here we set $\hbar = 1$.)

$$\begin{aligned} H = &\int \mathrm{d}^3 r \left\{ \frac{1}{2m} \nabla \psi_\alpha^\dagger(\mathbf{r}) \cdot \nabla \psi_\alpha(\mathbf{r}) + U^{(1)}(\mathbf{r}) \psi_\alpha^\dagger(\mathbf{r}) \psi_\alpha(\mathbf{r}) \right\} \\ &+ \frac{1}{2} \int \mathrm{d}^3 r \, \mathrm{d}^3 r' \psi_\alpha^\dagger(\mathbf{r}) \psi_\beta^\dagger(\mathbf{r}') U^{(2)}(\mathbf{r}, \mathbf{r}') \psi_\beta(\mathbf{r}') \psi_\alpha(\mathbf{r}). \end{aligned} \tag{12.4}$$

Summation over spin indices α and β is understood in (12.4). For the moment, let us omit spin to simplify the notation. Then

$$\psi(\mathbf{r}) = \sum_i \phi_i(\mathbf{r}) a_i. \tag{12.5}$$

We can write the density at a position $\mathbf{r}_0$ as

$$n(\mathbf{r}_0) = \int \mathrm{d}^3 r \psi^\dagger(\mathbf{r})\psi(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}_0) = \psi^\dagger(\mathbf{r}_0)\psi(\mathbf{r}_0). \tag{12.6}$$

The total particle number N is simply the integral of the density

$$N = \int \mathrm{d}^3 r n(\mathbf{r}) = \int \mathrm{d}^3 r \psi^\dagger(\mathbf{r})\psi(\mathbf{r}). \tag{12.7}$$

If we substitute (12.5) in (12.7), we obtain

$$N = \int \mathrm{d}^3 r \left(\sum_i \phi_i^*(\mathbf{r}) a_i\right)\left(\sum_j \phi_j(\mathbf{r}) a_j\right) = \sum_{ij} \langle \phi_i \mid \phi_j \rangle a_i^\dagger a_j. \tag{12.8}$$

By $\langle \phi_i \mid \phi_j \rangle = \delta_{ij}$, this reduces to

$$N = \sum_i a_i^\dagger a_i, \tag{12.9}$$

so that $\hat{n}_i = a_i^\dagger a_i$ is the number operator for the state i and $\hat{N} = \sum_i \hat{n}_i$ is the total number operator. It simply counts the number of particles.

12.1.1 Schrödinger Equation

The Schrödinger equation of the many particle wave function $\Psi(1, 2, \ldots, N)$ is ($\hbar \equiv 1$)

$$i\hbar \frac{\partial}{\partial t}\Psi = H\Psi. \tag{12.10}$$

We can write the time dependent solution $\Psi(t)$ as

$$\Psi(t) = \mathrm{e}^{-\mathrm{i}Ht}\Psi_\mathrm{H}, \tag{12.11}$$

where Ψ_H is time independent. If F is some operator whose matrix element between two states $\Psi_n(t)$ and $\Psi_m(t)$ is defined as

$$F_{nm}(t) = \langle \Psi_n(t)|F|\Psi_m(t)\rangle, \tag{12.12}$$

we can write $|\Psi_m(t)\rangle = \mathrm{e}^{-\mathrm{i}Ht}|\Psi_{\mathrm{H}m}\rangle$ and $\langle \Psi_n(t)| = \langle \Psi_{\mathrm{H}n}|\mathrm{e}^{\mathrm{i}Ht}$. Then $F_{nm}(t)$ can be rewritten

$$F_{nm}(t) = \langle \Psi_{\mathrm{H}n}|F(t)|\Psi_{\mathrm{H}m}\rangle, \tag{12.13}$$

where $F_\mathrm{H}(t) = \mathrm{e}^{\mathrm{i}Ht}F\mathrm{e}^{-\mathrm{i}Ht}$. The process of going from (12.12) to (12.13) is a transformation from the *Schrödinger picture* (where the state vector $\Psi(t)$ depends on time but the operator F does not) to the *Heisenberg picture* (where Ψ_H is a time-independent state vector but $F_\mathrm{H}(t)$ is a time-dependent operator). The transformation from (to) Schrödinger picture to (from) Heisenberg picture can be summarized by

$$\Psi_{\mathrm{S}}(t) = \mathrm{e}^{-\mathrm{i}Ht}\Psi_{\mathrm{H}} \quad \text{and} \quad F_{\mathrm{H}}(t) = \mathrm{e}^{\mathrm{i}Ht}F_{\mathrm{S}}\mathrm{e}^{-\mathrm{i}Ht}. \tag{12.14}$$

From these equations and (12.10), it is clear that

$$\frac{\partial F_{\mathrm{H}}(t)}{\partial t} = \mathrm{i}\left[H, F_{\mathrm{H}}\right]. \tag{12.15}$$

12.1.2 Interaction Representation

Suppose that the Hamiltonian H can be divided into two parts H_0 and H', where H' represents the interparticle interactions. We can define the state vector $\Psi_{\mathrm{I}}(t)$ in the *interaction representation* as

$$\Psi_{\mathrm{I}}(t) = \mathrm{e}^{\mathrm{i}H_0 t}\Psi_{\mathrm{S}}(t). \tag{12.16}$$

Operate $\mathrm{i}\partial/\partial t$ on $\Psi_{\mathrm{I}}(t)$ and make use of the fact that $\Psi_{\mathrm{S}}(t)$ satisfies the Schrödinger equation. This gives

$$\mathrm{i}\frac{\partial \Psi_{\mathrm{I}}(t)}{\partial t} = H_{\mathrm{I}}(t)\Psi_{\mathrm{I}}(t), \tag{12.17}$$

where

$$H_{\mathrm{I}}(t) = \mathrm{e}^{\mathrm{i}H_0 t}H'\mathrm{e}^{-\mathrm{i}H_0 t}. \tag{12.18}$$

From (12.12) and the fact that $\Psi_{\mathrm{S}}(t) = \mathrm{e}^{-\mathrm{i}H_0 t}\Psi_{\mathrm{I}}(t)$ it is apparent that

$$F_{\mathrm{I}}(t) = \mathrm{e}^{\mathrm{i}H_0 t}F_{\mathrm{S}}\mathrm{e}^{-\mathrm{i}H_0 t}. \tag{12.19}$$

By explicit evaluation of $\frac{\partial F_{\mathrm{I}}}{\partial t}$ from (12.19), it is clear that

$$\frac{\partial F_{\mathrm{I}}}{\partial t} = \mathrm{i}\left[H_0, F_{\mathrm{I}}(t)\right]. \tag{12.20}$$

The interaction representation has a number of advantages for interacting systems; among them are:

1. All operators have the form of Heisenberg operators of the noninteracting system, i.e., (12.19).
2. Wave functions satisfy the Schrödinger equation with Hamiltonian $H_{\mathrm{I}}(t)$, i.e., (12.17).

Because operators satisfy commutation relations only for equal times, $H_{\mathrm{I}}(t_1)$ and $H_{\mathrm{I}}(t_2)$ do not commute if $t_1 \neq t_2$. Because of this, we cannot simply integrate the Schrödinger equation, (12.17), to obtain

$$\Psi_{\mathrm{I}}(t) \propto \mathrm{e}^{-\mathrm{i}\int^t H_{\mathrm{I}}(t')\mathrm{d}t'}. \tag{12.21}$$

Instead, we do the following:

1. Assume that $\Psi(t)$ is known at $t = t_0$.
2. Integrate the differential equation from t_0 to t.

This gives that

$$\Psi_{\rm I}(t) - \Psi_{\rm I}(t_0) = -{\rm i}\int_{t_0}^{t} {\rm d}t'\, H_{\rm I}(t')\Psi_{\rm I}(t'). \tag{12.22}$$

This is an integral equation for $\Psi_{\rm I}(t)$ that we can try to solve by iteration. Let us write

$$\Psi_{\rm I}(t) \equiv \Psi_{\rm I}^{(0)}(t) + \Psi_{\rm I}^{(1)}(t) + \cdots + \Psi_{\rm I}^{(n)}(t) + \cdots . \tag{12.23}$$

Here

$$\begin{aligned}
\Psi_{\rm I}^{(0)}(t) &= \Psi_{\rm I}(t_0),\\
\Psi_{\rm I}^{(1)}(t) &= -{\rm i}\textstyle\int_{t_0}^{t} {\rm d}t'\, H_{\rm I}(t')\Psi_{\rm I}^{(0)}(t'),\\
\Psi_{\rm I}^{(2)}(t) &= -{\rm i}\textstyle\int_{t_0}^{t} {\rm d}t'\, H_{\rm I}(t')\Psi_{\rm I}^{(1)}(t'),\\
&\vdots\\
\Psi_{\rm I}^{(n)}(t) &= -{\rm i}\textstyle\int_{t_0}^{t} {\rm d}t'\, H_{\rm I}(t')\Psi_{\rm I}^{(n-1)}(t').
\end{aligned} \tag{12.24}$$

This result can be expressed as

$$\Psi_{\rm I}(t) = S(t,t_0)\Psi_{\rm I}(t_0), \tag{12.25}$$

where $S(t,t_0)$ is the so-called S *matrix* is given by

$$\begin{aligned}
S(t,t_0) &= 1 - {\rm i}\textstyle\int_{t_0}^{t} {\rm d}t_1\, H_{\rm I}(t_1) + (-{\rm i})^2 \int_{t_0}^{t} {\rm d}t_1 \int_{t_0}^{t_1} {\rm d}t_2 H_{\rm I}(t_1)H_{\rm I}(t_2) + \cdots\\
&= \textstyle\sum_{n=0}^{\infty}(-{\rm i})^n \int_{t_0}^{t} {\rm d}t_1 \int_{t_0}^{t_1} {\rm d}t_2 \cdots \int_{t_0}^{t_{n-1}} {\rm d}t_n \Big[H_{\rm I}(t_1)H_{\rm I}(t_2)\cdots H_{\rm I}(t_n)\Big].
\end{aligned} \tag{12.26}$$

Let us look at the third term involving integration over t_1 and t_2

$$I_2 = \int_{t_0}^{t} {\rm d}t_1 \int_{t_0}^{t_1} {\rm d}t_2 H_{\rm I}(t_1)H_{\rm I}(t_2) = \frac{1}{2}I_2 + \frac{1}{2}I_2. \tag{12.27}$$

In the second $\frac{1}{2}I_2$, let us reverse the order of integration (see Fig. 12.1). We first integrated over t_2 from t_0 to t_1, then over t_1 from t_0 to t. Inverting the order gives

$$\int_{t_0}^{t} {\rm d}t_1 \int_{t_0}^{t_1} {\rm d}t_2 \Rightarrow \int_{t_0}^{t} {\rm d}t_2 \int_{t_2}^{t} {\rm d}t_1,$$

Therefore, we have

$$\frac{1}{2}I_2 = \frac{1}{2}\int_{t_0}^{t} {\rm d}t_1 \int_{t_0}^{t_1} {\rm d}t_2 H_{\rm I}(t_1)H_{\rm I}(t_2) = \frac{1}{2}\int_{t_0}^{t} {\rm d}t_2 \int_{t_2}^{t} {\rm d}t_1 H_{\rm I}(t_1)H_{\rm I}(t_2). \tag{12.28}$$

But t_1 and t_2 are dummy integration variables and we can interchange the names to get

$$\frac{1}{2}I_2 = \frac{1}{2}\int_{t_0}^{t} {\rm d}t_1 \int_{t_1}^{t} {\rm d}t_2 H_{\rm I}(t_2)H_{\rm I}(t_1). \tag{12.29}$$

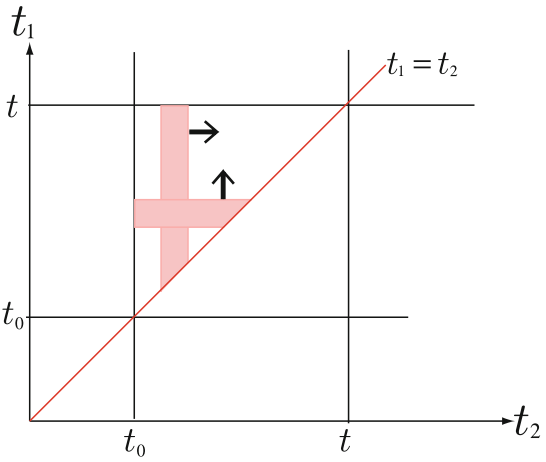


Fig. 12.1. Order of integration I_2 appearing in (12.27)

Adding this term to the $\frac{1}{2}I_2$ that was left in its original form gives

$$I_2 = \frac{1}{2}\int_{t_0}^{t} \mathrm{d}t_1 \int_{t_0}^{t_1} \mathrm{d}t_2 H_\mathrm{I}(t_1)H_\mathrm{I}(t_2) + \frac{1}{2}\int_{t_0}^{t} \mathrm{d}t_1 \int_{t_1}^{t} \mathrm{d}t_2 H_\mathrm{I}(t_2)H_\mathrm{I}(t_1). \quad (12.30)$$

We are integrating over a square of edge $\Delta t = t - t_0$ in the t_1t_2-plane. The second term, with $t_2 > t_1$, is just an integral on the lower triangle shown in Fig. 12.1. The first term, where $t_1 > t_2$, is an integral on the upper triangle. Therefore, we can combine the time integrals and write the limits of integration from t_0 to t.

$$I_2 = \frac{1}{2}\int_{t_0}^{t} \mathrm{d}t_1 \int_{t_0}^{t} \mathrm{d}t_2 \left[H_\mathrm{I}(t_1)H_\mathrm{I}(t_2)\theta(t_1 - t_2) + H_\mathrm{I}(t_2)H_\mathrm{I}(t_1)\theta(t_2 - t_1)\right]. \quad (12.31)$$

The thing we have to be careful about here, however, is that $H_\mathrm{I}(t_1)$ and $H_\mathrm{I}(t_2)$ do not necessarily commute. We can get around this difficulty by using the *time ordering operator* T. The product of functions $H_\mathrm{I}(t_j)$ that follows the operator T must have the largest t values on the left. In the first term of (12.31), $t_1 > t_2$, so we can write the integrand as

$$H_\mathrm{I}(t_1)H_\mathrm{I}(t_2) = \mathrm{T}\{H_\mathrm{I}(t_1)H_\mathrm{I}(t_2)\}.$$

In the second term, with $t_2 > t_1$, we may write

$$H_\mathrm{I}(t_2)H_\mathrm{I}(t_1) = \mathrm{T}\{H_\mathrm{I}(t_1)H_\mathrm{I}(t_2)\}.$$

Equation (12.31) can, thus, be rewritten as

$$I_2 = \frac{1}{2}\int_{t_0}^{t} \mathrm{d}t_1 \int_{t_0}^{t} \mathrm{d}t_2 \mathrm{T}\left\{H_\mathrm{I}(t_1)H_\mathrm{I}(t_2)\right\}. \quad (12.32)$$

For the general term we have

$$I_n = \int_{t_0}^{t} \mathrm{d}t_1 \int_{t_0}^{t_1} \mathrm{d}t_2 \cdots \int_{t_0}^{t_{n-1}} \mathrm{d}t_n \; H_{\mathrm{I}}(t_1)H_{\mathrm{I}}(t_2)\cdots H_{\mathrm{I}}(t_n)$$
$$= \int \mathrm{d}t_1 \mathrm{d}t_2 \cdots \mathrm{d}t_n \; H_{\mathrm{I}}(t_1)H_{\mathrm{I}}(t_2)\cdots H_{\mathrm{I}}(t_n) \text{ with } t \geq t_1 \geq t_2 \geq \cdots \geq t_n,$$

and it is not difficult to see that the same technique can be applied to give

$$I_n = \frac{1}{n!} \int_{t_0}^{t} \mathrm{d}t_1 \int_{t_0}^{t} \mathrm{d}t_2 \cdots \int_{t_0}^{t} \mathrm{d}t_n \mathrm{T}\{H_{\mathrm{I}}(t_1)H_{\mathrm{I}}(t_2)\cdots H_{\mathrm{I}}(t_n)\}. \tag{12.33}$$

Note that

$$\mathrm{T}\{H_{\mathrm{I}}(t_1)H_{\mathrm{I}}(t_2)\cdots H_{\mathrm{I}}(t_n)\} = H_{\mathrm{I}}(t_1)H_{\mathrm{I}}(t_2)\cdots H_{\mathrm{I}}(t_n) \text{ if } t_1 \geq t_2 \geq \cdots \geq t_n.$$

Making use of (12.33), the S matrix can be written in the compact form

$$S(t, t_0) = \mathrm{T}\left\{\mathrm{e}^{-\mathrm{i}\int_{t_0}^{t} H_{\mathrm{I}}(t')\mathrm{d}t'}\right\}, \tag{12.34}$$

where it is understood that in the nth term in the expansion of the exponential, (12.33) holds. We note that, at $t = 0$, the wave functions Ψ_{S}, Ψ_{I} coincide,

$$\Psi_{\mathrm{S}}(0) = \Psi_{\mathrm{I}}(0) = S(0, t_0)\Psi_{\mathrm{I}}(t_0) = S(0, t_0)\mathrm{e}^{\mathrm{i}H_0 t_0}\Psi_{\mathrm{S}}(t_0),$$

where we have used $\Psi_{\mathrm{I}}(t) = \mathrm{e}^{\mathrm{i}H_0 t}\Psi_{\mathrm{S}}(t)$ and $\Psi_{\mathrm{I}}(t) = S(t, t_0)\Psi_{\mathrm{I}}(t_0)$.

12.2 Adiabatic Approximation

Suppose that we multiply H_{I} by $\mathrm{e}^{-\beta|t|}$ where $\beta \geq 0$, and treat the resulting interaction as one that vanishes at $t = \pm\infty$. Then, the interaction is slowly turned on from $t = -\infty$ up to $t = 0$ and slowly turned off from $t = 0$ till $t = +\infty$. We can write $H(t = -\infty) = H_0$, the noninteracting Hamiltonian, and

$$\Psi_{\mathrm{I}}(t = -\infty) = \Psi_{\mathrm{H}}(t = -\infty) = \Phi_{\mathrm{H}}. \tag{12.35}$$

Here, Φ_{H} is the Heisenberg state vector of the noninteracting system. We know that eigenstates of the interacting system in the Heisenberg, Schrödinger, and interaction representation are related by

$$\Psi_{\mathrm{H}}(t) = \mathrm{e}^{\mathrm{i}Ht}\Psi_{\mathrm{S}}(t) \quad \text{and} \quad \Psi_{\mathrm{I}}(t) = \mathrm{e}^{\mathrm{i}H_0 t}\Psi_{\mathrm{S}}(t). \tag{12.36}$$

Therefore, at time t = 0,

$$\Psi_{\mathrm{I}}(t = 0) = \Psi_{\mathrm{H}}(t = 0) = \Psi_{\mathrm{H}}. \tag{12.37}$$

Henceforth, we will use Ψ_{H} to denote the state vector of the fully interacting system in the Heisenberg representation. We can express Ψ_{H} as

$$\Psi_{\rm H} = S(0,-\infty)\Phi_{\rm H}, \tag{12.38}$$

where S is the S matrix defined in (12.34). Because $\Psi_{\rm I}(t=0) = \Psi_{\rm H}$ we can write

$$\Psi_{\rm I}(t) = S(t,0)\Psi_{\rm H} = S(t,-\infty)\Phi_{\rm H}. \tag{12.39}$$

In the last step, we have used $S(t_2,-\infty) = S(t_2,t_1)S(t_1,-\infty)$. If we write $|\Psi_{\rm I}(t)\rangle = S(t,0)|\Psi_{\rm H}\rangle$ and $\langle\Psi_{\rm I}(t)| = \langle\Psi_{\rm H}|S^{-1}(t,0)$, then for some operator F

$$\langle\Psi_{\rm I}(t)|F_{\rm I}|\Psi_{\rm I}(t)\rangle = \langle\Psi_{\rm H}|S^{-1}(t,0)F_{\rm I}S(t,0)|\Psi_{\rm H}\rangle = \langle\Psi_{\rm H}(t)|F_{\rm H}|\Psi_{\rm H}(t)\rangle. \tag{12.40}$$

But this must equal $\langle\Psi_{\rm H}|F_{\rm H}|\Psi_{\rm H}\rangle$. Therefore, we have

$$F_{\rm H} = S^{-1}(t,0)F_{\rm I}S(t,0). \tag{12.41}$$

Now, look at the expectation value in the exact Heisenberg interacting state $\Psi_{\rm H}$ of the time ordered product of Heisenberg operators

$$\langle\Psi_{\rm H}|{\rm T}\{A_{\rm H}(t_1)B_{\rm H}(t_2)\cdots Z_{\rm H}(t_n)\}|\Psi_{\rm H}\rangle.$$

If we assume that the t_i's have been arranged in the order $t_1 \geq t_2 \geq t_3 \geq \cdots \geq t_n$, then we can write

$$\frac{\langle\Psi_{\rm H}|{\rm T}\{A_{\rm H}(t_1)B_{\rm H}(t_2)\cdots Z_{\rm H}(t_n)\}|\Psi_{\rm H}\rangle}{\langle\Psi_{\rm H}|\Psi_{\rm H}\rangle} = \frac{\langle\Phi_{\rm H}|S(\infty,0)S^{-1}(t_1,0)A_{\rm I}(t_1)S(t_1,0)S^{-1}(t_2,0)B_{\rm I}(t_2)S(t_2,0)S^{-1}(t_3,0)\cdots S^{-1}(t_n,0)Z_{\rm I}(t_n)S(t_n,0)S(0,-\infty)|\Phi_{\rm H}\rangle}{\langle\Phi_{\rm H}|S(\infty,0)S(0,-\infty)|\Phi_{\rm H}\rangle}. \tag{12.42}$$

But from $S(t_1,0) = S(t_1,t_2)S(t_2,0)$ we can see that

$$S(t_1,0)S^{-1}(t_2,0) = S(t_1,t_2). \tag{12.43}$$

Using this in (12.42) gives

$$\frac{\langle\Psi_{\rm H}|{\rm T}\{A_{\rm H}(t_1)B_{\rm H}(t_2)\cdots Z_{\rm H}(t_n)\}|\Psi_{\rm H}\rangle}{\langle\Psi_{\rm H}|\Psi_{\rm H}\rangle} = \frac{\langle\Phi_{\rm H}|S(\infty,t_1)A_{\rm I}(t_1)S(t_1,t_2)B_{\rm I}(t_2)S(t_2,t_3)\cdots Z_{\rm I}(t_n)S(t_n,-\infty)|\Phi_{\rm H}\rangle}{\langle\Phi_{\rm H}|S(\infty,-\infty)|\Phi_{\rm H}\rangle}. \tag{12.44}$$

We note that, in (12.44), the operators are in time-ordered form, i.e. $t_n \geq -\infty$, $t_1 \geq t_2$, $\infty \geq t_1$, so the operators

$$S(\infty,t_1)A_{\rm I}(t_1)S(t_1,t_2)B_{\rm I}(t_2)S(t_2,t_3)\cdots Z_{\rm I}(t_n)S(t_n,-\infty)$$

are chronologically ordered, and hence we can rewrite (12.44) as

$$\frac{\langle\Psi_{\rm H}|{\rm T}\{A_{\rm H}(t_1)B_{\rm H}(t_2)\cdots Z_{\rm H}(t_n)\}|\Psi_{\rm H}\rangle}{\langle\Psi_{\rm H}|\Psi_{\rm H}\rangle} = \frac{\langle\Phi_{\rm H}|{\rm T}\{S(\infty,-\infty)A_{\rm I}(t_1)B_{\rm I}(t_2)\cdots Z_{\rm I}(t_n)\}|\Phi_{\rm H}\rangle}{\langle\Phi_{\rm H}|S(\infty,-\infty)|\Phi_{\rm H}\rangle}. \tag{12.45}$$

12.3 Green's Function

We define the Green's function $G_{\alpha\beta}(x,x')$, where $x=\{\mathbf{r},t\}$ and α, β are spin indices, by

$$G_{\alpha\beta}(x,x') = -\mathrm{i}\frac{\langle\Psi_\mathrm{H}|\mathrm{T}\{\psi^\mathrm{H}_\alpha(x)\psi^{\mathrm{H}\dagger}_\beta(x')\}|\Psi_\mathrm{H}\rangle}{\langle\Psi_\mathrm{H}|\Psi_\mathrm{H}\rangle}. \tag{12.46}$$

Here, $\psi^\mathrm{H}_\alpha(x)$ is an operator (particle field operator) in the Heisenberg representation.

By using Eq.(12.45) in Eq.(12.46), we obtain

$$G_{\alpha\beta}(x,x') = -\mathrm{i}\frac{\langle\Phi_\mathrm{H}|\mathrm{T}\{S(\infty,-\infty)\psi^\mathrm{I}_\alpha(x)\psi^{\mathrm{I}\dagger}_\beta(x')\}|\Phi_\mathrm{H}\rangle}{\langle\Phi_\mathrm{H}|S(\infty,-\infty)|\Phi_\mathrm{H}\rangle}. \tag{12.47}$$

The operator $\psi^\mathrm{I}_\alpha(x)$ is now in the interaction representation. If we write out the expansion for $S(\infty,-\infty)$ in the numerator and are careful to keep the time ordering, we obtain

$$\begin{aligned} G_{\alpha\beta}(\mathbf{r},t,\mathbf{r}',t') = &-\tfrac{\mathrm{i}}{\langle S(\infty,-\infty)\rangle}\textstyle\sum_{n=0}^{\infty}\frac{(-\mathrm{i})^n}{n!}\int_{-\infty}^{\infty}\mathrm{d}t_1\mathrm{d}t_2\cdots\mathrm{d}t_n \\ &\times\langle\Phi_\mathrm{H}|\mathrm{T}\{\psi^\mathrm{I}_\alpha(\mathbf{r},t)\psi^{\mathrm{I}\dagger}_\beta(\mathbf{r}',t')H_\mathrm{I}(t_1)\cdots H_\mathrm{I}(t_n)\}|\Phi_\mathrm{H}\rangle. \end{aligned} \tag{12.48}$$

12.3.1 Averages of Time-Ordered Products of Operators

If $F_1(t)$ and $F_2(t')$ are Fermion operators, then by $\mathrm{T}\{F_1(t)F_2(t')\}$ we mean

$$\begin{aligned} \mathrm{T}\{F_1(t)F_2(t')\} &= F_1(t)F_2(t') \quad \text{if } t>t' \\ &= -F_2(t')F_1(t) \quad \text{if } t<t'. \end{aligned} \tag{12.49}$$

In other words, we need a minus sign for every permutation of one Fermion operator past another. For Bosons no minus sign is needed.

In $G_{\alpha\beta}$ we find the ground state average of products of time-ordered operators like $\mathrm{T}\{ABC\cdots\}$. Here, A, B, ... are field operators (or products of field operators). When the entire time ordered product is expressed as a product of $\psi^\dagger$'s and ψ's, it is useful to put the product in what is called *normal* form, in which all annihilation operators appear to the right of all creation operators. For example, the normal product of $\psi^\dagger(1)\psi(2)$ can be written

$$\mathrm{N}\{\psi^\dagger(1)\psi(2)\} = \psi^\dagger(1)\psi(2) \;\text{ while }\; \mathrm{N}\{\psi(1)\psi^\dagger(2)\} = -\psi^\dagger(2)\psi(1). \tag{12.50}$$

The difference between a T product and an N product is called a *pairing* or a *contraction*. For example, the difference in the T ordered product and the N product of AB is given by

$$\mathrm{T}(AB) - \mathrm{N}(AB) = A^\mathrm{c}B^\mathrm{c}. \tag{12.51}$$

We note that the contraction of a pair of operators is the anticommutator we omit when we formally reorder a T product of a pair of operators to get an N product. The contractions are c-numbers for the operators we are interested in.

12.3.2 Wick's Theorem

The Wick's theorem states that T product of operators $ABC\cdots$ can be expressed as the sum of all possible N products with all possible pairings. By this we mean that

$$\begin{aligned}
&\mathrm{T}(ABCD\cdots XYZ)\\
=\ &\mathrm{N}(ABCD\cdots XYZ)\\
&+\mathrm{N}(A^{\mathrm{c}}B^{\mathrm{c}}CD\cdots XYZ)+\mathrm{N}(A^{\mathrm{c}}BC^{\mathrm{c}}D\cdots XYZ)+\mathrm{N}(A^{\mathrm{c}}BCD^{\mathrm{c}}\cdots XYZ)\\
&+\cdots+\mathrm{N}(ABCD\cdots XY^{\mathrm{c}}Z^{\mathrm{c}})\\
&+\mathrm{N}(A^{\mathrm{c}}B^{\mathrm{c}}C^{\mathrm{a}}D^{\mathrm{a}}\cdots XYZ)+\cdots+\mathrm{N}(ABCD\cdots W^{\mathrm{c}}X^{\mathrm{c}}Y^{\mathrm{a}}Z^{\mathrm{a}})\\
&\vdots\\
&+\mathrm{N}(A^{\mathrm{c}}B^{\mathrm{c}}C^{\mathrm{a}}D^{\mathrm{a}}\cdots Y^{\mathrm{b}}Z^{\mathrm{b}})+\mathrm{N}(A^{\mathrm{c}}B^{\mathrm{a}}C^{\mathrm{c}}D^{\mathrm{a}}\cdots Y^{\mathrm{b}}Z^{\mathrm{b}})+\text{ All other pairings.}
\end{aligned}\tag{12.52}$$

In evaluating the ground state expectation value of (12.52) only the term in which every operator is paired with some other operator is nonvanishing since the normal products that contain unpaired operators must vanish (they annihilate excitations that are not present in the ground state). In the second and third lines on the right, in each term we bring two operators together by anticommuting, but neglecting the anticommutators, then replace the pair by its contraction, and finally take the N product of the remaining $n-2$ operators. We do this with all possible pairings so we obtain $\frac{n(n-1)}{2}$ terms, each term containing an N product of the $n-2$ remaining operators. In the fourth line on the right, we choose two pairs in all possible ways, replace them by their contractions, and leave in each term an N products of the $n-4$ remaining operators. We repeat the same procedure, and in the last line on the right, every operator is paired with some other operator in all possible ways leaving no unpaired operators. Only the completely contracted terms (last line on the right of (12.52)) give finite contributions in the ground-state expectation value. That is, we have

$$\begin{aligned}
&\langle\mathrm{T}(ABCD\cdots XYZ)\rangle_0\\
=\ &\langle\mathrm{T}(AB)\rangle\langle\mathrm{T}(CD)\rangle\cdots\langle\mathrm{T}(YZ)\rangle\pm\langle\mathrm{T}(AC)\rangle\langle\mathrm{T}(BD)\rangle\cdots\langle\mathrm{T}(YZ)\rangle\\
&\pm\text{ All other pairings.}
\end{aligned}\tag{12.53}$$

Here, we have used $A^{\mathrm{c}}B^{\mathrm{c}}=\mathrm{T}(AB)-\mathrm{N}(AB)$ and noted that $\langle\mathrm{N}(AB)\rangle=0$, so the ground state expectation value of $\langle A^{\mathrm{c}}B^{\mathrm{c}}\rangle=\langle\mathrm{T}(AB)\rangle$. Now let us return to the expansion of the Green's function. The first term in the sum over n in (12.48) is

$$G^{(0)}(\mathbf{r},t,\mathbf{r}',t') = -\frac{\mathrm{i}}{\langle S(\infty,-\infty)\rangle}\langle \Phi_{\mathrm{H}}^{*} \mid \mathrm{T}\{\psi_{\mathrm{I}}(\mathbf{r},t)\psi_{\mathrm{I}}^{\dagger}(\mathbf{r}',t')\} \mid \Phi_{\mathrm{H}}\rangle_0, \quad (12.54)$$

where, now, the operators $\psi(\mathbf{r},t)$ and $\psi^{\dagger}(\mathbf{r},t)$ are in the interaction representation. $G^{(0)}$ is the noninteracting Green's function (i.e., it is the Green function when $H' = 0$). Here we shall take the interaction to be given, in second-quantized form, by

$$H' = \frac{1}{2}\int \mathrm{d}^3 r_1 \mathrm{d}^3 r_2 \psi^{\dagger}(\mathbf{r}_1)\psi^{\dagger}(\mathbf{r}_2)U(\mathbf{r}_1-\mathbf{r}_2)\psi(\mathbf{r}_2)\psi(\mathbf{r}_1). \quad (12.55)$$

Now, introduce a function $V(x_1 - x_2) \equiv U(\mathbf{r}_1 - \mathbf{r}_2)\delta(t_1 - t_2)$ to write the first correction due to interaction as (let $x = \mathbf{r}, t$)

$$\begin{aligned}\delta G^{(1)}(x,x') = &-\tfrac{\mathrm{i}}{2\langle S(\infty,-\infty)\rangle}\int \mathrm{d}^4 x_1 \mathrm{d}^4 x_2 V(x_1-x_2)\\ &\times\langle \mathrm{T}\{\psi(x)\psi(x')\psi^{\dagger}(x_1)\psi^{\dagger}(x_2)\psi(x_2)\psi(x_1)\}\rangle_0.\end{aligned} \quad (12.56)$$

The time-ordered product of the six operators (3 ψ's and 3$\psi^{\dagger}$'s) can be written out by using (12.53)

$$\begin{aligned}&\langle \mathrm{T}\{\psi(x)\psi^{\dagger}(x')\psi^{\dagger}(x_1)\psi^{\dagger}(x_2)\psi(x_2)\psi(x_1)\}\rangle_0\\ &= \langle \mathrm{T}(\psi(x)\psi^{\dagger}(x_1))\rangle\langle \mathrm{T}(\psi^{\dagger}(x_2)\psi(x_2))\rangle\langle \mathrm{T}(\psi(x_1)\psi^{\dagger}(x'))\rangle\\ &\quad -\langle \mathrm{T}(\psi(x)\psi^{\dagger}(x_1))\rangle\langle \mathrm{T}(\psi^{\dagger}(x_2)\psi(x_1))\rangle\langle \mathrm{T}(\psi(x_2)\psi^{\dagger}(x'))\rangle \pm \text{ all other pairings.}\end{aligned} \quad (12.57)$$

But $\langle \mathrm{T}(\psi(x_i)\psi^{\dagger}(x_j))\rangle$ is proportional to $G^{(0)}(x_i,x_j)$. Therefore, the first term on the right hand side of (12.57) is proportional to

$$G^{(0)}(x,x_1)G^{(0)}(x_2,x_2)G^{(0)}(x_1,x'). \quad (12.58)$$

It is simpler to draw Feynman diagram for each of the possible pairings. There are six of them in $\delta G^{(1)}(x,x')$ because there are six ways to pair one $\psi^{\dagger}$ with one ψ. The diagrams are shown in Fig. 12.2. Note that x_1 and x_2 are always connected by an interaction line $V(x_1 - x_2)$. An electron propagates in from x and out to x'. At each x_1 and x_2 there must be one $G^{(0)}$ entering and one leaving.

In a standard book on many body theory, such as Fetter–Walecka(1971), Mahan(2000), and Abrikosov–Gorkov–Dzyaloshinskii(1963), one can find

1. Rules for constructing the Feynman diagrams for the n^{th} order correction and
2. Rules for writing down the analytic expression for $\delta G^{(n)}$ associated with each diagram.

Let us give one simple example of constructing diagrams. For the n^{th} order corrections, there are n interaction lines and $(2n+1)$ directed Green's functions, $G^{(0)}$'s. The rules for the nth-order corrections are as follows:

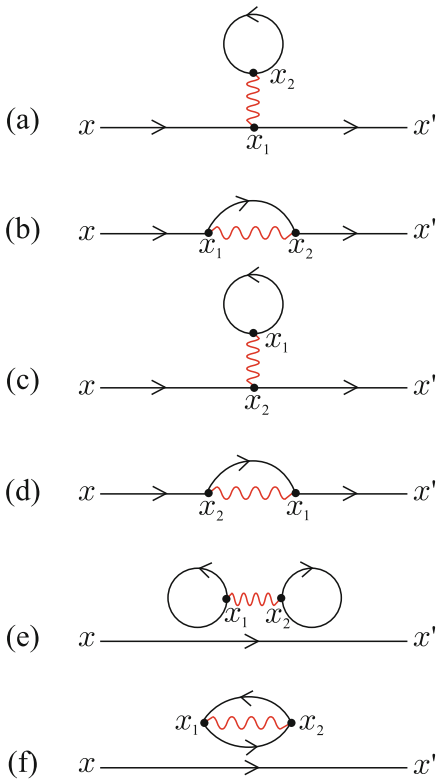


Fig. 12.2. Feynman diagrams in the first-order perturbation calculation

1. Form all *connected*, *topologically nonequivalent* diagrams containing $2n$ vertices and two external points. Two solid lines and one wavy line meet at each vertex.
2. With each solid line associate a Green's function $G^{(0)}(x, x')$ where x and x' are the coordinates of the initial and final points of the line.
3. With each wavy line associate $V(x - x') = U(\mathbf{r} - \mathbf{r}')\delta(t - t')$ for a wavy line connecting x and x'.
4. Integrate over the internal variables $\mathrm{d}^4 x_i = \mathrm{d}^3 r_i \, \mathrm{d}t_i$ for all vertex coordinates (and sum over all internal spin variables if spin is included).
5. Multiply by $i^n(-)^F$, where F is the number of closed Fermion loops.
6. Understand equal time $G^{(0)}$'s to mean, as $\delta \to 0^+$,

$$G^{(0)}(\mathbf{r}_1 t, \mathbf{r}_2 t) \to G^{(0)}(\mathbf{r}_1 t, \mathbf{r}_2 t + \delta).$$

The allowed diagrams contributing to $\delta G^{(2)}(x, x')$ are shown in Fig. 12.3.

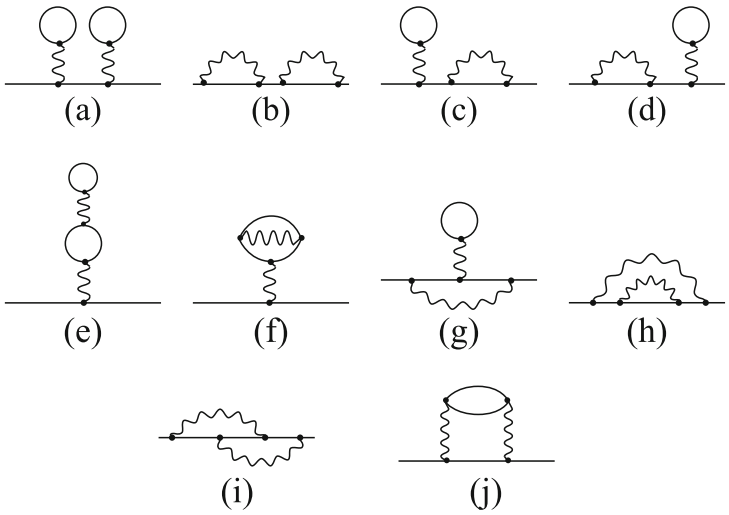


Fig. 12.3. Feynman diagrams in the second order perturbation calculation

12.3.3 Linked Clusters

In writing down the rules, we have only considered *linked* (or *connected*) *diagrams*, but diagrams (e) and (f) in Fig. 12.2 are unlinked diagrams. By this we mean that they fall into two separate pieces, one of which contains the coordinates x and x' of $G(x, x')$. It can be shown (see a standard many body text like Abrikosov–Gorkov–Dzyaloshinskii (1963).) that when the contributions from unlinked diagrams are included, they simply multiply the contribution from linked diagrams by a factor $\langle S(\infty, -\infty)\rangle$. Since this factor appears in denominator of $G_{\alpha\beta}(x, x')$ in (12.48), it simply cancels out. Furthermore, diagrams (a) and (c) in Fig. 12.2 are identical except for interchange of the dummy variables x_1 and x_2, and so too are (b) and (d). The rules for constructing diagrams for $\delta G^{(n)}(x, x')$ take this into account correctly and one can find the proof in standard many body texts mentioned above.

12.4 Dyson's Equations

If we look at the corrections to $G^{(0)}(x, x')$ we notice that for our linked cluster diagrams the corrections always begin with a $G^{(0)}(x, x_1)$, and this is followed by something called a *self energy* part. Look, for example, at the figures labeled (a) or (b) in Fig. 12.2 or (j) in Fig. 12.3. The final part of the diagram has another $G^{(0)}(x_n, x')$. Suppose we represent the general self energy by Σ. Then we can write

$$G = G^{(0)} + G^{(0)}\Sigma G. \tag{12.59}$$

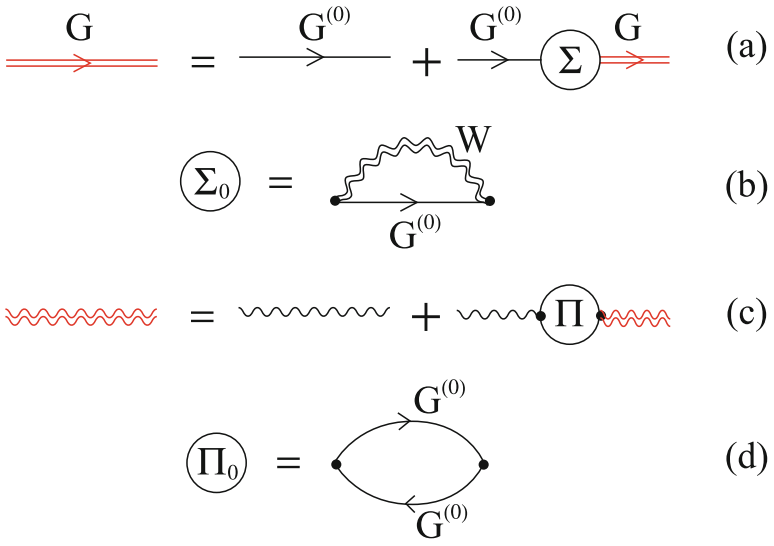


Fig. 12.4. Diagrammatic expressions of (**a**) Dyson equation $G = G^{(0)} + G^{(0)}\Sigma G$, (**b**) self energy Σ_0, (**c**) Dyson equation $W = V + V\Pi W$, (**d**) polarization part $\Pi_0 = G^{(0)}G^{(0)}$

This equation says that G is the sum of $G^{(0)}$ and $G^{(0)}$ followed by Σ which in turn can be followed by the exact G we are trying to determine. We can express (12.59) in diagrammatic terms as is shown in Fig. 12.4a. The simplest self-energy part that is of importance in the problem of electron interactions in a degenerate electron gas is Σ_0, where

$$\Sigma_0 = G^{(0)}W. \tag{12.60}$$

In diagrammatic terms this is expressed as shown in Fig. 12.4b, where the double wavy line is a screened interaction and we can write a Dyson equation for it by

$$W = V + V\Pi W. \tag{12.61}$$

The Π is called a *polarization part*; the simplest polarization part is

$$\Pi_0 = G^{(0)}G^{(0)}, \tag{12.62}$$

the diagrammatic expression of which is given in Fig. 12.4d. Of course, in (12.60) and (12.62) we could replace $G^{(0)}$ by the exact G to have a result that includes many terms of higher order. Approximating the self energy by the product of a Green's function G and an effective interaction W is often referred to as the **GW** *approximation* to the self-energy. The simplest GW approximation is the *random phase approximation* (RPA). In the RPA, the G is replaced by $G^{(0)}$ and W is the solution to (12.61) with (12.62) used for

the polarization part. This RPA approximation for W is exactly equivalent to $\frac{V(q)}{\varepsilon(q,\omega)}$, where $\varepsilon(q,\omega)$ is the Lindhard dielectric function. The key role of the electron self-energy in studying *electron–electron interactions* in a degenerate electron gas was initially emphasized by Quinn and Ferrell.[1] In their paper, the simplest GW approximation to Σ was used. $G^{(0)}$ was used for the Green's function and $\frac{V(q)}{\varepsilon(q,\omega)}$, the RPA screened interaction (equivalent to Lindhard screened interaction) was used for W.

12.5 Green's Function Approach to the Electron–Phonon Interaction

In this section we apply the Green's function formalism to the electron–phonon interaction. The Hamiltonian H is divided into three parts:

$$H = H_{\mathrm{e}} + H_{\mathrm{N}} + H_{\mathrm{I}}, \tag{12.63}$$

where

$$H_{\mathrm{e}} = \sum_i \left[\frac{p_i^2}{2m} + \sum_l U(\mathbf{r}_i - \mathbf{R}_l^0) \right], \tag{12.64}$$

$$H_{\mathrm{N}} = \sum_l \left[\frac{P_l^2}{2M} + \sum_{l>m} V(\mathbf{R}_l - \mathbf{R}_m) \right], \tag{12.65}$$

and

$$H_{\mathrm{I}} = \sum_{i>j} \frac{e^2}{r_{ij}} - \sum_{i,l} \mathbf{u}_l \cdot \nabla \mathbf{U}(\mathbf{r}_i - \mathbf{R}_l^0). \tag{12.66}$$

Here, $U(\mathbf{r}_i - \mathbf{R}_l)$ and $V(\mathbf{R}_l - \mathbf{R}_m)$ represent the interaction between an electron at $\mathbf{r}_i$ and an ion at $\mathbf{R}_l$ and the interaction potential of the ions with each other, respectively. Let us write $\mathbf{R}_l = \mathbf{R}_l^0 + \mathbf{u}_l$ for an ion where $\mathbf{R}_l^0$ is the equilibrium position of the ion and $\mathbf{u}_l$ is its atomic displacement. The electronic Hamiltonian H_{e} is simply a sum of one-electron operators, whose eigenfunctions and eigenvalues are the object of considerable investigation for energy band theorists. To keep the calculations simple, we shall assume that the effect of periodic potential can be approximated to sufficient accuracy for our purpose by the introduction of an effective mass. The nuclear or ionic Hamiltonian H_{N} has already been analyzed in normal modes in earlier chapters. It should be pointed out that the normal modes of (12.65) are not the usual sound waves. The reason for this is that $V(\mathbf{R}_l - \mathbf{R}_m)$ is a "bare" ion–ion interaction, for a pair of ions sitting in a uniform background of negative charge, not the true interaction which is screened by the conduction electrons. We can express (12.64)–(12.66) in the usual second quantized notation as

[1] J.J. Quinn, R.A. Ferrell, Phys. Rev **112**, 812 (1958).

$$H_{\mathrm{e}} = \sum_k \frac{\hbar^2 k^2}{2m^*} c_k^\dagger c_k, \tag{12.67}$$

$$H_{\mathrm{N}} = \sum_\alpha \hbar\omega_\alpha \left(b_\alpha^\dagger b_\alpha + \frac{1}{2} \right), \tag{12.68}$$

and

$$H_{\mathrm{I}} = \sum_{k,k',q} \frac{4\pi e^2}{\Omega q^2} c_{k+q}^\dagger c_{k'-q}^\dagger c_{k'} c_k + \sum_{k,\alpha,\mathbf{G}} \gamma(\alpha, \mathbf{G})(b_\alpha - b_{-\alpha}^\dagger) c_{\mathbf{k}+\mathbf{q}+\mathbf{G}}^\dagger c_k. \tag{12.69}$$

The c_k and b_α are the destruction operators for an electron in state $\mathbf{k}$ and a phonon in state $\alpha = (\mathbf{q}, \mu)$, respectively.[2] The creation and annihilation operators satisfy the usual commutation (phonon operators) or anticommutation (electron operators) rules. The coupling constant $\frac{4\pi e^2}{\Omega q^2}$ is simply the Fourier transform of the Coulomb interation, and $\gamma(\alpha, \mathbf{G})$ is given by

$$\gamma(\alpha, \mathbf{G}) = -\mathrm{i}(\mathbf{q} + \mathbf{G}) \underline{\varepsilon}_\alpha \left(\frac{\hbar N}{2M\omega_\alpha} \right)^{1/2} U(\mathbf{q} + \mathbf{G}), \tag{12.70}$$

where $U(\mathbf{q} + \mathbf{G})$ is the Fourier transform of $U(\mathbf{r} - \mathbf{R})$. For simplicity we shall limit ourselves to normal processes (*i.e.*, $\mathbf{G} = 0$), and take $U(\mathbf{r} - \mathbf{R})$ as the Coulomb interaction $-\frac{Ze^2}{|\mathbf{r}-\mathbf{R}|}$ between an electron of charge $-e$ and an ion of charge Ze. With these simplifications $\gamma(\alpha, \mathbf{G})$ reduces to

$$\gamma(q) = \mathrm{i} \frac{4\pi Z e^2}{q} \left(\frac{\hbar N}{2M\omega_q} \right)^{1/2} \tag{12.71}$$

for the interaction of electrons with a longitudinal wave, and zero for interaction with a transverse wave. Furthermore, when we make these assumptions, the longitudinal modes of the "bare" ions all have the frequency

$$\omega_{\mathbf{q}} = \Omega_{\mathrm{p}}, \tag{12.72}$$

where $\Omega_{\mathrm{p}} = \left(\frac{4\pi Z^2 e^2 N}{M} \right)^{1/2}$ is the plasma frequency of the ions.

We want to treat H_{I} as a perturbation. The brute-force application of perturbation theory is plagued by divergence difficulties. The divergence arise from the long range of the Coulomb interaction, and are reflected in the behavior of the coupling constants as q tends to zero. We know that in the solid, the Coulomb field of a given electron is screened because of the response of all the other electrons in the medium. This screening can be taken into account by

[2] We should really be careful to include the spin state in describing the electrons. We will omit the spin index for simplicity of notation, but the state k should actually be understood to represent a given wave vector and spin: $k \equiv (\mathbf{k}, \sigma)$.

pertubation theory, but it requires summing certain classes of terms to infinite order. This is not very difficult to do if one makes use of Green's functions and Feynman diagrams. Before we discuss these, we would like to give a very quanlitative sketch of why a straightforward perturbation approach must be summed to infinite order.

Suppose we introduce a static positive point charge in a degenerate electron gas. In vacuum the point charge would set up a potential Φ_0. In the electron gas the point charge attracts electrons, and the electron cloud around it contributes to the potential set up in the medium. Suppose that we can define a *polarizability factor* α such that a potential Φ acting on the electron gas will distort the electron distribution in such a way that the potential set up by the distortion is $\alpha\Phi$. We can then apply a perturbation approach to the potential Φ_0. Φ_0 distorts the electron gas: the distortion sets up a potential $\Phi_1 = \alpha\Phi_0$. But Φ_1 further distorts the medium and this further distortion sets up a potential $\Phi_2 = \alpha\Phi_1$, *etc.* such that $\Phi_{n+1} = \alpha\Phi_n$. The total potential Φ set up by the point charge in the electron gas is

$$\begin{aligned}\Phi = \Phi_0 + \Phi_1 + \Phi_2 + \cdots &= \Phi_0(1 + \alpha + \alpha^2 + \cdots) \\ &= \Phi_0(1-\alpha)^{-1}. \end{aligned} \tag{12.73}$$

We see that we must sum the straightforward perturbation theory to infinite order. It is usually much simpler to apply "self-consistent" perturbation theory. In this approach one simply says that Φ_0 will ultimately set up some self-consistent field Φ. Now, the field acting on the electron gas and polarizing it is not Φ_0 but the full self-consistent field Φ. Therefore, the polarization contribution to the full potential should be $\alpha\Phi$: this gives

$$\Phi = \Phi_0 + \alpha\Phi, \tag{12.74}$$

which is the same result obtained by summing the infinite set of perturbation contributions in (12.73).

We want to use some simple Feynman propagation functions or Green's functions, so we will give a very quick definition of what we must know to use them. If we have the Schrödinger equation

$$\mathrm{i}\hbar\frac{\partial\Psi}{\partial t} = H\Psi, \tag{12.75}$$

and we know $\Psi(t_1)$, we can determine Ψ at a later time from the equation

$$\Psi(x_2, t_2) = \int d^3x_1 G_0(x_2, t_2; x_1, t_1)\Psi(x_1, t_1). \tag{12.76}$$

By substitution, one can show that G_0 satisfies the differential equation

$$\left[\mathrm{i}\hbar\frac{\partial}{\partial t_2} - H(x_2)\right] G_0(2,1) = \mathrm{i}\hbar\delta(t_2 - t_1)\delta(x_2 - x_1), \tag{12.77}$$

where $(2,1)$ denotes $(x_2,t_2;x_1,t_1)$. One can easily show that $G_0(2,1)$ can be expressed in terms of the stationary states of H. That is, if

$$Hu_n = E_n u_n, \tag{12.78}$$

then $G_0(2,1)$ can be shown to be

$$\begin{aligned} G_0(2,1) &= \sum_n u_n(x_2)u_n^*(x_1)\mathrm{e}^{-\mathrm{i}E_n t_{21}/\hbar}, \text{ if } t_{21}>0, \\ &= 0 \qquad\qquad\qquad \text{otherwise .} \end{aligned} \tag{12.79}$$

If we are considering a system of many Fermions, we can take into account the exclusion principle in a very simple way. We simply subtract from (12.79) the summation over all states of energy less than the Fermi energy E_F.

$$\begin{aligned} G_0(2,1) &= \sum_{E_n>E_\mathrm{F}} u_n(x_2)u_n^*(x_1)\mathrm{e}^{-\mathrm{i}E_n t_{21}/\hbar}, \text{ if } t_{21}>0, \\ &= -\sum_{E_n<E_\mathrm{F}} u_n(x_2)u_n^*(x_1)\mathrm{e}^{-\mathrm{i}E_n t_{21}/\hbar}, \text{ if } t_{21}<0. \end{aligned} \tag{12.80}$$

We always represent a Fermion propagator by a directed solid line. A negative (relative to the last filled state E_F) energy Fermion propagates backward in time. This corresponds to the propagation of a hole in a normally filled state. For free electrons the functions $u_n(x)$ are plane waves. We are often interested in $G_0(q,\omega)$, the Fourier transform of $G_0(2,1)$:

$$G_0(2,1) = \int \frac{\mathrm{d}^3q\mathrm{d}\omega}{(2\pi)^4} G_0(q,\omega)\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{x}_{21}-\mathrm{i}\omega t_{21}}. \tag{12.81}$$

The single particle propagator $G_0(q,\omega)$ for a system of free electrons is

$$G_0(q,\omega) = \frac{\mathrm{i}}{\omega - E(q)(1-\mathrm{i}\delta)}, \tag{12.82}$$

where $E(q) = \frac{\hbar^2}{2m}(q^2-k_\mathrm{F}^2)$ is the energy measured relative to the Fermi energy and takes on both positive and negative values. k_F is the Fermi wave number, and δ is a positive infinitesimal.

In the language of second quantization $G_0(2,1)$ can also be expressed as the ground state expectation value of the time ordered product of two electron field operators

$$G_0(2,1) = \langle \mathrm{GS} \mid T\{\Psi(2)\Psi^\dagger(1)\} \mid \mathrm{GS}\rangle. \tag{12.83}$$

In this expression, $\Psi(2) = \Psi(x_2,t_2)$ is the electron field operator and $\Psi^\dagger(2)$ is its conjugate. These operators satisfy the usual Fermion anticommutation relations. T is the chronological operator. It should be pointed out that people often define $G_0(2,1)$ with an additional factor of i on the right hand side of

(12.83). This arbitrariness in defining the propagation functions is compensated for by slight differences in the rules for calculating the amplitudes of the Feynman diagrams which appear in perturbation theory.

We can also define a propagation function for the instantaneous Coulomb interaction $\frac{e^2}{r_{21}}\delta(t_{21})$ between electrons at two points in space time. We shall use i times the Fourier transform of $\frac{e^2}{r_{21}}\delta(t_{21})$ as the bare Coulomb propagator $V(q,\omega)$

$$V(q,\omega) = \frac{4\pi e^2 \mathrm{i}}{\Omega q^2}. \tag{12.84}$$

If we define the phonon field operator $\Phi(x)$ by the equation

$$\Phi(x) = \sum_{\mathbf{q}} \gamma(\mathbf{q}) \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{x}} (b^\dagger_{\mathbf{q}} - b_{-\mathbf{q}}) \tag{12.85}$$

then we can define the space time representation for the phonon propagator in the usual way

$$P_0(2,1) = \mathrm{i}\langle \mathrm{GS} \mid T\{\Phi(x_2,t_2)\Phi^\dagger(x_1,t_1)\} \mid \mathrm{GS}\rangle \tag{12.86}$$

where $\Phi(x_2,t_2) = \mathrm{e}^{-\mathrm{i}Ht_2}\Phi(x_2)\mathrm{e}^{\mathrm{i}Ht_2}$. The Fourier transform of (12.86) is $P_0(q,\omega)$, the wave vector frequency space representation of P_0. For the phonon system of (12.71) and (12.72), $P_0(q,\omega)$ is given by

$$P_0(q,\omega) = \frac{2\mathrm{i}\Omega_\mathrm{p}\gamma^2(q)}{\omega^2 - \Omega_\mathrm{p}^2}. \tag{12.87}$$

It is quite convenient to use Feynman diagrams to keep track of the various terms in perturbation theory. The rules for constructing diagrams are quite simple. Each electron in an excited state is represented by a solid line directed upward. Each hole in a normally filled state is represented by a solid line directed downward. The instantaneous Coulomb interaction is represented by a horizontal dotted line connecting the two particles undergoing a virtual scattering, and propagation of a phonon is represented by a wavy line.

Consider the scattering of two electrons. In vacuum they can scatter by the exchange of one virtual photon (Coulomb line) in only one way, which is shown in Fig. 12.5. Now consider the Coulomb interaction in the medium. The set of

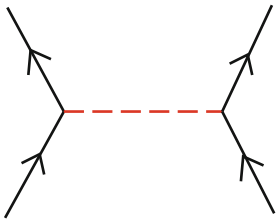


Fig. 12.5. Diagrammatic expression of the exchange of virtual photon

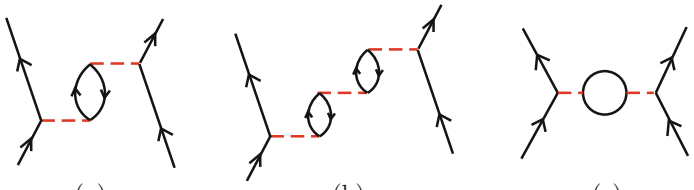


Fig. 12.6. Diagrammatic expressions of representative polarization parts in pair approximation

diagrams, of which Fig. 12.6a and b are representative, are additional processes which can not occur in the absence of the polarizable medium. In Fig. 12.6c, the circle represents any possible part of a diagram which is connected to the remainder by two Coulomb interaction lines only. All such parts of a general diagram are called *polarizable parts*, because they obviously represent the response or screening of the polarizable medium. The effective Coulomb interaction between two particles should be the sum of all the possible polarization parts (the bare intraction can be thought of as the zeroth order polarization part). Actually we can not sum all the possible polarization parts, but we can sum the class of which Fig. 12.6a and b are representative, that is, the chain of bubbles. The approximation of replacing the effective interaction by the sum of all bubble graph is called the *pair approximation*. Before looking at the sum we will write down the rules for calculating the amplitude associated with a given Feynman diagram which appears in perturbation theory. The amplitude for a given diagram contains a product of

(i) a propagation function $G_0(k,\omega)$ for each internal electron–hole line of wave vector $\mathbf{k}$ and frequency ω.
(ii) a propagation function $P_0(k,\omega)$ for each phonon line of wave vector $\mathbf{q}$ and frequency ω.
(iii) a propagation function $V(q,\omega)$ for each Coulomb line of wave vector $\mathbf{q}$.
(iv) a factor (-1) for each closed loop.
(v) $(-\mathrm{i}/\hbar)^n$ for the nth-order term in perturbation theory.
(vi) delta functions conserving energy, momentum, and spin at each vertex.
(vii) Finally, we must integrate over the wave vectors and frequencies of all internal lines.

The set of diagrams we would like to sum in order to obtain the effective Coulomb propagator $W(\mathbf{q},\omega)$ can easily be seen to be the solution of the equation given pictorially by Fig. 12.7. This equation can be written

$$W(q,\omega) = V(q,\omega) - V(q,\omega)Q_0(q,\omega)W(q,\omega), \tag{12.88}$$

Fig. 12.7. Diagrammatic expression of Dyson equation for the effective Coulomb propagator $W = V + V\Pi W$

where

$$Q_0(q,\omega) = 2(-1)\int \frac{\mathrm{d}^3k_1\mathrm{d}\omega_1\mathrm{d}^3k_2\mathrm{d}\omega_2}{(2\pi)^8} G_0(k_1,\omega_1)G_0(k_2,\omega_2)\delta(\omega_1-\omega_2+\omega) \times\delta(\mathbf{k}_1-\mathbf{k}_2+\mathbf{q}) \tag{12.89}$$

is the propagation function for the electron–hole pair. The factor of two is introduced to account for the two possible spin orientations. Using the electron propagation functions defined by (12.82) and integrating gives

$$Q_0(q,\omega) = \frac{-2\mathrm{i}}{(2\pi)^3}\int_{k<k_\mathrm{F}} \mathrm{d}^3k \left[\frac{1}{E(\mathbf{k}+\mathbf{q})-E(\mathbf{k})+\omega} + \frac{1}{E(\mathbf{k}+\mathbf{q})-E(\mathbf{k})-\omega}\right]. \tag{12.90}$$

The solution of (12.88) is simply

$$W = \frac{V}{1+VQ_0} \tag{12.91}$$

and using (12.90) one can easily see that $1+VQ_0$ is just the Lindhard dielectric function $\varepsilon(q,\omega)$. This dielectric function is discussed at some length in the previous chapter, and the reader is referred to Lindhard's paper[3] for a complete treatment. For our purposes we must note two things: first $\varepsilon(q,\omega)$ is complex, the imaginary part being proportional to the number of electrons which can be excited to an unoccupied state by addition of a momentum $\hbar\mathbf{q}$ whose energy change is equal to $\hbar\omega$. The second point is that for zero frequency $\varepsilon(q,0)$ is given by

$$\varepsilon(q) = 1 + F\left(\frac{q}{2k_\mathrm{F}}\right)\frac{k_\mathrm{s}^2}{q^2}, \tag{12.92}$$

where k_s is the Fermi–Thomas screening parameter and k_F is the Fermi wave number. The function $F\left(\frac{q}{2k_\mathrm{F}}\right)$ is the function sketched in Fig. 11.12. $F(x)$ is equal to unity for x equal to zero, approaches zero as x approaches infinity, and has logarithmic singularity in slope at $x = 1$.

Now, let us return to our "model solid" which contains longitudinal phonons as well as electrons. Two electrons can scatter via the virtual exchange of phonons. In fact, anywhere a Coulomb interaction line has

[3] J. Lindhard, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **28**, 8 (1954); ibid., **27**, 15 (1953).

appeared previously, a phonon line may equally well appear. If we call $D_0(\mathbf{q},\omega)$ the sum of $P_0(\mathbf{q},\omega)$ and $V(\mathbf{q},\omega)$ we can just replace V and W by D_0 and D in Eq.(12.88). $D(\mathbf{q},\omega)$ then represents the propagation function for the total interaction, i.e., the sum of the Coulomb interaction and the interaction due to virtual exchange of phonons. It is apparent that

$$D(\mathbf{q},\omega) = \frac{D_0(\mathbf{q},\omega)}{1 + D_0(\mathbf{q},\omega)Q_0(\mathbf{q},\omega)}. \tag{12.93}$$

By substituting into (12.92) the expressions for the propagation functions and using the fact that $1 + VQ_0 = \varepsilon(q,\omega)$, one can obtain

$$D(\mathbf{q},\omega) = \frac{4\pi e^2 \mathrm{i}}{q^2\left[\varepsilon(q,\omega) - \Omega_\mathrm{p}^2/\omega^2\right]}. \tag{12.94}$$

The propagation function D as a pole at

$$\omega^2 = \frac{\Omega_\mathrm{p}^2}{\varepsilon(q,\omega)}. \tag{12.95}$$

The solutions of this equation are our "renormalized" phonons. From the long wavelength, zero frequency dielectric constant we get the approximate solution

$$\omega_q = \frac{\Omega_\mathrm{p}}{k_\mathrm{s}} q. \tag{12.96}$$

For most metals, $\frac{\Omega_\mathrm{p}}{k_\mathrm{s}}$ is within about 15–20% of the velocity of longitudinal sound waves. If we look at the derivative of ω_q^2 with respect to q we see a logarithmic singularity at $q = 2k_\mathrm{F}$. This is responsible for the Kohn effect, which has been observed by neutron scattering. If we take account of the imaginary as well as the real part of the dielectric constant, the solution of (12.95) has both real and imaginary part. If we write $\omega = \omega_1 + \mathrm{i}\omega_2$, then ω_2 turns out to be

$$\omega_2 \approx \frac{\pi}{4}\frac{k_\mathrm{s}^2}{k_\mathrm{s}^2 + q^2}\frac{c_\mathrm{s}}{v_\mathrm{F}}\omega_1, \tag{12.97}$$

where c_s is the velocity of sound and v_F the Fermi velocity. The coefficient of attenuation of the sound wave (due to excitation of conduction electrons) is simply $\frac{\omega_2}{c_\mathrm{s}}$. This result agrees with the more standard calculations of the attenuation coefficient.

Finally, if we wish to define the effective interaction between electrons due to virtual exchange of phonons, or the *effective phonon propagator*, we can simply subtract from $D(q,\omega)$ that part which contains no phonons, namely $W(q,\omega)$. If we call the resultant effective phonon propagator $P(q,\omega)$, we obtain

$$P(q,\omega) = \frac{2\mathrm{i}\omega_q \mid \gamma_q^{\mathrm{eff}} \mid^2}{\omega^2 - \omega_q^2}, \tag{12.98}$$

where ω_q is given by (12.96) and

$$\mid \gamma_q^{\text{eff}} \mid = \frac{4\pi Z e^2}{q\varepsilon(q,\omega_q)} \left(\frac{N}{2M\omega_q}\right)^{1/2}. \tag{12.99}$$

Replacing $\varepsilon(q,\omega_q)$ by its long wavelength, zero frequency limit $\left[1+\frac{k_s^2}{q^2}\right]$, reduces (12.99) to the result of Bardeen and Pines[4] for the effective electron–electron interaction.

12.6 Electron Self Energy

The Dyson equation for the Green's function can be written

$$G(k,\omega) = G^{(0)}(k,\omega) + G^{(0)}(k,\omega)\Sigma(k,\omega)G(k,\omega). \tag{12.100}$$

Dividing by $GG^{(0)}$ gives

$$\Sigma(k,\omega) = [G^{(0)}(k,\omega)]^{-1} - [G(k,\omega)]^{-1}. \tag{12.101}$$

The energy of a quasiparticle can be written

$$E_p = \varepsilon_p + \Sigma(p,\omega) \mid_{\omega=E_p}. \tag{12.102}$$

E_p and ε_p, the kinetic energy, are usually measured relative to E_{F}, the Fermi energy. Knowing how $\Sigma(p,\omega)$ depends on p, ω, and r_{s} allows one to calculate almost all the properties of an electron gas that are of interest. Some results of interest are worth mentioning.

(1) $\Sigma(p,E_p)$ has both a real and an imaginary part.

$$\Sigma(p,E_p) = \Sigma_1(p,E_p) + \mathrm{i}\Sigma_2(p,E_p). \tag{12.103}$$

The imaginary part is related to the lifetime of the quasiparticle excitation.

(2) The spectral function $A(p,\omega)$ is defined by

$$A(p,\omega) = \frac{-2\Sigma_2(p,\omega)}{[\omega - \varepsilon_p - \Sigma_1(p,\omega)]^2 + [\Sigma_2(p,\omega)]^2}. \tag{12.104}$$

For noninteracting electrons $A(p,\omega)$ has a δ function singularity at $\omega = \varepsilon_p$, the energy of the excitation. The δ function is broadened by $\Sigma_2(p,\omega)$. If $A(p,\omega)$ has a pole in a region where $\Sigma_2(p,\omega)$ is zero, the strength of the pole is decreased by a *renormalization factor* $Z(p)$.

$$A(p,\omega) = 2\pi Z(p)\delta\left(\omega - \varepsilon_p - \Sigma_1(p,\omega)\right) \tag{12.105}$$

[4] John Bardeen, David Pines, Phys. Rev. **99**, 1140 (1955).

and

$$Z(p) = \left[\frac{1}{1 - \frac{\partial}{\partial\omega}\Sigma_1(p,\omega)}\right]_{\omega=E_p}. \tag{12.106}$$

(3) The quasiparticle excitations at the Fermi surface have an effective mass m^* given by

$$\frac{m}{m^*} = \left[\frac{1 + \frac{\partial\Sigma_1(k,\omega)}{\partial\varepsilon_k}}{1 - \frac{\partial\Sigma_1(k,\omega)}{\partial\omega}}\right]_{\substack{k=k_F \\ \omega=0}}. \tag{12.107}$$

(4) Properties like the spin susceptibility, the specific heat, the compressibility, and the ground state energy can be evaluated from a knowledge of $\Sigma(k,\omega)$. But, we do not have time to go through these in any detail.
(5) The self-energy approach leads very naturally to an understanding of the Landau theory of a Fermi liquid. We will describe a very brief and intuitive explanation of the theory.

12.7 Quasiparticle Interactions and Fermi Liquid Theory

Instead of describing the interacting ground state and excited states of an electron gas, we can think of simply describing how many quasiparticles are present in some excited state. Let us start by noting that if we begin with a filled Fermi sphere of noninteracting electrons (i.e., the Sommerfeld model) and adiabatically turn on the electron–electron interaction, we will generate at $t = 0$ the exact interacting ground state. Now consider the noninteracting state described by a filled Fermi sphere plus one electron of momentum p outside the Fermi sphere (or one hole of momentum p inside the Fermi sphere). When interactions are adiabatically turned on, this is a single quasiparticle state. The energy of this *quasielectron* (or *quasihole*) is written by

$$E_{\mathbf{p}} = \varepsilon_{\mathbf{p}} + \Sigma(\mathbf{p}, E_{\mathbf{p}}). \tag{12.108}$$

If $E_{\mathbf{p}}$ is much larger than $\Sigma_2(\mathbf{p}, E_{\mathbf{p}})$, then the quasiparticles have long lifetimes. It is much simpler to describe a state by saying how many quasielectrons and quasiholes are present. Then, the energy of the state can be written as

$$E = E_0 + \sum_{\mathbf{p}\sigma} \delta n_{\mathbf{p}\sigma} E_{\mathbf{p}\sigma} + \frac{1}{2} \sum_{\substack{\mathbf{p},\,\mathbf{p}' \\ \sigma,\,\sigma'}} f_{\sigma\sigma'}(\mathbf{p}, \mathbf{p}') \delta n_{n\sigma} \delta n_{\mathbf{p}'\sigma'}. \tag{12.109}$$

The first term on the right is the ground state energy, the second is the quasiparticle energy $E_{\mathbf{p}\sigma}$ multiplied by the quasiparticle distribution function, and the third represents the interactions of the quasiparticles with one another. $\Sigma(\mathbf{p}, E_{\mathbf{p}})$ represents the interaction of a quasiparticle of momentum $\mathbf{p}$ with the ground state of the interacting electron gas. But if the electron gas is

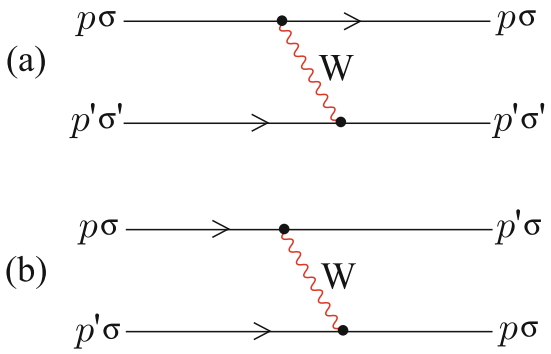


Fig. 12.8. Diagrammatic representation of quasiparticle scattering of (**a**) zero momentum transfer and (**b**) of finite momentum transfer

not in its ground state, there are quasielectrons and quasiholes present that change the energy of the quasiparticle of momentum $\mathbf{p}$. We can get a simple picture of the *Fermi liquid* interaction between quasiparticles by considering the Feynman diagrams that describe the scattering processes that take a pair of quasiparticles in states $(\mathbf{p}, \sigma)$ and $(\mathbf{p}', \sigma')$ from this initial state to an equivalent final state. These processes are represented in diagrammatic terms in Fig. 12.8a and b. Here, the interaction W (denoted by wavy lines) will be taken as the RPA screened interaction. In the first term (a) $W(\mathbf{q}, \omega)$ corresponds to zero momentum transfer since $\mathbf{p} \to \mathbf{p}$ and $\mathbf{p}' \to \mathbf{p}'$. This term is exactly zero since the Coulomb interaction is cancelled by the interaction with the uniform background of positive charge at $\mathbf{q} = 0$. The second term (b) gives the same final state as the initial state only if $\sigma = \sigma'$. Then $W(\mathbf{q}, \omega)$ is $W(\mathbf{p} - \mathbf{p}', 0)$ since the momentum transfer is $\mathbf{p} - \mathbf{p}'$ and there is no change in energy.

Of course, higher order processes in the effective interaction could be important, but we will ignore them to get the simplified picture. We take Landau's $f_{\sigma\sigma'}(\mathbf{p}, \mathbf{p}')$ to be equal to

$$f_{\sigma\sigma'}(\mathbf{p}, \mathbf{p}') = \begin{cases} \frac{4\pi e^2}{(\mathbf{p}-\mathbf{p}')^2 \varepsilon(\mathbf{p}-\mathbf{p}', 0)} & \text{if } \sigma = \sigma' \\ 0 & \text{if } \sigma \neq \sigma'. \end{cases} \tag{12.110}$$

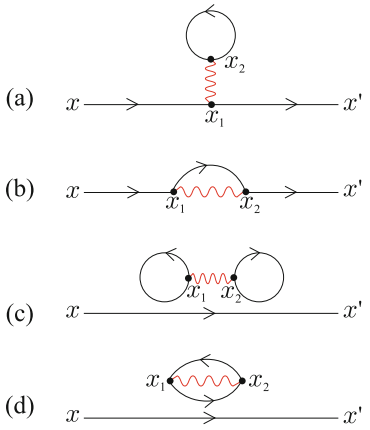
With this simple approximation a rather good estimate of m^* (and hence of the electronic specific heat) can be obtained. Results for the spin susceptibility are not quite as good, and the estimate of the interaction contribution to the compressibility is poor. One important effect that is omitted is the effect of spin fluctuations (in addition to charge density fluctuations) and another is the local field corrections to the RPA. In order to get a more thorough understanding of these ideas, one needs to read advanced texts on many body theory.

Problems

12.1. Show explicitly that

$$\int_{t_0}^{t} \mathrm{d}t_1 \int_{t_0}^{t_1} \mathrm{d}t_2 \int_{t_0}^{t_3} \mathrm{d}t_3 \; H_{\mathrm{I}}(t_1)H_{\mathrm{I}}(t_2)H_{\mathrm{I}}(t_3)$$
$$= \frac{1}{3!} \int_{t_0}^{t} \mathrm{d}t_1 \int_{t_0}^{t} \mathrm{d}t_2 \int_{t_0}^{t} \mathrm{d}t_3 \mathrm{T}\{H_{\mathrm{I}}(t_1)H_{\mathrm{I}}(t_2)H_{\mathrm{I}}(t_3)\}.$$

12.2. The complete first-order contributions to $G(x, x')$ are shown in the figure.



(a) Write each term $\delta G^{(1)}(x, y)$ out in terms of noninteraction two particle Green's function $G^{(0)}(x, y)$ and the interaction $V(x_1 - x_2) \equiv U(\mathbf{r}_1 - \mathbf{r}_2)\delta(t_1 - t_2)$. Here, $x = (\mathbf{r}, t)$ and one may omit the spin to simplify the notation.

(b) Let us introduce the Fourier transform $G_{\alpha\beta}(k)$ of the $G_{\alpha\beta}(x, y)$ as follows:

$$G_{\alpha\beta}(x, y) = \frac{1}{(2\pi)^4} \int \mathrm{d}^4 k \, \mathrm{e}^{\mathrm{i}k\cdot(x-y)} G_{\alpha\beta}(k)$$
$$G^{(0)}_{\alpha\beta}(x, y) = \frac{1}{(2\pi)^4} \int \mathrm{d}^4 k \, \mathrm{e}^{\mathrm{i}k\cdot(x-y)} G^{(0)}_{\alpha\beta}(k),$$

where $k = (\mathbf{k}, \omega)$, $\mathrm{d}^4 k \equiv \mathrm{d}^3 k \, \mathrm{d}\omega$, and $k \cdot x \equiv \mathbf{k} \cdot \mathbf{x} - \omega t$. In addition, for the interaction given by $V(x_1 - x_2) \equiv U(\mathbf{r}_1 - \mathbf{r}_2)\delta(t_1 - t_2)$ we can write

$$V_(x, x')_{\alpha\alpha',\beta\beta'} = \frac{1}{(2\pi)^4} \int \mathrm{d}^4 k \, \mathrm{e}^{\mathrm{i}k\cdot(x-x')} V(k)_{\alpha\alpha',\beta\beta'}$$
$$= \frac{1}{(2\pi)^3} \int \mathrm{d}^3 x \, \mathrm{e}^{\mathrm{i}k\cdot(\mathbf{x}-\mathbf{x}')} U(\mathbf{k})_{\alpha\alpha',\beta\beta'} \delta(t - t'),$$

where $V(k)_{\alpha\alpha',\beta\beta'} = U(\mathbf{k})_{\alpha\alpha',\beta\beta'} = \frac{1}{(2\pi)^3} \int \mathrm{d}^3 x \, \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}} U(\mathbf{k})_{\alpha\alpha',\beta\beta'}$ is the spatial Fourier transform of the interparticle potential. Express each

term obtained in part (a) in terms of $G^{(0)}_{\alpha\beta}(k)$ and $V(k)$ in the momentum space.

12.3. By definition the noninteracting fermion Green's function is given by

$$G^{(0)}_{\alpha\beta}(\mathbf{x}t, \mathbf{x}'t') = -\mathrm{i}\langle \Phi | \mathrm{T}\{\psi_{\mathrm{I}\alpha}(\mathbf{x}t)\psi^{\dagger}_{\mathrm{I}\beta}(\mathbf{x}'t')\}\Phi\rangle,$$

the noninteracting ground state vector is taken to be normalized. Show that

$$G^{(0)}_{\alpha\beta}(\mathbf{k}, \omega) = \delta_{\alpha\beta}\left[\frac{\theta(k - k_{\mathrm{F}})}{\omega - \hbar^{-1}\varepsilon_k + \mathrm{i}\eta} + \frac{\theta(k_{\mathrm{F}} - k)}{\omega - \hbar^{-1}\varepsilon_k - \mathrm{i}\eta}\right].$$

12.4. Let us define the phonon field operator $\Phi(x)$ by

$$\Phi(x) = \sum_{\mathbf{q}} \gamma(\mathbf{q})\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{x}}(b^{\dagger}_{\mathbf{q}} - b_{-\mathbf{q}}),$$

where $\gamma(q) = \mathrm{i}\frac{4\pi Ze^2}{q}\left(\frac{\hbar N}{2M\omega_q}\right)^{1/2}$. Then we can define the phonon propagator by $P_0(2,1) = \mathrm{i}\langle \mathrm{GS} \mid T\{\Phi(x_2, t_2)\Phi^{\dagger}(x_1, t_1)\} \mid \mathrm{GS}\rangle$ where $\Phi(x_2, t_2) = \mathrm{e}^{-\mathrm{i}Ht_2}\Phi(x_2)\mathrm{e}^{\mathrm{i}Ht_2}$.

(a) Take the Fourier transform of $P_0(2,1)$ to obtain $P_0(q, \omega)$, the wavevector frequency space representation of $P_0(2,1)$.

(b) Show that $P_0(q, \omega)$ can be written as

$$P_0(q, \omega) = \frac{2\mathrm{i}\Omega_{\mathrm{p}}\gamma^2(q)}{\omega^2 - \Omega^2_{\mathrm{p}}},$$

where $\Omega_{\mathrm{p}} = \left(\frac{4\pi Z^2e^2N}{M}\right)^{1/2}$ is the plasma frequency of the ions.

12.5. Let us consider a Dyson equation given by

$$W(q, \omega) = V(q, \omega) - V(q, \omega)Q_0(q, \omega)W(q, \omega),$$

where

$$Q_0(q, \omega) = 2(-1)\int \frac{\mathrm{d}^3k_1\mathrm{d}\omega_1\mathrm{d}^3k_2\mathrm{d}\omega_2}{(2\pi)^8}$$
$$G_0(k_1, \omega_1)G_0(k_2, \omega_2)\delta(\omega_1 - \omega_2 + \omega)\delta(\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{q})$$

is the propagation function for the electron–hole pair.

(a) Show that $Q_0(q, \omega)$ can be written as

$$Q_0(q, \omega) = \frac{-2\mathrm{i}}{(2\pi)^3}\int_{k<k_{\mathrm{F}}} \mathrm{d}^3k$$
$$\left[\frac{1}{E(\mathbf{k}+\mathbf{q}) - E(\mathbf{k}) + \omega} + \frac{1}{E(\mathbf{k}+\mathbf{q}) - E(\mathbf{k}) - \omega}\right].$$

Note that the single particle propagator $G_0(q, \omega)$ is given by

$$G_0(q, \omega) = \frac{\mathrm{i}}{\omega - E(q)(1 - \mathrm{i}\delta)}.$$

(b) Show that the solution of the Dyson equation given above is simply $W = \frac{V}{1+VQ_0}$.
(c) Show that $1 + VQ_0$ is the same as the Lindhard dielectric function $\varepsilon(q, \omega)$.

Summary

In this chapter, we study Green's function method – a formal theory of many body interactions. Green's function is defined in terms of a matrix element of time ordered Heisenberg operators in the exact interacting ground state. We then introduce the interaction representation of the state functions of many particle states and write the Green's function in terms of time ordered products of interaction operators. Wick's theorem is introduced to write the exact Green's function as a perturbation expansion involving only pairings of field operators in the interaction representation. Dyson equations for Green's function and the screened interaction are illustrated and Fermi liquid picture of quasiparticle interactions is also discussed.

The Hamiltonian H of a many particle system can be divided into two parts H_0 and H', where H' represents the interparticle interactions given, in second quantized form, by

$$H' = \frac{1}{2}\int \mathrm{d}^3 r_1 \mathrm{d}^3 r_2 \psi^\dagger(\mathbf{r}_1)\psi^\dagger(\mathbf{r}_2)U(\mathbf{r}_1-\mathbf{r}_2)\psi(\mathbf{r}_2)\psi(\mathbf{r}_1).$$

Particle density at a position $\mathbf{r}_0$ and the total particle number N are written, respectively, as $n(\mathbf{r}_0)=\psi^\dagger(\mathbf{r}_0)\psi(\mathbf{r}_0)$; $N=\int \mathrm{d}^3 r \psi^\dagger(\mathbf{r})\psi(\mathbf{r})$.

The Schrödinger equation of the many particle wave function $\Psi(1,2,\ldots,N)$ is ($\hbar \equiv 1$) $\mathrm{i}\hbar\frac{\partial}{\partial t}\Psi = H\Psi$, where $\Psi(t)=\mathrm{e}^{-\mathrm{i}Ht}\Psi_\mathrm{H}$. Here, Ψ_H is time independent. The state vector $\Psi_\mathrm{I}(t)$ and operator $F_\mathrm{I}(t)$ in the *interaction representation* are written as

$$\Psi_\mathrm{I}(t)=\mathrm{e}^{\mathrm{i}H_0 t}\Psi_\mathrm{S}(t);\ F_\mathrm{I}(t)=\mathrm{e}^{\mathrm{i}H_0 t}F_\mathrm{S}\mathrm{e}^{-\mathrm{i}H_0 t}.$$

The equation of motion for $F_\mathrm{I}(t)$ is $\frac{\partial F_\mathrm{I}}{\partial t} = \mathrm{i}\,[H_0, F_\mathrm{I}(t)]$ and the solution for $F_\mathrm{I}(t)$ can be expressed as $\Psi_\mathrm{I}(t)=S(t,t_0)\Psi_\mathrm{I}(t_0)$, where $S(t,t_0)$ is the S *matrix* given by

$$S(t,t_0)=\mathrm{T}\left\{\mathrm{e}^{-\mathrm{i}\int_{t_0}^{t} H_\mathrm{I}(t')\mathrm{d}t'}\right\}.$$

The eigenstates of the interacting system in the Heisenberg, Schrödinger, and interaction representation are related by

$$\Psi_\mathrm{H}(t)=\mathrm{e}^{\mathrm{i}Ht}\Psi_\mathrm{S}(t) \quad \text{and} \quad \Psi_\mathrm{I}(t)=\mathrm{e}^{\mathrm{i}H_0 t}\Psi_\mathrm{S}(t).$$

At time t=0, $\Psi_\mathrm{I}(t=0)=\Psi_\mathrm{H}(t=0)=\Psi_\mathrm{H}$. Ψ_H is the state vector of the fully interacting system in the Heisenberg representation: $\Psi_\mathrm{H}=S(0,-\infty)\Phi_\mathrm{H}$.

The Green's function $G_{\alpha\beta}(x,x')$ is defined, in terms of ψ_α^H and $\psi_\beta^{\mathrm{H}\dagger}$, by

$$G_{\alpha\beta}(x,x') = -\mathrm{i}\frac{\langle\Psi_\mathrm{H}|\mathrm{T}\{\psi_\alpha^\mathrm{H}(x)\psi_\beta^{\mathrm{H}\dagger}(x')\}|\Psi_\mathrm{H}\rangle}{\langle\Psi_\mathrm{H}|\Psi_\mathrm{H}\rangle},$$

where $x=\{\mathbf{r},t\}$ and α, β are spin indices.

In *normal product* of operators, all annihilation operators appear to the right of all creation operators: for example,

$$\mathrm{N}\{\psi^\dagger(1)\psi(2)\} = \psi^\dagger(1)\psi(2) \;\text{ while }\; \mathrm{N}\{\psi(1)\psi^\dagger(2)\} = -\psi^\dagger(2)\psi(1).$$

Pairing or a *contraction* is the difference between a T product and an N product: $\mathrm{T}(AB) - \mathrm{N}(AB) = A^{\mathrm{c}}B^{\mathrm{c}}$. The Wick's theorem states that T product of operators $ABC\cdots$ can be expressed as the sum of all possible N products with all possible pairings.

Dyson's equations for the interacting Green's function G and the screened interaction W are written as $G = G^{(0)} + G^{(0)}\Sigma G;\; W = V + V\Pi W$. Here, Σ and Π denote the self energy and polarization part, and the simplest of which are given, respectively, by $\Sigma_0 = G^{(0)}W;\; \Pi_0 = G^{(0)}G^{(0)}$. In the RPA, the G is replaced by $G^{(0)}$ and W is exactly equivalent to $\frac{V(q)}{\varepsilon(q,\omega)}$, where $\varepsilon(q,\omega)$ is the Lindhard dielectric function.

The Hamiltonian H of a system with the electron–phonon interaction is divided into three parts: $H = H_{\mathrm{e}} + H_{\mathrm{N}} + H_{\mathrm{I}}$, where

$$H_{\mathrm{e}} = \sum_k \frac{\hbar^2 k^2}{2m^*} c_k^\dagger c_k,\;\; H_{\mathrm{N}} = \sum_\alpha \hbar\omega_\alpha (b_\alpha^\dagger b_\alpha + \frac{1}{2}),$$

and $H_{\mathrm{I}} = \sum_{k,k',q} \frac{4\pi e^2}{\Omega q^2} c_{k+q}^\dagger c_{k'-q}^\dagger c_{k'} c_k + \sum_{k,\alpha,\mathbf{G}} \gamma(\alpha,\mathbf{G})(b_\alpha - b_{-\alpha}^\dagger) c_{\mathbf{k+q+G}}^\dagger c_k$. Once we know $\Psi(t_1)$ of the Schrödinger equation, $\mathrm{i}\hbar\frac{\partial\Psi}{\partial t} = H\Psi$, we have

$$\Psi(x_2,t_2) = \int \mathrm{d}^3 x_1 G_0(x_2,t_2;x_1,t_1)\Psi(x_1,t_1).$$

For free electrons, $G_0(q,\omega)$ is the Fourier transform of $G_0(2,1)$:

$$G_0(q,\omega) = \frac{\mathrm{i}}{\omega - E(q)(1-\mathrm{i}\delta)}.$$

For a "model solid" containing longitudinal phonons as well as electrons, two electrons can scatter via the virtual exchange of phonons and the total interaction, i.e. the sum of the Coulomb interaction and the interaction due to virtual exchange of phonons, is given, in terms of bare interaction D_0 and polarization Q_0, by $D(\mathbf{q},\omega) = \frac{D_0(\mathbf{q},\omega)}{1+D_0(\mathbf{q},\omega)Q_0(\mathbf{q},\omega)}$.

The Dyson equation for the Green's function can be written

$$G(k,\omega) = G^{(0)}(k,\omega) + G^{(0)}(k,\omega)\Sigma(k,\omega)G(k,\omega).$$

The electron self energy is $\Sigma(k,\omega) = [G^{(0)}(k,\omega)]^{-1} - [G(k,\omega)]^{-1}$ and the energy of a quasiparticle is written as $E_p = \varepsilon_p + \Sigma(p,\omega)\;|_{\omega=E_p}$. $\Sigma(\mathbf{p},E_{\mathbf{p}})$ represents the interaction of a quasiparticle of momentum $\mathbf{p}$ with the ground state of the interacting electron gas. The energy of the state is written as

$$E = E_0 + \sum_{\mathbf{p}\sigma} \delta n_{\mathbf{p}\sigma} E_{\mathbf{p}\sigma} + \frac{1}{2} \sum_{\substack{\mathbf{p},\,\mathbf{p}' \\ \sigma,\,\sigma'}} f_{\sigma\sigma'}(\mathbf{p},\mathbf{p}')\delta n_{n\sigma}\delta n_{\mathbf{p}'\sigma'}.$$

The first term on the right is the ground-state energy, the second is the quasiparticle energy $E_{\mathbf{p}\sigma}$ multiplied by the quasiparticle distribution function, and the third represents the interactions of the quasiparticles with one another.

13

Semiclassical Theory of Electrons

13.1 Bloch Electrons in a dc Magnetic Field

In the presence of an electric field $\mathbf{E}$ and a dc magnetic field $\mathbf{B}$, the equation of motion of a Bloch electron in k-space takes the form

$$\hbar\dot{\mathbf{k}} = -e\mathbf{E} - \frac{e}{c}\mathbf{v} \times \mathbf{B}. \tag{13.1}$$

Here, $\mathbf{v} = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon(\mathbf{k})$ is the velocity of the Bloch electron whose energy $\varepsilon(\mathbf{k})$ is an arbitrary function of wave vector $\mathbf{k}$. In deriving (13.1), we noted that no interband transitions were allowed, and that when $\mathbf{k}$ became equal to a value on the boundary of the Brillouin zone this value of $\mathbf{k}$ was identical to the value on the opposite side of the Brillouin zone separated from it by a reciprocal lattice vector $\mathbf{K}$.

Equation (13.1) can be obtained from an effective Hamiltonian

$$\mathcal{H} = \varepsilon(\frac{\mathbf{p}}{\hbar} + \frac{e}{\hbar c}\mathbf{A}) - e\phi, \tag{13.2}$$

where $\varepsilon(\mathbf{k})$ is the energy as a function of $\mathbf{k}$ in the absence of the magnetic field and $\mathbf{B} = \nabla \times \mathbf{A}$. Hamilton's equations give, since $\mathbf{p} = \hbar\mathbf{k} - \frac{e}{c}\mathbf{A}$,

$$v_x = \frac{\partial \mathcal{H}}{\partial p_x} = \frac{1}{\hbar}\frac{\partial \varepsilon}{\partial k_x}, \tag{13.3}$$

$$-\dot{p}_x = \frac{\partial \mathcal{H}}{\partial x} = \nabla_{\mathbf{k}}\varepsilon \cdot \frac{\partial}{\partial x}\left(\frac{e}{\hbar c}\mathbf{A}\right) - e\frac{\partial \phi}{\partial x} = \frac{e}{c}\left(\mathbf{v} \cdot \frac{\partial \mathbf{A}}{\partial x}\right) - e\frac{\partial \phi}{\partial x}. \tag{13.4}$$

But we also know that $\dot{p}_x = \hbar\dot{k}_x - \frac{e}{c}\dot{A}_x$, or

$$\dot{p}_x = \hbar\dot{k}_x - \frac{e}{c}\frac{\partial A_x}{\partial t} - \frac{e}{c}\left(v_x\frac{\partial A_x}{\partial x} + v_y\frac{\partial A_x}{\partial y} + v_z\frac{\partial A_x}{\partial z}\right). \tag{13.5}$$

Equating the $\dot{p}_x$ from (13.4) with that from (13.5) gives

$$\hbar \dot{k}_x = -eE_x - \frac{e}{c}\left(\mathbf{v} \times \mathbf{B}\right)_x . \tag{13.6}$$

Since the equation of motion, (13.1) or (13.6), is derived from Hamilton's equations, (13.3) and (13.4), using the effective Hamiltonian (13.2), $\mathbf{p}$ and $\mathbf{r}$ must be canonically conjugate coordinates.

13.1.1 Energy Levels of Bloch Electrons in a Magnetic Field

Onsager determined the energy levels of electrons in a dc magnetic field by noting that

$$\hbar \dot{\mathbf{k}} = -\frac{e}{c}\mathbf{v} \times \mathbf{B} \tag{13.7}$$

could be written as

$$\dot{\mathbf{k}}_\perp = -\frac{e}{\hbar c}|\mathbf{v}_\perp|B. \tag{13.8}$$

Here, $\mathbf{v}_\perp$ is the component of $\mathbf{v}$ perpendicular to $\mathbf{B}$ and $\mathbf{k}_\perp$ is perpendicular to both $\mathbf{B}$ and $\mathbf{v}$. Integrating (13.7) gives

$$\mathbf{k}_\perp = \frac{e\mathbf{B}}{\hbar c} \times \mathbf{r}_\perp + \text{ constant}. \tag{13.9}$$

We can choose the origin of $\mathbf{k}$ and $\mathbf{r}$ space such that the constant vanishes for the electron of interest. Thus, the orbit in real space (by this we mean the periodic part of the motion in $\mathbf{r}$-space that is perpendicular to $\mathbf{B}$) will be exactly the same shape as the orbit in k-space except that it is rotated by 90° and scaled by a factor $\frac{eB}{\hbar c}$. This factor $\frac{eB}{\hbar c}$ is called l_0^{-2}, where l_0 is the *magnetic length*.

Let us choose $\mathbf{B}$ to define the z-direction. Then $\dot{k}_z = 0$ and k_z is a constant of the motion. Now, look at the time rate of change of the energy

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} = \nabla_k \varepsilon \cdot \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = \hbar \mathbf{v} \cdot \left(-\frac{e}{\hbar c}\mathbf{v} \times \mathbf{B}\right). \tag{13.10}$$

This is clearly zero since $\boldsymbol{v}$ is perpendicular to $\boldsymbol{v} \times \boldsymbol{B}$, meaning that ε is a constant of the motion also. Thus, the orbit of a particle in $\boldsymbol{k}$-space is the intersection of a constant energy surface $\varepsilon(\boldsymbol{k}) = \varepsilon$ and a plane of constant k_z (see Fig. 13.1).

Let us look at the different kinds of orbits that are possible by considering the intersections of a plane $k_z = 0$ with a number of different energy surfaces for a simple model $\varepsilon(\boldsymbol{k})$ (see, for example, Fig. 13.2). The orbits can be classified as

- Closed electron orbits like A and B
- Closed hole orbits like C
- Open orbits like D

Often one simply repeats the Brillouin zone a number of times to show how the pieces of hole orbits or the open orbits look as illustrated in Fig. 13.3.

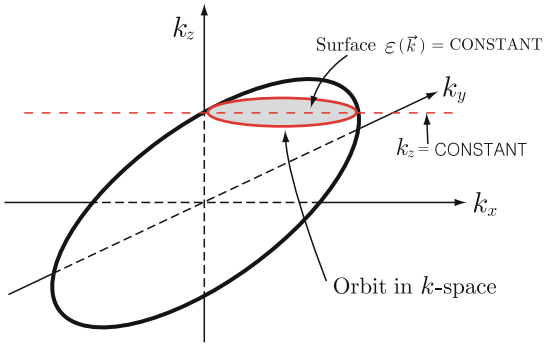


Fig. 13.1. A constant energy surface $\varepsilon(\boldsymbol{k}) = \varepsilon$ and the orbit of a particle in k-space

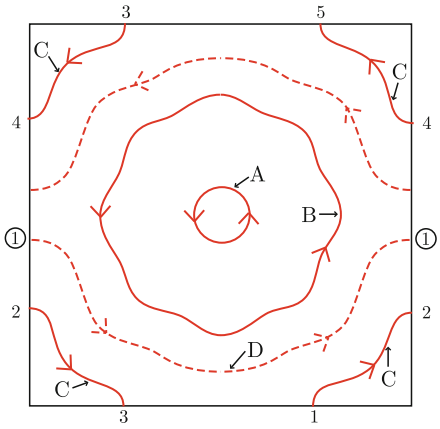


Fig. 13.2. Different kinds of orbits of a particle in k-space

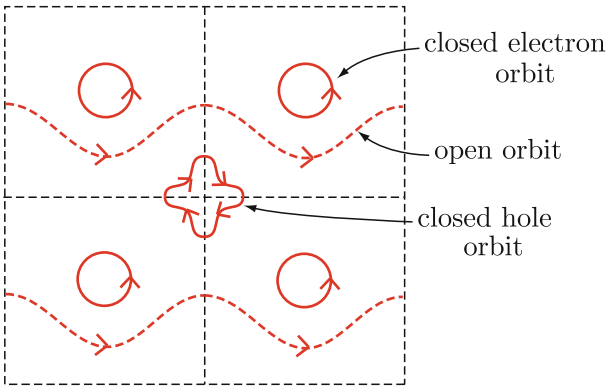


Fig. 13.3. Repeated zone scheme of the orbits in k-space

13.1.2 Quantization of Energy

For closed orbits in $\boldsymbol{k}$-space, the motion is periodic. The real space orbits in the direction perpendicular to $\boldsymbol{B}$ will also be periodic. Because $\boldsymbol{p}$ and $\boldsymbol{r}$ are canonically conjugate coordinates we can apply the *Bohr–Sommerfeld quantization condition*

$$\oint \mathbf{p}_\perp \cdot \mathrm{d}\mathbf{r}_\perp = 2\pi\hbar(n+\gamma), \tag{13.11}$$

where γ is a constant satisfying $0 \le \gamma \le 1$, and $n = 0, 1, 2, \ldots$. We can use $\mathbf{p}_\perp = \hbar\mathbf{k}_\perp - \frac{e}{c}\mathbf{A}_\perp$ and the fact that $\mathbf{k}_\perp(t) = \frac{e\mathbf{B}}{\hbar c} \times \mathbf{r}_\perp(t)$. Then

$$\oint \mathbf{p}_\perp \cdot \mathrm{d}\mathbf{r}_\perp = \frac{e}{c}\mathbf{B} \cdot \oint \mathbf{r}_\perp \times \mathrm{d}\mathbf{r}_\perp - \frac{e}{c}\oint \mathbf{A} \cdot \mathrm{d}\mathbf{r}_\perp .$$

But $\oint \mathbf{r}_\perp \times \mathrm{d}\mathbf{r}_\perp$ is just twice the area of the orbit as is illustrated in Fig. 13.4. Furthermore, $\oint \mathbf{A} \cdot \mathrm{d}\mathbf{r}_\perp = \int_{\text{Surface}} \nabla \times \mathbf{A} \cdot \mathrm{d}\mathbf{S} = B \times$ (area of orbit). Therefore, we obtain

$$\oint \mathbf{p}_\perp \cdot \mathrm{d}\mathbf{r}_\perp = \frac{e}{c} B\,\mathcal{A} = 2\pi\hbar(n+\gamma), \tag{13.12}$$

where $\mathcal{A}$ is the area of the orbit. $\mathcal{A}$ depends on energy ε and on k_z. We know that $\mathcal{A}(\varepsilon, k_z)$ is proportional to the area $\mathcal{S}(\varepsilon, k_z)$ of the orbit in $\boldsymbol{k}$-space, from (13.9), with $\mathcal{S}(\varepsilon, k_z) = \left(\frac{eB}{\hbar c}\right)^2 \mathcal{A}(\varepsilon, k_z)$. Thus, the quantization condition can be written

$$\mathcal{S}(\varepsilon, k_z) = \frac{2\pi e B}{\hbar c}(n+\gamma). \tag{13.13}$$

Example For free electrons $\varepsilon = \frac{\hbar^2 k^2}{2m}$. The area $\mathcal{S}(\varepsilon, k_z)$ is equal to $\pi k_\perp^2$, where $k_\perp^2 + k_z^2 = k^2$. Therefore,

$$\mathcal{S}(\varepsilon, k_z) = \pi\left(\frac{2m\varepsilon}{\hbar^2} - k_z^2\right). \tag{13.14}$$

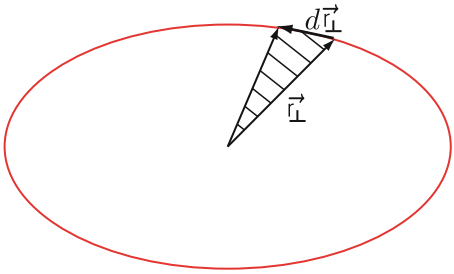


Fig. 13.4. $\mathbf{r}_\perp \times \mathrm{d}\mathbf{r}_\perp$ is twice the area of the triangle within the orbit

Setting this result equal to $\frac{2\pi eB}{\hbar c}(n+\gamma)$ and solving for energy ε gives

$$\varepsilon = \frac{\hbar^2 k_z^2}{2m} + \hbar\omega_c(n+\gamma), \tag{13.15}$$

where $\omega_c = \frac{eB}{mc}$ is the *cyclotron frequency.*

13.1.3 Cyclotron Effective Mass

In the absorption of radiation direct transitions between energy levels (Landau levels) occur. If we make a transition from $\varepsilon_n(k_z)$ to $\varepsilon_{n+1}(k_z)$, we can write

$$\begin{aligned} \mathcal{S}(\varepsilon_n(k_z), k_z) &= \frac{2\pi eB}{\hbar c}(n+\gamma), \\ \mathcal{S}(\varepsilon_{n+1}(k_z), k_z) &= \frac{2\pi eB}{\hbar c}(n+1+\gamma). \end{aligned} \tag{13.16}$$

We define the energy difference $\varepsilon_{n+1}(k_z) - \varepsilon_n(k_z)$ as $\hbar\omega_c$, where $\omega_c(\varepsilon, k_z)$ is the cyclotron frequency associated with the orbit $\{\varepsilon, k_z\}$. By subtracting the first equation of (13.16) from the second we can obtain

$$[\varepsilon_{n+1}(k_z) - \varepsilon_n(k_z)] \frac{\partial S(\varepsilon, k_z)}{\partial \varepsilon} = \frac{2\pi eB}{\hbar c}$$

and from this, we obtain

$$\omega_c(\varepsilon, k_z) = \frac{2\pi eB}{\hbar^2 c}\left[\frac{\partial S(\varepsilon, k_z)}{\partial \varepsilon}\right]^{-1} = \frac{eB}{m^* c} \tag{13.17}$$

or

$$m^*(\varepsilon, k_z) = \frac{\hbar^2}{2\pi}\frac{\partial S(\varepsilon, k_z)}{\partial \varepsilon}. \tag{13.18}$$

13.1.4 Velocity Parallel to B

Consider two orbits that have different values of k_z separated by Δk_z. Then, for the same ε_n we have

$$S\left[\varepsilon_n(k_z + \Delta k_z), k_z + \Delta k_z\right] - S\left[\varepsilon_n(k_z), k_z\right] = 0 \tag{13.19}$$

because both orbits have cross-sectional area equal to $\frac{2\pi eB}{\hbar c}(n+\gamma)$. We can write (13.19) as

$$\frac{\partial S}{\partial \varepsilon}\frac{\partial \varepsilon_n(k_z)}{\partial k_z} + \frac{\partial S}{\partial k_z} = 0. \tag{13.20}$$

But $\frac{\partial \varepsilon}{\partial k_z} = \hbar v_z$ and $\frac{\partial S}{\partial \varepsilon} = \frac{2\pi m^*}{\hbar^2}$ giving

$$v_z(\varepsilon, k_z) = -\frac{\hbar}{2\pi m^*(\varepsilon, k_z)}\frac{\partial S(\varepsilon, k_z)}{\partial k_z}. \tag{13.21}$$

Example For the free electron gas model, we have $S = \pi\left(\frac{2m^*\varepsilon}{\hbar^2} - k_z^2\right)$. Therefore, $\frac{\partial S}{\partial k_z} = -2\pi k_z$ and, hence,

$$v_z(\varepsilon, k_z) = -\frac{\hbar}{2\pi m^*}(-2\pi k_z) = \frac{\hbar k_z}{m^*}.$$

13.2 Magnetoresistance

The study of the change in resistivity of a metal as a function of the strength of an applied magnetic field is very useful in understanding certain properties of the Fermi surface of a metal or semiconductor. The standard geometry for *magnetoresistance* measurements is shown in Fig. 13.5. Current flows only in the x direction. Usually $\mathbf{B}$ is in the z direction and the *transverse magnetoresistance* is defined as

$$\frac{R(B_z) - R(0)}{R(0)} = \Delta R(B_z). \tag{13.22}$$

Sometimes people also study the case where $\mathbf{B}$ is parallel to $\mathbf{E}$ and measure the *longitudinal magnetoresistance*.

It might seem surprising that anything of interest arises from studying the magnetoresistance, because, as we described, for the simple free electron model $\Delta R(B_z) = 0$. This resulted from the equation

$$\begin{aligned} j_x &= \sigma_{xx}E_x + \sigma_{xy}E_y \\ j_y &= -\sigma_{xy}E_x + \sigma_{xx}E_y = 0. \end{aligned} \tag{13.23}$$

Combining these gives

$$j_x = \frac{\sigma_{xx}^2 + \sigma_{xy}^2}{\sigma_{xx}}E_x. \tag{13.24}$$

But for free electrons

$$\begin{aligned} \sigma_{xx} &= \frac{\sigma_0}{1 + \omega_c^2\tau^2}, \\ \sigma_{xy} &= -\frac{\omega_c\tau\sigma_0}{1 + \omega_c^2\tau^2}, \end{aligned} \tag{13.25}$$

where $\sigma_0 = \frac{n_0 e^2 \tau}{m}$ is the dc conductivity. Using (13.25) in (13.24) gives $j_x = \sigma_0 E_x$, independent of $\mathbf{B}$ so that the magnetoresistance vanishes.

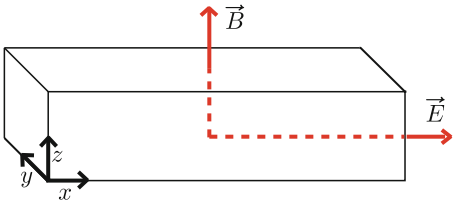


Fig. 13.5. Standard geometry for magnetoresistance measurements

Experimental Results

Before discussing other models than the simple one band free electron model, let us discuss briefly the experimental results. The following types of behaviors are common:

1. The magnetoresistance is nonzero, but it saturates at very high magnetic fields at a value that is several times larger than the zero field resistance.
2. The magnetoresistance does not saturate, but continues to increase with increasing B in all directions.
3. The magnetoresistance saturates in some crystal directions but does not saturate in other directions.

Simple metals like Na, Li, In, and Al belong to the type (1). Semimetals like Bi and Sb belong to type (2). The noble metals (Cu, Ag, and Au), Mn, Zn, Cd, Ga, Sn, and Pb belong to type (3). One can obtain some understanding of magnetoresistance by using a two-band model.

13.3 Two-Band Model and Magnetoresistance

Let us consider two simple parabolic bands with mass, cyclotron frequency, charge, concentration, and collision time given by m_i, ω_{ci}, e_i, n_i, and τ_i, respectively, where $i = 1$ or 2. Each band has a conductivity $\underline{\sigma}_i$, and the total current is simply the sum of $\mathbf{j}_1$ and $\mathbf{j}_2$

$$\mathbf{j}_{\mathrm{T}} = (\underline{\sigma}_1 + \underline{\sigma}_2) \cdot \mathbf{E}. \tag{13.26}$$

But

$$\underline{\sigma}_i = \frac{n_i e_i^2 \tau_i / m_i}{1 + \omega_{ci}^2 \tau_i^2} \begin{pmatrix} 1 & \omega_{ci}\tau_i & 0 \\ -\omega_{ci}\tau_i & 1 & 0 \\ 0 & 0 & 1 + \omega_{ci}^2 \tau_i^2 \end{pmatrix}. \tag{13.27}$$

Note that we are taking $\omega_{ci} = \frac{e_i B}{m_i c}$ which is negative for an electron; this is why the σ_{xy} has a plus sign. At very high magnetic fields $|\omega_{ci}\tau_i| \gg 1$ for both types of carriers. Therefore, we can drop the 1 in $1 + \omega_{ci}^2 \tau_i^2$:

$$\underline{\sigma}_i \simeq \frac{n_i e_i c}{B} \begin{pmatrix} \dfrac{1}{\omega_{ci}\tau_i} & 1 & 0 \\ -1 & \dfrac{1}{\omega_{ci}\tau_i} & 0 \\ 0 & 0 & \omega_{ci}\tau_i \end{pmatrix}, \tag{13.28}$$

and

$$\underline{\sigma}_{\mathrm{T}} \simeq \frac{c}{B} \begin{pmatrix} \dfrac{n_1 e_1}{\omega_{c1}\tau_1} + \dfrac{n_2 e_2}{\omega_{c2}\tau_2} & n_1 e_1 + n_2 e_2 & 0 \\ -(n_1 e_1 + n_2 e_2) & \dfrac{n_1 e_1}{\omega_{c1}\tau_1} + \dfrac{n_2 e_2}{\omega_{c2}\tau_2} & 0 \\ 0 & 0 & n_1 e_1 \omega_{c1}\tau_1 + n_2 e_2 \omega_{c2}\tau_2 \end{pmatrix}, \tag{13.29}$$

Now, suppose that $n_1 = n_2 = n$ and $e_1 = -e_2 = e$. This corresponds to a semimetal with an equal number of electrons and holes. Then, (13.29) reduces to

$$\underline{\sigma} \simeq \frac{nec}{B} \begin{pmatrix} \frac{1}{|\omega_{c1}\tau_1|} + \frac{1}{|\omega_{c2}\tau_2|} & 0 & 0 \\ 0 & \frac{1}{|\omega_{c1}\tau_1|} + \frac{1}{|\omega_{c2}\tau_2|} & 0 \\ 0 & 0 & |\omega_{c1}\tau_1| + |\omega_{c2}\tau_2| \end{pmatrix},$$

The Hall field vanishes since $\sigma_{xy} = 0$ and

$$j_x = \sigma_{xx} E_x = \frac{nec}{B} \left[\frac{1}{\frac{eB}{m_1 c}\tau_1} + \frac{1}{\frac{eB}{m_2 c}\tau_2} \right] E_x. \tag{13.30}$$

The resistivity is the ratio of E_x to j_x giving

$$\rho = \frac{B^2}{nec^2} \left(\frac{|\mu_1||\mu_2|}{|\mu_1| + |\mu_2|} \right), \tag{13.31}$$

where $\mu_i = \frac{e_i \tau_i}{m_i}$ is the mobility of the ith type. Thus, we find for equal numbers of electrons and holes the magnetoresistance does not saturate, but continues to increase as B^2. The arguments can be generalized to two bands described by energy surfaces $\varepsilon_i(\mathbf{k})$, but we will not bother with that much detail. If $n_e \neq n_h$, $\sigma_{xy} \simeq -\frac{(n_e - n_h)ec}{B}$ while $\sigma_{xx} = \left(\frac{n_e}{|\omega_{ce}|\tau_e} + \frac{n_h}{|\omega_{ch}|\tau_h} \right) \frac{ec}{B}$. For $|\omega_{ci}\tau_i| \gg 1$, $\sigma_{xy} \gg \sigma_{xx}$ and $\rho = \left[\frac{\sigma_{xx}^2 + \sigma_{xy}^2}{\sigma_{xx}} \right]^{-1} \simeq \left[\frac{\sigma_{xy}^2}{\sigma_{xx}} \right]^{-1}$. This saturates to a constant because $\sigma_{xy} \propto B^{-1}$ while $\sigma_{xx} \propto B^{-2}$.

Influence of Open Orbits

For the sake of concreteness we will first consider a model which is extremely simple and has open orbits to see what happens. Suppose there is a section of the Fermi surface with energy given by

$$\varepsilon(\mathbf{k}) = \frac{\hbar^2}{2m}(k_x^2 + k_z^2), \tag{13.32}$$

i.e. $\varepsilon(\mathbf{k})$ is independent of k_y as is shown in Fig. 13.6. Again take the magnetic field in the z-direction. Then the orbits are all open orbits, and run parallel to the cylinder axis. Note that

$$v_x = \frac{\hbar k_x}{m}, \quad v_y = 0, \quad \text{and} \quad v_z = \frac{\hbar k_z}{m}. \tag{13.33}$$

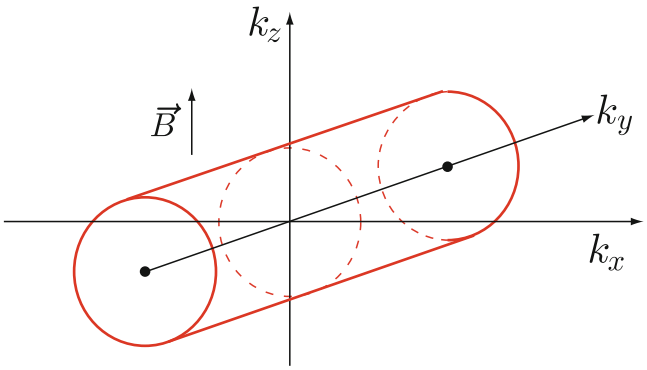


Fig. 13.6. A model energy surface with open orbits

Look at equations of motion in the presence of B_z and $\mathbf{E} = (E_x, E_y, 0)$:

$$\begin{aligned}\hbar\dot{k}_x &= -e(E_x + \frac{1}{c}v_y B_z) = -eE_x,\\ \hbar\dot{k}_y &= -e(E_y - \frac{1}{c}v_x B_z),\\ \hbar\dot{k}_z &= 0.\end{aligned} \tag{13.34}$$

The equation of motion for $\hbar\dot{k}_x$ can be written as

$$\dot{v}_x = -\frac{eE_x}{m}. \tag{13.35}$$

We have completely neglected collisions so far; they can be added by simply writing

$$\dot{v}_x + \frac{v_x}{\tau} = -\frac{eE_x}{m}. \tag{13.36}$$

Then in the steady-state we have

$$v_x = -\frac{eE_x\tau}{m}. \tag{13.37}$$

If n_o is the number of open orbit states per unit volume, then

$$\underline{\sigma}_{\mathrm{Open}} = \frac{n_\mathrm{o}e^2\tau}{m}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}. \tag{13.38}$$

Here, we have used $\sigma_{yy} = \sigma_{xy} = \sigma_{yx} = 0$; this is correct because j_x depends only on E_x and v_y must be zero since the mass in the y direction is infinite.

Now, suppose there is another piece of Fermi surface that contains n_c electrons per unit volume all in closed orbit states. We can approximate the contribution of these electrons to the conductivity by

$$\underline{\sigma}_{\mathrm{Closed}} = \frac{n_\mathrm{c}e^2\tau}{m}\begin{pmatrix}\frac{1}{\omega_\mathrm{c}^2\tau^2} & \frac{1}{\omega_\mathrm{c}\tau}\\ -\frac{1}{\omega_\mathrm{c}\tau} & \frac{1}{\omega_\mathrm{c}^2\tau^2}\end{pmatrix}. \tag{13.39}$$

The total conductivity is simply the sum of $\underline{\sigma}_{\rm Open}$ and $\underline{\sigma}_{\rm Closed}$:

$$\underline{\sigma}_{\rm T} = \frac{e^2\tau}{m}\begin{pmatrix} n_{\rm o} + \dfrac{n_{\rm c}}{\omega_{\rm c}^2\tau^2} & \dfrac{n_{\rm c}}{\omega_{\rm c}\tau} \\ -\dfrac{n_{\rm c}}{\omega_{\rm c}\tau} & \dfrac{n_{\rm c}}{\omega_{\rm c}^2\tau^2} \end{pmatrix}. \tag{13.40}$$

Let $n_{\rm o}$, the concentration of open orbit electrons, be equal to a number S times $n_{\rm c}$, the concentration of closed orbit electrons. Then

$$\begin{aligned} j_x &= \frac{n_{\rm c}e^2\tau}{m}\left\{(S+\frac{1}{\omega_{\rm c}^2\tau^2})E_x + \frac{1}{\omega_{\rm c}\tau}E_y\right\}, \\ j_y &= \frac{n_{\rm c}e^2\tau}{m}\left\{-\frac{1}{\omega_{\rm c}\tau}E_x + \frac{1}{\omega_{\rm c}^2\tau^2}E_y\right\} \end{aligned} \quad . \tag{13.41}$$

We can investigate two different cases:

1. In the standard geometry j_x is nonzero but j_y is zero.
2. In the standard geometry j_y is nonzero but j_x is zero.

Case 1:

$j_y = 0$ implies that

$$E_y = \omega_{\rm c}\tau E_x. \tag{13.42}$$

Therefore, j_x is given by

$$j_x = \frac{n_{\rm c}e^2\tau}{m}\left(S + \frac{1}{\omega_{\rm c}^2\tau^2} + 1\right)E_x. \tag{13.43}$$

The magnetoresistivity is $\frac{E_x}{j_x}$ giving

$$\rho = \frac{m/(n_{\rm c}e^2\tau)}{S + \frac{1}{\omega_{\rm c}^2\tau^2} + 1} \longrightarrow \frac{m/(n_{\rm c}e^2\tau)}{S+1} \text{ as } B \to \infty. \tag{13.44}$$

Thus, in this geometry the magnetoresistance saturates as B tends to infinity.

Case 2:

$j_x = 0$ implies that

$$E_x = -\frac{1}{\omega_{\rm c}\tau(S + \frac{1}{\omega_{\rm c}^2\tau^2})}E_y. \tag{13.45}$$

This means that

$$j_y = \frac{n_{\rm c}e^2\tau}{m}\left[\frac{1}{\omega_{\rm c}^2\tau^2 S + 1} + \frac{1}{\omega_{\rm c}^2\tau^2}\right]E_y. \tag{13.46}$$

But the magnetoresistivity is $\frac{E_y}{j_y}$ and it is given by

$$\rho = \frac{m}{n_c e^2 \tau} \frac{\omega_c^2 \tau^2 (\omega_c^2 \tau^2 S + 1)}{\omega_c^2 \tau^2 (S+1) + 1} \longrightarrow \frac{m}{n_c e^2 \tau} \frac{S}{S+1} \omega_c^2 \tau^2 \text{ as } B \to \infty. \tag{13.47}$$

Since ρ is proportional to ω_c^2, the magnetoresistance does not saturate but increases as B^2 as long as S is finite.

13.4 Magnetoconductivity of Metals

We consider an electron gas for which the energy is an arbitrary function of $\mathbf{k}$. We introduce a uniform dc magnetic field $\mathbf{B}_0$, and an ac electric field $\mathbf{E}$ of the form

$$\mathbf{E} \propto \mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}\mathbf{q}\cdot\mathbf{r}}. \tag{13.48}$$

The Boltzmann equation is

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f - \frac{e}{\hbar} \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}_0 \right) \cdot \nabla_k f = -\frac{f - \bar{f}_0}{\tau}. \tag{13.49}$$

As we shall see later, some care must be taken in the collision term $-\frac{f-\bar{f}_0}{\tau}$, to choose $\bar{f}_0$ to be the proper local equilibrium distribution toward which the electrons relax.[1] For now, we shall just put $\bar{f}_0 = f_0$, the actual thermal equilibrium function for the system. This gives the *conduction current* correctly, but omits a *diffusion current* which is actually present. We put $f = f_0 + f_1$ and then the Boltzmann equation becomes

$$\mathrm{i}\omega f_1 - \mathrm{i}\mathbf{q} \cdot \mathbf{v} f_1 - e\mathbf{E} \cdot \mathbf{v} \frac{\partial f_0}{\partial \varepsilon} - \frac{e}{\hbar c} (\mathbf{v} \times \mathbf{B}_0) \cdot \nabla_k f_1 + \frac{f_1}{\tau} = 0. \tag{13.50}$$

Here, we have used the fact that

$$\nabla_k f_0 = \frac{\partial f_0}{\partial \varepsilon} \nabla_k \varepsilon = \hbar \mathbf{v} \frac{\partial f_0}{\partial \varepsilon}, \tag{13.51}$$

and have linearized with respect to the ac field. We can write the Boltzmann equation as

$$(1 + \mathrm{i}\omega\tau - \mathrm{i}\mathbf{q} \cdot \mathbf{v}\tau) f_1 - \frac{e\tau}{\hbar c} (\mathbf{v} \times \mathbf{B}_0) \cdot \nabla_k f_1 = e\tau \mathbf{E} \cdot \mathbf{v} \frac{\partial f_0}{\partial \varepsilon}. \tag{13.52}$$

From the equation of motion, we remember, when the ac fields are $\mathbf{E} = 0 = \mathbf{B}$, that

$$\hbar \dot{\mathbf{k}} = -\frac{e}{c} \mathbf{v} \times \mathbf{B}_0, \tag{13.53}$$

[1] See, for eample, M.P. Greene, H.J. Lee, J.J. Quinn, S. Rodriguez, Phys. Rev. **177**, 1019 (1969).

and we were able to show that

1. The orbit of an electron in **k** space is along the intersection of a surface of constant energy and a plane of constant k_z.
2. The motion in **k** space is periodic either because the orbit is closed, or because an open orbit is actually periodic in **k** space also.

We introduce a parameter s with the dimension of time which describes the position of an electron on its orbit of constant energy and constant k_z. By this, we mean that if $s = 0$ is a point on the orbit, $s = T$ corresponds to the same point, where T is the period. The equation of motion can be written

$$\hbar \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}s} = -\frac{e}{c}\mathbf{v} \times \mathbf{B}_0, \tag{13.54}$$

and the rate of change of ε is

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}s} = \nabla_k \varepsilon \cdot \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}s} = \hbar \mathbf{v} \cdot \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}s} = 0.$$

Now, consider $\frac{\partial f_1}{\partial s}$ as

$$\frac{\partial f_1}{\partial s} = \nabla_k f_1 \cdot \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}s} = -\frac{e}{\hbar c}(\mathbf{v} \times \mathbf{B}_0) \cdot \nabla_k f_1. \tag{13.55}$$

This is exactly one of the terms in our Boltzmann equation which can be written

$$\frac{\partial f_1}{\partial s} + (\frac{1}{\tau} + \mathrm{i}\omega - \mathrm{i}\mathbf{q} \cdot \mathbf{v}) f_1 = e\mathbf{E} \cdot \mathbf{v} \frac{\partial f_0}{\partial \varepsilon}. \tag{13.56}$$

Closed Orbits:

Let us consider closed orbits first. We can write

$$\hbar \frac{d\mathbf{k}_\perp}{ds} = -\frac{e}{c}\mathbf{v}_\perp \times \mathbf{B}_0, \tag{13.57}$$

where $\mathbf{v}_\perp$ is the component of $\mathbf{v}$ perpendicular to $\mathbf{B}_0$. Let k_N and k_T be the normal and tangential components of $\mathbf{k}_\perp$ as shown in Fig. 13.7. Then

$$\hbar \frac{\mathrm{d}k_\mathrm{T}}{\mathrm{d}s} = \frac{e}{c} v_\perp B_0, \tag{13.58}$$

because $\mathbf{v}_\perp$ is in the direction of $\mathbf{k}_\mathrm{N}$. Solving for $\mathrm{d}s$ gives

$$\mathrm{d}s = \frac{\hbar c}{eB_0} \frac{\mathrm{d}k_\mathrm{T}}{v_\perp}. \tag{13.59}$$

Thus, the period is given by

$$T(\varepsilon, k_z) = \frac{2\pi}{\omega_\mathrm{c}(\varepsilon, k_z)} = \frac{\hbar c}{eB_0} \oint \frac{\mathrm{d}k_\mathrm{T}}{v_\perp}. \tag{13.60}$$

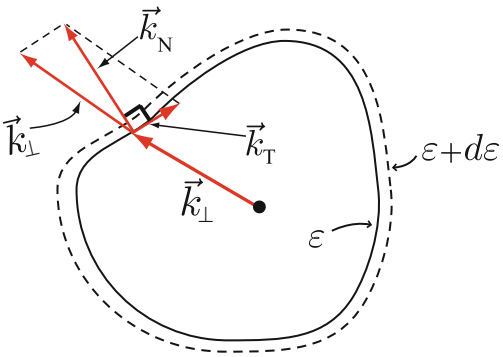


Fig. 13.7. A closed orbit in **k** space

From now on, we shall use the independent variables ε, k_z, and s to describe the location of an electron in **k** space.

$$\mathrm{d}\varepsilon = \nabla_k \varepsilon \cdot \mathrm{d}\mathbf{k}_{\mathrm{N}} = \hbar v_{\perp} \mathrm{d}k_{\mathrm{N}}, \tag{13.61}$$

and hence

$$\mathrm{d}k_{\mathrm{N}} = \frac{\mathrm{d}\varepsilon}{\hbar v_{\perp}}. \tag{13.62}$$

Now, the volume element in **k** space is given by

$$\mathrm{d}^3 k = \mathrm{d}k_z \, \mathrm{d}k_{\mathrm{N}} \, \mathrm{d}k_{\mathrm{T}} = \mathrm{d}k_z \frac{\mathrm{d}\varepsilon}{\hbar v_{\perp}} \frac{e v_{\perp}}{\hbar c} B_0 \mathrm{d}s,$$

hence

$$\mathrm{d}^3 k = \frac{eB_0}{\hbar^2 c} \mathrm{d}k_z \mathrm{d}\varepsilon \mathrm{d}s. \tag{13.63}$$

Now, let us return to the differential equation given by (13.56)

$$\frac{\partial f_1}{\partial s} + \left(\frac{1}{\tau} + \mathrm{i}\omega - \mathrm{i}\mathbf{q} \cdot \mathbf{v} \right) f_1 = e\mathbf{E} \cdot \mathbf{v} \frac{\partial f_0}{\partial \varepsilon}. \tag{13.64}$$

Multiply by

$$\mathrm{e}^{\int_0^s \mathrm{d}t' \left[\frac{1}{\tau} + \mathrm{i}\omega - \mathrm{i}\mathbf{q} \cdot \mathbf{v}(t') \right]}. \tag{13.65}$$

Then, we have

$$\begin{aligned} &\mathrm{e}^{\int_0^s \mathrm{d}t' \left[\frac{1}{\tau} + \mathrm{i}\omega - \mathrm{i}\mathbf{q} \cdot \mathbf{v}(t') \right]} \left\{ \frac{\partial f_1(s)}{\partial s} + \left[\frac{1}{\tau} + \mathrm{i}\omega - \mathrm{i}\mathbf{q} \cdot \mathbf{v}(s) \right] f_1(s) \right\} \\ &= e\mathbf{E} \cdot \mathbf{v}(s) \frac{\partial f_0}{\partial \varepsilon} \mathrm{e}^{\int_0^s \mathrm{d}t' \left[\frac{1}{\tau} + \mathrm{i}\omega - \mathrm{i}\mathbf{q} \cdot \mathbf{v}(t') \right]}. \end{aligned} \tag{13.66}$$

Notice that the left-hand side can be simplified to write

$$\frac{\partial}{\partial s} \left[f_1 \mathrm{e}^{\int_0^s \mathrm{d}t' \left(\frac{1}{\tau} + \mathrm{i}\omega - \mathrm{i}\mathbf{q} \cdot \mathbf{v}(t') \right)} \right] = e\mathbf{E} \cdot \mathbf{v}(s) \frac{\partial f_0}{\partial \varepsilon} \mathrm{e}^{\int_0^s \mathrm{d}t' \left(\frac{1}{\tau} + \mathrm{i}\omega - \mathrm{i}\mathbf{q} \cdot \mathbf{v}(t') \right)}. \tag{13.67}$$

Integrate and get

$$f_1(s)\mathrm{e}^{\int_0^s \mathrm{d}t'\left(\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}(t')\right)} = \int_{-\infty}^{s} \mathrm{d}t\, e\mathbf{E}\cdot\mathbf{v}(t)\frac{\partial f_0}{\partial\varepsilon}\mathrm{e}^{\int_0^t \mathrm{d}t'\left(\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}(t')\right)}, \quad (13.68)$$

or

$$f_1(s) = \int_{-\infty}^{s} \mathrm{d}t\, e\mathbf{E}\cdot\mathbf{v}(t)\frac{\partial f_0}{\partial\varepsilon}\mathrm{e}^{-\int_t^s \mathrm{d}t'\left(\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}(t')\right)}. \quad (13.69)$$

Note that the lower limit $t=-\infty$ is chosen in the integration over t for a very important reason. $f_1(k_z,\varepsilon,s)$ must be a periodic function of s with period T. Consider the function $f_1(s+T)$ for an arbitrary lower limit (LL)

$$f_1(s+T) = \int_{\mathrm{LL}}^{s+T} \mathrm{d}t\, e\mathbf{E}\cdot\mathbf{v}(t)\frac{\partial f_0}{\partial\varepsilon}\mathrm{e}^{-\int_t^{s+T} \mathrm{d}t'\left(\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}(t')\right)}. \quad (13.70)$$

Now, let $w=t-T$, so that $t=s+T \longrightarrow w=s$ and $t=\mathrm{LL} \longrightarrow w=\mathrm{LL}-T$. We know $v(t)$ is a periodic function of t with period T. Likewise, if we let $t'=w'+T$, we can get

$$f_1(s+T) = \int_{\mathrm{LL}-T}^{s} \mathrm{d}w\, e\mathbf{E}\cdot\mathbf{v}(w)\frac{\partial f_0}{\partial\varepsilon}\mathrm{e}^{-\int_w^s \mathrm{d}w'\left[\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}(w')\right]}.$$

It is obvious $f_1(s+T)=f_1(s)$ if $\mathrm{LL}-T=\mathrm{LL}$. This is valid for $\mathrm{LL}=-\infty$ as we have chosen. Now, look at the exponential

$$\int_t^s \mathrm{d}t'\left(\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}(t')\right) = \left(\frac{1}{\tau}+\mathrm{i}\omega\right)(s-t)-\mathrm{d}\mathbf{q}\cdot\left[\mathbf{R}(\varepsilon,k_z,s)-\mathbf{R}(\varepsilon,k_z,t)\right], \quad (13.71)$$

so that

$$f_1(\varepsilon,k_z,s) = \int_{-\infty}^{s} \mathrm{d}s'\, e\mathbf{E}\cdot\mathbf{v}(s')\frac{\partial f_0}{\partial\varepsilon}\mathrm{e}^{-\left(\frac{1}{\tau}+\mathrm{i}\omega\right)(s-s')+\mathrm{i}\mathbf{q}\cdot\left[\mathbf{R}(\varepsilon,k_z,s)-\mathbf{R}(\varepsilon,k_z,s')\right]}. \quad (13.72)$$

The current density is given by

$$\mathbf{j}(\mathbf{r},t) = \frac{2}{(2\pi)^3}\int(-e)\mathbf{v}f_1\,\mathrm{d}^3k. \quad (13.73)$$

Or, substituting the result of (13.63) for d^3k, we have

$$\mathbf{j}(\mathbf{r},t) = -\frac{2e}{(2\pi)^3}\frac{eB_0}{\hbar^2 c}\int \mathrm{d}\varepsilon\mathrm{d}k_z\mathrm{d}s\mathbf{v}f_1(\varepsilon,k_z,s). \quad (13.74)$$

Now, substitute the solution $f_1(\varepsilon,k_z,s)$ given by (13.72) to obtain

$$\mathbf{j}(\mathbf{r},t) = -\frac{2e^2B_0}{(2\pi)^3\hbar^2 c}\int_0^\infty \mathrm{d}\varepsilon\int_{-\infty}^{\infty}\mathrm{d}k_z\int_0^{T(\varepsilon,k_z)}\mathrm{d}s\,\mathbf{v} \times\int_{-\infty}^{s} ds'\, e\mathbf{E}\cdot\mathbf{v}(s')\frac{\partial f_0}{\partial\varepsilon}\,\mathrm{e}^{-\left(\frac{1}{\tau}+\mathrm{i}\omega\right)(s-s')+\mathrm{i}\mathbf{q}\cdot\left[\mathbf{R}(\varepsilon,k_z,s)-\mathbf{R}(\varepsilon,k_z,s')\right]}. \quad (13.75)$$

We define the conductivity tensor by $\mathbf{j} = \underline{\sigma} \cdot \mathbf{E}$. Then, it is apparent that

$$\underline{\sigma} = \frac{2e^3 B_0}{(2\pi)^3 \hbar^2 c} \int_0^\infty \mathrm{d}\varepsilon \left(-\frac{\partial f_0}{\partial \varepsilon}\right) \int_{-\infty}^{\infty} \mathrm{d}k_z \int_0^{T(\varepsilon,k_z)} \mathrm{d}s\, \mathbf{v}(\varepsilon,k_z,s) \mathrm{e}^{-(\frac{1}{\tau}+\mathrm{i}\omega)s + \mathrm{i}\mathbf{q}\cdot\mathbf{R}(\varepsilon,k_z,s)} \times \int_{-\infty}^{s} \mathrm{d}s'\, \mathbf{v}(\varepsilon,k_z,s') \mathrm{e}^{(\frac{1}{\tau}+\mathrm{i}\omega)s' - \mathrm{i}\mathbf{q}\cdot\mathbf{R}(\varepsilon,k_z,s')}. \tag{13.76}$$

We assume that τ depends only on ε. Now, look at the function $\mathbf{v}(\varepsilon, k_z, s)$ $\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{R}(\varepsilon,k_z,s)}$. The position vector $\mathbf{R}(\varepsilon, k_z, s)$ consists of two parts:

1. A periodic part $\mathbf{R}_\mathrm{p}(\varepsilon, k_z, s)$:

$$\mathbf{R}_\mathrm{p}\left[\varepsilon, k_z, s + T(\varepsilon, k_z)\right] = \mathbf{R}_\mathrm{p}\left(\varepsilon, k_z, s\right). \tag{13.77}$$

2. A nonperiodic part or secular part $\mathbf{R}_\mathrm{s}(\varepsilon, k_z, s)$:

$$\mathbf{R}_\mathrm{s}(\varepsilon, k_z, s) = \mathbf{v}_\mathrm{s}(\varepsilon, k_z) s. \tag{13.78}$$

Remember that the variable s is time. The $\mathbf{v}_\mathrm{s}$ is the average value of $\mathbf{v}(\varepsilon, k_z, s)$ over one period, and written as

$$\mathbf{v}_\mathrm{s}(\varepsilon, k_z) = \frac{1}{T(\varepsilon, k_z)} \int_t^{t+T} \mathbf{v}(\varepsilon, k_z, t')\, \mathrm{d}t'. \tag{13.79}$$

Then, we have, since $\mathbf{R} = \mathbf{R}_\mathrm{p} + \mathbf{v}_\mathrm{s} s$,

$$\mathbf{v}(\varepsilon, k_z, s) \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{R}(\varepsilon,k_z,s)} = \mathbf{v}(\varepsilon, k_z, s) \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{v}_\mathrm{s}(\varepsilon,k_z)s} \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{R}_\mathrm{p}(\varepsilon,k_z,s)}. \tag{13.80}$$

Thus, for $\underline{\sigma}$ we can write

$$\underline{\sigma} = \frac{2e^3 B_0}{(2\pi)^3 \hbar^2 c} \int_0^\infty \mathrm{d}\varepsilon \left(-\frac{\partial f_0}{\partial \varepsilon}\right) \int_{-\infty}^{\infty} \mathrm{d}k_z \int_0^{T(\varepsilon,k_z)} \mathrm{d}s\, \mathbf{v}(\varepsilon,k_z,s) \mathrm{e}^{-[\frac{1}{\tau}+\mathrm{i}\omega+\mathrm{i}\mathbf{q}\cdot\mathbf{v}_\mathrm{s}(\varepsilon,k_z)]s} \times \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{R}_\mathrm{p}(\varepsilon,k_z,s)} \int_{-\infty}^{s} \mathrm{d}s'\, \mathbf{v}(\varepsilon,k_z,s') \mathrm{e}^{[\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}_\mathrm{s}(\varepsilon,k_z)]s'} \mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{R}_\mathrm{p}(\varepsilon,k_z,s')}. \tag{13.81}$$

We now expand the periodic function $\mathbf{v}(\varepsilon, k_z, s)\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{R}_\mathrm{p}(\varepsilon,k_z,s)}$ in Fourier series in s. Let $\omega_\mathrm{c}(\varepsilon, k_z) = \frac{2\pi}{T(\varepsilon,k_z)}$. Then

$$\mathbf{v}(\varepsilon, k_z, s) \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{R}_\mathrm{p}(\varepsilon,k_z,s)} = \sum_{n=-\infty}^{\infty} \mathbf{v}_n(\varepsilon, k_z) \mathrm{e}^{\mathrm{i}n\omega_\mathrm{c} s}. \tag{13.82}$$

Obviously, the Fourier coefficients $\mathbf{v}_n(\varepsilon, k_z)$ are given by

$$\mathbf{v}_n(\varepsilon, k_z) = \frac{\omega_\mathrm{c}(\varepsilon, k_z)}{2\pi} \int_0^{2\pi/\omega_\mathrm{c}} \mathrm{d}s\, \mathbf{v}(\varepsilon, k_z, s) \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{R}_\mathrm{p}(\varepsilon,k_z,s) - \mathrm{i}n\omega_\mathrm{c} s}. \tag{13.83}$$

We substitute the Fourier expansions, (13.82), into the expression for $\underline{\sigma}$ to obtain

$$\underline{\sigma} = \frac{2e^3 B_0}{(2\pi)^3 \hbar^2 c} \int_0^\infty \mathrm{d}\varepsilon \left(-\frac{\partial f_0}{\partial \varepsilon}\right) \int_{-\infty}^{\infty} \mathrm{d}k_z \int_0^{2\pi/\omega_\mathrm{c}} \mathrm{d}s\, \mathrm{e}^{-[\frac{1}{\tau}+\mathrm{i}\omega+\mathrm{i}\mathbf{q}\cdot\mathbf{v}_\mathrm{s}(\varepsilon,k_z)]s} \times \sum_{n=-\infty}^{\infty} \mathbf{v}_n(\varepsilon, k_z) \mathrm{e}^{\mathrm{i}n\omega_\mathrm{c} s} \int_{-\infty}^{s} \mathrm{d}s'\, \mathrm{e}^{[\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}_\mathrm{s}(\varepsilon,k_z)]s'} \times \sum_{n'=-\infty}^{\infty} \mathbf{v}^*_{n'}(\varepsilon, k_z) \mathrm{e}^{-\mathrm{i}n'\omega_\mathrm{c} s'}. \tag{13.84}$$

First perform the integration over s' to obtain

$$\int_{-\infty}^{s} \mathrm{d}s' \, \mathrm{e}^{[\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}_\mathrm{s}(\varepsilon,k_z)-\mathrm{i}\omega_\mathrm{c}n']s'} = \frac{\mathrm{e}^{[\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}_\mathrm{s}-in'\omega_\mathrm{c}]s}}{\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}_\mathrm{s}-\mathrm{i}n'\omega_\mathrm{c}}. \tag{13.85}$$

Thus, we have

$$\begin{aligned}\underline{\sigma} = {} & \frac{2e^3B_0}{(2\pi)^3\hbar^2c}\int_0^\infty \mathrm{d}\varepsilon\left(-\frac{\partial f_0}{\partial\varepsilon}\right)\int_{-\infty}^{\infty}\mathrm{d}k_z \\ & \times\int_0^{\frac{2\pi}{\omega_\mathrm{c}}}\mathrm{d}s\sum_{n,n'=-\infty}^{\infty}\frac{\mathbf{v}_n(\varepsilon,k_z)\mathbf{v}^*_{n'}(\varepsilon,k_z)\mathrm{e}^{\mathrm{i}(n-n')\omega_\mathrm{c}s}}{\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}_\mathrm{s}-\mathrm{i}n'\omega_\mathrm{c}}.\end{aligned} \tag{13.86}$$

However, $\int_0^{2\pi/\omega_\mathrm{c}}\mathrm{d}s\,\mathrm{e}^{\mathrm{i}(n-n')\omega_\mathrm{c}s} = \frac{2\pi}{\omega_\mathrm{c}}\delta_{nn'}$. Thus, we have

$$\underline{\sigma}=\frac{2e^3B_0}{(2\pi)^3\hbar^2c}\int_0^\infty \mathrm{d}\varepsilon\left(-\frac{\partial f_0}{\partial\varepsilon}\right)\int_{-\infty}^{\infty}\mathrm{d}k_z\frac{2\pi}{\omega_\mathrm{c}(\varepsilon,k_z)}\sum_{n=-\infty}^{\infty}\frac{\mathbf{v}_n(\varepsilon,k_z)\mathbf{v}^*_n(\varepsilon,k_z)}{\frac{1}{\tau}+\mathrm{i}\omega-\mathrm{i}\mathbf{q}\cdot\mathbf{v}_\mathrm{s}-\mathrm{i}n\omega_\mathrm{c}}. \tag{13.87}$$

This can be rewritten

$$\begin{aligned}\underline{\sigma} = {} & \frac{2e^3B_0}{(2\pi)^3\hbar^2c}\int_0^\infty \mathrm{d}\varepsilon\left(-\frac{\partial f_0}{\partial\varepsilon}\right)\tau\int_{-\infty}^{\infty}\mathrm{d}k_z T(\varepsilon,k_z) \\ & \times\sum_{n=-\infty}^{\infty}\frac{\mathbf{v}_n(\varepsilon,k_z)\mathbf{v}^*_n(\varepsilon,k_z)}{1+\mathrm{i}\tau[\omega-\mathbf{q}\cdot\mathbf{v}_\mathrm{s}-\frac{2\pi n}{T(\varepsilon,k_z)}]}.\end{aligned} \tag{13.88}$$

This expression is valid even for the case of open orbits. For closed orbits it is customary to define

$$\omega_\mathrm{c}(\varepsilon,k_z) = \frac{eB_0}{m^*(\varepsilon,k_z)c} = \frac{2\pi}{T(\varepsilon,k_z)}$$

where m^* is the *cyclotron effective mass*. Then, $\underline{\sigma}$ can be written

$$\begin{aligned}\underline{\sigma} = {} & \frac{e^2}{2\pi^2\hbar^2}\int_0^\infty \mathrm{d}\varepsilon\left(-\frac{\partial f_0}{\partial\varepsilon}\right)\tau(\varepsilon)\int_{-\infty}^{\infty}\mathrm{d}k_z m^*(\varepsilon,k_z) \\ & \times\sum_{n=-\infty}^{\infty}\frac{\mathbf{v}_n(\varepsilon,k_z)\mathbf{v}^*_n(\varepsilon,k_z)}{1+\mathrm{i}\tau(\varepsilon)[\omega-\mathbf{q}\cdot\mathbf{v}_\mathrm{s}-n\omega_\mathrm{c}(\varepsilon,k_z)]}.\end{aligned} \tag{13.89}$$

At very low temperatures the integration over energy just picks out the Fermi energy because $-\frac{\partial f_0}{\partial\varepsilon}$ acts just like a δ function, and we have

$$\underline{\sigma} = \frac{e^2}{2\pi^2\hbar^2}\tau(\varepsilon_\mathrm{F})\int_{\mathrm{F.S.}}\mathrm{d}k_z\, m^*(k_z)\sum_{n=-\infty}^{\infty}\frac{\mathbf{v}_n(\varepsilon_\mathrm{F},k_z)\mathbf{v}^*_n(\varepsilon_\mathrm{F},k_z)}{1+\mathrm{i}\tau(\varepsilon_\mathrm{F})[\omega-\mathbf{q}\cdot\mathbf{v}_\mathrm{s}-n\omega_\mathrm{c}(\varepsilon_\mathrm{F},k_z)]}, \tag{13.90}$$

where all quantities are evaluated on the Fermi surface.

13.4.1 Free Electron Model

For the free electron model $m^*(k_z) = m$ is a constant independent of k_z. We shall assume that τ is also constant. The energy $\varepsilon(\mathbf{k})$ and velocity $\mathbf{v}$ are, respectively, given by

$$\begin{aligned} \varepsilon(\mathbf{k}) &= \frac{\hbar^2 k^2}{2m}, \\ \mathbf{v} &= \frac{\hbar \mathbf{k}}{m}, \end{aligned} \tag{13.91}$$

and hence,

$$\begin{aligned} v_z &= \frac{\hbar k_z}{m}, \\ v_\perp = |\mathbf{v}_\perp| &= \sqrt{v^2 - v_z^2} = \frac{\hbar}{m}\left(\tfrac{2m\varepsilon}{\hbar^2} - k_z^2\right)^{1/2}. \end{aligned}$$

We shall choose s, the time along the orbit, such that

$$\begin{aligned} v_x &= v_\perp \cos\omega_c s, \\ v_y &= v_\perp \sin\omega_c s. \end{aligned} \tag{13.92}$$

Thus, for $\mathbf{v}(\varepsilon_F, k_z, s)$ we have

$$\mathbf{v}(\varepsilon, k_z, s) = \frac{\hbar}{m}\left(\sqrt{\frac{2m\varepsilon}{\hbar^2} - k_z^2}\cos\omega_c s, \sqrt{\frac{2m\varepsilon}{\hbar^2} - k_z^2}\sin\omega_c s, k_z\right). \tag{13.93}$$

The periodic part of the position vector is given by

$$\begin{aligned} \mathbf{R}_p(\varepsilon, k_z, s) &= \int \mathbf{v}_\perp(\varepsilon, k_z, s)\mathrm{d}s, \\ &= \frac{v_\perp}{\omega_c}\left(\sin\omega_c s, -\cos\omega_c s, 0\right). \end{aligned} \tag{13.94}$$

Thus, we have

$$\mathrm{i}\mathbf{q}\cdot\mathbf{R}_p(\varepsilon, k_z, s) = \frac{\mathrm{i}v_\perp(\varepsilon, k_z)}{\omega_c}\left(q_x \sin\omega_c s - q_y \cos\omega_c s\right). \tag{13.95}$$

There is no loss in generality incurred by choosing the vector $\mathbf{q}$ to lie in the $y-z$ plane, i.e. $q_x = 0$. Thus

$$\mathrm{i}\mathbf{q}\cdot\mathbf{R}_p(\varepsilon, k_z, s) = -\frac{\mathrm{i}v_\perp(\varepsilon, k_z)}{\omega_c} q_y \cos\omega_c s. \tag{13.96}$$

Now, let us evaluate the Fourier coefficients $\mathbf{v}_n$ in (13.83)

$$\mathbf{v}_n(\varepsilon, k_z) = \frac{\omega_c(\varepsilon, k_z)}{2\pi}\int_0^{2\pi/\omega_c} \mathrm{d}s\, \mathbf{v}(\varepsilon, k_z, s)\mathrm{e}^{-\frac{\mathrm{i}v_\perp(\varepsilon,k_z)q_y}{\omega_c}\cos\omega_c s - \mathrm{i}n\omega_c s}. \tag{13.97}$$

Let $w' = \frac{q_y v_\perp}{\omega_c}$ and $x = \omega_c s$ to have

$$\mathbf{v}_n(\varepsilon, k_z) = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}x\, \mathbf{v}(\varepsilon, k_z, x) \mathrm{e}^{-\mathrm{i}w' \cos x - \mathrm{i}nx}. \tag{13.98}$$

Now, we use the relation

$$\mathrm{e}^{\mathrm{i}w' \sin x} = \sum_{l=-\infty}^{\infty} J_l(w') \mathrm{e}^{\mathrm{i}lx}. \tag{13.99}$$

One can easily see that

$$\mathrm{e}^{-\mathrm{i}w' \cos x} = \mathrm{e}^{\mathrm{i}w' \sin(x + \frac{3\pi}{2})} = \sum_{l=-\infty}^{\infty} J_l(w') \mathrm{e}^{\mathrm{i}lx} \mathrm{e}^{\mathrm{i}l\frac{3\pi}{2}} = \sum_{l=-\infty}^{\infty} (-\mathrm{i})^l J_l(w') \mathrm{e}^{\mathrm{i}lx}. \tag{13.100}$$

Thus, we have

$$\begin{aligned} \mathbf{v}_n(\varepsilon, k_z) &= \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}x\, \mathbf{v}(\varepsilon, k_z, x) \mathrm{e}^{-\mathrm{i}nx} \sum_{l=-\infty}^{\infty} (-\mathrm{i})^l J_l(w') \mathrm{e}^{ilx}, \\ &= \sum_{l=-\infty}^{\infty} (-\mathrm{i})^l J_l(w') \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}x\, \mathbf{v}(\varepsilon, k_z, x) \mathrm{e}^{\mathrm{i}(l-n)x}. \end{aligned} \tag{13.101}$$

Now, let us investigate the individual components of $\mathbf{v}_n$.

$$\begin{aligned} v_{n_x}(\varepsilon, k_z) &= \sum_{l=-\infty}^{\infty} (-\mathrm{i})^l J_l(w') \frac{1}{2\pi} v_\perp \int_0^{2\pi} \mathrm{d}x\, \cos x\, \mathrm{e}^{\mathrm{i}(l-n)x}, \\ &= \sum_{l=-\infty}^{\infty} (-\mathrm{i})^l J_l(w') \frac{v_\perp}{2} \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}x \left[\mathrm{e}^{\mathrm{i}(l-n+1)x} + \mathrm{e}^{\mathrm{i}(l-n-1)x} \right], \\ &= \frac{v_\perp}{2} \sum_{l=-\infty}^{\infty} (-\mathrm{i})^l J_l(w') \left[\delta_{l,n-1} + \delta_{l,n+1} \right], \\ &= \frac{(-\mathrm{i})^{n-1}}{2} v_\perp \left[J_{n-1}(w') - J_{n+1}(w') \right]. \end{aligned} \tag{13.102}$$

In an analogous way, we can obtain

$$v_{n_y}(\varepsilon, k_z) = \frac{(-\mathrm{i})^{n-1}}{2\mathrm{i}} v_\perp \left[J_{n-1}(w') + J_{n+1}(w') \right], \tag{13.103}$$

and

$$v_{n_z}(\varepsilon, k_z) = (-\mathrm{i})^n v_z J_n(w'). \tag{13.104}$$

Here, we recall some properties of Bessel functions:

$$J_n(z) = \sum_{m=0}^{\infty} (-1)^m \left(\frac{z}{2}\right)^{2m+n} \frac{1}{m!(m+n)!}, \tag{13.105}$$

$$J_{-n}(z) = (-1)^n J_n(z), \tag{13.106}$$

$$\lim_{z\to 0} J_n(z) = \frac{1}{n!}\left(\frac{z}{2}\right)^n, \tag{13.107}$$

$$J_n'(z) = \frac{\mathrm{d}}{\mathrm{d}z} J_n(z) = \frac{1}{2}\left[J_{n-1}(z) - J_{n+1}(z)\right], \tag{13.108}$$

and

$$\frac{n}{z} J_n(z) = \frac{1}{2}\left[J_{n-1}(z) + J_{n+1}(z)\right]. \tag{13.109}$$

Using some of these equations we can write the vector $\mathbf{v}_n$ as

$$\mathbf{v}_n(\varepsilon, k_z) = (-\mathrm{i})^n \begin{pmatrix} \mathrm{i}v_\perp J_n'(w') \\ (n\omega_\mathrm{c}/q_y)\, J_n(w') \\ v_z J_n(w') \end{pmatrix}. \tag{13.110}$$

If we take the zero temperature limit, we have $\left(-\frac{\partial f_0}{\partial \varepsilon}\right) = \delta(\varepsilon - \zeta)$ and, hence, we have

$$\underline{\sigma} = \frac{e^2\tau m}{2\pi^2\hbar^2} \int_{-k_\mathrm{F}}^{k_\mathrm{F}} \mathrm{d}k_z \sum_{n=-\infty}^{\infty} \frac{\mathbf{v}_n(\varepsilon_\mathrm{F}, k_z)\mathbf{v}_n^*(\varepsilon_\mathrm{F}, k_z)}{1 - \mathrm{i}\tau[n\omega_\mathrm{c}(\varepsilon_\mathrm{F}, k_z) + \mathbf{q}\cdot\mathbf{v}_\mathrm{s} - \omega]}. \tag{13.111}$$

For free electrons, the secular part of the velocity is simply v_z, thus

$$\begin{aligned} \mathbf{q}\cdot\mathbf{v}_\mathrm{s} &= q_z v_z = q_z v_\mathrm{F}\cos\theta, \\ \mathrm{d}k_z &= \frac{m}{\hbar}\mathrm{d}v_z = \frac{m v_\mathrm{F}}{\hbar}\mathrm{d}(\cos\theta), \\ v_\perp &= \sqrt{v_\mathrm{F}^2 - v_z^2} = v_\mathrm{F}\sin\theta, \\ w' &= \frac{q_y v_\perp}{\omega_\mathrm{c}} = \frac{q_y v_\mathrm{F}}{\omega_\mathrm{c}}\sin\theta = w\sin\theta. \end{aligned} \tag{13.112}$$

By substituting these in (13.111), we obtain

$$\underline{\sigma} = \tfrac{3\sigma_0}{2} \textstyle\sum_{n=-\infty}^{\infty} \int_{-1}^{1} \mathrm{d}(\cos\theta) \times \dfrac{\begin{pmatrix} \mathrm{i}\sin\theta J_n'(w\sin\theta) \\ \frac{n}{w} J_n(w\sin\theta) \\ \cos\theta J_n(w\sin\theta) \end{pmatrix} \left(-\mathrm{i}\sin\theta J_n'(w\sin\theta), \frac{n}{w} J_n(w\sin\theta), \cos\theta J_n(w\sin\theta)\right)}{1 - \mathrm{i}\tau[n\omega_\mathrm{c}(\varepsilon_\mathrm{F}, k_z) - \omega + q_z v_\mathrm{F}\cos\theta]}, \tag{13.113}$$

where $J_n'(x) = \mathrm{d}J_n(x)/\mathrm{d}x$ and σ_0 is the dc conductivity given by $\sigma_0 = \frac{n_0 e^2 \tau}{m}$. In addition, $\sin\theta J_n'(w\sin\theta) = \frac{\partial}{\partial w} J_n(w\sin\theta)$. Hence, we can rewrite

(13.113) as

$$\underline{\sigma} = \frac{3\sigma_0}{2}\sum_{n=-\infty}^{\infty}\int_{-1}^{1}\mathrm{d}(\cos\theta) \times \frac{\begin{pmatrix} \mathrm{i}\frac{\partial}{\partial w}J_n(w\sin\theta) \\ \frac{n}{w}J_n(w\sin\theta) \\ \cos\theta J_n(w\sin\theta)\end{pmatrix}\left(-\mathrm{i}\frac{\partial}{\partial w}J_n(w\sin\theta), \frac{n}{w}J_n(w\sin\theta), \cos\theta J_n(w\sin\theta)\right)}{1-\mathrm{i}\tau[n\omega_\mathrm{c}(\varepsilon_\mathrm{F},k_z)-\omega+q_z v_\mathrm{F}\cos\theta]}. \quad (13.114)$$

This result was first obtained by Cohen, Harrison, and Harrison.[2]

13.4.2 Propagation Parallel to B_0

To acquaint ourselves with the properties of $\underline{\sigma}$, let us first evaluate it for the case $\mathbf{q} \parallel \mathbf{B}_0$, i.e. $\mathbf{q} = (0, 0, q)$. In the limit $q_y \to 0$, $w \to 0$ and, hence, we have

$$\begin{aligned} \frac{n}{w}J_n(w\sin\theta) &\longrightarrow \frac{1}{2}\sin\theta(\delta_{n,1}+\delta_{n,-1}), \\ \mathrm{i}\frac{\partial}{\partial w}J_n(w\sin\theta) &\longrightarrow \frac{\mathrm{i}}{2}\sin\theta(\delta_{n,1}-\delta_{n,-1}), \\ \cos\theta J_n(w\sin\theta) &\longrightarrow \cos\theta\delta_{n,0}. \end{aligned} \quad (13.115)$$

It is easy to see that $\sigma_{xz} = \sigma_{zx} = \sigma_{yz} = \sigma_{zy} = 0$ because of the δ functions involved. The nonvanishing components of $\underline{\sigma}$ are $\sigma_{xx} = \sigma_{yy}$, $\sigma_{xy} = -\sigma_{yx}$ and σ_{zz}. They can easily be evaluated to be written

$$\sigma_{zz} = \frac{3}{2}\sigma_0\int_{-1}^{1}\frac{\mathrm{d}(\cos\theta)\,\cos^2\theta}{1+\mathrm{i}\omega\tau-\mathrm{i}q_z v_\mathrm{F}\tau\cos\theta}, \quad (13.116)$$

and

$$\sigma_\pm = \sigma_{xx} \mp \mathrm{i}\sigma_{xy} = \frac{3}{4}\sigma_0\int_{-1}^{1}\frac{\mathrm{d}(\cos\theta)\,\sin^2\theta}{1+\mathrm{i}(\omega\mp\omega_\mathrm{c})\tau-\mathrm{i}q_z v_\mathrm{F}\tau\cos\theta}. \quad (13.117)$$

Notice that when $q_z \to 0$ the integral in (13.116) becomes $\int_{-1}^{1}\mathrm{d}(\cos\theta)\,\cos^2\theta = \frac{2}{3}$ so that, in that case, we have

$$\sigma_{zz} = \frac{\sigma_0}{1+\mathrm{i}\omega\tau}. \quad (13.118)$$

The $q_z \to 0$ limit of the integral in (13.117) becomes $\int_{-1}^{1}\mathrm{d}(\cos\theta)\,\sin^2\theta = \frac{4}{3}$ so that

$$\sigma_\pm = \frac{\sigma_0}{1+\mathrm{i}(\omega\mp\omega_\mathrm{c})\tau}. \quad (13.119)$$

[2] M.H. Cohen, M.J. Harrison, W.A. Harrison, Phys. Rev. **117**, 937 (1960).

13.4.3 Propagation Perpendicular to B_0

We consider the case $\mathbf{q} \perp \mathbf{B}_0$, i.e. $\mathbf{q} = (0, q, 0)$ to write (13.114) as

$$\underline{\sigma} = \frac{3\sigma_0}{2} \sum_{n=-\infty}^{\infty} \int_{-1}^{1} \mathrm{d}(\cos\theta) \times \frac{\begin{pmatrix} \mathrm{i}\frac{\partial}{\partial w} J_n(w\sin\theta) \\ \frac{n}{w} J_n(w\sin\theta) \\ \cos\theta J_n(w\sin\theta) \end{pmatrix} \left(-\mathrm{i}\frac{\partial}{\partial w} J_n(w\sin\theta), \frac{n}{w} J_n(w\sin\theta), \cos\theta J_n(w\sin\theta)\right)}{1 - \mathrm{i}\tau[n\omega_\mathrm{c}(\varepsilon_\mathrm{F}, k_z) - \omega]}. \tag{13.120}$$

We define the following functions

$$\begin{aligned} c_n(w) &= \frac{1}{2} \int_{-1}^{1} \mathrm{d}(\cos\theta)\, \cos^2\theta J_n^2(w\sin\theta), \\ s_n(w) &= \frac{1}{2} \int_{-1}^{1} \mathrm{d}(\cos\theta)\, \sin^2\theta [J_n'(w\sin\theta)]^2, \\ g_n(w) &= \frac{1}{2} \int_{-1}^{1} \mathrm{d}(\cos\theta)\, J_n^2(w\sin\theta). \end{aligned} \tag{13.121}$$

It is obvious that the angular integrations appearing in σ_{xz}, σ_{zx}, σ_{yz}, and σ_{zy} vanish because the integrands are odd functions of $\cos\theta$. For the nonvanishing components we obtain

$$\sigma_{xx} = 3\sigma_0 \sum_{n=-\infty}^{\infty} \frac{s_n(w)}{1 - \mathrm{i}\tau(n\omega_\mathrm{c} - \omega)}, \tag{13.122}$$

$$\sigma_{yy} = \frac{3\sigma_0}{w^2} \sum_{n=-\infty}^{\infty} \frac{n^2 g_n(w)}{1 - \mathrm{i}\tau(n\omega_\mathrm{c} - \omega)}, \tag{13.123}$$

$$\sigma_{xy} = -\sigma_{yx} = \frac{3\sigma_0 \mathrm{i}}{2w} \sum_{n=-\infty}^{\infty} \frac{n g_n'(w)}{1 - \mathrm{i}\tau(n\omega_\mathrm{c} - \omega)}, \tag{13.124}$$

and

$$\sigma_{zz} = 3\sigma_0 \sum_{n=-\infty}^{\infty} \frac{c_n(w)}{1 - \mathrm{i}\tau(n\omega_\mathrm{c} - \omega)}. \tag{13.125}$$

13.4.4 Local vs. Nonlocal Conduction

What we have been studying is the nonlocal theory of the electrical conductivity of a solid. It is worth emphasizing once again what is meant by nonlocal conduction, and in which case the nonlocal theory reduces to a local theory. The result we have obtained is

$$\mathbf{j}(\mathbf{q}, \omega) = \underline{\sigma}(\mathbf{q}, \omega) \cdot \mathbf{E}(\mathbf{q}, \omega). \tag{13.126}$$

It is easy to show that this expression corresponds to the relation

$$\mathbf{j}(\mathbf{r},t) = \int \underline{K}(\mathbf{r}-\mathbf{r}',t-t')\cdot\mathbf{E}(\mathbf{r}',t')\mathrm{d}^3r'\mathrm{d}t'. \tag{13.127}$$

Let us take the Fourier transforms of each side by multiplying by $\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}-\mathrm{i}\omega t}$ and integrating

$$\begin{aligned}&\int \mathbf{j}(\mathbf{r},t)\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}-\mathrm{i}\omega t}\mathrm{d}^3r\mathrm{d}t\\ &\quad=\int \underline{K}(\mathbf{r}-\mathbf{r}',t-t')\cdot\mathbf{E}(\mathbf{r}',t')\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}'-\mathrm{i}\omega t'}\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')-\mathrm{i}\omega(t-t')}\mathrm{d}^3r'\mathrm{d}t'\mathrm{d}^3r\mathrm{d}t.\end{aligned}$$

The left-hand side is simply $\mathbf{j}(\mathbf{q},\omega)$. The right-hand side can be simplified by letting $\mathbf{r}-\mathbf{r}'=\mathbf{x}$ and $t-t'=s$

$$\mathbf{j}(\mathbf{q},\omega) = \underbrace{\int \underline{K}(\mathbf{x},s)\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{x}-\mathrm{i}\omega s}\mathrm{d}^3x\mathrm{d}s}_{\underline{\sigma}(\mathbf{q},\omega)}\cdot\underbrace{\int \mathbf{E}(\mathbf{r}',t')\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}'-\mathrm{i}\omega t'}\mathrm{d}^3r'\mathrm{d}t'}_{\mathbf{E}(\mathbf{q},\omega)}. \tag{13.128}$$

This is just the relation we gave in (13.126) if

$$\underline{\sigma}(\mathbf{q},\omega) = \int \underline{K}(\mathbf{x},s)\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{x}-\mathrm{i}\omega s}\mathrm{d}^3x\mathrm{d}s$$

or

$$\underline{K}(\mathbf{r}-\mathbf{r}',t-t') = \frac{1}{(2\pi)^4}\int \underline{\sigma}(\mathbf{q},\omega)\mathrm{e}^{\mathrm{i}\omega(t-t')-\mathrm{i}\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')}\mathrm{d}^3q\mathrm{d}\omega. \tag{13.129}$$

Consider for a moment what would happen if $\underline{\sigma}(\mathbf{q},\omega)$ were independent of $\mathbf{q}$. In that case, we have

$$\begin{aligned}\underline{K}(\mathbf{r}-\mathbf{r}',\omega) &= \frac{1}{(2\pi)^3}\int \underline{\sigma}(\omega)\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')}\mathrm{d}^3q,\\ &= \underline{\sigma}(\omega)\,\delta(\mathbf{r}-\mathbf{r}').\end{aligned} \tag{13.130}$$

Thus, we have

$$\mathbf{j}(\mathbf{r},\omega) = \underline{\sigma}(\omega)\cdot\mathbf{E}(\mathbf{r},\omega). \tag{13.131}$$

This is just Ohm's law in the local theory, in which $\mathbf{j}(\mathbf{r})$ depends only on the electric field at the same point $\mathbf{r}$. Thus, the local theory is the special case of the general nonlocal theory, in which the $\mathbf{q}$ dependence of $\underline{\sigma}$ is unimportant. By looking at the expressions we derived one can see that $\underline{\sigma}$ is essentially independent of $\mathbf{q}$ if

1. $ql \ll 1$, in the absence of a magnetic field, where $l = v_\mathrm{F}\tau$ is the electron mean free path.
2. $q_\perp r_\mathrm{c} \ll 1$ or $q_\perp l_0 \ll 1$, and $q_z l_0 \ll 1$, in the presence of a magnetic field. Here r_c is the *radius of the cyclotron orbit.*

13.5 Quantum Theory of Magnetoconductivity of an Electron Gas

The evaluation of $\underline{\sigma}(\mathbf{q},\omega,B_0)$ for a quantum mechanical system is very similar to our evaluation of the wave vector and frequency dependent conductivity in the absence of the field B_0. We will give a very brief summary of the technique here.[3]

The zero-order Hamiltonian for an electron in the presence of a vector potential $\mathbf{A}_0 = (0, xB_0, 0)$ is given by

$$H_0 = \frac{1}{2m}\left[p_x^2 + (p_y + \frac{e}{c}B_0 x)^2 + p_z^2\right]. \tag{13.132}$$

The eigenfunctions and eigenvalues of H_0 can be written as

$$\begin{aligned} |\nu\rangle = |nk_yk_z\rangle &= \frac{1}{L}\mathrm{e}^{\mathrm{i}k_y y+\mathrm{i}k_z z}u_n\left(x+\frac{\hbar k_y}{m\omega_\mathrm{c}}\right), \\ \varepsilon_\nu = \varepsilon_{n,k_y,k_z} &= \frac{\hbar^2 k_z^2}{2m} + \hbar\omega_\mathrm{c}\left(n+\frac{1}{2}\right). \end{aligned} \tag{13.133}$$

Perturbing self-consistent electromagnetic fields $\mathbf{E}(\mathbf{r},t)$ and $\mathbf{B}(\mathbf{r},t)$ are assumed to be of the form $\mathrm{e}^{\mathrm{i}\omega t-\mathrm{i}\mathbf{q}\cdot\mathbf{r}}$. These fields can be derived from the potentials $\mathbf{A}(\mathbf{r},t)$ and $\phi(\mathbf{r},t)$:

$$\begin{aligned} \mathbf{E} &= -\frac{1}{c}\dot{\mathbf{A}} - \nabla\phi = -\frac{i\omega}{c}\mathbf{A} + \mathrm{i}\mathbf{q}\phi, \\ \mathbf{B} &= \nabla\times\mathbf{A} = -\mathrm{i}\mathbf{q}\times\mathbf{A}. \end{aligned} \tag{13.134}$$

As in the Lindhard case, the theory can be shown to be gauge invariant (we will not prove it here but it is done in the references listed earlier). Therefore, we can take a gauge in which the scalar potential $\phi = 0$. Then, we write the linearized (in $\mathbf{A}$) Hamiltonian as

$$H = H_0 + H_1, \tag{13.135}$$

where H_0 is given by (13.132) and H_1 is the perturbing part

$$H_1 = \frac{e}{2c}\left(\mathbf{v}_0\cdot\mathbf{A} + \mathbf{A}\cdot\mathbf{v}_0\right). \tag{13.136}$$

Here, $\mathbf{v}_0 = \frac{1}{m}\left(\mathbf{p} + \frac{e}{c}\mathbf{A}_0\right)$ is the velocity operator in the presence of the field $\mathbf{A}_0$. From here on, one can simply follow the steps we carried out in evaluating $\underline{\sigma}(\mathbf{q},\omega,B_0=0)$. We use

$$\begin{aligned} H_0|\nu\rangle &= \varepsilon_\nu|\nu\rangle, \\ \rho_0|\nu\rangle &= f_0(\varepsilon_\nu)|\nu\rangle. \end{aligned} \tag{13.137}$$

[3] For details one is referred to the references by J.J. Quinn, S. Rodriguez, Phys. Rev. **128**, 2480 (1962) and M.P. Greene, H.J. Lee, J.J. Quinn, S. Rodriguez, Phys. Rev. **177**, 1019 (1969).

The perturbation is given by (13.136) and use that

$$\mathbf{A}(\mathbf{r},t) = \mathbf{A}(\mathbf{q},\omega)\mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}\mathbf{q}\cdot\mathbf{r}}.$$

The resulting expression for $\mathbf{j}(\mathbf{q},\omega)$ can be written (for the collisionless limit where $\tau \to \infty$)

$$\mathbf{j}(\mathbf{q},\omega) = -\frac{\omega_\mathrm{p}^2}{4\pi c}\left[\underline{\mathbf{1}} + \underline{\mathbf{I}}(\mathbf{q},\omega)\right]\cdot\mathbf{A}(\mathbf{q},\omega). \tag{13.138}$$

Here, $\omega_\mathrm{p}^2 = \frac{4\pi n_0 e^2}{m}$ and $n_0 = \frac{N}{V}$. The symbol $\underline{\mathbf{1}}$ stands for the unit tensor and

$$\underline{\mathbf{I}}(\mathbf{q},\omega) = \frac{m}{N}\sum_{\nu\nu'}\frac{f_0(\varepsilon_{\nu'}) - f_0(\varepsilon_\nu)}{\varepsilon_{\nu'} - \varepsilon_\nu - \hbar\omega}\langle\nu'|\mathbf{V}(q)|\nu\rangle\langle\nu'|\mathbf{V}(q)|\nu\rangle^*. \tag{13.139}$$

The operator $\mathbf{V}(q)$ is given by

$$\mathbf{V}(q) = \frac{1}{2}\mathbf{v}_0\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}} + \frac{1}{2}\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}\mathbf{v}_0, \tag{13.140}$$

and $\mathbf{v}_0 = \frac{1}{m}\{\mathbf{p} + \frac{e}{c}\mathbf{A}_0\}$. The matrix elements of $\mathbf{V}(q)$ are given by

$$\begin{aligned}
\langle\nu'|V_z(q)|\nu\rangle &= \delta(k_y', k_y + q_y)\delta(k_z', k_z + q_z)\frac{\hbar}{m}\left(k_z + \frac{q_z}{2}\right)f_{n'n}(q_y),\\
\langle\nu'|V_y(q)|\nu\rangle &= \delta(k_y', k_y+q_y)\delta(k_z', k_z+q_z)\left\{\frac{\hbar q_y}{2m}f_{n'n}(q_y) + \left(\frac{\hbar\omega_\mathrm{c}}{2m}\right)^{1/2}X_{n'n}^{(+)}(q_y)\right\},\\
\langle\nu'|V_x(q)|\nu\rangle &= \delta(k_y', k_y + q_y)\delta(k_z', k_z + q_z)\left\{\mathrm{i}\left(\frac{\hbar\omega_\mathrm{c}}{2m}\right)^{1/2}X_{n'n}^{(-)}(q_y)\right\}
\end{aligned} \tag{13.141}$$

In these equations, we have taken $q_x = 0$; this can be done without loss of generality. The function $f_{n'n}(q_y)$ is the two-center harmonic oscillator integral:

$$f_{n'n}(q_y) = \int_{-\infty}^{\infty} u_{n'}(x + \frac{\hbar q_y}{m\omega_\mathrm{c}})u_n(x)\mathrm{d}x, \tag{13.142}$$

and

$$X_{n'n}^{(\pm)}(q_y) = (n+1)^{1/2}f_{n',n+1}(q_y) \pm n^{1/2}f_{n',n-1}(q_y). \tag{13.143}$$

The function $f_{n'n}$ also appears in another useful matrix element

$$\langle\nu'|\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{r}}|\nu\rangle = \delta(k_y', k_y + q_y)\delta(k_z', k_z + q_z)f_{n'n}(q_y). \tag{13.144}$$

The function $f_{n'n}$ can be evaluated in terms of associated Laguerre polynomials. For $n' \geq n$, we have

$$f_{n'n}(q) = \left(\frac{n!}{n'!}\right)^{1/2}\xi^{(n'-n)/2}\mathrm{e}^{-\xi/2}L_n^{n'-n}(\xi), \tag{13.145}$$

where $\xi = \frac{\hbar q^2}{2m\omega_c}$ and $L_n^\alpha(\xi)$ is the associated Laguerre polynomial of order n. For $n' < n$ we have $f_{n'n}(q) = (-1)^{n-n'} f_{nn'}(q)$. Some useful facts about the functions $f_{n'n}$ are

$$\begin{aligned}
f_{n'n}(q) &= i^{n-n'} \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{\mathrm{i}qx} u_{n'}(x) u_n(x), \\
f_{n'n}(-q) &= f_{nn'}(q) = (-1)^{n'-n} f_{n'n}(q), \\
\sum_{n=0}^{\infty} f_{n'n}^2(q) &= 1, \\
\sum_{n'=0}^{\infty} (n'-n) f_{n'n}^2(q) &= \xi, \\
\frac{\partial f_{n'n}}{\partial q} &= \left(\frac{\hbar}{2m\omega_c}\right)^{1/2} X_{n'n}^{(-)}(q), \\
(n'-n-\xi) f_{n'n}(q) &= \xi^{1/2} X_{n'n}^{(+)}(q).
\end{aligned} \tag{13.146}$$

13.5.1 Propagation Perpendicular to $\mathbf{B_0}$

For the case of $\mathbf{q} = (0, q, 0)$ we find that

$$\begin{aligned}
j_x(q,\omega) &= \sigma_{xx} E_x(q,\omega) + \sigma_{xy} E_y(q,\omega), \\
j_y(q,\omega) &= \sigma_{yx} E_x(q,\omega) + \sigma_{yy} E_y(q,\omega), \\
j_z(q,\omega) &= \sigma_{zz} E_z(q,\omega).
\end{aligned} \tag{13.147}$$

The nonvanishing components of $\underline{\sigma}$ can be evaluated and they are written as

$$\begin{aligned}
\sigma_{xx} &= \frac{\omega_p^2}{4\pi\mathrm{i}\omega} \left[1 - \frac{2m\omega_c}{\hbar} \frac{1}{N} \sum_{nk_yk_z\alpha}{}' f_0(\varepsilon_{nk_z}) \left(\frac{\partial f_{n+\alpha,n}}{\partial q} \right)^2 \frac{\alpha}{\alpha^2 - (\omega/\omega_c)^2} \right], \\
\sigma_{yy} &= \frac{\mathrm{i}m\omega_p^2 \omega}{2\pi\hbar\omega_c} \frac{1}{N} \sum_{nk_yk_z\alpha}{}' f_0(\varepsilon_{nk_z}) f_{n+\alpha,n}^2 \frac{\alpha}{\alpha^2 - (\omega/\omega_c)^2}, \\
\sigma_{xy} &= -\sigma_{yx} = \frac{\mathrm{i}\omega_c}{2\omega q} \frac{\partial(q^2 \sigma_{yy})}{\partial q}, \\
\sigma_{zz} &= \frac{\omega_p^2}{4\pi\mathrm{i}\omega} \left[1 - \frac{2\hbar}{m\omega_c} \frac{1}{N} \sum_{nk_yk_z\alpha}{}' f_0(\varepsilon_{nk_z}) k_z^2 f_{n+\alpha,n}^2 \frac{\alpha}{\alpha^2 - (\omega/\omega_c)^2} \right].
\end{aligned} \tag{13.148}$$

In these equations, the sum on α is to be performed from $-n$ to ∞ (because $0 \le n' = n + \alpha \le \infty$). The summations over n, k_y, k_z extend over all values of the quantum numbers for which $\varepsilon_{nk_z} \le \zeta$, where ζ is the chemical potential of the electron gas in the field B_0. This restriction is indicated by a prime following the summation sign.

The semiclassical limit can be obtained by replacing the sum over n by an integral. Remember that in general we can write

$$\sum_{nk_yk_z} \Longrightarrow 2\left(\frac{L}{2\pi}\right)^2 \sum_n \int_0^{\frac{m\omega_c L}{\hbar}} \mathrm{d}k_y \int \mathrm{d}k_z = \frac{m\omega_c}{\hbar}\frac{\Omega}{2\pi^2}\int \mathrm{d}k_z \sum_n, \quad (13.149)$$

where Ω is the volume of the sample $\Omega = L^3$. We define $n_0 = \frac{\rho}{\hbar\omega_c} - \frac{1}{2}$ and let $n = n_0 \sin^2\theta$. For zero temperature, we can integrate over θ from $\theta = 0$ to $\theta = \frac{\pi}{2}$ instead of summing over n. For $n_0 \gg 1$ it is not hard to see that the main contribution to the integrals comes from rather large values of n. For large n, we can approximate

$$f_{n+\alpha,n} \simeq J_\alpha\left[(4n+2\alpha+2)^{1/2}\xi^{1/2}\right] \simeq J_\alpha(w\sin\theta), \quad (13.150)$$

where $w = \frac{qv_F}{\omega_c}$. By substituting into the expressions for the components of $\underline{\sigma}$ we obtain

$$\sigma_{xx} = \frac{3\alpha}{4\pi \mathrm{i}\omega}\sum_{\alpha=-\infty}^{\infty}\frac{s_\alpha(w)}{1+\alpha\omega_c/\omega}, \quad (13.151)$$

where

$$s_\alpha(w) = \int_0^{\pi/2} \mathrm{d}\theta \sin^3\theta \left[J'_\alpha(w\sin\theta)\right]^2. \quad (13.152)$$

Equation (13.151) is the semiclassical expression we already obtained in (13.122) in the collisionless limit.

The quantum mechanical conductivity tensor can be written as the sum of a semiclassical term and a quantum oscillatory part

$$\underline{\sigma}(q,\omega) = \underline{\sigma}^{\mathrm{SC}}(q,\omega) + \underline{\sigma}^{\mathrm{QO}}(q,\omega), \quad (13.153)$$

where the semiclassical part $\underline{\sigma}^{\mathrm{SC}}$ has been given earlier. As an example of the quantum oscillatory part we give, without derivation, one example

$$\sigma_{zz}^{\mathrm{QO}} = \frac{3}{2}\delta_2\frac{\omega_p^2}{4\pi\mathrm{i}\omega}\left[1+3\frac{\omega^2}{\omega_c^2}\sum_{\alpha=-\infty}^{\infty}\frac{1}{\alpha^2-(\omega/\omega_c)^2}\left(1+w\frac{\partial}{\partial w}\right)c_\alpha(w)\right]. \quad (13.154)$$

Here, δ_2 is a quantum oscillatory function of the *de Haas–van Alphen* type and is given by

$$\delta_2 = \pi\left(\frac{k_B T}{\zeta_0}\right)\sqrt{\frac{\hbar\omega_c}{2\zeta_0}}\sum_{\nu=1}^{\infty}\frac{(-1)^\nu \nu^{-1/2}\sin(\frac{2\pi\nu\zeta_0}{\hbar\omega_c}-\frac{\pi}{4})}{\sinh(\frac{2\pi^2\nu k_B T}{\hbar\omega_c})}. \quad (13.155)$$

The function $c_\alpha(w)$, with $w = \frac{qv_F}{\omega_c}$, was defined by (13.121) in the discussion of the semiclassical conductivity. If $k_B T$ becomes large compared to $\hbar\omega_c$, the amplitude of the quantum oscillations becomes negligibly small and $\underline{\sigma}$ reduces

to the semiclassical result $\underline{\sigma}^{\mathrm{SC}}$ given previously. What the quantum mechanical conductivity tensor contains but what the semiclassical one does not is the quantum structure of the energy levels. This, of course, determines all the *quantum effects* like

1. *de Haas–van Alphen oscillations* in the magnetism
2. *Shubnikov–de Haas oscillations* in the resistivity
3. Quantum oscillations in acoustic attenuation, etc.

Problems

13.1. Take direction of current flow to make an angle θ with x axis as is shown in Fig. 13.8. First, transform to $x' - y'$ frame. Then, put $j_{y'} = 0$, and check for what angles θ the magnetoresistance fails to saturate.

13.2. Consider a band for a simple cubic structure given by $E(\mathbf{k}) = E_0[\cos(k_x a) + \cos(k_y a) + \cos(k_z a)]$, where a is the lattice constant. Let an electron at rest ($k = 0$) at $t = 0$ feel a uniform external electric field $\boldsymbol{E}$ which is constant in time.

(a) Find the real space trajectory $[x(t), y(t), z(t)]$.
(b) Sketch the trajectory for $\boldsymbol{E}$ in a [120] direction.

13.3. Consider an electron in a state with a linear energy dispersion given by $E = \pm\hbar v_{\mathrm{F}}|\mathbf{k}|$, where $\mathbf{k}$ is a two-dimensional wave vector. (It occurs in the low-energy states in graphene – a single layer of graphite.)

(a) When a dc magnetic field $\mathbf{B}$ is applied perpendicular to the graphene layer, write down the area $\mathcal{S}(\varepsilon)$ and sketch $\mathcal{S}(\varepsilon_n)$ for various values of n.
(b) Solve for the quantized energies ε_n and plot the resulting ε_n for $-5\hbar\omega_{\mathrm{c}} \leq \varepsilon_n \leq 5\hbar\omega_{\mathrm{c}}$.
(c) What can you say about the effective mass of the particle in a graphene subject to the magnetic field $\mathbf{B}$?

13.4. The energy of an electron in a particular band of a solid is given by

$$\varepsilon(k_x, k_y, k_z) = \frac{\hbar^2 k_x^2}{2m_x} + \frac{\hbar^2 k_y^2}{2m_y},$$

where $-\frac{\pi}{a} < k_i < \frac{\pi}{a}$ is the first Brillouin zone of a simple cubic lattice.

(a) Determine $v_i(\mathbf{k})$ for $i = x, y$, and z.
(b) Show that $\hbar k_i(t) = \sqrt{2m_i\varepsilon}\cos\omega_{\mathrm{c}}t$ where $i = x$ and y for a dc magnetic field $\mathbf{B}_0$ in the z-direction.
(c) Determine ω_{c} in terms of m_i, B_0, etc.

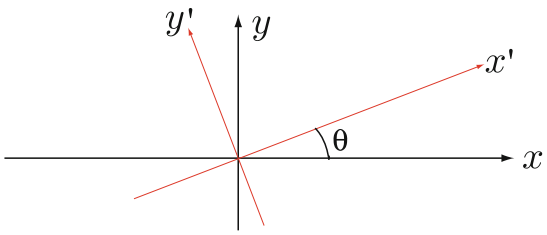


Fig. 13.8. A simple geometry of current flow

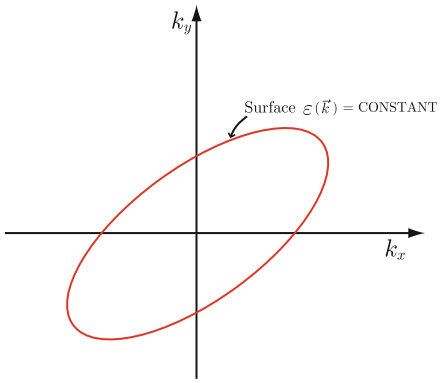


Fig. 13.9. A constant energy surface $\varepsilon(\mathbf{k})$ in a two-dimensional system

13.5. Consider two-dimensional electrons with a linear dispersion given by $E = \hbar v_{\mathrm{F}}|\mathbf{k}|$, where $\mathbf{k}$ is a two-dimensional wave vector. Now apply a dc magnetic field B perpendicular to the system. We shall assume that τ is constant.

(a) Write down the $\mathbf{v}(\varepsilon_{\mathrm{F}}, s)$ and the periodic part of the position vector $\mathbf{R}_{\mathrm{p}}(\varepsilon, s)$.
(b) Evaluate the Fourier coefficients $\mathbf{v}_n(\varepsilon)$, and discuss the conductivity tensor $\underline{\sigma}$ defined by $\mathbf{j} = \underline{\sigma} \cdot \mathbf{E}$.

13.6. Consider an electron in a two-dimensional system subject to a dc magnetic field B perpendicular to the system. The constant energy surface of the particle is shown in in Fig. 13.9.

(a) Sketch the orbit of the particle in real space.
(b) Sketch the velocity $v_y(t)$ as a function of t.

Summary

In this chapter, we study behaviors of Bloch electrons in the presence of a dc magnetic field. Energy levels and possible trajectories of electrons are discussed, and simple two band model of magnetoresistance is illustrated including the effect of collisions. General expression of semiclassical magnetoconductivity tensor is derived by solving the Boltzmann equation of the distribution function, and the results are applied to the case of free electrons. The relationship between the local and nonlocal descriptions are discussed. Finally, quantum mechanical theory of magnetoconductivity tensor is described and quantum oscillatory behavior in magnetoconductivity of Bloch electrons is compared with its semiclassical counterpart.

In the presence of an electric field $\mathbf{E}$ and a dc magnetic field $\mathbf{B}(=\nabla\times\mathbf{A})$, an effective Hamiltonian is given by $\mathcal{H}=\varepsilon(\frac{\mathbf{p}}{\hbar}+\frac{e}{\hbar c}\mathbf{A})-e\phi$, where $\varepsilon(\mathbf{k})$ is the energy as a function of $\mathbf{k}$ in the absence of $\mathbf{B}$. The equation of motion of a Bloch electron in k-space takes the form

$$\hbar\dot{\mathbf{k}}=-e\mathbf{E}-\frac{e}{c}\mathbf{v}\times\mathbf{B}.$$

Here, $\mathbf{v}=\frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon(\mathbf{k})$ is the velocity of the Bloch electron whose energy $\varepsilon(\mathbf{k})$ is an arbitrary function of wave vector $\mathbf{k}$. The orbit in real space will be exactly the same shape as the orbit in k-space except that it is rotated by 90° and scaled by a factor $\frac{eB}{\hbar c}$: $\mathbf{k}_\perp=\frac{e\mathbf{B}}{\hbar c}\times\mathbf{r}_\perp$. The factor $\frac{eB}{\hbar c}$ is l_0^{-2}, where l_0 is the *magnetic length*. The orbit of a particle in $\boldsymbol{k}$-space is the intersection of a constant energy surface $\varepsilon(\boldsymbol{k})=\varepsilon$ and a plane of constant k_z:

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}t}=\nabla_k\varepsilon\cdot\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t}=\hbar\mathbf{v}\cdot\left(-\frac{e}{\hbar c}\mathbf{v}\times\mathbf{B}\right)=0.$$

The area of the orbit $\mathcal{A}(\varepsilon,k_z)$ in real space is proportional to the area $\mathcal{S}(\varepsilon,k_z)$ of the orbit in $\boldsymbol{k}$-space: $\mathcal{S}(\varepsilon,k_z)=\left(\frac{eB}{\hbar c}\right)^2\mathcal{A}(\varepsilon,k_z)$. The area $\mathcal{S}(\varepsilon,k_z)$ is quantized by $\mathcal{S}(\varepsilon,k_z)=\frac{2\pi eB}{\hbar c}(n+\gamma)$ and the cyclotron effective mass is given by $m^*(\varepsilon,k_z)=\frac{\hbar^2}{2\pi}\frac{\partial S(\varepsilon,k_z)}{\partial\varepsilon}$. The Bloch electron velocity parallel to the magnetic field becomes

$$v_z(\varepsilon,k_z)=-\frac{\hbar}{2\pi m^*(\varepsilon,k_z)}\frac{\partial S(\varepsilon,k_z)}{\partial k_z}.$$

The transverse magnetoresistance is defined by $\frac{R(B_z)-R(0)}{R(0)}=\Delta R(B_z)$. The simple free electron model gives $\Delta R(B_z)=0$, which is different from the experimental results.

The current density is given by $\mathbf{j}(\mathbf{r},t)=\frac{2}{(2\pi)^3}\int(-e)\mathbf{v}f_1\,\mathrm{d}^3k$. In the presence of a uniform dc magnetic field $\mathbf{B}_0$, the semiclassical magnetoconductivity of an electron gas is written as

$$\underline{\sigma}=\frac{e^2}{2\pi^2\hbar^2}\tau(\varepsilon_\mathrm{F})\int_{\mathrm{F.S.}}\mathrm{d}k_z\,m^*(k_z)\sum_{n=-\infty}^{\infty}\frac{\mathbf{v}_n(\varepsilon_\mathrm{F},k_z)\mathbf{v}_n^*(\varepsilon_\mathrm{F},k_z)}{1+\mathrm{i}\tau(\varepsilon_\mathrm{F})[\omega-\mathbf{q}\cdot\mathbf{v}_\mathrm{s}-n\omega_\mathrm{c}(\varepsilon_\mathrm{F},k_z)]},$$

where $\mathbf{v}_n(\varepsilon, k_z)$ is defined by

$$\mathbf{v}_n(\varepsilon, k_z) = \frac{\omega_\mathrm{c}(\varepsilon, k_z)}{2\pi} \int_0^{2\pi/\omega_\mathrm{c}} \mathrm{d}s\, \mathbf{v}(\varepsilon, k_z, s) \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{R}_\mathrm{p}(\varepsilon,k_z,s) - \mathrm{i}n\omega_\mathrm{c}s}.$$

Here, $R_\mathrm{p}(\varepsilon, k_z, s)$ denotes the periodic part of the position vector in real space.

For the free electron model $m^*(k_z) = m$ is a constant independent of k_z and the periodic part of the position vector is given by $\mathbf{R}_\mathrm{p}(\varepsilon, k_z, s) = \frac{v_\perp}{\omega_\mathrm{c}} (\sin\omega_\mathrm{c}s, -\cos\omega_\mathrm{c}s, 0)$. For the propagation $\mathbf{q} \perp \mathbf{B}_0$, i.e. $\mathbf{q} = (0, q, 0)$, the nonvanishing components of semiclassical conductivity $\underline{\sigma}$ are

$$\sigma_{xx} = 3\sigma_0 \sum_{n=-\infty}^{\infty} \frac{s_n(w)}{1 - \mathrm{i}\tau(n\omega_\mathrm{c} - \omega)};\ \sigma_{yy} = \frac{3\sigma_0}{w^2} \sum_{n=-\infty}^{\infty} \frac{n^2 g_n(w)}{1 - \mathrm{i}\tau(n\omega_\mathrm{c} - \omega)};$$

$$\sigma_{xy} = -\sigma_{yx} = \frac{3\sigma_0 \mathrm{i}}{2w} \sum_{n=-\infty}^{\infty} \frac{n g_n'(w)}{1 - \mathrm{i}\tau(n\omega_\mathrm{c} - \omega)};\ \sigma_{zz} = 3\sigma_0 \sum_{n=-\infty}^{\infty} \frac{c_n(w)}{1 - \mathrm{i}\tau(n\omega_\mathrm{c} - \omega)}.$$

Here, $c_n(w) = \frac{1}{2}\int_{-1}^{1} \mathrm{d}(\cos\theta)\, \cos^2\theta J_n^2(w\sin\theta)$, $s_n(w) = \frac{1}{2}\int_{-1}^{1} \mathrm{d}(\cos\theta)\, \sin^2\theta\, [J_n'(w\sin\theta)]^2$, and $g_n(w) = \frac{1}{2}\int_{-1}^{1} \mathrm{d}(\cos\theta)\, J_n^2(w\sin\theta)$.

In the presence of a vector potential $\mathbf{A}_0 = (0, xB_0, 0)$, the electronic states are described by $H_0 = \frac{1}{2m}\left[p_x^2 + (p_y + \frac{e}{c}B_0 x)^2 + p_z^2\right]$. The eigenfunctions and eigenvalues of H_0 can be written as

$$|\nu\rangle = |n k_y k_z\rangle = \frac{1}{L}\mathrm{e}^{\mathrm{i}k_y y + \mathrm{i}k_z z} u_n\left(x + \frac{\hbar k_y}{m\omega_\mathrm{c}}\right),$$

$$\varepsilon_\nu = \varepsilon_{n,k_y,k_z} = \frac{\hbar^2 k_z^2}{2m} + \hbar\omega_\mathrm{c}(n + \frac{1}{2}).$$

The quantum mechanical version of the nonvanishing components of $\underline{\sigma}$ are given, for the case of $\mathbf{q} = (0, q, 0)$, by

$$\sigma_{xx} = \frac{\omega_\mathrm{p}^2}{4\pi\mathrm{i}\omega}\left[1 - \frac{2m\omega_\mathrm{c}}{\hbar}\frac{1}{N}{\sum_{n k_y k_z \alpha}}' f_0(\varepsilon_{nk_z})\left(\frac{\partial f_{n+\alpha,n}}{\partial q}\right)^2 \frac{\alpha}{\alpha^2 - (\omega/\omega_\mathrm{c})^2}\right],$$

$$\sigma_{yy} = \frac{\mathrm{i}m\omega_\mathrm{p}^2\omega}{2\pi\hbar\omega_\mathrm{c}}\frac{1}{N}{\sum_{n k_y k_z \alpha}}' f_0(\varepsilon_{nk_z}) f_{n+\alpha,n}^2 \frac{\alpha}{\alpha^2 - (\omega/\omega_\mathrm{c})^2},$$

$$\sigma_{xy} = -\sigma_{yx} = \frac{\mathrm{i}\omega_\mathrm{c}}{2\omega q}\frac{\partial(q^2\sigma_{yy})}{\partial q},$$

$$\sigma_{zz} = \frac{\omega_\mathrm{p}^2}{4\pi\mathrm{i}\omega}\left[1 - \frac{2\hbar}{m\omega_\mathrm{c}}\frac{1}{N}{\sum_{n k_y k_z \alpha}}' f_0(\varepsilon_{nk_z}) k_z^2 f_{n+\alpha,n}^2 \frac{\alpha}{\alpha^2 - (\omega/\omega_\mathrm{c})^2}\right],$$

where $f_{n'n}(q_y)$ is the two-center harmonic oscillator integral:

$$f_{n'n}(q_y) = \int_{-\infty}^{\infty} u_{n'}(x + \frac{\hbar q_y}{m\omega_c}) u_n(x) \mathrm{d}x$$

and

$$X_{n'n}^{(\pm)}(q_y) = (n+1)^{1/2} f_{n',n+1}(q_y) \pm n^{1/2} f_{n',n-1}(q_y).$$

The quantum mechanical conductivity tensor is the sum of a semiclassical term and a quantum oscillatory part:

$$\underline{\sigma}(q,\omega) = \underline{\sigma}^{\mathrm{SC}}(q,\omega) + \underline{\sigma}^{\mathrm{QO}}(q,\omega).$$

14

Electrodynamics of Metals

14.1 Maxwell's Equations

There are two aspects of the electrodynamics of metals. The first is linear response theory and the second is the problem of boundary conditions. We have already discussed linear response theory in some detail. Its application to waves in an infinite medium is fairly straightforward. The problem of boundary conditions is usually much more involved. We shall cover some examples of each type in the rest of this chapter.

We consider an electromagnetic disturbance with space–time dependence of the form $\mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}\mathbf{q}\cdot\mathbf{r}}$. Maxwell's equations can be written

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \tag{14.1}$$

and

$$\nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{j}_{\mathrm{T}} + 4\pi\nabla \times \mathbf{M}_{\mathrm{s}}. \tag{14.2}$$

In (14.2) $\mathbf{j}_{\mathrm{T}}$ is the total current in the system; it includes any external currents and the diamagnetic response currents in the medium. The term M_{s} is the spin magnetization in the case of a system containing spins. Equation (14.1) can be written

$$\mathbf{B} = \boldsymbol{\xi} \times \mathbf{E}, \tag{14.3}$$

where $\boldsymbol{\xi} = \frac{c\mathbf{q}}{\omega}$. Therefore, the magnetic induction $\mathbf{B}$ can be eliminated from (14.2):

$$\boldsymbol{\xi} \times (\boldsymbol{\xi} \times \mathbf{E}) + \mathbf{E} = \frac{4\pi\mathrm{i}}{\omega}\mathbf{j}_{\mathrm{T}} + \frac{4\pi}{c}\boldsymbol{\xi} \times \mathbf{M}_{\mathrm{s}},$$

or

$$\boldsymbol{\xi}(\boldsymbol{\xi} \cdot \mathbf{E}) - \xi^2\mathbf{E} + \mathbf{E} = \frac{4\pi\mathrm{i}}{\omega}\mathbf{j}_{\mathrm{T}} + \frac{4\pi}{c}\boldsymbol{\xi} \times \mathbf{M}_{\mathrm{s}}. \tag{14.4}$$

Normally, the total current $\mathbf{j}_{\mathrm{T}}$ can be written as

$$\mathbf{j}_{\mathrm{T}} = \mathbf{j}_0 + \mathbf{j}_{\mathrm{ind}}, \tag{14.5}$$

where $\mathbf{j}_0$ is some external current and $\mathbf{j}_{\rm ind}$ is the current induced (in the electron gas) by the self-consistent field. The spin magnetization $\mathbf{M}_{\rm s}$ and the induced current $\mathbf{j}_{\rm ind}$ are found in terms of the self-consistent fields $\mathbf{E}$ and $\mathbf{B}$ from linear response theory. For example $\mathbf{j}_{\rm ind}$ might simply be the electron current density

$$\mathbf{j}_{\rm e} = \underline{\sigma} \cdot \mathbf{E}, \tag{14.6}$$

and the spin magnetization $\mathbf{M}_{\rm s}$ will be some similar function of $\mathbf{B}$

$$\mathbf{M}_{\rm s} = \underline{\alpha} \cdot \mathbf{B}. \tag{14.7}$$

For the moment, let us ignore the effect of spin to drop the term $\mathbf{M}_{\rm s}$. Then, (14.4) can be solved for $\mathbf{j}_{\rm T}$.

$$\mathbf{j}_{\rm T} = \underline{\Gamma} \cdot \mathbf{E}, \tag{14.8}$$

where

$$\underline{\Gamma} = \frac{{\rm i}\omega}{4\pi} \left\{ (\xi^2 - 1)\underline{\mathbf{1}} - \boldsymbol{\xi}\boldsymbol{\xi} \right\}. \tag{14.9}$$

If we choose $q_x = 0$ (as we did in linear response theory) $\underline{\Gamma}$ can be written

$$\underline{\Gamma} = \frac{{\rm i}c^2}{4\pi\omega} \begin{pmatrix} q_y^2 + q_z^2 - \frac{\omega^2}{c^2} & 0 & 0 \\ 0 & q_z^2 - \frac{\omega^2}{c^2} & -q_y q_z \\ 0 & -q_y q_z & q_y^2 - \frac{\omega^2}{c^2} \end{pmatrix}. \tag{14.10}$$

Notice that $\underline{\Gamma}$ is diagonal for propagation parallel or perpendicular to the dc magnetic field (which we take to be in the z-direction).

14.2 Skin Effect in the Absence of a dc Magnetic Field

Consider a semi-infinite metal to fill the space $z > 0$ and vacuum the space $z < 0$. Let us consider the propagation of an electromagnetic wave parallel to the z-axis. Electromagnetic radiation is a self-sustaining oscillation of any medium in which it propagates. Therefore, we need no external "driving" current $\mathbf{j}_0$, and the total current is simply the electronic current

$$\mathbf{j}_{\rm e}(\mathbf{q}, \omega) = \underline{\sigma}(\mathbf{q}, \omega) \cdot \mathbf{E}(\mathbf{q}, \omega). \tag{14.11}$$

But Maxwell's equations require that $\mathbf{j}_{\rm T} = \underline{\Gamma} \cdot \mathbf{E}$, and we have just seen that $\mathbf{j}_{\rm T} = \mathbf{j}_{\rm e}$. Therefore, the electromagnetic waves must be solutions of the secular equation

$$| \underline{\Gamma} - \underline{\sigma} | = 0, \tag{14.12}$$

which can be written

$$\begin{vmatrix} -\frac{c^2q^2}{\omega^2}+\varepsilon(\mathbf{q},\omega) & 0 & 0 \\ 0 & -\frac{c^2q^2}{\omega^2}+\varepsilon(\mathbf{q},\omega) & 0 \\ 0 & 0 & \varepsilon(\mathbf{q},\omega) \end{vmatrix} = 0. \tag{14.13}$$

Here, we have introduced the dielectric function

$$\underline{\varepsilon}(\mathbf{q},\omega) = \underline{\mathbf{1}} - \frac{4\pi \mathrm{i}}{\omega}\underline{\sigma}(\mathbf{q},\omega), \tag{14.14}$$

and we have assumed that $\underline{\varepsilon}$ is diagonal (this is true for an electron gas in the absence of a dc magnetic field). The transverse electromagnetic waves which can propagate in the medium are solutions of the equation

$$c^2q^2 = \omega^2\varepsilon(\mathbf{q},\omega). \tag{14.15}$$

In addition, there is a longitudinal wave which is the solution of the equation

$$\varepsilon(\mathbf{q},\omega) = 0. \tag{14.16}$$

Normal Skin Effect

In the absence of a dc magnetic field, the local theory of conduction gives

$$\sigma = \frac{\sigma_0}{1+\mathrm{i}\omega\tau} = \frac{\omega_\mathrm{p}^2\tau/4\pi}{1+\mathrm{i}\omega\tau}. \tag{14.17}$$

Therefore, we have

$$\varepsilon(\mathbf{q},\omega) = 1 - \frac{\mathrm{i}\omega_\mathrm{p}^2\tau/\omega}{1+\mathrm{i}\omega\tau} \tag{14.18}$$

or

$$\varepsilon(\mathbf{q},\omega) = \frac{1+(\omega^2-\omega_\mathrm{p}^2)\tau^2}{1+\omega^2\tau^2} - \mathrm{i}\frac{\omega_\mathrm{p}^2\tau^2}{\omega\tau(1+\omega^2\tau^2)} \tag{14.19}$$

Usually in a good metal $\omega_\mathrm{p} \simeq 10^{16}/\mathrm{s}$, a frequency in the ultraviolet. Therefore, in the optical or infrared range $\omega_\mathrm{p} \gg \omega$. The parameter τ can be as small as 10^{-14}s or as large as 10^{-9}s in very pure metals at very low temperatures. Let us first consider the case $\omega\tau \gg 1$. Then, since $\omega_\mathrm{p} \gg \omega$, we have

$$\varepsilon(\mathbf{q},\omega) \simeq -\frac{\omega_\mathrm{p}^2}{\omega^2}. \tag{14.20}$$

Substituting this result into the wave equation $c^2q^2 = \omega^2\varepsilon(q,\omega)$ gives

$$q = \pm\mathrm{i}\frac{\omega_\mathrm{p}}{c} = \pm\frac{\mathrm{i}}{\delta}. \tag{14.21}$$

We choose the well-behaved solution $q = -\frac{\mathrm{i}}{\delta}$ so that the field in the metal is of the form

$$E(z,t) = E_0 \mathrm{e}^{\mathrm{i}\omega t - z/\delta}. \tag{14.22}$$

What we find is that electromagnetic waves do not propagate in the metal (for frequencies lower than ω_p), and that the electric field in the solid drops off exponentially with distance from the surface. The distance $\delta = \frac{c}{\omega_\mathrm{p}}$ is called the *normal skin depth*.

If $\omega\tau \ll 1$, (this is usually true at rf frequencies, even at low temperatures with pure materials) we have

$$\varepsilon(\mathbf{q},\omega) \simeq 1 - \mathrm{i}\frac{\omega_\mathrm{p}^2\tau}{\omega} \sim -\mathrm{i}\frac{\omega_\mathrm{p}^2\omega\tau}{\omega^2} \text{ when } \omega_\mathrm{p} \gg \omega. \tag{14.23}$$

The solution of the wave equation is given by

$$q = \pm\frac{\omega_\mathrm{p}}{c}\left(\frac{\omega\tau}{2}\right)^{1/2}(1-\mathrm{i}), \tag{14.24}$$

so that the field E is of the form

$$E(z,t) = E_0 \mathrm{e}^{\mathrm{i}\omega t}\mathrm{e}^{-(\mathrm{i}+1)\frac{\omega_\mathrm{p}}{c}\left(\frac{\omega\tau}{2}\right)^{1/2}z}. \tag{14.25}$$

Thus, the skin depth is given by

$$\delta = \frac{c}{\omega_\mathrm{p}}\left(\frac{2}{\omega\tau}\right)^{1/2}. \tag{14.26}$$

If the mean free path l is much greater than the skin depth, $l \gg \delta$, then the local theory is not valid. In good metals at low temperatures, it turns out that $l \simeq v_\mathrm{F}\tau \simeq 10^7$nm and $\delta \simeq 10$nm, so that $l \gg \delta$, and we must use the nonlocal theory.

Anomalous Skin Effect

The normal skin effect was derived under the assumption that the $\mathbf{q}$ dependence of $\underline{\sigma}$ was unimportant. Remember that this assumption is valid if $ql = qv_\mathrm{F}\tau \ll 1$. We have found that the electric field varies like $\mathrm{e}^{-z/\delta}$. If δ turns out to be smaller than $l = v_\mathrm{F}\tau$, our initial assumption was certainly incorrect. The skin depth δ is of the order of

$$\delta \simeq \frac{c}{\omega_\mathrm{p}}. \tag{14.27}$$

Therefore, if

$$\frac{\omega c}{\omega_\mathrm{p} v_\mathrm{F}} < \omega\tau, \tag{14.28}$$

the theory is inconsistent because the field $E(z)$ changes appreciably over a mean free path l contradicting the assumption that the $\mathbf{q}$ dependence of $\underline{\sigma}$ can be neglected. The theory for this case in which the $\mathbf{q}$ dependence of $\underline{\sigma}$ must

be included is called the theory of the *anomalous skin effect*. In the nonlocal theory, we can write, (we take the y-axis to be perpendicular to the metal's surface),

$$\mathbf{j}_{\mathrm{e}}(y) = \int \mathrm{d}y' \underline{\sigma}(y, y') \cdot \mathbf{E}(y'), \tag{14.29}$$

which is true for an infinite medium. However, we have to take into account the surface of the metal here. We shall do this by using the formalism for the infinite medium, and imposing appropriate boundary conditions, namely, the method of images.

14.3 Azbel–Kaner Cyclotron Resonance

The theory of the anomalous skin effect in the presence of a dc magnetic field aligned parallel to the surface is the theory of *Azbel–Kaner cyclotron resonance* in metals. We shall present a brief treatment of this effect, and leave the problem of the anomalous skin effect in the absence of a dc magnetic field as an exercise.

Let us choose a Cartesian coordinate system with the y-axis normal to the surface, and the z-axis parallel to the dc magnetic field (see Fig. 14.1). For a polarization in which $\mathbf{E}(\mathbf{r}, t)$ is parallel to the z-axis, the wave equation can be written, since $\mathbf{q}$ is along the y-direction,

$$\left(-q^2 + \frac{\omega^2}{c^2}\right)\mathbf{E}(q, \omega) = \frac{4\pi \mathrm{i}\omega}{c^2}\mathbf{j}_{\mathrm{T}}(q, \omega). \tag{14.30}$$

This comes from the Fourier transform of the wave equation. For the case of self-sustaining oscillations of an infinite medium, we would set j_{T} equal to

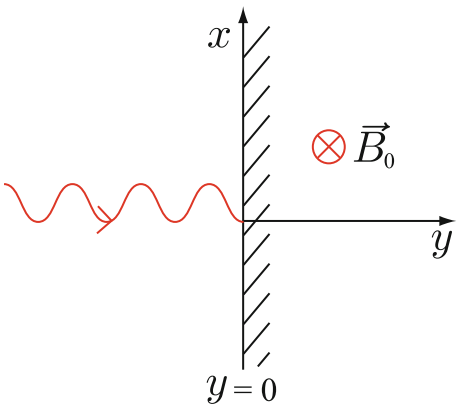


Fig. 14.1. The coordinate system for an electromagnetic wave propagating parallel to the y-axis with the surface at $y = 0$. A dc magnetic field $\mathbf{B}_0$ is parallel to the z-axis

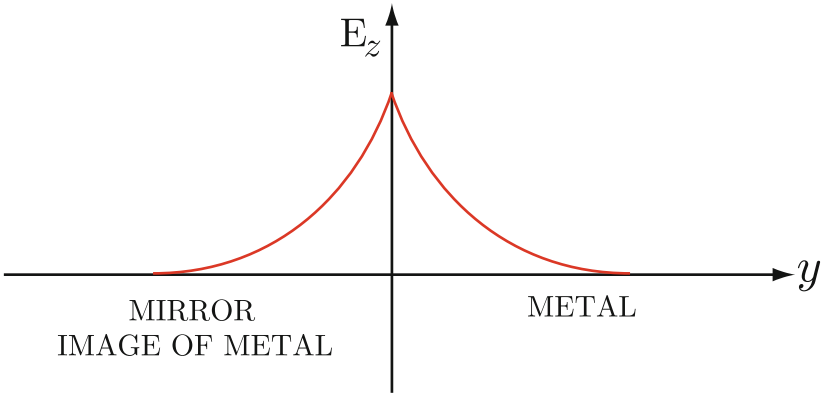


Fig. 14.2. A semi-infinite medium in terms of infinite medium picture. An electric field $E_z(y)$ in the metal is shown near the surface $y = 0$

the induced electron current given by $\sigma_{zz}(q,\omega)\mathrm{E}(q,\omega)$. For the semi-infinite medium, however, one must exercise some care to account for the boundary conditions. The electric field in the metal will decay in amplitude with distance from the surface $y = 0$. There is a discontinuity in the first derivative of $\mathbf{E}(\mathbf{r},t)$ at $y = 0$. Actually the term $-q^2\mathbf{E}$ in the wave equation came from making the assumption that $\mathbf{E}(y)$ was of the form $\mathrm{e}^{-\mathrm{i}qy}$. One can use this "infinite medium" picture by replacing the vacuum by the mirror image of the metal as shown in Fig. 14.2. The fictitious surface current $\mathbf{j}_0 \propto \delta(y)$ must be introduced to properly take account of the boundary conditions. By putting $\mathbf{j}_\mathrm{e}(q,\omega) = \sigma_{zz}(q,\omega)\mathbf{E}(q,\omega)$, we can solve the wave equation for $\mathbf{E}(q,\omega)$. The fictitious surface current sheet of density $\mathbf{j}_0$ is very simply related to the magnetic field at the surface

$$j_0(y) = \frac{c}{2\pi}H(0)\delta(y) \text{ or } j_0(q) = \frac{cH(0)}{2\pi}. \tag{14.31}$$

Solving (14.30) for $E(q,\omega)$ gives

$$E(q,\omega) = \frac{2\mathrm{i}\omega H(0)/c}{-q^2 + \frac{\omega^2}{c^2} - \frac{4\pi\mathrm{i}\omega}{c^2}\sigma_{zz}(q,\omega)}. \tag{14.32}$$

By substituting this result into the Fourier transform

$$E(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}q\, \mathrm{e}^{-\mathrm{i}qy}E(q,\omega), \tag{14.33}$$

and the definition of the surface impedance $\mathcal{Z}$

$$\mathcal{Z} = \frac{4\pi}{c}\frac{E(0)}{H(0)}, \tag{14.34}$$

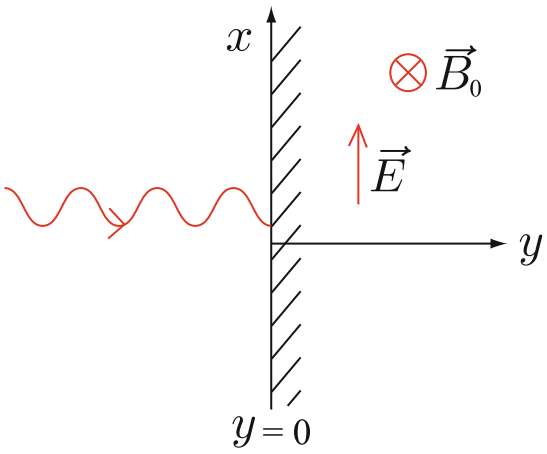


Fig. 14.3. The coordinate system for an electromagnetic wave propagating parallel to the y-axis to the metal surface ($y = 0$) for the case of polarization in the x-direction. A dc magnetic field $\mathbf{B}_0$ is parallel to the z-axis

we can easily obtain the electric field as a function of position and the surface impedance of the metal. In (14.34), $E(0)$ and $H(0)$ are the electric field and the magnetic field at the surface, respectively.

For the case of a wave polarized in the x-direction instead of the z-direction, σ_{zz} is replaced by

$$\sigma_{\mathrm{T}} = \sigma_{xx} + \frac{\sigma_{xy}^2}{\sigma_{yy}}. \tag{14.35}$$

This is equivalent to assuming that electrons are specularly reflected at the boundary $y = 0$. Figure 14.3 shows the coordinate system for a wave propagating perpendicular to the boundary of the metal with polarization in the x-direction normal to the dc magnetic field. One can see that although $\mathbf{E}(y)$ is continuous, its first derivative is not. Therefore, in defining the Fourier transform of $\frac{\partial^2 \mathbf{E}(y)}{\partial y^2}$ we must add terms to take account of these continuities.

$$\int_{-\infty}^{\infty} \mathrm{d}y\, \mathrm{e}^{\mathrm{i}qy} \frac{\partial^2 \mathbf{E}(y)}{\partial y^2} = \Delta \mathbf{E}_0' - \mathrm{i}q \Delta \mathbf{E}_0 - q^2 \mathbf{E}(q), \tag{14.36}$$

where

$$\Delta \mathbf{E}_0' = \left(\frac{\partial \mathbf{E}(y)}{\partial y} \right)_{y=0^+} - \left(\frac{\partial \mathbf{E}(y)}{\partial y} \right)_{y=0^-}, \tag{14.37}$$

and

$$\Delta \mathbf{E}_0 = \mathbf{E}(0^+) - \mathbf{E}(0^-). \tag{14.38}$$

For the case of specular reflection we take $\Delta \mathbf{E}_0 = 0$ and $\Delta \mathbf{E}_0' = 2\mathbf{E}'(0^+)$. This adds a constant term to the wave equation

$$\left(-q^2+\frac{\omega^2}{c^2}\right)\mathbf{E}(q,\omega)=-\Delta\mathbf{E}_0'+\frac{4\pi\mathrm{i}\omega}{c^2}\mathbf{j}_\mathrm{e}(q,\omega). \tag{14.39}$$

This added term can equally well be thought of as a fictitious surface current $\mathbf{j}_0(y)$ given by

$$\mathbf{j}_0(y)=-\frac{c^2\Delta\mathbf{E}_0'}{4\pi\mathrm{i}\omega}\delta(y). \tag{14.40}$$

So that the infinite medium result (or infinite medium wave equation) can be used if we take for $\mathbf{j}_\mathrm{T}(q,\omega)$

$$\mathbf{j}_\mathrm{T}(q,\omega)=\mathbf{j}_\mathrm{e}(q,\omega)+\mathbf{j}_0(q). \tag{14.41}$$

Actually, the results just derived are valid in the absence of a magnetic field as well as in the presence of a dc magnetic field. In the absence of a magnetic field the conductivity tensor is given by

$$\underline{\sigma}(q,\omega)=\frac{\omega_\mathrm{p}^2}{4\pi\mathrm{i}\omega}\left\{\underline{\mathbf{1}}+\underline{\mathbf{I}}(q,\omega)\right\}, \tag{14.42}$$

where

$$\underline{\mathbf{I}}(q,\omega)=\frac{m}{N}\sum_{\mathbf{k}\mathbf{k}'}\frac{f_0(\varepsilon_{\mathbf{k}'})-f_0(\varepsilon_\mathbf{k})}{\varepsilon_{\mathbf{k}'}-\varepsilon_\mathbf{k}-\hbar\omega}\langle\mathbf{k}'|\mathbf{V}_q|\mathbf{k}\rangle\langle\mathbf{k}'|\mathbf{V}_q|\mathbf{k}\rangle^*. \tag{14.43}$$

14.4 Azbel–Kaner Effect

If we use the Cohen–Harrison–Harrison expression for $\sigma_{zz}(q,\omega)$, (13.125), we have

$$\sigma_{zz}=3\sigma_0\sum_{n=-\infty}^{\infty}\frac{c_n(w)}{1-\mathrm{i}\tau(n\omega_\mathrm{c}-\omega)}, \tag{14.44}$$

where

$$c_n(w)=\frac{1}{2}\int_{-1}^{1}\mathrm{d}(\cos\theta)\,\cos^2\theta J_n^2(w\sin\theta). \tag{14.45}$$

Since $w\gg 1$ (i.e., we assume $w\equiv\frac{qv_\mathrm{F}}{\omega_\mathrm{c}}\gg 1$ for the values of q of interest in this problem) we can replace $J_n(w\sin\theta)$ by its asymptotic value for large argument.

$$\lim_{z\to\infty}J_n(z)\approx\sqrt{\frac{2}{\pi z}}\cos[z-\left(n+\frac{1}{2}\right)\frac{\pi}{2}]. \tag{14.46}$$

Substituting (14.46) into the expression for $c_n(w)$ (14.45) gives

$$c_n(w)\approx\frac{1}{4w}. \tag{14.47}$$

Therefore, we have

$$\sigma_{zz} \approx \frac{3\omega_{\mathrm{p}}^2}{4\pi \mathrm{i}\tilde{\omega}} \frac{1}{4w} \sum_{n=-\infty}^{\infty} \frac{1}{1 + n\omega_{\mathrm{c}}/\tilde{\omega}}. \tag{14.48}$$

Making use of the fact that

$$\sum_{n=-\infty}^{\infty} \frac{1}{1 + n\omega_{\mathrm{c}}/\omega} = \frac{\pi\omega}{\omega_{\mathrm{c}}} \cot \frac{\pi\omega}{\omega_{\mathrm{c}}} \tag{14.49}$$

and $-\mathrm{i} \cot x = \coth \mathrm{i}x$, one can easily see that

$$\sigma_{zz} \approx \frac{3\omega_{\mathrm{p}}^2}{16 q v_{\mathrm{F}}} \coth \left[\frac{\pi(1 + \mathrm{i}\omega\tau)}{\omega_{\mathrm{c}}\tau} \right]. \tag{14.50}$$

This can be rewritten

$$\sigma_{zz} \approx -\frac{3\mathrm{i}\omega_{\mathrm{p}}^2}{16 q v_{\mathrm{F}}} \cot \left[\frac{\pi(\omega - \mathrm{i}/\tau)}{\omega_{\mathrm{c}}} \right]. \tag{14.51}$$

In the limit $\omega\tau \gg 1$, this function has sharp peaks at $\omega \simeq n\omega_{\mathrm{c}}$. In this last expression, we have substituted $\omega - \mathrm{i}/\tau$ for ω. In the limit $\omega\tau \gg 1$, σ_{zz} shows periodic oscillations as a function of ω_{c}. These oscillations also show up in the surface impedance.

In using the Cohen–Harrison–Harrison expression for σ_{zz}, we have obviously omitted *quantum oscillations*. By using the quantum mechanical expression for $\underline{\sigma}$, for example $\underline{\sigma}^{\mathrm{QO}}(q, \omega)$ given by (13.154), one can easily obtain the quantum oscillations of the surface impedance.

14.5 Magnetoplasma Waves

We have seen that if we omit spin magnetization, the Maxwell equations for a wave of the form $\mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}\mathbf{q}\cdot\mathbf{r}}$ can be written

$$\mathbf{j}_{\mathrm{T}} = \underline{\Gamma}(\mathbf{q}, \omega) \cdot \mathbf{E}. \tag{14.52}$$

The total current is usually the sum of some external current and the induced electron current. If one is interested in the self-sustaining oscillations of the system, one wants the external driving current to be equal to zero. Then the electron current is given by $\mathbf{j}_{\mathrm{e}} = \underline{\sigma} \cdot \mathbf{E}$, and this is the only current. Thus, $\mathbf{j}_{\mathrm{T}} = \underline{\Gamma} \cdot \mathbf{E} = \underline{\sigma} \cdot \mathbf{E}$, so that we have

$$| \underline{\Gamma} - \underline{\sigma} | = 0 \tag{14.53}$$

gives the dispersion relation for the normal modes of the system.

In the absence of a dc magnetic field (14.53) reduces to

$$(\omega^2 \varepsilon - c^2 q^2)^2 \varepsilon = 0. \tag{14.54}$$

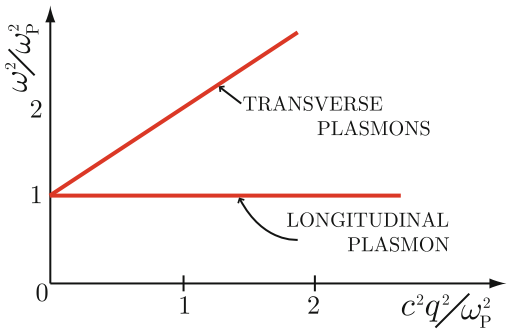


Fig. 14.4. The dispersion curves of transverse and longitudinal plasmon modes in the absence of a dc magnetic field

In the local (collisionless) theory, the dielectric function is given by

$$\varepsilon \approx 1 - \frac{\omega_\mathrm{p}^2}{\omega^2}, \tag{14.55}$$

so the normal modes are two degenerate transverse modes of frequency

$$\omega^2 = \omega_\mathrm{p}^2 + c^2 q^2, \tag{14.56}$$

and a longitudinal mode of frequency

$$\omega = \omega_\mathrm{p}. \tag{14.57}$$

Figure 14.4 shows the dispersion curves of transverse and longitudinal plasmon modes in the absence of a dc magnetic field. There are no propagating modes for frequencies ω smaller than the plasma frequency ω_p.

Now, consider the normal modes of the system in the presence of a dc magnetic field. We choose the z-axis parallel to the magnetic field, and let the wave vector $\mathbf{q}$ lie in the $y-z$ plane. The secular equation can be written

$$\begin{vmatrix} \varepsilon_{xx} - \xi^2 & \varepsilon_{xy} & 0 \\ -\varepsilon_{xy} & \varepsilon_{xx} - \xi_z^2 & \xi_y \xi_z \\ 0 & \xi_y \xi_z & \varepsilon_{zz} - \xi_y^2 \end{vmatrix} = 0. \tag{14.58}$$

In writing down (14.58) we have introduced $\boldsymbol{\xi} = \frac{c\mathbf{q}}{\omega}$, and now assume a local theory of conductivity in which the nonvanishing elements of $\underline{\varepsilon}$ are

$$\begin{aligned} \varepsilon_{xx}(\omega) &= \varepsilon_{yy}(\omega) = 1 - \frac{\omega_\mathrm{p}^2}{\omega^2 - \omega_\mathrm{c}^2}, \\ \varepsilon_{xy}(\omega) &= -\varepsilon_{yx}(\omega) = -\mathrm{i}\frac{\omega_\mathrm{p}^2 \omega_\mathrm{c}/\omega}{\omega^2 - \omega_\mathrm{c}^2}, \\ \varepsilon_{zz}(\omega) &= 1 - \frac{\omega_\mathrm{p}^2}{\omega^2}. \end{aligned} \tag{14.59}$$

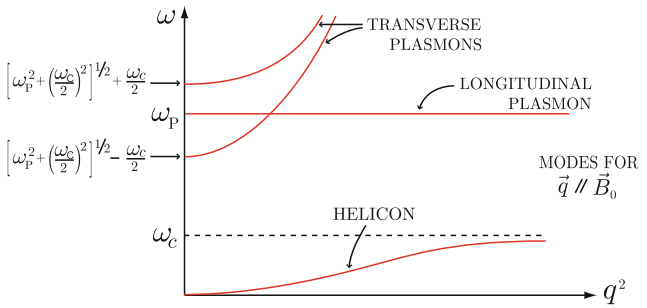


Fig. 14.5. The dispersion curves of magnetoplasma modes in a metal when the wave propagates in the z-direction parallel to the dc magnetic field $\mathbf{B}_0$

Because the dielectric constant is independent of $\mathbf{q}$, the secular equation turns out to be rather simple. It is a quadratic equation in q^2

$$\alpha q^4 + \beta q^2 + \gamma = 0, \tag{14.60}$$

where

$$\begin{aligned} \alpha &= \varepsilon_{xx}(\omega)\sin^2\theta + \varepsilon_{zz}(\omega)\cos^2\theta, \\ \beta &= -\omega^2\left\{\varepsilon_{xx}(\omega)\varepsilon_{zz}(\omega)(1+\cos^2\theta) + [\varepsilon_{xy}^2(\omega) + \varepsilon_{xx}^2(\omega)]\sin^2\theta\right\}, \\ \gamma &= \omega^4\left[\varepsilon_{xx}^2(\omega) + \varepsilon_{xy}^2(\omega)\right]\varepsilon_{zz}. \end{aligned} \tag{14.61}$$

Here, θ is the angle between the direction of propagation and the direction of the dc magnetic field. For $\theta = 0$ (14.60) reduces to

$$\varepsilon_{zz}(\omega)\left\{[q^2 - \omega^2\varepsilon_{xx}(\omega)]^2 + \omega^4\varepsilon_{xy}^2(\omega)\right\} = 0. \tag{14.62}$$

The roots can easily be plotted; there are four roots as are shown in Fig. 14.5. The longitudinal plasmon $\omega = \omega_\mathrm{p}$ is the solution of $\varepsilon_{zz}(\omega) = 0$. The two transverse plasmons start out at $q = 0$ as $\omega = \left[\omega_\mathrm{p}^2 + (\omega_\mathrm{c}/2)^2\right]^{1/2} \pm \frac{\omega_\mathrm{c}}{2}$. At very large q they are just light waves, but there is a difference in the phase velocity for the two different (circular) polarizations. Their difference in phase velocity is responsible for the Faraday effect–the rotation of the plane of polarization in a plane polarized wave. The low frequency mode is the well-known *helicon*. For small values of q it begins as

$$\omega = \frac{\omega_\mathrm{c} c^2 q^2}{\omega_\mathrm{p}^2}, \tag{14.63}$$

and it asymptotically approaches $\omega = \omega_\mathrm{c}$ for large q.

For $\theta = \frac{\pi}{2}$ (14.60) reduces to

$$[q^2 - \omega^2\varepsilon_{zz}(\omega)]\left\{q^2\varepsilon_{xx}(\omega) - \omega^2[\varepsilon_{xx}^2(\omega) + \varepsilon_{xy}^2(\omega)]\right\} = 0. \tag{14.64}$$

The mode corresponding to $q^2 = \omega^2\varepsilon_{zz}(\omega)$ is a transverse plasmon of frequency $\omega = \left[\omega_\mathrm{p}^2 + c^2q^2\right]^{1/2}$. The helicon mode appears no longer. The other

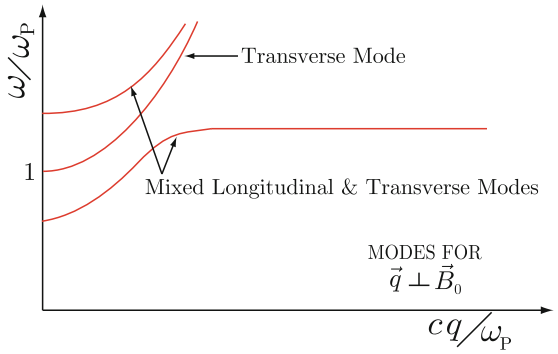


Fig. 14.6. The dispersion curves of magnetoplasma modes in a metal when the wave propagates in the z-direction normal to the dc magnetic field $\mathbf{B}_0$

two modes have mixed longitudinal and transverse character. They start at

$$\omega = \left[\omega_{\mathrm{p}}^2 + \left(\frac{\omega_{\mathrm{c}}}{2}\right)^2\right]^{1/2} \pm \frac{\omega_{\mathrm{c}}}{2} \tag{14.65}$$

for $q = 0$. For very large q one mode is nearly transverse and varies as $\omega \approx cq$ while the other approaches the finite asymptotic limit $\omega = \sqrt{\omega_{\mathrm{p}}^2 + \omega_{\mathrm{c}}^2}$. The roots of (14.64) are sketched in Fig. 14.6. For an arbitrary angle of propagation[1] the helicon mode has a frequency (we assume $\omega_{\mathrm{p}} \gg \omega_{\mathrm{c}}$)

$$\omega \simeq \frac{\omega_{\mathrm{c}} c^2 q^2 \cos\theta}{\omega_{\mathrm{p}}^2 + c^2 q^2} \left(1 + \frac{\mathrm{i}}{\omega_{\mathrm{c}} \tau \cos\theta}\right). \tag{14.66}$$

Here, we have included the damping due to collisions. For very large values of $\frac{cq}{\omega}$, the two finite frequency modes are sometimes referred to as the *hybrid-magnetoplasma modes*. Their frequencies are

$$\omega_{\pm}^2 = \frac{1}{2}\left(\omega_{\mathrm{p}}^2 + \omega_{\mathrm{c}}^2\right) + \left[\frac{1}{4}\left(\omega_{\mathrm{p}}^2 + \omega_{\mathrm{c}}^2\right)^2 - \omega_{\mathrm{p}}^2 \omega_{\mathrm{c}}^2 \cos^2\theta\right]^{1/2}. \tag{14.67}$$

For propagation at an arbitrary angle we can think of the four modes as coupled magnetoplasma modes. The $\omega_{\pm}$ modes described above for very large values of q are obviously the coupled helicon and longitudinal plasmon.

14.6 Discussion of the Nonlocal Theory

By considering the $\mathbf{q}$ dependence of the conductivity one can find a number of interesting effects that have been omitted from the local theory (as well as the quantitative changes in the dispersion relation which are to be expected). Among them are:

[1] M.A. Lampert, J.J. Quinn, S. Tosima, Phys. Rev. **152**, 661 (1966).

1. *Landau Damping and Doppler Shifted Cyclotron Resonance*

   Suppose an electromagnetic wave of frequency ω and wave number q propagates inside a metal. To absorb energy from the electromagnetic wave, the component of the velocity of an electron along the applied dc magnetic field must satisfy $n\omega_c + q_z v_z = \omega$ for some integral values of n. When $n\omega_c - q_z v_F < \omega < n\omega_c + q_z v_F$, there are electrons capable of direct absorption of energy from the electromagnetic field and we have cyclotron damping even in the absence of collisions. However, if this condition is not satisfied, then, for example, when $\mid \omega - \omega_c \mid > q_z v_F$, there does not exist any electron with $v_z = (\omega_c - \omega)/q_z$, which would resonantly absorb energy from the wave. For $n = 0$, this effect is usually known as *Landau damping*. Then, we have $-q_z v_F < \omega < q_z v_F$ or $-v_F < v_{phase} < v_F$. It corresponds to having a phase velocity v_{phase} parallel to $\mathbf{B}_0$ equal to the velocity of some electrons in the solid, i.e., $-v_F < v_z < v_F$. These electrons will ride the wave and thus absorb power from it resulting in collisionless damping of the wave. For $n \neq 0$, the effect is usually called *Doppler shifted cyclotron resonance*, because the effective frequency seen by the moving electron is $\omega_{eff} = \omega - q_z v_z$ and it is equal to n times the cyclotron resonance frequency ω_c.

2. *Bernstein Modes or Cyclotron Modes*

   These are the modes of vibration in an electron plasma, which occur only when $\underline{\sigma}$ has a $\mathbf{q}$-dependence. They are important in plasma physics, where they are known as *Bernstein modes*. In solid-state physics, they are known as *nonlocal waves* or *cyclotron waves*. These modes start out at $\omega = n\omega_c$ for $q = 0$. They propagate perpendicular to the dc magnetic field, and depend for their existence (even at very long wavelengths) on the $\mathbf{q}$ dependence of $\underline{\sigma}$.

3. *Quantum Waves*

   These are waves which arise from the gigantic quantum oscillations in $\underline{\sigma}$. These quantum effects depend, of course, on the $\mathbf{q}$ dependence of $\underline{\sigma}$.

14.7 Cyclotron Waves

We will give only one example of the new kind of wave that can occur when the $\mathbf{q}$ dependence of $\underline{\sigma}$ is taken into account. We consider the magnetic field in the z-direction and the wave vector $\mathbf{q}$ in the y-direction. The secular equation for wave propagation is the familiar

$$\begin{vmatrix} \varepsilon_{xx} - \xi^2 & \varepsilon_{xy} & 0 \\ -\varepsilon_{xy} & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} - \xi^2 \end{vmatrix} = 0. \tag{14.68}$$

This secular equation reduces to a 2×2 matrix and a 1×1 matrix. For the polarization with $\mathbf{E}$ parallel to the z-axis we are interested in the 1×1 matrix. The Lorentz force couples the x-y motions giving the 2×2 matrix for the other polarization. For the simple case of the 1×1 matrix we have

$$\frac{c^2 q^2}{\omega^2} = 1 - \frac{4\pi \mathrm{i}}{\omega} \sigma_{zz}, \tag{14.69}$$

where, in the collisionless limit (i.e. $\tau \to \infty$),

$$\sigma_{zz} = \frac{3\mathrm{i}\omega_\mathrm{p}^2}{4\pi} \sum_{n=0}^{\infty} \frac{c_n(w)}{1+\delta_{n0}} \frac{2\omega}{(n\omega_\mathrm{c})^2 - \omega^2} \tag{14.70}$$

with

$$c_n(w) = \int_0^1 \mathrm{d}(\cos\theta)\, \cos^2\theta J_n^2(w \sin\theta). \tag{14.71}$$

If we let $\frac{\omega}{\omega_\mathrm{c}} = a$, we can write

$$\frac{4\pi \mathrm{i}}{\omega} \sigma_{zz} = -\frac{6\omega_\mathrm{p}^2}{\omega_\mathrm{c}^2} \left[\frac{c_0/2}{-a^2} + \frac{c_1}{1-a^2} + \cdots + \frac{c_n}{n^2 - a^2} + \cdots \right]. \tag{14.72}$$

Let us look at the long wavelength limit where $\frac{q v_\mathrm{F}}{\omega_\mathrm{c}} \ll 1$. Remember that for small x

$$J_n(x) = \frac{1}{n!} \left(\frac{x}{2}\right)^n \left[1 - \frac{(x/2)^2}{1 \cdot (n+1)} + \cdots \right]. \tag{14.73}$$

We keep terms to order w^2. Because $c_n \propto J_n^2$, $c_n \propto w^{2n}$. Therefore, if we retain only terms of order w^2, we can drop all terms but the first two. Then, we have

$$J_0^2(x) \approx 1 - \frac{x^2}{2} \text{ and } J_1^2(x) \approx \frac{x^2}{4}. \tag{14.74}$$

Substituting (14.74) in c_0 and c_1 yields

$$c_0(w) \approx \int_0^1 \mathrm{d}(\cos\theta)\, \cos^2\theta \left[1 - \frac{1}{2} w^2 \sin^2\theta \right] = \frac{1}{3} - \frac{w^2}{15}, \tag{14.75}$$

and for c_1 we find

$$c_1(w) \approx \int_0^1 \mathrm{d}(\cos\theta)\, \cos^2\theta \frac{w^2}{4} \sin^2\theta = \frac{w^2}{30}. \tag{14.76}$$

Substituting these results into the secular equation, (14.69), gives

$$\frac{c^2 q^2}{\omega_\mathrm{c}^2 a^2} \simeq 1 - \frac{\omega_\mathrm{p}^2}{\omega_\mathrm{c}^2 a^2} \left[1 - \frac{w^2/5}{1-a^2} \right]. \tag{14.77}$$

This is a simple quadratic equation in a^2, where $a = \frac{\omega}{\omega_\mathrm{c}}$. The general solution is

$$\omega^2 = \frac{1}{2}\left(\omega_{\rm p}^2 + \omega_{\rm c}^2 + c^2q^2\right) \pm \sqrt{\frac{1}{4}\left(\omega_{\rm p}^2 + \omega_{\rm c}^2 + c^2q^2\right)^2 - \omega_{\rm c}^2\left(\omega_{\rm p}^2 + c^2q^2 - \frac{\omega_{\rm p}^2 w^2}{5}\right)}. \tag{14.78}$$

For $q \to 0$, the two roots are

$$\omega^2 = \frac{1}{2}\left(\omega_{\rm p}^2 + \omega_{\rm c}^2\right) \pm \frac{1}{2}\left(\omega_{\rm p}^2 - \omega_{\rm c}^2\right) = \begin{cases} \omega_{\rm p}^2, \\ \omega_{\rm c}^2. \end{cases} \tag{14.79}$$

If $\omega_{\rm p} \gg \omega_{\rm c}$, the lower root can be obtained quite well by setting

$$\left[1 - \frac{w^2/5}{1-a^2}\right] = 0,$$

which gives

$$\left(\frac{\omega}{\omega_{\rm c}}\right)^2 = 1 - \frac{w^2}{5}. \tag{14.80}$$

Actually going back to (14.72)

$$\frac{4\pi{\rm i}}{\omega}\sigma_{zz} = -\frac{6\omega_{\rm p}^2}{\omega_{\rm c}^2}\left[\frac{c_0/2}{-a^2} + \frac{c_1}{1-a^2} + \cdots + \frac{c_n}{n^2-a^2} + \cdots\right],$$

it is not difficult to see that, for $\omega_{\rm p} \gg n\omega_{\rm c}$, there must be a solution at $\omega^2 = n^2\omega_{\rm c}^2 + {\rm O}(q^{2n})$. We do this by setting $c_n = \alpha_n w^{2n}$ for $n \geq 1$. If $\omega_{\rm p} \gg n\omega_{\rm c}$, then the solutions are given, approximately, by

$$\left[\frac{-c_0/2}{a^2} + \frac{c_1}{1-a^2} + \cdots + \frac{c_n}{n^2-a^2} + \cdots\right] \simeq 0.$$

Let us assume a solution of the form $a^2 = n^2 + \Delta$, where $\Delta \ll n$. Then the above equation can be written

$$\left[\frac{-\frac{1}{3}(1-\frac{w^2}{5})}{n^2} + \frac{\alpha_1 w^2}{1^2-n^2} + \cdots + \frac{\alpha_{n-1}w^{2(n-1)}}{(n-1)^2-n^2} + \frac{\alpha_n w^{2n}}{\Delta} + \cdots\right] \simeq 0.$$

Solving for Δ gives $\Delta \simeq 3n^2\alpha_n w^{2n}$. Thus, we have a solution of the form

$$\left(\frac{\omega}{\omega_{\rm c}}\right)^2 = n^2 + {\rm O}(q^{2n}). \tag{14.81}$$

14.8 Surface Waves

There are many kinds of surface waves in solids–plasmons, magnetoplasma waves, magnons, acoustic phonons, optical phonons etc. In fact, we believe that every bulk wave has associated with it a surface wave. To give some

feeling for surface waves, we shall consider one simple case; surface plasmons in the absence of a dc magnetic field.

We consider a metal of dielectric function ε_1 to fill the space $z > 0$, and an insulator of dielectric constant ε_0 the space $z < 0$. The wave equation which describes propagation in the $y - z$ plane is given by

$$\begin{pmatrix} \varepsilon - \xi^2 & 0 & 0 \\ 0 & \varepsilon - \xi_z^2 & \xi_y \xi_z \\ 0 & \xi_y \xi_z & \varepsilon - \xi_y^2 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = 0. \tag{14.82}$$

Here, $\boldsymbol{\xi} = \frac{c\mathbf{q}}{\omega}$ and $\xi^2 = \xi_y^2 + \xi_z^2$. For the dielectric $\varepsilon = \varepsilon_0$, a constant, and we find only two transverse waves of frequency $\omega = \frac{cq}{\sqrt{\varepsilon_0}}$ for the bulk modes. For the metal (to be referred to as medium 1) there are one longitudinal plasmon of frequency $\omega = \omega_\mathrm{p}$ and two transverse plasmons of frequency $\omega = \sqrt{\omega_\mathrm{p}^2 + c^2 q^2}$ as the bulk modes. Here we are assuming that $\varepsilon_1 = 1 - \frac{\omega_\mathrm{p}^2}{\omega^2}$ is the dielectric function of the metal.

To study the surface waves, we consider ω and q_y to be given real numbers and solve the wave equation for q_z. For the transverse waves in the metal, we have

$$q_z^2 = \frac{\omega^2 - \omega_\mathrm{p}^2}{c^2} - q_y^2. \tag{14.83}$$

In the insulator, we have

$$q_z^2 = \varepsilon_0 \frac{\omega^2}{c^2} - q_y^2. \tag{14.84}$$

The $q_z = 0$ lines are indicated in Fig. 14.7 as solid lines. Notice that in region III of Fig. 14.7, $q_z^2 < 0$ in both the metal and the insulator. This is the region of interest for surface waves excitations, because negative q_z^2 implies that q_z itself is imaginary. Solving for $q_z^{(1)}$ (value of q_z in the metal) and $q_z^{(2)}$ (value of q_z in the insulator), when ω and q_y are such that we are considering region III, gives

$$\begin{aligned} q_z^{(1)} &= \pm \mathrm{i}(\omega_\mathrm{p}^2 + q_y^2 - \omega^2)^{1/2} = \pm \mathrm{i}\alpha_1, \\ q_z^{(0)} &= \pm \mathrm{i}(q_y^2 - \varepsilon_0 \omega^2)^{1/2} = \pm \mathrm{i}\alpha_0. \end{aligned} \tag{14.85}$$

This defines α_0 and α_1, which are real and positive. The wave in the metal must be of the form $\mathrm{e}^{\pm\alpha_1 z}$ and in the insulator of the form $\mathrm{e}^{\pm\alpha_0 z}$. To have solutions well behaved at $z \to +\infty$ in the metal and $z \to -\infty$ in the insulator we must choose the wave of the proper sign. Doing so gives

$$\begin{aligned} \mathbf{E}^{(1)}(\mathbf{r}, t) &= \mathbf{E}^{(1)} \mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}q_y y - \alpha_1 z}, \\ \mathbf{E}^{(0)}(\mathbf{r}, t) &= \mathbf{E}^{(0)} \mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}q_y y + \alpha_0 z}. \end{aligned} \tag{14.86}$$

The superscripts 1 and 0 refer, respectively, to the metal and dielectric. The boundary conditions at the plane $z = 0$ are the standard ones of continuity

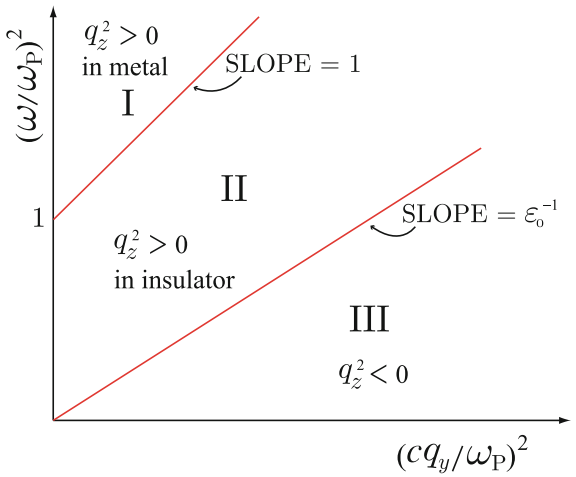


Fig. 14.7. The $\omega^2 - q_y^2$ plane for the waves near the interface of a metal and an insulator at $z = 0$. The solid lines show the region where $q_z = 0$ in the solid (line separating regions I and II) and in the dielectric (line separating II and III). In region III $q_z^2 < 0$ in both media, therefore excitations in this region are localized at the surface

of the tangential components of **E** and **H**, and of the normal components of **D** and **B**. By applying the boundary conditions (remembering that $\mathbf{B} = \mathrm{i}\frac{c}{\omega}\nabla \times \mathbf{E} = \frac{c}{\omega}(q_y E_z - q_z E_y, q_z E_x, -q_y E_x)$ we find that

1. For the independent polarization with $E_y = E_z = 0$, but $E_x \neq 0$ there are no solutions in region III.
2. For the polarization with $E_x = 0$, but $E_y \neq 0 \neq E_z$, there is a dispersion relation

$$\frac{\varepsilon_1}{\alpha_1} + \frac{\varepsilon_0}{\alpha_0} = 0. \tag{14.87}$$

If we substitute for α_0 and α_1, (14.87) becomes

$$c^2 q_y^2 = \frac{\omega^2 \varepsilon_0 \varepsilon_1(\omega)}{\varepsilon_0 + \varepsilon_1(\omega)} = \frac{\varepsilon_0 \omega^2 (\omega_P^2 - \omega^2)\left[\omega^2(\varepsilon_0 - 1) + \omega_P^2\right]}{(\omega_P^2 - \omega^2)^2 - \omega^4 \varepsilon_0^2}. \tag{14.88}$$

For very large q_y the root is approximately given by the zero of the denominator, viz $\omega = \frac{\omega_P}{\sqrt{1+\varepsilon_0}}$. For small values of q_y it goes as $\omega = \frac{c}{\sqrt{\varepsilon_0}} q_y$. Figure 14.8 shows the dispersion curve of the surface plasmon.

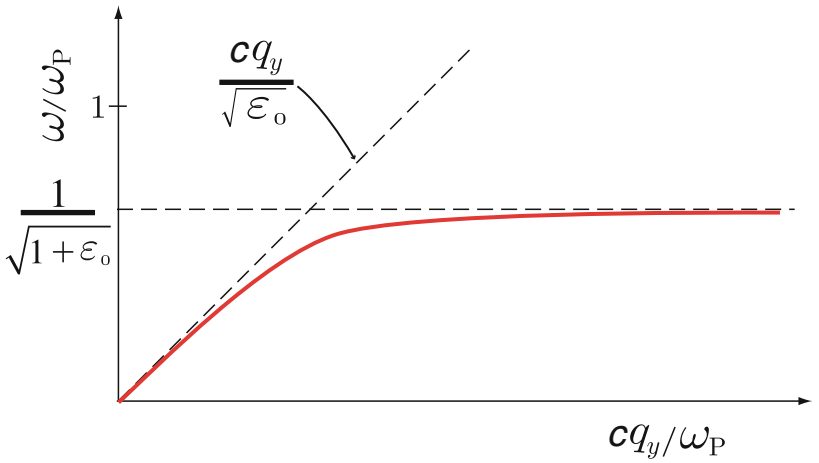


Fig. 14.8. Dispersion relation of surface plasmon

14.9 Magnetoplasma Surface Waves

In the presence of a dc magnetic field $\mathbf{B}_0$ oriented at an arbitrary angle to the surface, the problem of surface plasma waves becomes much more complicated.[2] We will discuss only the nonretarded limit of $cq \gg \omega$.

Let the metal or semiconductor be described by a dielectric function

$$\varepsilon_{ij}(\omega) = \varepsilon_{\rm L}\delta_{ij} - \frac{\omega_{\rm p}^2}{\omega^2(\omega^2-\omega_{\rm c}^2)}\left[\omega^2\delta_{ij} - \omega_{{\rm c}_i}\omega_{{\rm c}_j} - {\rm i}\,\omega\omega_{{\rm c}_k}\epsilon_{ijk}\right], \qquad (14.89)$$

where $\varepsilon_{\rm L}$ is the background dielectric constant, $\omega_{\rm c} = \frac{eB_0}{mc}$, $\omega_{{\rm c}_x} = \frac{eB_{0x}}{mc}$, and $\epsilon_{ijk} = +1(-1)$ if ijk is an even (odd) permutation of 123, and zero otherwise. Let the insulator have dielectric constant ε'. The wave equation is given by

$$\begin{pmatrix} \varepsilon_{xx} - q^2/\omega^2 & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} - q_z^2/\omega^2 & \varepsilon_{yz} + q_y q_z/\omega^2 \\ \varepsilon_{zx} & \varepsilon_{zy} + q_y q_z/\omega^2 & \varepsilon_{zz} - q_y^2/\omega^2 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = 0. \qquad (14.90)$$

In the nonretarded limit ($cq \gg \omega$) the off-diagonal elements ε_{xy}, ε_{yx}, ε_{xz}, ε_{zx} can be neglected and (14.90) can be approximated (we put $c = 1$) by

$$(q^2 - \omega^2\varepsilon_{xx})\left[\varepsilon_{zz}q_z^2 + \varepsilon_{yy}q_y^2 + (\varepsilon_{yz} + \varepsilon_{zy})q_y q_z\right] \approx 0. \qquad (14.91)$$

The surface magneto-plasmon solution arises from the second factor. Solving for q_z in the metal we find

[2] A summary of magnetoplasma surface wave results in semiconductors is reviewed by Quinn and Chiu in *P*olaritons, edited by E. Bernstein, F. DeMartini, Pergamon, New York (1971), p. 259.

$$\frac{q_z^{(1)}}{q_y} = -\mathrm{i}\sqrt{\frac{\varepsilon_{yy}}{\varepsilon_{zz}} - \left(\frac{\varepsilon_{yz}+\varepsilon_{zy}}{2\varepsilon_{zz}}\right)^2} - \frac{\varepsilon_{yz}+\varepsilon_{zy}}{2\varepsilon_{zz}}. \tag{14.92}$$

The superscript 1 refers to the metal. In the dielectric (superscript 0) $q_z^{(0)} = +\mathrm{i}q_y$. The eigenvectors are

$$\begin{aligned}
\mathbf{E}^{(1)}(\mathbf{r},t) &\simeq \left(0, E_y^{(1)}, -\frac{E_y^{(1)} q_z^{(1)}}{q_y}\right) \mathrm{e}^{i\omega t - \mathrm{i}q_y y - \mathrm{i}q_z^{(1)} z}, \\
\mathbf{B}^{(1)}(\mathbf{r},t) &\simeq \frac{c}{\omega}\left(q_y E_z^{(1)} - q_z^{(1)} E_y^{(1)}, 0, 0\right) \mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}q_y y - \mathrm{i}q_z^{(1)} z} \approx 0, \\
\mathbf{E}^{(0)}(\mathbf{r},t) &\simeq \left(0, E_y^{(0)}, -\frac{E_y^{(0)} q_z^{(0)}}{q_y}\right) \mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}q_y y - \mathrm{i}q_z^{(0)} z}, \\
\mathbf{B}^{(0)}(\mathbf{r},t) &\simeq \frac{c}{\omega}\left(q_y E_z^{(0)} - q_z^{(0)} E_y^{(0)}, 0, 0\right) \mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}q_y y - \mathrm{i}q_z^{(0)} z} \approx 0.
\end{aligned} \tag{14.93}$$

The dispersion relation obtained from the standard boundary conditions is

$$+\,\mathrm{i}\varepsilon' = \frac{q_z^{(1)}}{q_y}\varepsilon_{zz} + \varepsilon_{zy}. \tag{14.94}$$

With $\varepsilon_{\mathrm{L}} = \varepsilon'$, (14.94) simplifies to

$$\omega^2 - \omega_{\mathrm{c}}^2 + (\omega \pm \omega_{\mathrm{c}_x})^2 - \frac{\omega_{\mathrm{p}}^2}{\varepsilon_{\mathrm{L}}} = 0, \tag{14.95}$$

where the $\pm$ signs correspond to propagation in the $\pm y$-directions, respectively. For the case $\mathbf{B}_0 = 0$, this gives $\omega = \frac{\omega_{\mathrm{p}}}{\sqrt{2\varepsilon_{\mathrm{L}}}}$. For $\mathbf{B}_0 \perp \mathbf{x}$, we have

$$\omega = \sqrt{\frac{\omega_{\mathrm{c}}^2 + \omega_{\mathrm{p}}^2/\varepsilon_{\mathrm{L}}}{2}},$$

and with $\mathbf{B}_0 \parallel \mathbf{x}$ we obtain

$$\omega = \frac{1}{2}\sqrt{\omega_{\mathrm{c}}^2 + \frac{2\omega_{\mathrm{p}}^2}{\varepsilon_{\mathrm{L}}}} \mp \frac{\omega_{\mathrm{c}}}{2},$$

where the two roots correspond to propagations in the $\pm y$-directions, respectively.

14.10 Propagation of Acoustic Waves

Now, we will try to give a very brief summary of propagation of acoustic waves in metals. Our discussion will be based on a very simple model introduced by Quinn and Rodriguez.[3] The model treats the ions completely classically. The metal is considered to consist of

[3] J.J. Quinn, S. Rodriguez, Phys. Rev. **128**, 2487 (1962).

1. A lattice (Bravais lattice for simplicity) of positive ions of mass M and charge ze.
2. An electron gas with n_0 electrons per unit volume.
3. In addition to electromagnetic forces, there are short range forces between the ions which we represent by two 'unrenormalized' elastic constants C_ℓ and C_t.
4. The electrons encounter impurities and defects, and have a collision time τ associated with their motion.

First, let us investigate the classical equation of motion of the lattice. Let $\boldsymbol{\xi}(\mathbf{r}, t)$ be the displacement field of the ions. Then we have

$$M\frac{\partial^2 \boldsymbol{\xi}}{\partial t^2} = C_\ell \nabla(\nabla \cdot \boldsymbol{\xi}) - C_\mathrm{t} \nabla \times (\nabla \times \boldsymbol{\xi}) + ze\mathbf{E} + \frac{ze}{c}\dot{\boldsymbol{\xi}} \times (\mathbf{B}_0 + \mathbf{B}) + \mathbf{F}. \quad (14.96)$$

The forces appearing on the right hand side of (14.96) are

1. The short range "elastic" forces (the first two terms).
2. The Coulomb interaction of the charge ze with the self-consistent electric field produced by the ionic motion (the third term).
3. The Lorentz force on the moving ion in the presence of the dc magnetic field $\mathbf{B}_0$ and the self-consistent ac field $\mathbf{B}$. The term $\frac{ze}{c}\dot{\boldsymbol{\xi}} \times \mathbf{B}$ is always very small compared to $ze\mathbf{E}$, and we shall neglect it (the fourth term).
4. The *collision drag* force $\mathbf{F}$ exerted by the electrons on the ions (the last term).

The force $\mathbf{F}$ results from the fact that in a collision with the lattice, the electron motion is randomized, not in the laboratory frame of reference, but in a frame of reference moving with the local ionic velocity. Picture the collisions as shown in Fig. 14.9. Here, $\langle \mathbf{v} \rangle$ is the average electron velocity (at point $\mathbf{r}$ where the impurity is located) just before collision. Just after collision $\langle \mathbf{v} \rangle_\mathrm{final} \approx 0$ in the moving system, or $\langle \mathbf{v} \rangle_\mathrm{final} \approx \dot{\boldsymbol{\xi}}$ in the laboratory. Thus the momentum

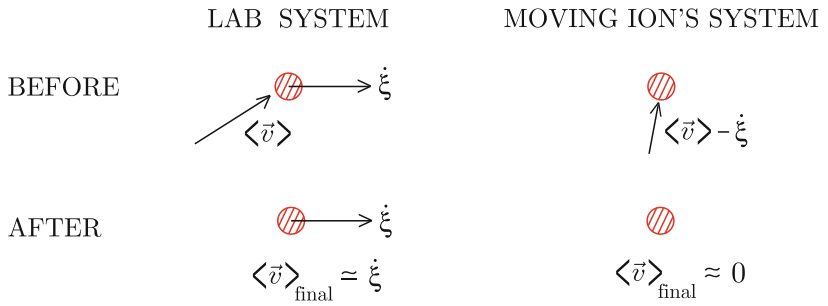


Fig. 14.9. Schematic of electron–impurity collision in the laboratory frame and in the coordinate system moving with the local ionic velocity. In the latter system a typical electron has velocity $\langle v \rangle - \dot{\xi}$ before collision and zero afterward. In the lab system, the corresponding velocities are $\langle v \rangle$ before collision and $\dot{\xi}$ afterwards

imparted to the positive ion must be $\Delta\mathbf{p} = m(\langle\mathbf{v}\rangle - \dot{\boldsymbol{\xi}})$. This momentum is imparted to the lattice per electron collision; since there are z electrons per atom and $\frac{1}{\tau}$ collisions per second for each electron, it is apparent that

$$\mathbf{F} = \frac{z}{\tau} m(\langle\mathbf{v}\rangle - \dot{\boldsymbol{\xi}}). \tag{14.97}$$

We can use the fact that the electronic current $\mathbf{j}_\mathrm{e}(\mathbf{r}) = -n_0 e\langle\mathbf{v}(\mathbf{r})\rangle$ to write

$$\mathbf{F} = -\frac{zm}{n_0 e\tau}(\mathbf{j}_\mathrm{e} + n_0 e\dot{\boldsymbol{\xi}}). \tag{14.98}$$

But the ionic current density is $\mathbf{j}_\mathrm{I} = n_0 e\dot{\boldsymbol{\xi}}$ so that

$$\mathbf{F} = -\frac{zm}{n_0 e\tau}(\mathbf{j}_\mathrm{e} + \mathbf{j}_\mathrm{I}). \tag{14.99}$$

The self-consistent electric field $\mathbf{E}$ appearing in the equation of motion, (14.96), is determined from the Maxwell equations, which can be written

$$\mathbf{j}_\mathrm{T} = \Gamma(\mathbf{q}, \omega) \cdot \mathbf{E}. \tag{14.100}$$

Let us consider $\mathbf{j}_\mathrm{T}$. It consists of the ionic current $\mathbf{j}_\mathrm{I}$, the electronic current $\mathbf{j}_\mathrm{e}$, and any external driving current $\mathbf{j}_0$. For considering the normal modes of the system (and the acoustic waves are normal modes) we set the external driving current $\mathbf{j}_0$ equal to zero and look for self-sustaining modes. Perhaps, if we have time, we can discuss the theory of direct electromagnetic generation of acoustic waves; in that case $\mathbf{j}_0$ is a "fictitious surface current" introduced to satisfy the boundary conditions in a finite solid (quite similar to the discussion given in our treatment of the Azbel–Kaner effect.). For the present, we consider the normal modes of an infinite medium. In that case, $\mathbf{j}_0 = 0$ so that $\mathbf{j}_\mathrm{T} = \mathbf{j}_\mathrm{I} + \mathbf{j}_\mathrm{e}$. The electronic current would be simply $\mathbf{j}_\mathrm{e} = \underline{\sigma} \cdot \mathbf{E}$ except for the effect of "collision drag" and diffusion. These two currents arise from the fact that the correct collision term in the Boltzmann equation must be

$$\left(\frac{\partial f}{\partial t}\right)_\mathrm{c} = -\frac{f - \bar{f}_0}{\tau}, \tag{14.101}$$

where $\bar{f}_0$ differs from the overall equilibrium distribution function f_0 in two respects:

1. $\bar{f}_0$ depends on the electron kinetic energy measured in the coordinates system of the moving lattice.
2. The chemical potential ζ appearing in $\bar{f}_0$ is not ζ_0, the actual chemical potential of the solid, but a local chemical potential $\zeta(\mathbf{r}, t)$ which is determined by the condition

$$\int \mathrm{d}^3k \left[f - \bar{f}_0\right] = 0, \tag{14.102}$$

i.e., the local equilibrium density at point $\mathbf{r}$ must be the same as the nonequilibrium density.

We can expand $\bar{f}_0$ as follows:

$$\bar{f}_0(\mathbf{k},\mathbf{r},t) = f_0(\mathbf{k}) + \frac{\partial f_0}{\partial \varepsilon}\left\{-m\mathbf{v}_k\cdot\dot{\boldsymbol{\xi}} + \zeta_1(\mathbf{r},t)\right\}, \tag{14.103}$$

where $\zeta_1(\mathbf{r},t) = \zeta(\mathbf{r},t) - \zeta_0$. Because of these two changes, instead of $\mathbf{j}_\mathrm{e} = \underline{\sigma}\cdot\mathbf{E}$, we have

$$\mathbf{j}_\mathrm{e}(\mathbf{q},\omega) = \underline{\sigma}(\mathbf{q},\omega)\cdot\left[\mathbf{E} - \frac{m\dot{\boldsymbol{\xi}}}{e\tau}\right] + e\underline{\mathbf{D}}\cdot\nabla n, \tag{14.104}$$

where

$$\underline{\mathbf{D}} = \frac{\underline{\sigma}}{e^2 g(\zeta_0)(1+\mathrm{i}\omega\tau)} \tag{14.105}$$

is the *diffusion tensor*. In (14.105), $g(\zeta_0)$ is the density of states at the Fermi surface and $n(\mathbf{r},t) = n_0 + n_1(\mathbf{r},t)$ is the electron density at point $(\mathbf{r},t)$. The electron density is determined from the distribution function f, which must be solved for. However, at all but the very highest ultrasonic frequencies, $n(\mathbf{r},t)$ can be determined accurately from the condition of charge neutrality.

$$\rho_\mathrm{e}(\mathbf{r},t) + \rho_\mathrm{I}(\mathbf{r},t) = 0, \tag{14.106}$$

where $\rho_\mathrm{e}(\mathbf{r},t) = -en_1(\mathbf{r},t)$ and ρ_I can be determined from the equation of continuity $i\omega\rho_\mathrm{I} - \mathrm{i}\mathbf{q}\cdot\mathbf{j}_\mathrm{I} = 0$. Using these results, we find

$$\mathbf{j}_\mathrm{e}(\mathbf{q},\omega) = \underline{\sigma}(\mathbf{q},\omega)\cdot\left[\mathbf{E} - \frac{\mathrm{i}\omega m}{e\tau}\boldsymbol{\xi} + \frac{n_0}{eg(\zeta_0)(1+\mathrm{i}\omega\tau)}\mathbf{q}(\mathbf{q}\cdot\boldsymbol{\xi})\right]. \tag{14.107}$$

If we define a tensor $\underline{\Delta}$ by

$$\underline{\Delta} = \frac{n_0 e\mathrm{i}\omega}{\sigma_0}\left\{\underline{\mathbf{1}} - \frac{1}{3}\,\frac{q^2 l^2}{\mathrm{i}\omega\tau(1+\mathrm{i}\omega\tau)}\hat{\mathbf{q}}\hat{\mathbf{q}}\right\}, \tag{14.108}$$

where $\hat{\mathbf{q}} = \frac{\mathbf{q}}{|\mathbf{q}|}$, we can write

$$\mathbf{j}_\mathrm{e}(\mathbf{q},\omega) = \underline{\sigma}(\mathbf{q},\omega)\cdot\mathbf{E}(\mathbf{q},\omega) - \underline{\sigma}(\mathbf{q},\omega)\cdot\underline{\Delta}(\mathbf{q},\omega)\cdot\boldsymbol{\xi}(\mathbf{q},\omega). \tag{14.109}$$

We can substitute (14.109) into the relation $\mathbf{j}_\mathrm{e} + \mathbf{j}_\mathrm{I} = \underline{\Gamma}\cdot\mathbf{E}$, and solve for the self-consistent field $\mathbf{E}$ to obtain

$$\mathbf{E}(\mathbf{q},\omega) = \left[\underline{\Gamma} - \underline{\sigma}\right]^{-1}(\mathrm{i}\omega ne\,\underline{\mathbf{1}} - \underline{\sigma}\cdot\underline{\Delta})\cdot\boldsymbol{\xi}. \tag{14.110}$$

Knowing $\mathbf{E}$, we also know $\mathbf{j}_\mathrm{e}$ and hence $\mathbf{F}$, (14.98) in terms of the ionic displacement $\boldsymbol{\xi}$. Thus, every term on the right-hand side of (14.96), the equation of motion of the ions can be expressed in terms of $\boldsymbol{\xi}$. The equation of motion is thus of the form

$$\underline{\mathbf{T}}(\mathbf{q},\omega)\cdot\boldsymbol{\xi}(\mathbf{q},\omega) = 0, \tag{14.111}$$

where $\underline{\mathbf{T}}$ is a very complicated tensor. The nontrivial solutions are determined from the secular equation

$$\det \mid \underline{\mathbf{T}}(\mathbf{q}, \omega) \mid = 0. \tag{14.112}$$

The roots of this secular equation give the frequencies of the sound waves (two transverse and one longitudinal modes) as a function of $\mathbf{q}$, $\mathbf{B}_0$, τ, etc. Actually the solutions $\omega(\mathbf{q})$ have both a real and imaginary parts; the real part determines the velocity of sound and the imaginary part the attenuation of the wave.

Here, we do not go through the details of the calculation outlined earlier. We will discuss special cases and attempt to give a qualitative feeling for the kinds of effects one can observe.

14.10.1 Propagation Parallel to $\mathbf{B_0}$

For propagation parallel to the dc magnetic field, it is convenient to introduce circularly polarized transverse waves with

$$\begin{aligned} \xi_{\pm} &= \xi_x \pm \mathrm{i}\xi_y, \\ \sigma_{\pm} &= \sigma_{xx} \mp \mathrm{i}\sigma_{xy}. \end{aligned} \tag{14.113}$$

We also introduce the parameter $\beta = \frac{c^2 q^2}{4\pi\omega\sigma_0} = \frac{c^2 q^2}{\omega_{\mathrm{p}}^2 \omega\tau}$. Then the nonvanishing components of $\underline{\Gamma}$ are $\Gamma_{xx} = \Gamma_{yy} = \mathrm{i}\beta\sigma_0$ and $\Gamma_{zz} = -\frac{\mathrm{i}\omega}{4\pi}$. Define the resistivity tensor $\underline{\mathbf{R}}$ by

$$\underline{\mathbf{R}} = \underline{\sigma}^{-1}. \tag{14.114}$$

Then $R_{\pm} = \sigma_{\pm}^{-1}$ and $R_{zz} = \sigma_{zz}^{-1}$. The secular equation $\mid \underline{\mathbf{T}} \mid = 0$ reduces to two simple equations:

$$\omega_{\pm}^2 = s_{\mathrm{t}}^2 q^2 \mp \frac{ze\omega B_0}{Mc} + \frac{zm\mathrm{i}\omega}{M\tau} \frac{(1 - \mathrm{i}\beta)(\sigma_0 R_{\pm} - 1)}{1 - \mathrm{i}\beta\sigma_0 R_{\pm}} \tag{14.115}$$

for the circularly polarized transverse waves, and

$$\omega^2 = s_{\mathrm{l}}^2 q^2 + \frac{zm\mathrm{i}\omega}{M\tau} \left(\sigma_0 R_{xx} - 1 - \frac{q^2 l^2/3}{1 + \omega^2\tau^2} \right). \tag{14.116}$$

for the longitudinal waves. In (14.115) and (14.116), s_{t} and s_{l} are the speeds of transverse and longitudinal acoustic waves given, respectively, by

$$s_{\mathrm{t}} = \sqrt{\frac{C_{\mathrm{t}}}{M}} \text{ and } s_{\mathrm{l}} = \sqrt{\frac{zm}{3M} \frac{v_{\mathrm{F}}^2}{1 + \omega^2\tau^2} + \frac{C_\ell}{M}}. \tag{14.117}$$

From these results, we observe that

1. ω has both real and imaginary parts. The real part gives the frequency and hence velocity as a function of B_0. The imaginary part gives the acoustic attenuation as a function of ql, $\omega_{\mathrm{c}}\tau$, B_0, etc.

2. For longitudinal waves, if we use the semiclassical result for σ_{zz}, ω is completely independent of B_0.
3. In the case where the quantum mechanical result for σ_{zz} is used, both the velocity and attenuation display quantum oscillations of the de Haas van Alphen type.
4. For shear waves, the right and left circular polarizations have slightly different velocity and attenuation. This leads to a rotation of the plane of polarization of a linearly polarized wave. This is the *acoustic analogue of the Faraday effect*.
5. $R_\pm$ does depend on the magnetic field, and the acoustic wave shows a fairly abrupt increase in attenuation as the magnetic field is lowered below $\omega_c = qv_F$. This effect is called *Doppler shifted cyclotron resonances; DSCR*.
6. The helicon wave solution actually appears in (14.115), so that the equation for $\omega_\pm$ actually describes *helicon–phonon coupling*.

14.10.2 Helicon–Phonon Interaction

Look at (14.115), the dispersion relation of the circularly polarized shear waves propagating parallel to $\mathbf{B}_0$:

$$\omega_\pm^2 = s_t^2 q^2 \mp \frac{ze\omega B_0}{Mc} + \frac{zmi\omega}{M\tau}\,\frac{(1-i\beta)(\sigma_0 R_\pm - 1)}{1 - i\beta\sigma_0 R_\pm}. \tag{14.118}$$

In the local limit, where $\underline{\sigma}$ is shown in (13.116) and (13.117), we have

$$\sigma_0 R_\pm \simeq 1 + i\omega\tau \mp i\omega_c\tau. \tag{14.119}$$

Remember that $\beta \simeq \frac{c^2q^2}{\omega\tau\omega_p^2}$. Therefore, $1 - i\beta\sigma_0 R_\pm$ can be written

$$1 - i\beta\sigma_0 R_\pm \simeq 1 - i\frac{c^2q^2}{\omega\tau\omega_p^2}[1 + i\omega\tau \mp i\omega_c\tau]. \tag{14.120}$$

Let us assume $\omega_c\tau \gg 1$, $\omega_c \gg \omega$, and $\beta \ll 1$. Then, we can write that

$$1 - i\beta\sigma_0 R_\pm \approx 1 \mp \frac{\omega_H}{\omega}, \tag{14.121}$$

where $\omega_H = \frac{\omega_c c^2 q^2}{\omega_p^2}\left(1 - \frac{i}{\omega_c\tau}\right)$ is the *helicon frequency*. Substituting this in (14.118) gives

$$\omega_\pm^2 - s_t^2 q^2 \simeq \mp\omega\Omega_c \pm \frac{\Omega_c\omega^2}{\omega \mp \omega_H}, \tag{14.122}$$

where $\Omega_c = \frac{zeB_0}{Mc}$ is the ionic cyclotron frequency. Equation (14.122) can be rewritten as

$$(\omega - s_t q)(\omega + s_t q)(\omega \mp \omega_H) \simeq \omega\omega_H\Omega_c. \tag{14.123}$$

The dispersion curves are illustrated in Fig. 14.10. The helicon and transverse sound wave of the same polarization are strongly coupled by the term on the right-hand side of (14.123), when their phase velocities are almost equal. The solid lines depict the coupled helicon–phonon modes.

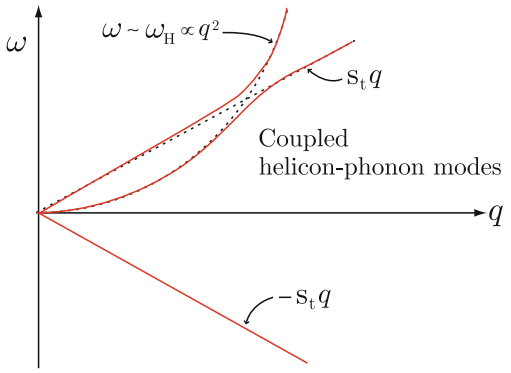


Fig. 14.10. Schematic of the roots of Eq.(14.123). The region of strongly coupled helicon–phonon modes for circularly polarized acoustic shear waves

14.10.3 Propagation Perpendicular to $\mathbf{B_0}$

For propagation to the dc magnetic field, the resistivity tensor has the following nonvanishing elements:

$$R_{xx} = \frac{\sigma_{yy}}{\sigma_{xx}\sigma_{yy} + \sigma_{xy}^2}, R_{yy} = \frac{\sigma_{xx}}{\sigma_{xx}\sigma_{yy} + \sigma_{xy}^2},$$
$$R_{xy} = -R_{yx} = \frac{\sigma_{xy}}{\sigma_{xx}\sigma_{yy} + \sigma_{xy}^2}, R_{zz} = \sigma_{zz}^{-1}. \tag{14.124}$$

The secular equation $|\underline{\mathbf{T}}| = 0$ again reduces to a 2×2 matrix and a 1×1 matrix, which can be written

$$\begin{pmatrix} \omega^2 - A_{xx} & -A_{xy} \\ -A_{yx} & \omega^2 - A_{yy} \end{pmatrix} \begin{pmatrix} \xi_x \\ \xi_y \end{pmatrix} = 0 \tag{14.125}$$

and

$$\left(\omega^2 - A_{zz}\right) \xi_z = 0, \tag{14.126}$$

where

$$A_{xx} = \frac{C_\mathrm{t}}{M} q^2 + \frac{zmi\omega}{M\tau} \frac{(1 - \mathrm{i}\beta)(\sigma_0 R_{xx} - 1)}{1 - \mathrm{i}\beta\sigma_0 R_{xx}},$$
$$A_{yy} = \frac{C_\ell}{M} q^2 + \frac{zmq^2 v_\mathrm{F}^2}{3M(1 + \omega^2\tau^2)}$$
$$+ \frac{zmi\omega}{M\tau} \left\{ \sigma_0 R_{yy} - 1 - \frac{\mathrm{i}\beta\sigma_0^2 R_{xy}^2}{1 - \mathrm{i}\beta\sigma_0 R_{xx}} - \frac{q^2 l^2}{3(1 + \omega^2\tau^2)} \right\},$$
$$A_{xy} = -A_{yx} = \frac{zmi\omega}{M\tau} \left\{ \frac{(1 - \mathrm{i}\beta)\sigma_0 R_{xy}}{1 - \mathrm{i}\beta\sigma_0 R_{xx}} - \omega_\mathrm{c}\tau \right\},$$
$$A_{zz} = \frac{C_\mathrm{t}}{M} q^2 + \frac{zmi\omega}{M\tau} \frac{(1 - \mathrm{i}\beta)(\sigma_0 R_{zz} - 1)}{1 - \mathrm{i}\beta\sigma_0 R_{zz}}. \tag{14.127}$$

The velocity and attenuation of sound can display several different types of oscillatory behavior as a function of applied magnetic field. Here we mention very briefly each of them.

1. *Cyclotron resonances* When $\omega = n\omega_c$, for propagation perpendicular to $\mathbf{B}_0$, the components of the conductivity tensor become very large. This gives rise to absorption peaks.
2. *de Haas–van Alphen type oscillations* Because the conductivity involves sums $\sum_{n,k_y,s}$ over quantum mechanical energy levels, as is shown in (13.81), the components of the conductivity tensor display de Haas–van Alphen type oscillations exactly as the magnetization, free energy, etc. One small difference is that instead of being associated with extremal orbits $\overline{v}_z = 0$, these oscillations in acoustic attenuation are associated with orbits for which $\bar{v}_z = s$.
3. *Geometric resonances* Due to the matrix elements $\langle \nu' | e^{i\mathbf{q}\cdot\mathbf{r}} | \nu \rangle$ which behave like Bessel function in the semiclassical limit, we find oscillations associated with $J_{n'-n}(q_\perp v_F/\omega_c)$ for propagation perpendicular to $\mathbf{B}_0$. The physical origin is associated with matching the cyclotron orbit diameter to multiples of the acoustic wavelength. Figure 14.11 shows the schematic of geometric resonances.
4. *Giant quantum oscillations* These result from the quantum nature of the energy levels together with "resonance" due to vanishing of the energy denominator in $\underline{\sigma}$. The physical picture and feeling for the "giant" nature of the oscillations can easily be obtained from consideration of
   (a) Energy conservation and momentum conservation in the transition $E_n(k_z) + \hbar\omega_{\mathbf{q}} \longrightarrow E_{n'}(k_z + q_z)$.
   (b) The Pauli exclusion principle.

Suppose that we had a uniform field $\mathbf{B}_0$ parallel to the z-axis. Then, with usual choice of gauge, our states are $|nk_yk_z\rangle$, with energies given by

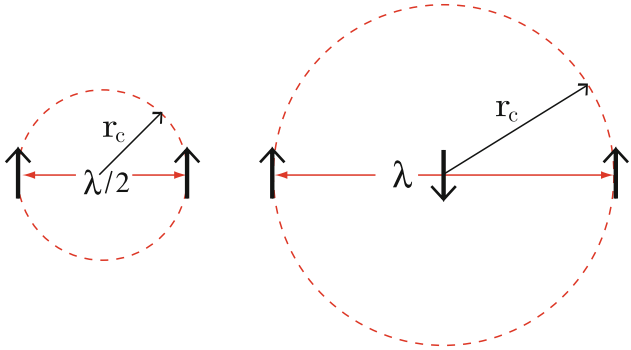


Fig. 14.11. Schematic of the origin of the geometric resonances in ultrasonic attenuation

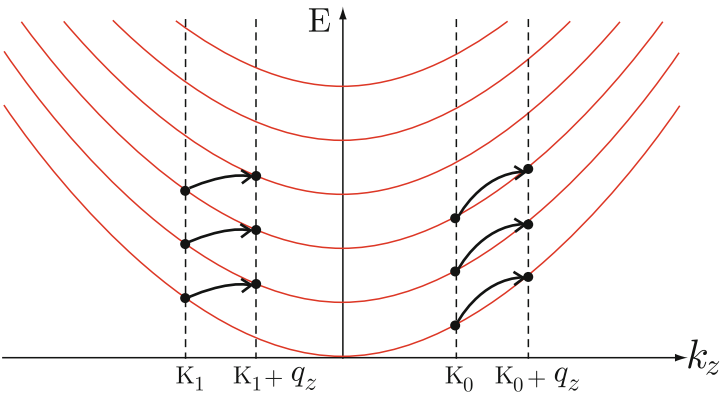


Fig. 14.12. Schematic of the transitions giving rise to giant quantum oscillations

$$E_n(k_z) = \hbar\omega_c \left(n + \frac{1}{2}\right) + \frac{\hbar^2 k_z^2}{2m}. \tag{14.128}$$

Now, we can do spectroscopy with these electrons, and have them absorb radiation. Thus, suppose that an electron absorbs a phonon of energy $\hbar\omega$ and momentum $\hbar q_z$. Then, energy conservation gives

$$E_{n'}(k_z + q_z) - E_n(k_z) = \hbar\omega_{q_z}. \tag{14.129}$$

Thus, only electrons with k_z given, with $\alpha = n' - n$, by

$$k_z = \frac{m}{\hbar q_z}(\omega \mp \alpha\omega_c) + \frac{q_z}{2} \tag{14.130}$$

will undergo transitions between different Landau levels. Let us call this value of k_z the parameter K_α. Then only electrons with $k_z = K_\alpha$ can make the transition $(n, k_z) \longrightarrow (n + \alpha, k_z + q_z)$ and absorb energy $\hbar\omega$. Figure 14.12 shows a schematic picture of the transitions giving rise to giant quantum oscillations.

To satisfy the exclusion principle $E_n(k_z) < \zeta$ and $E_{n+\alpha}(k_z + q_z) = E_n(k_z) + \hbar\omega > \zeta$. For $\omega_c \gg \omega$ this occurs only when the initial and final states are right at the Fermi surface. Then the absorption is "gigantic"; otherwise it is zero. The velocity as well as the attenuation displays these quantum oscillations. The oscillations, in principle, are infinitely sharp, but actually they are broadened out due to the fact that the Landau levels themselves are not perfectly sharp, and various other things. However, the oscillations are actually quite sharp, and so the amplitudes are much larger than the widths of absorption peaks.

Problems

14.1. Evaluate $\underline{\sigma}$ given, for zero temperature, by

$$\underline{\sigma}(q,\omega) = \frac{\omega_{\rm p}^2}{4\pi {\rm i}\omega}\left\{\underline{\mathbf{1}} + \underline{\mathbf{I}}(q,\omega)\right\},$$

where

$$\underline{\mathbf{I}}(q,\omega) = \frac{m}{N}\sum_{\mathbf{k}\mathbf{k}'}\frac{f_0(\varepsilon_{\mathbf{k}'}) - f_0(\varepsilon_{\mathbf{k}})}{\varepsilon_{\mathbf{k}'} - \varepsilon_{\mathbf{k}} - \hbar\omega}\langle\mathbf{k}'|\mathbf{V}_q|\mathbf{k}\rangle\langle\mathbf{k}'|\mathbf{V}_q|\mathbf{k}\rangle^*,$$

and use it to determine $E(y)$ and $\mathcal{Z}$. This is the problem of the anomalous skin effect, with the specular reflection boundary condition, in the absence of a dc magnetic field.

14.2. Derive the result for a helicon wave in a metal propagating at an angle θ to the direction of the applied magnetic field along the z-axis.

$$\omega \simeq \frac{\omega_{\rm c}c^2q^2\cos\theta}{\omega_{\rm p}^2 + c^2q^2}\left(1 + \frac{\rm i}{\omega_{\rm c}\tau\cos\theta}\right).$$

One may assume that $\omega_{\rm p} \gg \omega_{\rm c}$.

14.3. Investigate the case of helicon–plasmon coupling in a degenerate semiconductor in which $\omega_{\rm p}$ and $\omega_{\rm c}$ are of the same order-of-magnitude. Take $\omega_{\rm c}\tau \gg 1$, but let the angle θ and $\omega\tau$ be arbitrary. Study ω as a function of B_0, the applied magnetic field.

14.4. Evaluate σ_{xx}, σ_{xy}, and σ_{yy} from the Cohen–Harrison–Harrison result for propagatin perpendicular to $\mathbf{B}_0$ in the limit that $w = qv_{\rm F}/\omega_{\rm c} \ll 1$. Calculate to order w^2. See if any modes exist (at cyclotron harmonics) for the wave equation $\xi^2 = \varepsilon_{xx} + \frac{\varepsilon_{xy}^2}{\varepsilon_{yy}}$ where $\xi = cq/\omega$.

14.5. Consider a semi-infinite metal of dielectric function ε_1 to fill the space $z > 0$, and an insulator of dielectric constant ε_0 in the space $z < 0$.

(a) Show that the dispersion relation of the surface plasmon for the polarization with $E_x = 0$ and $E_y \neq 0 \neq E_z$ is written by

$$\frac{\varepsilon_1}{\alpha_1} + \frac{\varepsilon_0}{\alpha_0} = 0,$$

where α_0 and α_1 are the decay constants in the insulator and metal, respectively.

(b) Sketch $\frac{\omega}{\omega_{\rm p}}$ as a function of $\frac{cq_y}{\omega_{\rm p}}$ for the surface plasmon excitation.

14.6. Assume a vacuum–metal interface at $z = 0$ ($z < 0$ is vacuum), and let the electric charge density ρ appearing in Maxwell's equations vanish both inside and outside the metal. (This does not preclude a surface charge density concentrated in the plane $z = 0$.)

(a) Solve the wave equation in both vacuum and solid by assuming

$$\begin{aligned}\mathbf{E}_{\mathrm{v}}(\mathbf{r},t) &= \mathbf{E}_{\mathrm{v}}\mathrm{e}^{\mathrm{i}\omega t-\mathrm{i}q_y y+\alpha_{\mathrm{v}}z},\\ \mathbf{E}_{\mathrm{m}}(\mathbf{r},t) &= \mathbf{E}_{\mathrm{m}}\mathrm{e}^{\mathrm{i}\omega t-\mathrm{i}q_y y-\alpha_{\mathrm{m}}z},\end{aligned}$$

where q and ω are given and α_{v} and α_{m} must be determined.

(b) Apply the usual boundary conditions at $z = 0$ to determine the dispersion relation (ω vs. q) for surface waves.

(c) Sketch ω^2 as a function of q^2c^2 assuming $\omega\tau \gg 1$.

Summary

In this chapter, we study electromagnetic behavior of waves in metals. The linear response theory and Maxwell's equations are combined to obtain the condition of self-sustaining oscillations in metals. Both normal skin effect and Azbel–Kaner cyclotron resonance are discussed, and dispersion relations of plasmon modes and magnetoplasma modes are illustrated. Nonlocal effects in the wave dispersions are also pointed out, and behavior of cyclotron waves is considered as an example of the nonlocal behavior of the modes. General dispersion relation of the surface waves in the metal–insulator interface is derived by imposing standard boundary conditions, and the magnetoplasma surface waves are illustrated. Finally, we briefly discussed propagation of acoustic waves in metals.

The wave equation in metals, in the present of the total current $\mathbf{j}_{\mathrm{T}}$ $(= \mathbf{j}_0 + \mathbf{j}_{\mathrm{ind}})$, is written as

$$\mathbf{j}_{\mathrm{T}} = \underline{\Gamma} \cdot \mathbf{E},$$

where $\underline{\Gamma} = \frac{i\omega}{4\pi}\left\{(\xi^2 - 1)\underline{\mathbf{1}} - \boldsymbol{\xi}\boldsymbol{\xi}\right\}$. Here, the spin magnetization is neglected and $\boldsymbol{\xi} = \frac{c\mathbf{q}}{\omega}$. The $\mathbf{j}_0$ and $\mathbf{j}_{\mathrm{ind}}$ denote, respectively, some external current and the induced current $\mathbf{j}_{\mathrm{e}} = \underline{\sigma} \cdot \mathbf{E}$ by the self-consistent field $\mathbf{E}$.

For a system consisting of a semi-infinite metal filling the space $z > 0$ and vacuum in the space $z < 0$ and in the absence of $\mathbf{j}_0$, the wave equation reduces to $[\underline{\sigma}(\mathbf{q}, \omega) - \underline{\Gamma}(\mathbf{q}, \omega)] \cdot \mathbf{E} = 0$, and the electromagnetic waves are solutions of the secular equation $\mid \underline{\Gamma} - \underline{\sigma} \mid = 0$. The dispersion relations of the transverse and longitudinal electromagnetic waves propagating in the medium are given, respectively, by

$$c^2 q^2 = \omega^2 \varepsilon(\mathbf{q}, \omega) \text{ and } \varepsilon(\mathbf{q}, \omega) = 0.$$

In the range $\omega_{\mathrm{p}} \gg \omega$ and for $\omega\tau \gg 1$, the local theory of conduction $(ql \ll 1)$ gives a well-behaved field, inside the metal, of the form

$$E(z, t) = E_0 \mathrm{e}^{\mathrm{i}\omega t - z/\delta},$$

where $q = -\mathrm{i}\frac{\omega_{\mathrm{p}}}{c} = -\frac{\mathrm{i}}{\delta}$. The distance $\delta = \frac{c}{\omega_{\mathrm{p}}}$ is called the *normal skin depth*. If $l \gg \delta$, the local theory is not valid. The theory for this case, in which the $\mathbf{q}$ dependence of $\underline{\sigma}$ must be included, explains the *anomalous skin effect*.

In the absence of a dc magnetic field, the condition of the collective modes reduces to

$$(\omega^2 \varepsilon - c^2 q^2)^2 \varepsilon = 0.$$

Using the local (collisionless) theory of the dielectric function $\varepsilon \approx 1 - \frac{\omega_{\mathrm{p}}^2}{\omega^2}$, we have two degenerate transverse modes of frequency $\omega^2 = \omega_{\mathrm{p}}^2 + c^2 q^2$, and a longitudinal mode of frequency $\omega = \omega_{\mathrm{p}}$.

In the presence of a dc magnetic field along the z-axis and $\mathbf{q}$ in the y-direction, the secular equation for wave propagation is given by

$$\begin{vmatrix} \varepsilon_{xx} - \xi^2 & \varepsilon_{xy} & 0 \\ -\varepsilon_{xy} & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} - \xi^2 \end{vmatrix} = 0.$$

For the polarization with $\mathbf{E}$ parallel to the z-axis we have

$$\frac{c^2 q^2}{\omega^2} = 1 - \frac{4\pi \mathrm{i}}{\omega} \sigma_{zz}(\mathbf{q}, \omega),$$

where $\sigma_{zz}(\mathbf{q}, \omega)$ is the nonlocal conductivity. For $\omega_\mathrm{p} \gg n\omega_\mathrm{c}$ and in the limit $q \to 0$, we obtain the cyclotron waves given by $\omega^2 = n^2\omega_\mathrm{c}^2 + \mathrm{O}(q^{2n})$. They propagate perpendicular to the dc magnetic field, and depend for their existence on the $\mathbf{q}$ dependence of $\underline{\sigma}$.

For a system consisting of a metal of dielectric function ε_1 filling the space $z > 0$ and an insulator of dielectric constant ε_0 in the space $z < 0$, the waves localized near the interface ($z = 0$) are written as

$$\mathbf{E}^{(1)}(\mathbf{r}, t) = \mathbf{E}^{(1)} \mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}q_y y - \alpha_1 z},$$
$$\mathbf{E}^{(0)}(\mathbf{r}, t) = \mathbf{E}^{(0)} \mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}q_y y + \alpha_0 z}.$$

The superscripts 1 and 0 refer, respectively, to the metal and dielectric. The boundary conditions at the plane $z = 0$ are the standard ones of continuity of the tangential components of $\mathbf{E}$ and $\mathbf{H}$, and of the normal components of $\mathbf{D}$ and $\mathbf{B}$. For the polarization with $E_x = 0$, but $E_y \neq 0 \neq E_z$, the dispersion relation of the surface plasmon is written as

$$\frac{\varepsilon_1}{\alpha_1} + \frac{\varepsilon_0}{\alpha_0} = 0,$$

where $\alpha_1 = (\omega_\mathrm{p}^2 + q_y^2 - \omega^2)^{1/2}$ and $\alpha_0 = (q_y^2 - \varepsilon_0 \omega^2)^{1/2}$.

The classical equation of motion of the ionic displacement field $\boldsymbol{\xi}(\mathbf{r}, t)$ in a metal is written as

$$M \frac{\partial^2 \boldsymbol{\xi}}{\partial t^2} = C_\ell \nabla(\nabla \cdot \boldsymbol{\xi}) - C_\mathrm{t} \nabla \times (\nabla \times \boldsymbol{\xi}) + ze\mathbf{E} + \frac{ze}{c} \dot{\boldsymbol{\xi}} \times (\mathbf{B}_0 + \mathbf{B}) + \mathbf{F}.$$

Here, C_ℓ and C_t are elastic constants, and the *collision drag* force $\mathbf{F}$ is $\mathbf{F} = -\frac{zm}{n_0 e\tau}(\mathbf{j}_\mathrm{e} + \mathbf{j}_\mathrm{I})$, where the ionic current density is $\mathbf{j}_\mathrm{I} = n_0 e \dot{\boldsymbol{\xi}}$. The self-consistent electric field $\mathbf{E}$ is determined from the Maxwell equations $\mathbf{j}_\mathrm{T} = \Gamma(\mathbf{q}, \omega) \cdot \mathbf{E}$:

$$\mathbf{E}(\mathbf{q}, \omega) = [\underline{\Gamma} - \underline{\sigma}]^{-1} \left(\mathrm{i}\omega ne\, \underline{\mathbf{1}} - \underline{\sigma} \cdot \underline{\Delta}\right) \cdot \boldsymbol{\xi}.$$

Here, a tensor $\underline{\Delta}$ is defined by $\underline{\Delta} = \frac{n_0 e \mathrm{i}\omega}{\sigma_0} \left\{ \underline{\mathbf{1}} - \frac{1}{3} \frac{q^2 l^2}{\mathrm{i}\omega\tau(1 + \mathrm{i}\omega\tau)} \hat{\mathbf{q}}\hat{\mathbf{q}} \right\}$, where $\hat{\mathbf{q}} = \frac{\mathbf{q}}{|\mathbf{q}|}$. The equation of motion is thus of the form $\underline{\mathbf{T}}(\mathbf{q}, \omega) \cdot \boldsymbol{\xi}(\mathbf{q}, \omega) = 0$, where $\underline{\mathbf{T}}$ is a very complicated tensor. The normal modes of an infinite medium are determined from the secular equation

$$\det \mid \underline{\mathbf{T}}(\mathbf{q}, \omega) \mid = 0.$$

The solutions $\omega(\mathbf{q})$ have both a real and imaginary parts; the real part determines the velocity of sound and the imaginary part the attenuation of the wave.

15

Superconductivity

15.1 Some Phenomenological Observations of Superconductors

Superconductors are materials that behave as normal metals at high temperatures ($T > T_c$; however, below T_c they have the following properties:

1. The dc resistivity vanishes.
2. They are perfect diamagnets; by this we mean that any magnetic field that is present in the bulk of the sample when $T > T_c$ is expelled when T is lowered through the transition temperature. This is called the *Meissner effect*.
3. The electronic properties can be understood by assuming that an energy gap 2Δ exists in the electronic spectrum at the Fermi energy.

Some common superconducting elements and their transition temperatures are given in Table 15.1.

Resistivity

A plot of $\rho(T)$, the resistivity versus temperature T, looks like the diagram shown in Fig. 15.1. Current flows in superconductor without dissipation. Persistent currents in superconducting rings have been observed to circulate without decaying for years. There is a critical current density j_c which, if exceeded, will cause the superconductor to go into the normal state. The ac current response is also dissipationless if the frequency ω satisfies $\omega < \frac{\Delta}{\hbar}$, where Δ is an energy of the order of $k_B T_c$.

Thermoelectric Properties

Superconducting materials are usually poor thermal conductors. In normal metals an electric current is accompanied by a *thermal current* that is associated with the *Peltier effect*. No Peltier effect occurs in superconductors; the current carrying electrons appear to carry no entropy.

Table 15.1. Transition temperatures of some selected superconducting elements

Elements	Al	Sn	Hg	In	FCC La	HCP La	Nb	Pb
T_c (K)	1.2	3.7	4.2	3.4	6.6	4.9	9.3	7.2

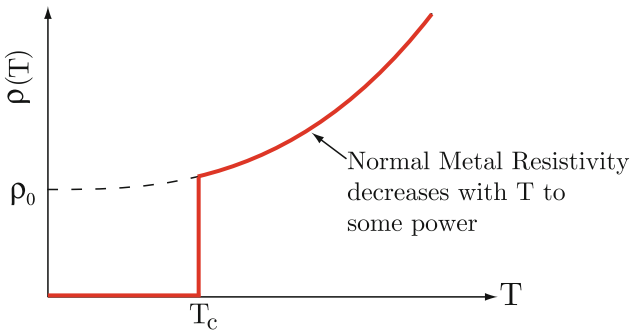


Fig. 15.1. Temperature dependence of the resistivity of typical superconducting metals

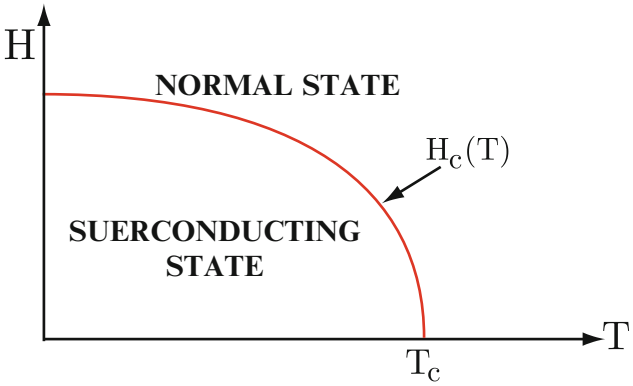


Fig. 15.2. Temperature dependence of the critical magnetic field of a typical superconducting material

Magnetic Properties

There is a critical magnetic field $H_c(T)$, which depends on temperature. When H is above $H_c(T)$, the material is in the normal state; when $H < H_c(T)$ it is superconducting. A plot of $H_c(T)$ versus T is sketched in Fig. 15.2.

In a *type I superconductor*, the magnetic induction B must vanish in the bulk of the superconductor for $H < H_c(T)$. But we have

$$B = H + 4\pi M = 0 \quad \text{for} \quad H < H_c(T), \tag{15.1}$$

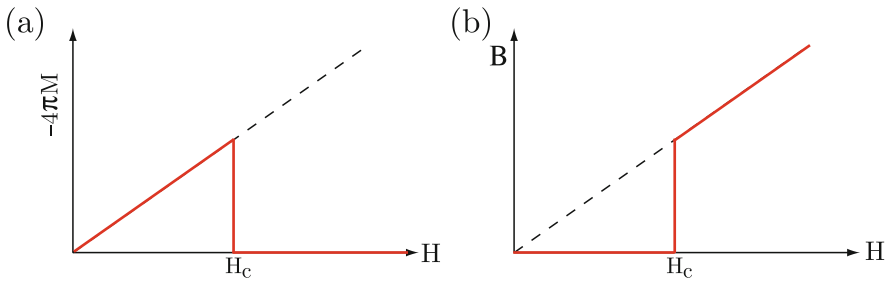


Fig. 15.3. Magnetic field dependence of the magnetization M and magnetic induction B of a type I superconducting material

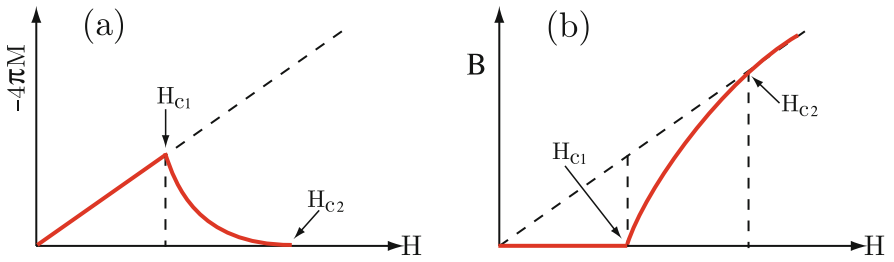


Fig. 15.4. Magnetic field dependence of the magnetization M and magnetic induction B of a type II superconducting material

which implies that

$$M = -\frac{H}{4\pi} \quad \text{for} \quad H < H_{\mathrm{c}}(T). \tag{15.2}$$

This behavior is illustrated in Fig. 15.3.

In a *type II superconductor*, the magnetic field starts to penetrate the sample at an applied field H_{c1} lower than the H_{c}. The Meissner effect is incomplete yet until at H_{c2}. The $\mathbf{B}$ approaches $\mathbf{H}$ only at an upper critical field H_{c2}. Figure 15.4 shows the magnetic field dependence of the magnetization, $-4\pi M$, and the magnetic induction B in a type II superconducting material. Between H_{c2} and H_{c1} flux penetrates the superconductor giving a *mixed state* consisting of superconductor penetrated by threads of the material in its normal state or flux lines. Abrikosov showed that the mixed state consists of vortices each carrying a single flux $\Phi = \frac{hc}{2e}$. These vortices are arranged in a regular two-dimensional array.

Specific Heat

The specific heat shows a jump at T_{c} and decays exponentially with an energy Δ of the order of $k_{\mathrm{B}}T_{\mathrm{c}}$ as $\mathrm{e}^{-\Delta/k_{\mathrm{B}}T_{\mathrm{c}}}$ below T_{c}, as is shown in Fig. 15.5. There is

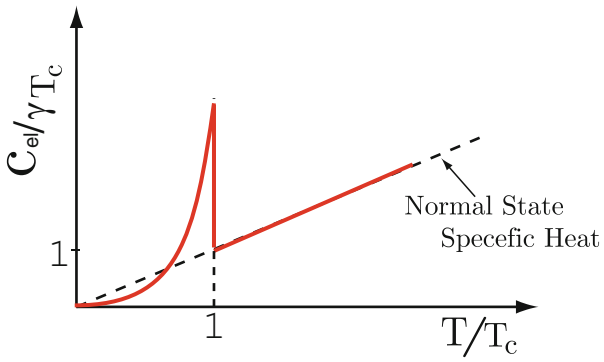


Fig. 15.5. Temperature dependence of the specific heat of a typical superconducting material

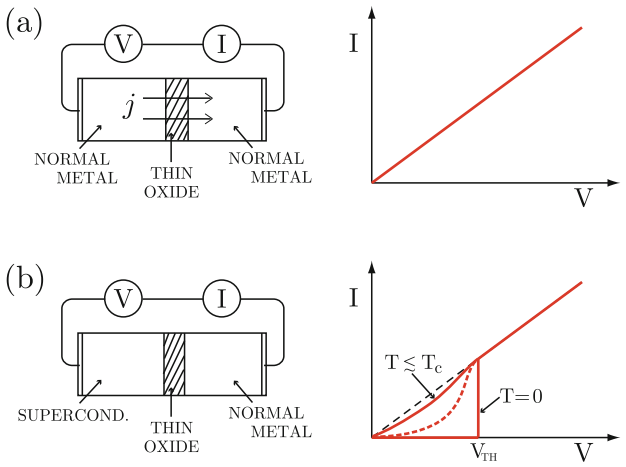


Fig. 15.6. Tunneling current behavior for (**a**) a normal metal–oxide–normal metal structure and (**b**) a superconductor–oxide–normal metal structure

a second order phase transition (constant entropy, constant volume, no latent heat) with discontinuity in the specific heat.

Tunneling Behavior

If one investigates tunneling through a thin oxide, in the case of two normal metals, one obtains a linear current–potential difference curve, as is sketched in Fig. 15.6a. For a superconductor–oxide–normal metal structure, a very different behavior of the tunneling current versus potential difference is obtained. Fig. 15.6b shows the tunneling current–potential difference curve of a superconductor–oxide–normal metal structure.

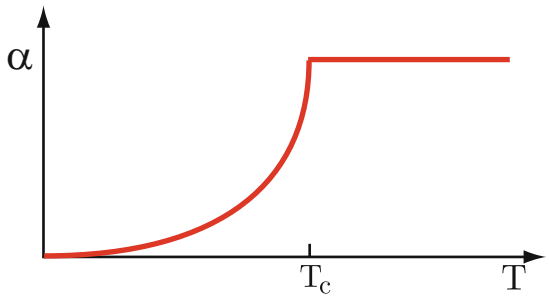


Fig. 15.7. Temperature dependence of the damping constant of low frequency sound waves in a superconducting material

Acoustic Attenuation

For $T < T_c$ and $\omega < 2\Delta$, there is no attenuation of sound due to electron excitation. In Fig. 15.7, the damping constant α of low frequency sound waves in a superconductor is sketched as a function of temperature.

15.2 London Theory

Knowing the experimental properties of superconductors, London introduced a phenomenological theory that can be described as follows:

1. The superconducting material contains two fluids below T_c. $\frac{n_S(T)}{n}$ is the fraction of the electron fluid that is in the *super fluid state*. $\frac{n_N(T)}{n} = \left[1 - \frac{n_S(T)}{n}\right]$ is the fraction in the *normal state*. The total density of electrons in the superconducting material is $n = n_N + n_S$.
2. Both the normal fluid and super fluid respond to external fields, but the superfluid is dissipationless while the normal fluid is not. We can write the electrical conductivities for the normal and super fluids as follows:

$$\begin{aligned} \sigma_N &= \frac{n_N e^2 \tau_N}{m}, \\ \sigma_S &= \frac{n_S e^2 \tau_S}{m}, \end{aligned} \tag{15.3}$$

   but $\tau_S \to \infty$ giving $\sigma_S \to \infty$.
3. $$\begin{aligned} n_S(T) &\to n \text{ as } T \to 0, \\ n_S(T) &\to 0 \text{ as } T \to T_c. \end{aligned}$$
4. To explain the Meissner effect, London proposed the *London equation*

$$\nabla \times \mathbf{j} + \frac{n_S e^2}{mc}\mathbf{B} = 0. \tag{15.4}$$

How does the London equation arise? Let us consider the equation of motion of a super fluid electron, which is dissipationless, in an electric field **E** that is momentarily present in the superconductor:

$$m\frac{\mathrm{d}v_{\mathrm{S}}}{\mathrm{d}t} = -eE$$

where v_{S} is the mean velocity of the super fluid electron caused by the field **E**. But the current density **j** is simply

$$\mathbf{j} = -n_{\mathrm{S}}e\mathbf{v}_{\mathrm{S}}. \tag{15.5}$$

Notice that this gives the relation

$$\frac{\mathrm{d}\mathbf{j}}{\mathrm{d}t} = -n_{\mathrm{S}}e\frac{\mathrm{d}v_{\mathrm{S}}}{\mathrm{d}t} = \frac{n_{\mathrm{S}}e^2}{m}\mathbf{E}. \tag{15.6}$$

Equation (15.6) describes the dynamics of collisionless electrons in a perfect conductor, which cannot sustain an electric field in stationary conditions. Now, from Faraday's induction law, we have

$$\nabla \times \mathbf{E} = -\frac{1}{c}\dot{\mathbf{B}}. \tag{15.7}$$

Combining this with (15.6) gives us

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\nabla \times \mathbf{j} + \frac{n_{\mathrm{S}}e^2}{mc}\mathbf{B}\right] = 0. \tag{15.8}$$

The solution of (15.8) is that

$$\nabla \times \mathbf{j} + \frac{n_{\mathrm{S}}e^2}{mc}\mathbf{B} = \text{ constant.}$$

Because in the bulk of a superconductor the magnetic induction **B** must be zero, London proposed that for superconductors, the "constant" had to be zero and $\mathbf{j} = -\frac{n_{\mathrm{s}}e^2}{mc}\mathbf{A}$ [called the London gauge] giving (15.4). The London equation implies that, in stationary conditions, a superconductor cannot sustain a magnetic field in its interior, but only within a narrow surface layer. If we use the relation

$$\nabla \times \mathbf{B} = \frac{4\pi}{c}\mathbf{j}, \tag{15.9}$$

(This is the Maxwell equation for $\nabla \times \mathbf{B}$ in stationary conditions without the displacement current $\frac{1}{c}\dot{\mathbf{E}}$.), we can obtain

$$\nabla \times (\nabla \times \mathbf{B}) = \frac{4\pi}{c}\nabla \times \mathbf{j} = -\frac{4\pi}{c}\frac{n_{\mathrm{S}}e^2}{mc}\mathbf{B}. \tag{15.10}$$

But, $\nabla \times (\nabla \times \mathbf{B}) = \nabla(\nabla \cdot \mathbf{B}) - \nabla^2 \mathbf{B}$ giving

$$\nabla^2 \mathbf{B} = \frac{4\pi n_S e^2}{mc^2}\mathbf{B} \equiv \frac{1}{\Lambda_L^2}\mathbf{B}. \tag{15.11}$$

The solutions of (15.11) show a magnetic field decaying exponentially with a characteristic length Λ. One can also obtain the relation $\nabla^2 \dot{\mathbf{j}} = \frac{4\pi n_S e^2}{mc^2}\mathbf{j}$. The quantity $\Lambda_L = \sqrt{\frac{mc^2}{4\pi n_S e^2}}$ is called the *London penetration depth*. For typical semiconducting materials, $\Lambda_L \sim 10 - 10^2$nm. If we have a thin superconducting film filling the space $-a < z < 0$ as shown in Fig. 15.8a, then the magnetic field $\mathbf{B}$ parallel to the superconductor surface has to fall off inside the superconductor from B_0, the value outside, as

$$B(z) = B_0 \mathrm{e}^{-|z|/\Lambda_L} \text{ near the surface } z = 0$$

and

$$B(z) = B_0 \mathrm{e}^{-|z+a|/\Lambda_L} \text{ near the surface } z = -a.$$

Figure 15.8b shows the schematic of the flux penetration in the superconducting film. The flux penetrates only a distance $\Lambda_L \leq 10^2$nm. One can show that it is impossible to have a magnetic field $\mathbf{B}$ normal to the superconductor surface but homogeneous in the $x - y$ plane.

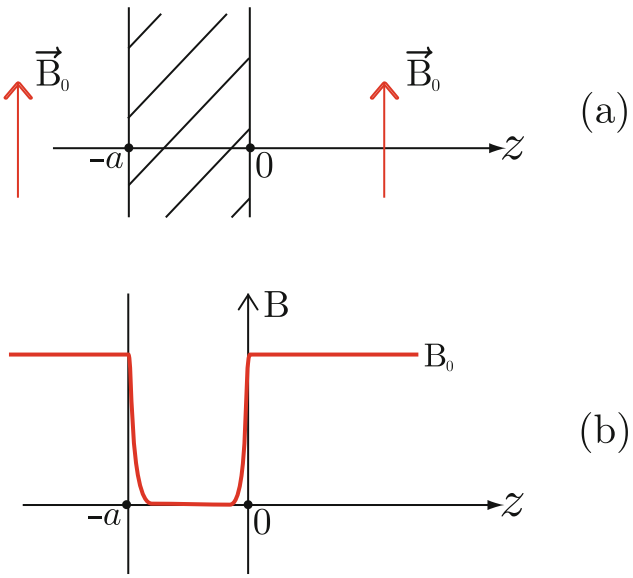


Fig. 15.8. A superconducting thin film (**a**) and the magnetic field penetration (**b**)

15.3 Microscopic Theory–An Introduction

In the early 1950s Frölich suggested that the attractive part of the electron–phonon interaction was responsible for superconductivity predicting the isotopic effect. The *isotope effect*, the dependence of T_c on the mass of the elements making up the lattice was discovered experimentally independent of Frölich's work, but it was in complete agreement with it. Both Frölich, and later Bardeen, attempted to describe superconductivity in terms of an electron self-energy associated with virtual exchange of phonons. Both attempts failed. In 1957, Bardeen, Cooper, and Schrieffer (BCS) produced the first correct microscopic theory of superconductivity.[1] The critical idea turned out to be the *pair correlations* that became manifest in a simple little paper by L.N. Cooper.[2]

Let us consider electrons in a simple metal described by the Hamiltonian $H = H_0 + H_{\rm ep}$, where H_0 and H_{ep} are, respectively the unperturbed Hamiltonian for a Bravais lattice and the interaction Hamiltonian of the electrons with the screened ions. Here we neglect the effect of the periodic part of the stationary lattice to write H_0 by

$$H_0 = \sum_{\mathbf{k},\sigma} \varepsilon_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + \sum_{\mathbf{q},\mathbf{s}} \hbar\omega_{\mathbf{q},\mathbf{s}} a^\dagger_{\mathbf{q},\mathbf{s}} a_{\mathbf{q},\mathbf{s}},$$

where σ and $\mathbf{s}$ denote, respectively, the spin of the electrons and the three dimensional polarization vector of the phonons, and $a_{\mathbf{q},\mathbf{s}}$ annihilates a phonon of wave vector q and polarization $\mathbf{s}$, and $c_{\mathbf{k}\sigma}$ annihilates an electron of wave vector $\mathbf{k}$ and spin σ. We will show the basic ideas leading to the microscopic theory of superconductivity.

15.3.1 Electron–Phonon Interaction

The electron–phonon interaction can be expressed as

$$H_{\rm ep} = \sum_{\mathbf{k},\mathbf{q},\sigma} M_{\mathbf{q}} \left(a^\dagger_{-\mathbf{q}} + a_{\mathbf{q}} \right) c^\dagger_{\mathbf{k}+\mathbf{q}\sigma} c_{\mathbf{k}\sigma}, \tag{15.12}$$

where $M_{\mathbf{q}}$ is the electron–phonon matrix element defined, in a simple model discussed earlier, by

$$M_{\mathbf{q}} = \mathrm{i} \sqrt{\frac{N\hbar}{2M\omega_{\mathbf{q}}}} \mid \mathbf{q} \mid V_{\mathbf{q}}.$$

Here $V_{\mathbf{q}}$ is the Fourier transform of the potential due to a single ion at the origin, and the phonon spectrum is assumed isotropic for simplicity. (In this case, only the longitudinal modes of $\mathbf{s}$ parallel to $\mathbf{q}$ give finite contribution

[1] J. Bardeen, L.N. Cooper, J.R. Schrieffer, Phys. Rev. **108**, 1175 (1957).
[2] L.N. Cooper, Phys. Rev. **104**, 1189 (1956).

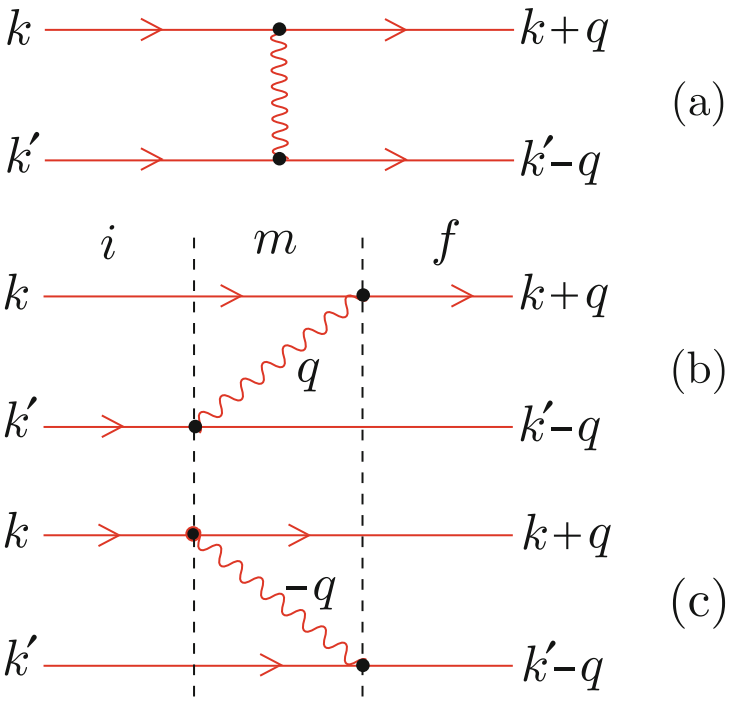


Fig. 15.9. Electron–phonon interaction (**a**) Effective electron–electron interaction through virtual exchange of phonons (**b**) and (**c**): Two possible intermediate states in the effective electron–electron interaction

to $H_{\rm ep}$.) This $H_{\rm ep}$ can give rise to an *effective electron–electron interaction* associated with virtual exchange of phonons as denoted in Fig. 15.9a. The figure shows that an electron polarizes the lattice and another electron interacts with the polarized lattice. There are two possible intermediate states in this process as shown in Fig. 15.9b and c. In Fig. 15.9b the initial energy is $E_i = \varepsilon_{\mathbf{k}} + \varepsilon_{\mathbf{k}'}$ and the intermediate state energy is $E_m = \varepsilon_{\mathbf{k}} + \varepsilon_{\mathbf{k}'-\mathbf{q}} + \hbar\omega_{\mathbf{q}}$. In Fig. 15.9c the initial energy is the same, but the intermediate state energy is $E_m = \varepsilon_{\mathbf{k}+\mathbf{q}} + \varepsilon_{\mathbf{k}'} + \hbar\omega_{\mathbf{q}}$. We can write this interaction in the second order as

$$\sum_m \frac{\langle f \mid H_{\rm ep} \mid m\rangle\langle m \mid H_{\rm ep} \mid i\rangle}{E_i - E_m}. \tag{15.13}$$

This gives us the interaction part of the Hamiltonian as follows:

$$H' = \sum_{\substack{\mathbf{k}\mathbf{k}'\mathbf{q}\\ \sigma\sigma'}} \mid M_{\mathbf{q}} \mid^2 \left\{ \frac{\langle f|c^{\dagger}_{\mathbf{k}+\mathbf{q}\sigma} c_{\mathbf{k}\sigma} a_{\mathbf{q}}|m\rangle\langle m|c^{\dagger}_{\mathbf{k}'-\mathbf{q}\sigma'} c_{\mathbf{k}'\sigma'} a^{\dagger}_{\mathbf{q}}|i\rangle}{\varepsilon(\mathbf{k}') - [\varepsilon(\mathbf{k}'-\mathbf{q}) + \hbar\omega_{\mathbf{q}}]} + \frac{\langle f|c^{\dagger}_{\mathbf{k}'-\mathbf{q}\sigma'} c_{\mathbf{k}'\sigma'} a_{-\mathbf{q}}|m\rangle\langle m|c^{\dagger}_{\mathbf{k}+\mathbf{q}\sigma} c_{\mathbf{k}\sigma} a^{\dagger}_{-\mathbf{q}}|i\rangle}{\varepsilon(\mathbf{k}) - [\varepsilon(\mathbf{k}+\mathbf{q}) + \hbar\omega_{\mathbf{q}}]} \right\}. \tag{15.14}$$

One can take the Hamiltonian

$$\begin{aligned} H &= H_0 + H' \\ &= \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}\sigma} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + \sum_{\mathbf{q}} \hbar\omega_{\mathbf{q}} a^\dagger_{\mathbf{q}} a_{\mathbf{q}} + \sum_{\mathbf{k},\mathbf{q},\sigma} M_{\mathbf{q}} \left(a^\dagger_{-\mathbf{q}} + a_{\mathbf{q}} \right) c^\dagger_{\mathbf{k}+\mathbf{q}\sigma} c_{\mathbf{k}\sigma} \end{aligned} \tag{15.15}$$

and make a canonical transformation

$$H_S = \mathrm{e}^{-S} H \mathrm{e}^{S} \tag{15.16}$$

where the operator S is defined by

$$S = \sum_{\mathbf{k}\mathbf{q}\sigma} M_{\mathbf{q}} \left(\alpha a^\dagger_{-\mathbf{q}} + \beta a_{\mathbf{q}} \right) c^\dagger_{\mathbf{k}+\mathbf{q}\sigma} c_{\mathbf{k}\sigma} \tag{15.17}$$

to eliminate the $a_{\mathbf{q}}$ and $a^\dagger_{-\mathbf{q}}$ operators to lowest order. To do so, α and β in (15.17) must be chosen, respectively, as

$$\begin{aligned} \alpha &= \left[\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}+\mathbf{q}) - \hbar\omega_{\mathbf{q}} \right]^{-1} \\ \beta &= \left[\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}+\mathbf{q}) + \hbar\omega_{\mathbf{q}} \right]^{-1} . \end{aligned} \tag{15.18}$$

Then, the transformed Hamiltonian is

$$H_S = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}\sigma} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + \sum_{\mathbf{k}\sigma,\mathbf{k}'\sigma',\mathbf{q}} W(\mathbf{k},\mathbf{q}) c^\dagger_{\mathbf{k}+\mathbf{q}\sigma} c^\dagger_{\mathbf{k}'-\mathbf{q}\sigma'} c_{\mathbf{k}'\sigma'} c_{\mathbf{k}\sigma}, \tag{15.19}$$

where $W_{\mathbf{kq}}$ is defined by

$$W_{\mathbf{kq}} = \frac{\mid M_{\mathbf{q}} \mid^2 \hbar\omega_{\mathbf{q}}}{\left[\varepsilon(\mathbf{k}+\mathbf{q}) - \varepsilon(\mathbf{k}) \right]^2 - \left(\hbar\omega_{\mathbf{q}} \right)^2}. \tag{15.20}$$

Note that when $\Delta E = \varepsilon(\mathbf{k}+\mathbf{q}) - \varepsilon(\mathbf{k})$ is smaller than $\hbar\omega_{\mathbf{q}}$, $W_{\mathbf{kq}}$ is negative. This results in an effective electron–electron attraction.

15.3.2 Cooper Pair

Leon Cooper investigated the simple problem of a pair of electrons interacting in the presence of a Fermi sea of "spectator electrons". He took the pair to have total momentum $\mathbf{P} = 0$ and spin $S = 0$. The Hamiltonian is written as

$$H = \sum_{\boldsymbol{\ell},\sigma} \varepsilon_{\boldsymbol{\ell}} c^\dagger_{\boldsymbol{\ell}\sigma} c_{\boldsymbol{\ell}\sigma} - \frac{1}{2} V \sum_{\boldsymbol{\ell}\boldsymbol{\ell}'\sigma} c^\dagger_{\boldsymbol{\ell}'\sigma} c^\dagger_{-\boldsymbol{\ell}'\bar{\sigma}} c_{-\boldsymbol{\ell}\bar{\sigma}} c_{\boldsymbol{\ell}\sigma}, \tag{15.21}$$

where $\varepsilon_{\boldsymbol{\ell}} = \frac{\hbar^2 \ell^2}{2m}$, $\{ c_{\boldsymbol{\ell}\sigma}, c^\dagger_{\boldsymbol{\ell}'\sigma'} \} = \delta_{\boldsymbol{\ell}\boldsymbol{\ell}'} \delta_{\sigma\sigma'}$, and the strength of the interaction, V, is taken as a constant for a small region of $\mathbf{k}$-space close to the Fermi surface. The interaction term allows for pairscattering from $(\boldsymbol{\ell}\sigma, -\boldsymbol{\ell}\bar{\sigma})$ to $(\boldsymbol{\ell}'\sigma, -\boldsymbol{\ell}'\bar{\sigma})$.

Cooper took a variational trial function

$$\Psi = \sum_{\mathbf{k}} a_{\mathbf{k}} c^{\dagger}_{\mathbf{k}\sigma} c^{\dagger}_{-\mathbf{k}\bar{\sigma}} \mid G >, \tag{15.22}$$

where $\mid G\rangle$ is the Fermi sea of spectator electrons, $\mid G\rangle = \Pi^{k<k_{\mathrm{F}}}_{|\mathbf{k}|} c^{\dagger}_{\mathbf{k}\sigma} c^{\dagger}_{-\mathbf{k}\bar{\sigma}} \mid 0 >$. If we evaluate

$$\langle \Psi \mid H \mid \Psi \rangle = E, \tag{15.23}$$

we get

$$E = 2 \sum_{\boldsymbol{\ell}} \varepsilon_{\boldsymbol{\ell}} a^{*}_{\boldsymbol{\ell}} a_{\boldsymbol{\ell}} - V \sum_{\boldsymbol{\ell}\boldsymbol{\ell}'} a^{*}_{\boldsymbol{\ell}'} a_{\boldsymbol{\ell}}. \tag{15.24}$$

The coefficient $a_{\boldsymbol{\ell}}$ is determined by requiring $E\{a_{\boldsymbol{\ell}}\}$ to be minimum subject to the constraint $\sum_{\boldsymbol{\ell}} a^{*}_{\boldsymbol{\ell}} a_{\boldsymbol{\ell}} = 1$. This can be carried out using a Lagrange multiplier λ as follows:

$$\frac{\partial}{\partial a^{*}_{\mathbf{k}}} \left\{ 2 \sum_{\boldsymbol{\ell}} \varepsilon_{\boldsymbol{\ell}} a^{*}_{\boldsymbol{\ell}} a_{\boldsymbol{\ell}} - V \sum_{\boldsymbol{\ell}\boldsymbol{\ell}'} a^{*}_{\boldsymbol{\ell}'} a_{\boldsymbol{\ell}} - \lambda \sum_{\boldsymbol{\ell}} a^{*}_{\boldsymbol{\ell}} a_{\boldsymbol{\ell}} \right\} = 0. \tag{15.25}$$

This gives

$$2\varepsilon_{\mathbf{k}} a_{\mathbf{k}} - V \sum_{\boldsymbol{\ell}} a_{\boldsymbol{\ell}} - \lambda a_{\mathbf{k}} = 0. \tag{15.26}$$

This can be written

$$a_{\mathbf{k}} = \frac{V \sum_{\boldsymbol{\ell}} a_{\boldsymbol{\ell}}}{2\varepsilon_{\mathbf{k}} - \lambda}. \tag{15.27}$$

Define a constant $C = \sum_{\boldsymbol{\ell}} a_{\boldsymbol{\ell}}$. Then, we have

$$a_{\mathbf{k}} = \frac{VC}{2\varepsilon_{\mathbf{k}} - \lambda}. \tag{15.28}$$

Summing over $\mathbf{k}$ and using the fact $C = \sum_{\boldsymbol{\ell}} a_{\boldsymbol{\ell}}$ we have

$$C = VC \sum_{\mathbf{k}} \frac{1}{2\varepsilon_{\mathbf{k}} - \lambda}, \tag{15.29}$$

or

$$f(\lambda) = \sum_{\mathbf{k}} \frac{1}{2\varepsilon_{\mathbf{k}} - \lambda} = \frac{1}{V}. \tag{15.30}$$

The values of $\varepsilon_{\mathbf{k}}$ form a closely spaced quasi continuum extending from the energy E_{F} to roughly $E_{\mathrm{F}} + \hbar\omega_{\mathrm{D}}$ where ω_{D} is the Debye frequency. In Fig. 15.10

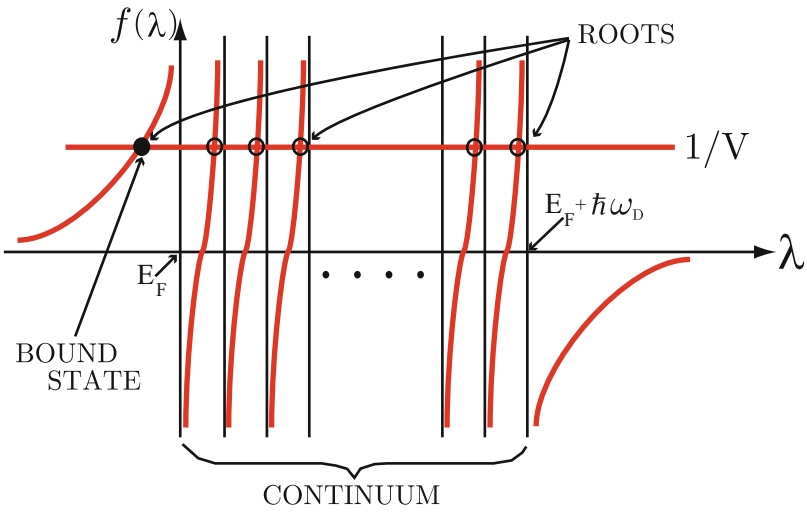


Fig. 15.10. Graphical solution of (15.30)

the function $f(\lambda)$ is displayed as a function of λ, and it shows the graphical solution of (15.30). Note that $f(\lambda)$ goes from $-\infty$ to ∞ every time λ crosses a value of $2\varepsilon_{\mathbf{k}}$ in the quasi continuum. If we take (15.26)

$$(2\varepsilon_{\mathbf{k}} - \lambda)\, a_{\mathbf{k}} - V \sum_{\boldsymbol{\ell}} a_{\boldsymbol{\ell}} = 0, \tag{15.31}$$

multiply by $a^*_{\mathbf{k}}$ and sum over $\mathbf{k}$, we obtain

$$\sum_{\mathbf{k}} (2\varepsilon_{\mathbf{k}} - \lambda)\, a^*_{\mathbf{k}} a_{\mathbf{k}} - V \sum_{\mathbf{k}\boldsymbol{\ell}} a^*_{\mathbf{k}} a_{\boldsymbol{\ell}} = 0. \tag{15.32}$$

This is exactly the same equation we obtained from writing

$$\langle \Psi \mid H - E \mid \Psi \rangle = 0, \tag{15.33}$$

if we take $\lambda = E$, the energy of the variational state Ψ. Thus, our equation for $f(\lambda) = \frac{1}{V}$ could be rewritten

$$\frac{1}{V} = \sum_{\mathbf{k}} \frac{1}{2\varepsilon_{\mathbf{k}} - E}. \tag{15.34}$$

Approximate the sum in (15.34) by an integral over the energy ε and write

$$\frac{1}{V} = \int_{E_{\mathrm{F}}}^{E_{\mathrm{F}} + \hbar\omega_{\mathrm{D}}} \frac{g(\varepsilon)\, \mathrm{d}\varepsilon}{2\varepsilon - E}. \tag{15.35}$$

Now, let us take $g(\varepsilon) \simeq g(E_{\mathrm{F}}) \equiv g$ in the region of integration to obtain

$$\frac{1}{V} = \frac{g}{2} \int_{E_{\mathrm{F}}}^{E_{\mathrm{F}}+\hbar\omega_{\mathrm{D}}} \frac{\mathrm{d}x}{x - E/2}. \tag{15.36}$$

Integrating (15.36) out gives

$$\frac{2}{gV} = \ln\left(\frac{E_{\mathrm{F}} + \hbar\omega_{\mathrm{D}} - \frac{E}{2}}{E_{\mathrm{F}} - \frac{E}{2}}\right), \tag{15.37}$$

or

$$\frac{E_{\mathrm{F}} + \hbar\omega_{\mathrm{D}} - \frac{E}{2}}{E_{\mathrm{F}} - \frac{E}{2}} = \mathrm{e}^{2/gV}.$$

For the case of weak coupling regime $\frac{2}{gV} \gg 1$ and $\mathrm{e}^{-2/gV} \ll 1$. This gives

$$E \simeq 2E_{\mathrm{F}} - 2\hbar\omega_{\mathrm{D}}\mathrm{e}^{-2/gV}. \tag{15.38}$$

The quantity $2\hbar\omega_{\mathrm{D}}\mathrm{e}^{-2/gV}$ is the *binding energy* of the Cooper pair. Notice that

1. One can get a bound state no matter how weakly attractive V is. The free electron gas is unstable with respect to the paired bound state.
2. This variational result, which predicts the binding energy proportional to $\mathrm{e}^{-2/gV}$, could not be obtained in perturbation theory.
3. The material with higher value of V would likely show higher T_{c}.

The BCS theory uses the idea of pairing to account of the most important correlations.

15.4 The BCS Ground State

Let us write the model Hamiltonian, (15.21) by

$$H = H_0 + H_1, \tag{15.39}$$

where

$$H_0 = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} \left(c^{\dagger}_{\mathbf{k}\uparrow} c_{\mathbf{k}\uparrow} + c^{\dagger}_{-\mathbf{k}\downarrow} c_{-\mathbf{k}\downarrow}\right) \tag{15.40}$$

and

$$H_1 = -V \sum_{\mathbf{k}\mathbf{k}'}{}' c^{\dagger}_{\mathbf{k}'\uparrow} c^{\dagger}_{-\mathbf{k}'\downarrow} c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow}. \tag{15.41}$$

Note that we have included in the interaction only the interaction of $\mathbf{k}\uparrow$ with $-\mathbf{k}\downarrow$. In our discussion of the totally noninteracting electron gas, we found it convenient to use a description in terms of quasielectrons and quasiholes, where a quasielectron was an electron with $|\mathbf{k}| > k_{\mathrm{F}}$ and a quasihole was the absence of an electron with $|\mathbf{k}| < k_{\mathrm{F}}$. We could define

$$\begin{aligned} d^{\dagger}_{\mathbf{k}\sigma} &= c_{\mathbf{k}\sigma} \quad \text{for } \mathbf{k} < \mathbf{k}_{\mathrm{F}}, \\ d_{\mathbf{k}\sigma} &= c^{\dagger}_{\mathbf{k}\sigma} \quad \text{for } \mathbf{k} < \mathbf{k}_{\mathrm{F}}. \end{aligned} \tag{15.42}$$

Then $d^{\dagger}_{\mathbf{k}\sigma}$ creates a "hole" and $d_{\mathbf{k}\sigma}$ annihilates a "hole". If we measure all energies $\varepsilon_{\mathbf{k}}$ relative to the Fermi energy, then H_0 can be written as

$$\begin{aligned} H_0 &= \textstyle\sum_{\mathbf{k},\sigma} \varepsilon_{\mathbf{k}} c^{\dagger}_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} \\ &= E_0 + \textstyle\sum_{|\mathbf{k}|>k_{\mathrm{F}},\sigma} \tilde{\varepsilon}_{\mathbf{k}} n_{\mathbf{k}\sigma} + \sum_{|\mathbf{k}|<k_{\mathrm{F}}} \mid \tilde{\varepsilon}_{\mathbf{k}} \mid (1 - n_{\mathbf{k}\sigma}), \end{aligned} \tag{15.43}$$

where $n_{\mathbf{k}\sigma} = c^{\dagger}_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma}$, $\tilde{\varepsilon}_{\mathbf{k}} = \varepsilon(\mathbf{k}) - E_{\mathrm{F}}$, and $E_0 \left(= \sum_{|\mathbf{k}|<k_{\mathrm{F}}} \varepsilon_{\mathbf{k}}\right)$ is the energy of the filled Fermi sphere. Because $c^{\dagger}_{\mathbf{k}\sigma}$ adds a momentum $\mathbf{k}$ and spin σ to the system while $d^{\dagger}_{\mathbf{k}\sigma}$ subtract $\mathbf{k}$ and σ (or adds $-\mathbf{k}$ and $\bar{\sigma}$), it is useful to introduce

$$\alpha^{\dagger}_{\mathbf{k}\sigma} = u_{\mathbf{k}} c^{\dagger}_{\mathbf{k}\sigma} + v_{-\mathbf{k}} c_{-\mathbf{k}\bar{\sigma}} \tag{15.44}$$

and its Hermitian conjugate

$$\alpha_{\mathbf{k}\sigma} = u_{\mathbf{k}} c_{\mathbf{k}\sigma} + v_{-\mathbf{k}} c^{\dagger}_{-\mathbf{k}\bar{\sigma}}. \tag{15.45}$$

The operator $\alpha^{\dagger}_{\mathbf{k}\sigma}$ adds momentum $\mathbf{k}$ and spin σ to the system. If $u_{\mathbf{k}} = 1$, $v_{\mathbf{k}} = 0$ for $|\mathbf{k}| > k_{\mathrm{F}}$ and $u_{\mathbf{k}} = 0$, $v_{\mathbf{k}} = 1$ for $|\mathbf{k}| < k_{\mathrm{F}}$, we have simply the noninteracting electron gas described in terms of $\alpha_{\mathbf{k}\sigma}$ and $\alpha^{\dagger}_{\mathbf{k}\sigma}$. We must have $u_{\mathbf{k}} = u_{-\mathbf{k}}$ and $v_{\mathbf{k}} = -v_{-\mathbf{k}}$. Also $u^2_{\mathbf{k}} + v^2_{\mathbf{k}} = 1$ in order to satisfy the anticommutation relations $\left[\alpha_{\mathbf{k}}, \alpha^{\dagger}_{\mathbf{k}'}\right]_+ = \delta_{\mathbf{k}\mathbf{k}'}$.

From (15.44) and (15.45) we have that

$$c^{\dagger}_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} = u^2_{\mathbf{k}} \alpha^{\dagger}_{\mathbf{k}\sigma} \alpha_{\mathbf{k}\sigma} + v^2_{\mathbf{k}} \alpha_{-\mathbf{k}\bar{\sigma}} \alpha^{\dagger}_{-\mathbf{k}\bar{\sigma}} + u_{\mathbf{k}} v_{\mathbf{k}} \left(\alpha^{\dagger}_{\mathbf{k}\sigma} \alpha^{\dagger}_{-\mathbf{k}\bar{\sigma}} + \alpha_{-\mathbf{k}\bar{\sigma}} \alpha_{\mathbf{k}\sigma}\right).$$

Hence

$$\begin{aligned} \textstyle\sum_{\mathbf{k},\sigma} \varepsilon_{\mathbf{k}} c^{\dagger}_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} &= \textstyle\sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} \left[u^2_{\mathbf{k}} \alpha^{\dagger}_{\mathbf{k}\uparrow} \alpha_{\mathbf{k}\uparrow} + v^2_{\mathbf{k}} (1 - \alpha^{\dagger}_{-\mathbf{k}\downarrow} \alpha_{-\mathbf{k}\downarrow})\right] \\ &= \textstyle\sum_{|\mathbf{k}|<k_{\mathrm{F}}} \varepsilon_{\mathbf{k}} + \sum_{\mathbf{k},\sigma} \mid \varepsilon_{\mathbf{k}} \mid \alpha^{\dagger}_{\mathbf{k}\sigma} \alpha_{\mathbf{k}\sigma}. \end{aligned}$$

Therefore, in terms of the $\alpha_{\mathbf{k}}$ and $\alpha^{\dagger}_{\mathbf{k}}$, the noninteracting Hamiltonian is written as

$$H_0 = E_0 + \sum_{\mathbf{k},\sigma} \mid \tilde{\varepsilon}_{\mathbf{k}} \mid \alpha^{\dagger}_{\mathbf{k}\sigma} \alpha_{\mathbf{k}\sigma}. \tag{15.46}$$

The ground state of the noninteracting electron gas (filled Fermi sphere) can be constructed by annihilating quasiholes in all states with $|\,\mathbf{k}\,| < k_{\mathrm{F}}$ and is given by

$$|\,\mathrm{GS}\rangle = \prod_{\mathbf{k}\sigma} \alpha_{\mathbf{k}\sigma}\alpha_{-\mathbf{k}\bar{\sigma}}\,|\,\mathrm{VAC}\rangle, \tag{15.47}$$

where $|\,\mathrm{VAC}\rangle$ is the true vacuum state. Using (15.44) and (15.45) gives

$$|\,\mathrm{GS}\rangle = \prod_{\mathbf{k}\sigma} v_{\mathbf{k}}^2 c_{-\mathbf{k}\bar{\sigma}}^{\dagger} c_{\mathbf{k}\sigma}^{\dagger}\,|\,\mathrm{VAC}\rangle. \tag{15.48}$$

But, for the noninteracting system $v_{\mathbf{k}} = 1$ if $|\mathbf{k}| < k_{\mathrm{F}}$ and zero otherwise so that

$$|\,\mathrm{GS}\rangle = \prod_{|\mathbf{k}|<k_{\mathrm{F}},\sigma} c_{-\mathbf{k}\bar{\sigma}}^{\dagger} c_{\mathbf{k}\sigma}^{\dagger}\,|\,\mathrm{VAC}\rangle. \tag{15.49}$$

For the interacting system we will use a slight generalization of the notation used above.

15.4.1 Bogoliubov–Valatin Transformation

For the Hamiltonian given in (15.40) and (15.41), we make the transformation (called Bogoliubov–Valatin transformation) defined by

$$\begin{aligned} \alpha_{\mathbf{k}} &= u_{\mathbf{k}} c_{k\uparrow} - v_{\mathbf{k}} c_{-\mathbf{k}\downarrow}^{\dagger} \\ \beta_{\mathbf{k}}^{\dagger} &= u_{\mathbf{k}} c_{-\mathbf{k}\downarrow}^{\dagger} + v_{\mathbf{k}} c_{\mathbf{k}\uparrow}, \end{aligned} \tag{15.50}$$

with Hermitian conjugates $\alpha_{\mathbf{k}}^{\dagger} = u_{\mathbf{k}} c_{\mathbf{k}\uparrow}^{\dagger} + v_{\mathbf{k}} c_{-\mathbf{k}\bar{\downarrow}}$ and $\beta_{\mathbf{k}} = u_{\mathbf{k}} c_{-\mathbf{k}\downarrow} + v_{\mathbf{k}} c_{\mathbf{k}\uparrow}^{\dagger}$.

Note that the up spin $\uparrow$ is associated with the index $\mathbf{k}$ and the down spin $\downarrow$ is associated with the index $-\mathbf{k}$. The operators $\alpha_{\mathbf{k}}^{\dagger}$ and $\alpha_{\mathbf{k}}$ create or destroy an excited state of the system, which is a correlated electron–hole pair. We take $u_{\mathbf{k}} = u_{-\mathbf{k}}$ and $v_{\mathbf{k}} = -v_{-\mathbf{k}}$; in addition $u_{\mathbf{k}}^2 + v_{\mathbf{k}}^2$ must be equal unity in order to satisfy the anticommutation relations:

$$\left[\alpha_{\mathbf{k}}, \alpha_{\mathbf{k}'}^{\dagger}\right]_+ = \left[\beta_{\mathbf{k}}, \beta_{\mathbf{k}'}^{\dagger}\right]_+ = \delta_{\mathbf{k}\mathbf{k}'};\;\; \left[\alpha_{\mathbf{k}}, \alpha_{\mathbf{k}'}\right]_+ = \left[\beta_{\mathbf{k}}, \beta_{\mathbf{k}'}\right]_+ = 0.$$

We can solve (15.50) for $c_{\mathbf{k}\uparrow}$ and $c_{-\mathbf{k}\downarrow}$ and their Hermitian conjugates

$$\begin{aligned} c_{\mathbf{k}\uparrow} &= u_{\mathbf{k}}\alpha_{\mathbf{k}} + v_{\mathbf{k}}\beta_{\mathbf{k}}^{\dagger}; \qquad c_{\mathbf{k}\uparrow}^{\dagger} = u_{\mathbf{k}}\alpha_{\mathbf{k}}^{\dagger} + v_{\mathbf{k}}\beta_{\mathbf{k}} \\ c_{-\mathbf{k}\downarrow}^{\dagger} &= u_{\mathbf{k}}\beta_{\mathbf{k}}^{\dagger} - v_{\mathbf{k}}\alpha_{\mathbf{k}}; \qquad c_{-\mathbf{k}\downarrow} = u_{\mathbf{k}}\beta_{\mathbf{k}} - v_{\mathbf{k}}\alpha_{\mathbf{k}}^{\dagger}. \end{aligned} \tag{15.51}$$

Note that $u_{\mathbf{k}}^2$ is the probability that a pair of states with opposite $\mathbf{k}$ and σ is unoccupied and $v_{\mathbf{k}}^2$ is the probability that it is occupied. Substituting (15.51) in (15.40) gives

$$H_0 = \sum_{\mathbf{k}} \tilde{\varepsilon}_{\mathbf{k}} \Big[u_{\mathbf{k}}^2 \alpha_{\mathbf{k}}^\dagger \alpha_{\mathbf{k}} + v_{\mathbf{k}}^2 \beta_{\mathbf{k}} \beta_{\mathbf{k}}^\dagger + u_{\mathbf{k}} v_{\mathbf{k}} (\alpha_{\mathbf{k}}^\dagger \beta_{\mathbf{k}}^\dagger + \beta_{\mathbf{k}} \alpha_{\mathbf{k}}) + u_{\mathbf{k}}^2 \beta_{\mathbf{k}}^\dagger \beta_{\mathbf{k}} + v_{\mathbf{k}}^2 \alpha_{\mathbf{k}} \alpha_{\mathbf{k}}^\dagger - u_{\mathbf{k}} v_{\mathbf{k}} (\beta_{\mathbf{k}}^\dagger \alpha_{\mathbf{k}}^\dagger + \alpha_{\mathbf{k}} \beta_{\mathbf{k}}) \Big]. \tag{15.52}$$

Let us put the operators in normal form using $\beta_{\mathbf{k}} \beta_{\mathbf{k}}^\dagger = 1 - \beta_{\mathbf{k}}^\dagger \beta_{\mathbf{k}}$. This gives

$$H_0 = \sum_{\mathbf{k}} \tilde{\varepsilon}_{\mathbf{k}} \Big[2v_{\mathbf{k}}^2 + (u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2)(\alpha_{\mathbf{k}}^\dagger \alpha_{\mathbf{k}} + \beta_{\mathbf{k}}^\dagger \beta_{\mathbf{k}}) + 2u_{\mathbf{k}} v_{\mathbf{k}} (\alpha_{\mathbf{k}}^\dagger \beta_{\mathbf{k}}^\dagger + \beta_{\mathbf{k}} \alpha_{\mathbf{k}}) \Big]. \tag{15.53}$$

If we do the same thing for the interaction part H_1 given by (15.41)

$$H_1 = -V \sum_{\mathbf{k}\mathbf{k}'}{}' c_{\mathbf{k}'\uparrow}^\dagger c_{-\mathbf{k}'\downarrow}^\dagger c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow}, \tag{15.54}$$

we obtain

$$\begin{aligned} H_1 &= -V \textstyle\sum_{\mathbf{k}\mathbf{k}'} (u_{\mathbf{k}'} \alpha_{\mathbf{k}'}^\dagger + v_{\mathbf{k}'} \beta_{\mathbf{k}'}) (u_{\mathbf{k}'} \beta_{\mathbf{k}'}^\dagger - v_{\mathbf{k}'} \alpha_{\mathbf{k}'}) (u_{\mathbf{k}} \beta_{\mathbf{k}} - v_{\mathbf{k}} \alpha_{\mathbf{k}}^\dagger) (u_{\mathbf{k}} \alpha_{\mathbf{k}} + v_{\mathbf{k}} \beta_{\mathbf{k}}^\dagger) \\ &= -V \textstyle\sum_{\mathbf{k}\mathbf{k}'} \big[u_{\mathbf{k}'} v_{\mathbf{k}'} u_{\mathbf{k}} v_{\mathbf{k}} (1 - \alpha_{\mathbf{k}'}^\dagger \alpha_{\mathbf{k}'} - \beta_{\mathbf{k}'}^\dagger \beta_{\mathbf{k}'}) (1 - \alpha_{\mathbf{k}}^\dagger \alpha_{\mathbf{k}} - \beta_{\mathbf{k}}^\dagger \beta_{\mathbf{k}}) \\ &\quad + u_{\mathbf{k}'} v_{\mathbf{k}'} (1 - \alpha_{\mathbf{k}'}^\dagger \alpha_{\mathbf{k}'} - \beta_{\mathbf{k}'}^\dagger \beta_{\mathbf{k}'}) (u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2) (\alpha_{\mathbf{k}}^\dagger \beta_{\mathbf{k}}^\dagger + \beta_{\mathbf{k}} \alpha_{\mathbf{k}}) \\ &\quad + 4^{\text{th}} \text{ order off-diagonal terms} \big]. \end{aligned} \tag{15.55}$$

When this expression is multiplied out and then put in normal form (with all annihilation operators on the right of all creation operators), the result can be written

$$H = H(0) + H(2) + H(4). \tag{15.56}$$

Here, $H(2n)$ contains terms involving products of $2n$ Fermion operators ($\alpha, \alpha^\dagger, \beta$, and $\beta^\dagger$). It is simple to evaluate $H(0)$:

$$H(0) = 2 \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} v_{\mathbf{k}}^2 - V \sum_{\mathbf{k}\mathbf{k}'} u_{\mathbf{k}} v_{\mathbf{k}} u_{\mathbf{k}'} v_{\mathbf{k}'}. \tag{15.57}$$

The terms in $H(2)$ can be written

$$\begin{aligned} H(2) = &\textstyle\sum_{\mathbf{k}} \left[\varepsilon_{\mathbf{k}} (u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2) + V (\sum_{\mathbf{k}'} 2u_{\mathbf{k}'} v_{\mathbf{k}'}) u_{\mathbf{k}} v_{\mathbf{k}} \right] (\alpha_{\mathbf{k}}^\dagger \alpha_{\mathbf{k}} + \beta_{\mathbf{k}}^\dagger \beta_{\mathbf{k}}) \\ &+ \textstyle\sum_{\mathbf{k}} \left[2u_{\mathbf{k}} v_{\mathbf{k}} \tilde{\varepsilon}_{\mathbf{k}} - (u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2) V \sum_{\mathbf{k}'} u_{\mathbf{k}'} v_{\mathbf{k}'} \right] (\alpha_{\mathbf{k}}^\dagger \beta_{\mathbf{k}}^\dagger + \beta_{\mathbf{k}} \alpha_{\mathbf{k}}). \end{aligned} \tag{15.58}$$

We will neglect the terms in $H(4)$; they contain interactions between the elementary excitations. $H(0) + H(2)$ is not exactly in the form we desire because of the term proportional to $(\alpha_{\mathbf{k}}^\dagger \beta_{\mathbf{k}}^\dagger + \beta_{\mathbf{k}} \alpha_{\mathbf{k}})$. We are still at liberty to choose $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$; we do this by requiring the coefficient of $(\alpha_{\mathbf{k}}^\dagger \beta_{\mathbf{k}}^\dagger + \beta_{\mathbf{k}} \alpha_{\mathbf{k}})$ to vanish. This gives us

$$2u_{\mathbf{k}} v_{\mathbf{k}} \tilde{\varepsilon}_{\mathbf{k}} = (u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2) V \sum_{\mathbf{k}'} u_{\mathbf{k}'} v_{\mathbf{k}'}. \tag{15.59}$$

Let us define Δ by

$$\Delta = V \sum_{\mathbf{k}'} u_{\mathbf{k}'} v_{\mathbf{k}'} \tag{15.60}$$

and use it in (15.59) after squaring both sides. This gives

$$4u_{\mathbf{k}}^2 v_{\mathbf{k}}^2 \tilde{\varepsilon}_{\mathbf{k}}^2 = \Delta^2 (u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2)^2. \tag{15.61}$$

We can eliminate $u_{\mathbf{k}}^2$ since we already know that $u_{\mathbf{k}}^2 + v_{\mathbf{k}}^2 = 1$. Doing so gives the quadratic equation in $v_{\mathbf{k}}^2$.

$$v_{\mathbf{k}}^4 [\tilde{\varepsilon}_{\mathbf{k}}^2 + \Delta^2] - v_{\mathbf{k}}^2 [\tilde{\varepsilon}_{\mathbf{k}}^2 + \Delta^2] + \frac{\Delta^2}{4} = 0. \tag{15.62}$$

We choose the solution of the form

$$v_{\mathbf{k}}^2 = \frac{1}{2}(1 - \xi_{\mathbf{k}}). \tag{15.63}$$

Here, $\xi_{\mathbf{k}} = \frac{\tilde{\varepsilon}_{\mathbf{k}}}{\sqrt{\tilde{\varepsilon}_{\mathbf{k}}^2 + \Delta^2}}$. Furthermore since $u_{\mathbf{k}}^2 = 1 - v_{\mathbf{k}}^2$, we find that

$$u_{\mathbf{k}}^2 = \frac{1}{2}(1 + \xi_{\mathbf{k}}). \tag{15.64}$$

But (15.60), the definition of Δ, can now be written

$$\Delta = \frac{1}{2} V \sum_{\mathbf{k}} \sqrt{1 - \xi_{\mathbf{k}}^2}. \tag{15.65}$$

With a little algebra, (15.65) becomes

$$\Delta = \frac{V}{2} \sum_{\mathbf{k}} \frac{\Delta}{\sqrt{\tilde{\varepsilon}_{\mathbf{k}}^2 + \Delta^2}}. \tag{15.66}$$

Thus, the equation for the energy gap Δ is given by

$$1 = \frac{V}{2} \sum_{\mathbf{k}} \frac{1}{\sqrt{\tilde{\varepsilon}_{\mathbf{k}}^2 + \Delta^2}}. \tag{15.67}$$

Now, replace the sum by an integral taking for the density of states $\frac{1}{2} g(E_{\mathrm{F}})$ of the pair. The $\frac{1}{2}$ results from the fact that only $\mathbf{k} \uparrow$ and $-\mathbf{k} \downarrow$ are coupled by the interaction. Then, (15.67) becomes

$$1 = \frac{V}{2} \frac{g(E_{\mathrm{F}})}{2} \int_{-\hbar\omega_q}^{\hbar\omega_q} \frac{\mathrm{d}\varepsilon}{\sqrt{\varepsilon^2 + \Delta^2}}. \tag{15.68}$$

Using $\int \frac{\mathrm{d}x}{\sqrt{x^2 + \Delta^2}} = \ln(x + \sqrt{x^2 + \Delta^2}) = \sinh^{-1}(x/\Delta)$, the result for Δ becomes

$$\Delta = 2\hbar\omega_q \mathrm{e}^{-\frac{2}{g(E_{\mathrm{F}})V}}. \tag{15.69}$$

If the interaction V is zero, the one-particle states of the system would be occupied up to $|\mathbf{k}| = k_{\mathrm{F}}$, and Δ agrees with the binding energy of a Cooper pair.

15.4.2 Condensation Energy

The condensation energy $\Delta E(\equiv E_{\mathrm{S}}^0 - E_{\mathrm{N}}^0)$ defined by the difference between the ground state energy in the normal state and the state with finite Δ is approximately given by

$$\Delta E \simeq -g(E_{\mathrm{F}})\frac{\Delta}{2} \times \frac{\Delta}{2} = -g(E_{\mathrm{F}})\frac{\Delta^2}{4}. \tag{15.70}$$

The ground-state wave function Ψ_0 of the superconducting system is the eigenfunction of the diagonalized BCS Hamiltonian, so that

$$\alpha_{\mathbf{k}} \mid \Psi_0\rangle = \beta_{\mathbf{k}}^{\dagger} \mid \Psi_0\rangle = 0. \tag{15.71}$$

One can obtain Ψ_0 by writing

$$\mid \Psi_0\rangle = \prod_{\mathbf{k}} \alpha_{\mathbf{k}}\beta_{\mathbf{k}} \mid \mathrm{VAC}\rangle. \tag{15.72}$$

This gives the normalized wave function

$$\mid \Psi_0\rangle = \prod_{\mathbf{k}} (u_{\mathbf{k}} + v_{\mathbf{k}} c_{\mathbf{k}}^{\dagger} c_{-\mathbf{k}}^{\dagger}) \mid \mathrm{VAC}\rangle, \tag{15.73}$$

which is the BCS variational wave function normalized so that $\langle \Psi_0 | \Psi_0 \rangle = 1$.

15.5 Excited States

From (15.58) we can see that

$$H(2) = \sum_{\mathbf{k}} E_{\mathbf{k}} (\alpha_{\mathbf{k}}^{\dagger}\alpha_{\mathbf{k}} + \beta_{\mathbf{k}}^{\dagger}\beta_{\mathbf{k}}), \tag{15.74}$$

where

$$E_{\mathbf{k}} = \tilde{\varepsilon}_{\mathbf{k}}(u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2) + 2\Delta u_{\mathbf{k}} v_{\mathbf{k}}. \tag{15.75}$$

Knowing that $u_{\mathbf{k}} = \frac{1}{\sqrt{2}}\sqrt{1 + \xi_{\mathbf{k}}}$ and $v_{\mathbf{k}} = \frac{1}{\sqrt{2}}\sqrt{1 - \xi_{\mathbf{k}}}$ allows us to obtain the energy of an individual quasiparticle

$$E_{\mathbf{k}} = \sqrt{\tilde{\varepsilon}_{\mathbf{k}}^2 + \Delta^2}, \tag{15.76}$$

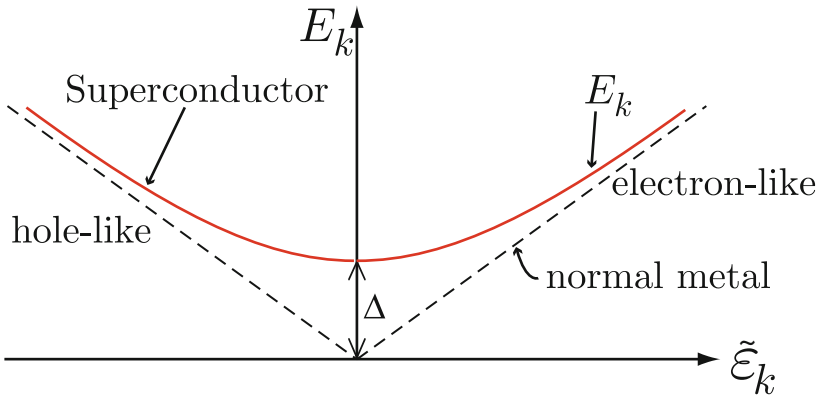


Fig. 15.11. Elementary excitations in a normal metal and in a superconductor

where $\tilde{\varepsilon}_{\mathbf{k}} = \frac{\hbar^2\mathbf{k}^2}{2m} - E_{\mathrm{F}}$, i.e., the energy is measured relative to the Fermi energy E_{F}. Thus, there is a gap Δ for the creation of elementary excitations $\alpha^\dagger_{\mathbf{k}} \mid \Psi_0\rangle$ or $\beta^\dagger_{\mathbf{k}} \mid \Psi_0\rangle$. The state $\alpha^\dagger_{\mathbf{k}} \mid \Psi_0\rangle$ is a quasiparticle state of wave vector $\mathbf{k}$, involving a superposition of an electron of wave vector $\mathbf{k}$ and a hole of wave vector $-\mathbf{k}$. In Fig. 15.11 quasiparticle energy spectra for a normal metal and for a superconductor are illustrated. The excitation spectrum has a gap Δ, which is known as the *gap parameter*. We notice that, since $\alpha^\dagger_{\mathbf{k}}$ and $\beta^\dagger_{\mathbf{k}}$ are linear combinations of single Fermion operators and always appear in pairs in the interaction Hamiltonian. Therefore, quasiparticles can be excited in pairs with the minimum excitation energy of 2Δ. The experimental gap should be 2Δ in experiments on absorption of electromagnetic radiation.

The density of quasiparticle states in the superconductor can be obtained using the quasiparticle dispersion relation $E_{\mathbf{k}}$.

$$g_{\mathrm{S}}(E) = \frac{1}{\Omega}\frac{\mathrm{d}N}{\mathrm{d}E} = \frac{2}{(2\pi)^3}\frac{\mathrm{d}}{\mathrm{d}k}\left(\frac{4\pi k^3}{3}\right)\frac{\mathrm{d}k}{\mathrm{d}E} = \frac{k^2}{\pi^2}\frac{1}{\mathrm{d}E/\mathrm{d}k}, \tag{15.77}$$

where $\frac{\mathrm{d}E}{\mathrm{d}k}$ is written, from (15.76), by

$$\frac{\mathrm{d}E}{\mathrm{d}k} = \frac{\mathrm{d}E}{\mathrm{d}\tilde{\varepsilon}}\frac{\mathrm{d}\tilde{\varepsilon}}{\mathrm{d}k} = \frac{\sqrt{E^2-\Delta^2}}{E}\frac{\hbar^2 k}{m}. \tag{15.78}$$

Substituting (15.78) in (15.79) gives

$$g_{\mathrm{S}}(E) = \frac{mk_{\mathrm{F}}}{\pi^2\hbar^2}\frac{E}{\sqrt{E^2-\Delta^2}} = g_{\mathrm{N}}(E_{\mathrm{F}})\frac{E}{\sqrt{E^2-\Delta^2}}, \tag{15.79}$$

where $g_{\mathrm{N}}(E) = \frac{mk}{\pi^2\hbar^2}$ is the density of states of the normal metal. Since we consider the quasiparticle energies close to the Fermi surface, we replaced $g_{\mathrm{N}}(E)$ by its value at the Fermi energy E_{F}. Notice that the density of quasiparticles

in the superconducting states shows a singularity at the energies $E = \pm\Delta$ measured with respect to the Fermi energy.

Essentially, all the other properties of a BCS superconductor can be evaluated knowing that

1. The ground-state energy given by $H(0)$, (15.57), is lower than the normal state energy ($E_{\rm N}^0 = \sum_{|\mathbf{k}|<k_{\rm F}} \tilde{\varepsilon}_{\mathbf{k}}$) by $-\frac{\Delta^2}{4} g(E_{\rm F})$, (15.70).
2. The energy of elementary excitations is given by $E_{\mathbf{k}} = \sqrt{\tilde{\varepsilon}_{\mathbf{k}}^2 + \Delta^2}$, (15.76).
3. The Fermi distribution function $n_{\mathbf{k}}$ is given by

$$f(E_{\mathbf{k}}) = n_{\mathbf{k}} = \frac{1}{\mathrm{e}^{E_{\mathbf{k}}/\Theta} + 1}.$$

Here, of course, $\tilde{\varepsilon}_{\mathbf{k}}$ appearing in $E_{\mathbf{k}}$ is measured relative to $E_{\rm F}$.

4. The BCS wave function is given by (15.73)

$$\mid \Psi_0\rangle = \prod_{\mathbf{k}} (u_{\mathbf{k}} + v_{\mathbf{k}} c_{\mathbf{k}}^\dagger c_{-\mathbf{k}}^\dagger) \mid \mathrm{VAC}\rangle.$$

One final example shows how to calculate the energy gap Δ as a function of temperature. We note that states $\mathbf{k}\uparrow$ and $-\mathbf{k}\downarrow$ are occupied statistically at finite temperatures. The Δ given in (15.69) was obtained under the assumption that $n_{\mathbf{k}} = 0$ at $T = 0$. But, at finite temperatures the Fermi distribution function should be understood as the occupation probability, and we expect $n_{\mathbf{k}} \neq 0$ and $\Delta = \Delta(T)$. To evaluate $\Delta(T)$ we need to extend (15.59) by writing

$$2u_{\mathbf{k}} v_{\mathbf{k}} \tilde{\varepsilon}_{\mathbf{k}} = (u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2) V \sum_{\mathbf{k}'} u_{\mathbf{k}'} v_{\mathbf{k}'} \left(1 - 2f(E_{\mathbf{k}'})\right). \tag{15.80}$$

This comes from keeping a term $-(\alpha_{\mathbf{k}'}^\dagger \alpha_{\mathbf{k}'} + \beta_{\mathbf{k}'}^\dagger \beta_{\mathbf{k}'})$ averaged at $T \neq 0$ in

$$\langle 1 - (\alpha_{\mathbf{k}'}^\dagger \alpha_{\mathbf{k}'} + \beta_{\mathbf{k}'}^\dagger \beta_{\mathbf{k}'})\rangle = 1 - n_{\mathbf{k}'} - n_{\mathbf{k}'} = 1 - 2f(E_{\mathbf{k}'}),$$

instead of just unity as was done in writing (15.58). Now, we define

$$\Delta(T) = V \sum_{\mathbf{k}} u_{\mathbf{k}} v_{\mathbf{k}} \left[1 - 2f(E_{\mathbf{k}})\right]. \tag{15.81}$$

Substituting (15.64) and (15.63) for $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ gives $1 = \frac{V}{2} \sum_{\mathbf{k}} \frac{1 - 2f(E_{\mathbf{k}})}{\sqrt{\tilde{\varepsilon}_{\mathbf{k}}^2 + \Delta^2(T)}}$, which reduces to

$$1 = \frac{V}{2} \frac{g(E_{\rm F})}{2} \int_{-\hbar\omega_{\rm D}}^{\hbar\omega_{\rm D}} \frac{\mathrm{d}\varepsilon}{\sqrt{\varepsilon^2 + \Delta^2(T)}} \left[1 - 2f(\sqrt{\varepsilon^2 + \Delta^2(T)})\right]. \tag{15.82}$$

At $T = 0$ this is the $T = 0$ gap equation, (15.68). As T increases from $T = 0$, $\Delta(T)$ would decrease from Δ_0, the zero temperature value. $\Delta(T)$ vanishes for $T \geq T_{\rm c}$, where $\Delta = 0$ is the only stable solution. The superconductivity disappears above $T_{\rm c}$. Now, (15.82) can be written

$$\frac{4}{g(E_\mathrm{F})V} = \int_{-\hbar\omega_\mathrm{D}}^{\hbar\omega_\mathrm{D}} \frac{\mathrm{d}\varepsilon}{\sqrt{\varepsilon^2+\Delta^2(T)}} - 2\int_{-\hbar\omega_\mathrm{D}}^{\hbar\omega_\mathrm{D}} \frac{\mathrm{d}\varepsilon}{\sqrt{\varepsilon^2+\Delta^2(T)}} f(\sqrt{\varepsilon^2+\Delta^2(T)}). \tag{15.83}$$

Since Δ becomes zero at $T = T_\mathrm{c}$, we can determine T_c by setting $\Delta = 0$ in (15.83). This gives

$$\frac{2}{g(E_\mathrm{F})V} = \int_0^{\hbar\omega_\mathrm{D}} \frac{\mathrm{d}\varepsilon}{\varepsilon} \tanh\frac{\varepsilon}{2\Theta_\mathrm{c}}, \tag{15.84}$$

where $\Theta_\mathrm{c} = k_\mathrm{B}T_\mathrm{c}$. Introducing the dimensionless variable $x = \frac{\varepsilon}{2\Theta_\mathrm{c}}$ we have

$$\begin{aligned} \tfrac{2}{g(E_\mathrm{F})V} &= \textstyle\int_0^{\hbar\omega_\mathrm{D}/2\Theta_\mathrm{c}} \frac{\mathrm{d}x}{x} \tanh x \\ &= \ln\tfrac{\hbar\omega_\mathrm{D}}{2\Theta_\mathrm{c}} \tanh\tfrac{\hbar\omega_\mathrm{D}}{2\Theta_\mathrm{c}} - \textstyle\int_0^{\hbar\omega_\mathrm{D}/2\Theta_\mathrm{c}} \ln x \,\mathrm{sech}^2 x \,\mathrm{d}x. \end{aligned} \tag{15.85}$$

Since $\eta \equiv \hbar\omega_\mathrm{D}/2\Theta_\mathrm{c} \gg 1$, in general for weak coupling superconductors, we may extend the upper limit of the integral to ∞ to have

$$\frac{2}{g(E_\mathrm{F})V} = \ln\frac{\hbar\omega_\mathrm{D}}{2\Theta_\mathrm{c}} \tanh\frac{\hbar\omega_\mathrm{D}}{2\Theta_\mathrm{c}} + \ln C, \tag{15.86}$$

where the constant $\ln C$ is given, in terms of Euler's constant γ:

$$\ln C = -\int_0^\infty \ln x \,\mathrm{sech}^2 x \,\mathrm{d}x = \gamma + \ln\frac{4}{\pi} \approx 0.81876.$$

Then, one can write

$$\Theta_\mathrm{c} \simeq 1.13\hbar\omega_\mathrm{D}\mathrm{e}^{-\frac{2}{gV}} \approx 0.57\Delta(0), \tag{15.87}$$

where ω is replaced by the Debye frequency ω_D. The Debye temperature Θ_D is much larger than the superconducting transition temperature Θ_c. Figure 15.12 sketches the $\Delta(T)$ obtained by numerical integration of (15.82).

15.6 Type I and Type II Superconductors

Correlations in superconductors involve electrons in a very limited range of values in momentum space. The range δp about the Fermi momentum p_F must be restricted to

$$\frac{p_\mathrm{F}^2}{2m} - \Delta \le \frac{(p_\mathrm{F}+\delta p)^2}{2m} \le \frac{p_\mathrm{F}^2}{2m} + \Delta. \tag{15.88}$$

This gives $|\,\delta p\,| \le \frac{\Delta}{v_\mathrm{F}}$. The spread of momentum δp leads to a *coherence length* in coordinate space $\xi_0 = \frac{\hbar}{\delta p} \sim \frac{\hbar v_\mathrm{F}}{\Delta}$. ξ_0 indicates the spatial range of the pair wave function. We distinguish *type I* and *type II* superconductors by whether

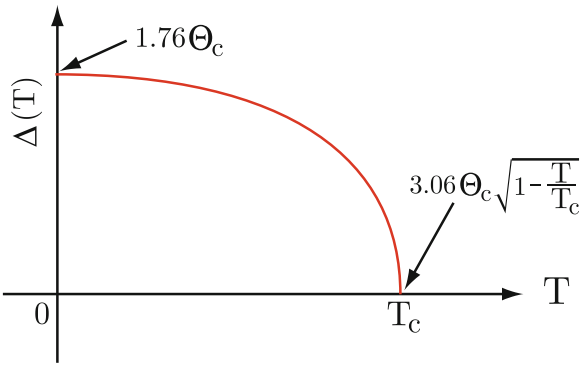


Fig. 15.12. Temperature dependence of the superconducting energy gap parameter $\Delta(T)$ in the weak coupling limit

the ratio of $\frac{\hbar v_{\rm F}}{\pi\Delta}(\simeq \xi_0)$ to $\Lambda_{\rm L} = \sqrt{\frac{mc^2}{4\pi n_{\rm S} e^2}}$, the London penetration depth, is large or small compared to unity. For example, we have $v_{\rm F} \simeq 10^8\,{\rm cm/s}$ and $T_{\rm c} \approx 0.57\Delta_0 \simeq 1.2{\rm K}$ for Al, and thus $\xi_0 \simeq 3.4 \times 10^3{\rm nm}$ and $\Lambda_{\rm L} \simeq 30{\rm nm}$ resulting $\frac{\xi_0}{\Lambda_{\rm L}} \sim 100$. But, for the case of Nb_3Sn we have $v_{\rm F} \simeq 10^6\,{\rm cm/s}$ and $T_{\rm c} \approx 0.57\Delta_0 \simeq 20{\rm K}$, and thus $\xi_0 \simeq 2\,{\rm nm}$ and $\Lambda_{\rm L} \simeq 200\,{\rm nm}$ resulting $\frac{\xi_0}{\Lambda_{\rm L}} \sim 10^{-2}$.

The London equation, (15.4), is written as

$$\nabla \times \mathbf{j} + \frac{n_{\rm S} e^2}{mc}\mathbf{B} = 0. \tag{15.89}$$

Using $\mathbf{B} = \nabla \times \mathbf{A}$ allows us to write

$$\mathbf{j}(\mathbf{r}) = -\frac{n_{\rm S} e^2}{mc}\mathbf{A}(\mathbf{r}). \tag{15.90}$$

This local relation between $\mathbf{j}$ and $\mathbf{A}$ is valid only for type II materials where $\Lambda_{\rm L}$ is much larger than ξ_0 and $\mathbf{A}(\mathbf{r})$ varies slowly on the scale of ξ_0. For type I materials, Pippard suggested a nonlocal relation between $\mathbf{j}$ and $\mathbf{A}$. Pippard equation is written as

$$\mathbf{j}(\mathbf{r}) = C\int \frac{\mathbf{A}(\mathbf{r}')\cdot\mathbf{R}}{R^4}\mathbf{R}{\rm e}^{-|\mathbf{R}|/\xi_0}{\rm d}^3 r', \tag{15.91}$$

where $\mathbf{R} = \mathbf{r} - \mathbf{r}'$. C is determined by requiring that slowly varying $\mathbf{A}(\mathbf{r})$ yields the London equation. Then $\mathbf{A}(\mathbf{r})$ comes outside the integral (and taking $\mathbf{A} \parallel \hat{z}$) and Eq.(15.91) reduces to

$$j_z(\mathbf{r}) = CA(\mathbf{r})\int_0^\infty \int_{-1}^1 \frac{R\cos\theta}{R^4} R\cos\theta {\rm e}^{-R/\xi_0} 2\pi d(\cos\theta) R^2 {\rm d}R = C\frac{4\pi}{3}\xi_0 A(\mathbf{r}). \tag{15.92}$$

We note that, by comparing (15.90) and (15.92), $C = -\frac{n_S e^2}{mc}\frac{3}{4\pi\xi_0}$, and picking $\xi_0 = \frac{\hbar v_F}{\pi\Delta_0}$ (at $T = 0$) gives excellent agreement with the microscopic theory. For the case $\Lambda_{\text{eff}} \ll \xi_0$, the vector potential $\mathbf{A}(\mathbf{r})$ is finite only in a surface layer and we can write

$$\mathbf{j}(\mathbf{r}) = -\frac{n_S e^2}{mc}\frac{\Lambda_{\text{eff}}}{\xi_0}\mathbf{A}(\mathbf{r}) \tag{15.93}$$

leading us to $\Lambda_{\text{eff}} \approx \Lambda_L(\frac{\xi_0}{\Lambda_L})^{1/3}$.

Flux Penetration

When a disc shaped type I superconductor is aligned perpendicular to an applied magnetic field $\mathbf{H}_0$, the magnetic field at the boundary of the sample (where the applied field is partially excluded) would become much greater than the magnitude of $\mathbf{H}_0$ (see Fig. 15.13). Then the sample starts to loose superconductivity at an applied field much below the critical field H_c forming a large number of normal and superconducting regions side by side, and the magnetic field energy gain is reduced significantly. The specimen is known to be in the *intermediate state*. It is a mixture of normal and superconducting regions due to geometric factors. The intermediate state has a domain structure that depends on competition between (1) superconducting condensation energy $\frac{1}{4}g(E_f)\Delta^2$, (2) magnetic field energy $\frac{H^2}{8\pi}$, (3) surface energy of N-S boundary (positive for type I material). In type II materials the surface energy turns out to be negative, and flux penetrates in single vortices each carrying one flux quantum (see Fig. 15.14). In alloys, if impurity scattering

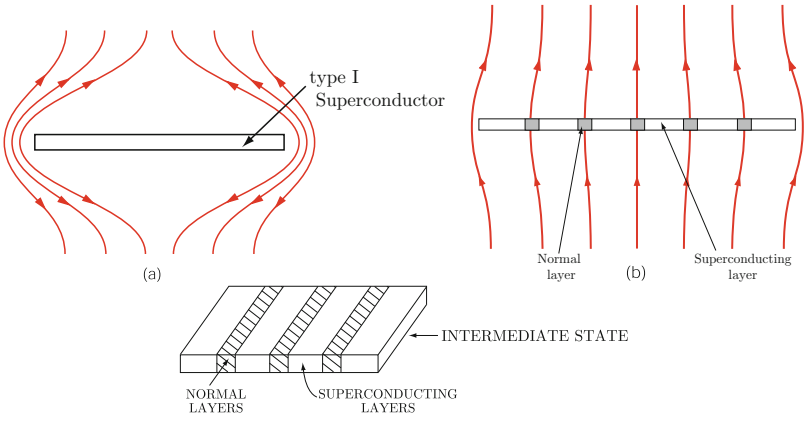


Fig. 15.13. Schematics of the intermediate state in a planar type I superconductor. It occurs when a planar sample is held perpendicular to an applied magnetic field as indicated in (**a**). Domain structure of normal and superconducting regions is formed as sketched in (**b**) and (**c**) by the magnified magnetic field due to geometric factors

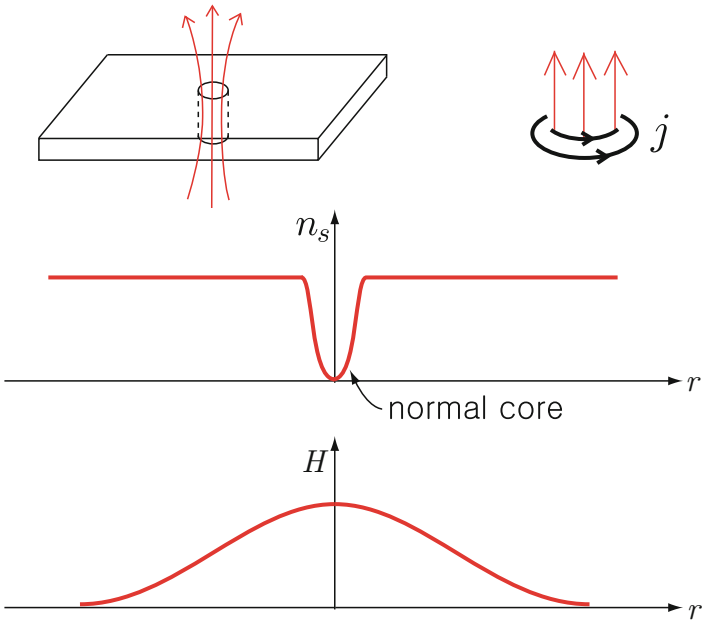


Fig. 15.14. Schematic representation of flux and field penetration in a type II superconductor

leads mean free path l, the Pippard relation replaces e^{-R/ξ_0} by $\mathrm{e}^{-R(\frac{1}{\xi_0}+\frac{1}{l})}$. In the extreme dirty limit of $\xi_0 \gg l$, relation between $\mathbf{j}(\mathbf{r})$ and $\mathbf{A}(\mathbf{r})$ becomes

$$\mathbf{j}(\mathbf{r}) = -\frac{n_S e^2}{mc}\frac{l}{\xi_0}\mathbf{A}(\mathbf{r}),$$

and the corresponding penetration depth $\Lambda_{\mathrm{eff}} = \Lambda_{\mathrm{L}}\sqrt{\xi_0/l}$ is increased greatly.

Problems

15.1. Demonstrate that the electronic contribution to the heat capacity of common intrinsic semiconductors shows the exponential temperature behavior at low temperatures.

15.2. Let's consider the equation of motion of a super fluid electron, which is dissipationless, in an electric field $\mathbf{E}$ that is momentarily present in the superconductor. That is, $m\frac{\mathrm{d}v_\mathrm{S}}{\mathrm{d}t} = -eE$, where v_S is the mean velocity of the super fluid electron caused by the field $\mathbf{E}$. In order to explain the Meissner effect, London proposed the London equation written as $\nabla \times \mathbf{j} + \frac{n_\mathrm{S} e^2}{mc}\mathbf{B} = 0$. Show that $\nabla^2 \mathbf{j} = \frac{1}{\Lambda^2}\mathbf{j}$ where $\Lambda_\mathrm{L} = \sqrt{\frac{mc^2}{4\pi n_\mathrm{S} e^2}}$ is the so-called the London penetration depth.

15.3. Assume $c_{\mathbf{k}\sigma}$ and $c^\dagger_{\mathbf{k}'\sigma'}$ satisfy standard Fermion anticommutation relations. Show that $\left[\alpha_\mathbf{k}, \alpha^\dagger_{\mathbf{k}'}\right]_+$ and $\left[\beta_\mathbf{k}, \beta^\dagger_{\mathbf{k}'}\right]_+$ each equal $\delta_{\mathbf{k}\mathbf{k}'}$ for $\alpha_\mathbf{k}$ and $\beta'_\mathbf{k}$ defined by the Bogoliubov–Valatin transformation

$$\begin{aligned}\alpha_\mathbf{k} &= u_\mathbf{k} c_{k\uparrow} - v_\mathbf{k} c^\dagger_{-\mathbf{k}\downarrow} \\ \beta^\dagger_\mathbf{k} &= u_\mathbf{k} c^\dagger_{-\mathbf{k}\downarrow} + v_\mathbf{k} c_{\mathbf{k}\uparrow}.\end{aligned}$$

15.4. Let us consider the interaction Hamiltonian H_1 given by

$$H_1 = -V \sum_{\mathbf{k}\mathbf{k}'}{}' c^\dagger_{\mathbf{k}'\uparrow} c^\dagger_{-\mathbf{k}'\downarrow} c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow}.$$

Use the Bogoliubov–Valatin transformation to show that H_1 can be written as

$$\begin{aligned}H_1 &= -V \textstyle\sum_{\mathbf{k}\mathbf{k}'} (u_{\mathbf{k}'}\alpha^\dagger_{\mathbf{k}'} + v_{\mathbf{k}'}\beta_{\mathbf{k}'})(u_{\mathbf{k}'}\beta^\dagger_{\mathbf{k}'} - v_{\mathbf{k}'}\alpha_{\mathbf{k}'})(u_\mathbf{k}\beta_\mathbf{k} - v_\mathbf{k}\alpha^\dagger_\mathbf{k})(u_\mathbf{k}\alpha_\mathbf{k} + v_\mathbf{k}\beta^\dagger_\mathbf{k}) \\ &= -V \textstyle\sum_{\mathbf{k}\mathbf{k}'} \big[u_{\mathbf{k}'} v_{\mathbf{k}'} u_\mathbf{k} v_\mathbf{k} (1 - \alpha^\dagger_{\mathbf{k}'}\alpha_{\mathbf{k}'} - \beta^\dagger_{\mathbf{k}'}\beta_{\mathbf{k}'})(1 - \alpha^\dagger_\mathbf{k}\alpha_\mathbf{k} - \beta^\dagger_\mathbf{k}\beta_\mathbf{k}) \\ &\quad + u_{\mathbf{k}'} v_{\mathbf{k}'} (1 - \alpha^\dagger_{\mathbf{k}'}\alpha_{\mathbf{k}'} - \beta^\dagger_{\mathbf{k}'}\beta_{\mathbf{k}'})(u^2_\mathbf{k} - v^2_\mathbf{k})(\alpha^\dagger_\mathbf{k}\beta^\dagger_\mathbf{k} + \beta_\mathbf{k}\alpha_\mathbf{k}) \\ &\quad + 4^{\text{th}} \text{ order off-diagonal terms } \big].\end{aligned}$$

15.5. Consider the condition given by

$$2u_\mathbf{k} v_\mathbf{k} \tilde{\varepsilon}_\mathbf{k} = (u^2_\mathbf{k} - v^2_\mathbf{k}) V \sum_{\mathbf{k}'} u_{\mathbf{k}'} v_{\mathbf{k}'}.$$

(a) Determine $u_\mathbf{k}$ and $v_\mathbf{k}$ satisfying the condition given above. Note that $u^2_\mathbf{k} + v^2_\mathbf{k} = 1$.
(b) Obtain the expression Δ defined by $\Delta = V \sum_{\mathbf{k}'} u_{\mathbf{k}'} v_{\mathbf{k}'}$.

Summary

In this chapter, we first briefly review some phenomenological observations of superconductivity and discuss a phenomenological theory by London. Then we introduce ideas of electron–phonon interaction and Cooper pairing to discuss microscopic theory by Bardeen, Cooper, and Schrieffer. The BCS ground state and excited states are discussed through Bogoliubov–Valatin transformation, and condensation energy and thermodynamic behavior of the superconducting energy gap are analyzed. Finally type I and type II superconductors are compared in terms of coherence length and London penetration depth.

The *Meissner effect* indicates that any magnetic field that is present in a bulk superconductor when $T > T_c$ is expelled when T is lowered through the transition temperature T_c. In a *type I superconductor*, the magnetic induction B vanishes in the bulk of the superconductor for $H < H_c(T)$. In a *type II superconductor*, the magnetic field starts to penetrate the sample at an applied field H_{c1} lower than the H_c. Between H_{c2} and H_{c1} flux penetrates the superconductor giving a *mixed state* consisting of superconductor penetrated by threads of the material in its normal state or flux lines. The mixed state consists of vortices each carrying a single flux $\Phi = \frac{hc}{2e}$.

The *London equation* is written as $\nabla \times \mathbf{j} + \frac{n_S e^2}{mc}\mathbf{B} = 0$, which implies that, in stationary conditions, a superconductor cannot sustain a magnetic field in its interior, but only within a narrow surface layer: $\nabla^2 \mathbf{B} = \frac{4\pi n_S e^2}{mc^2}\mathbf{B} \equiv \frac{1}{\Lambda_L^2}\mathbf{B}$. Here, the quantity $\Lambda_L = \sqrt{\frac{mc^2}{4\pi n_S e^2}}$ is called the *London penetration depth*.

The Hamiltonian of the electrons in a metal is written as

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}\sigma} c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + \sum_{\mathbf{k}\sigma,\mathbf{k}'\sigma',\mathbf{q}} W(\mathbf{k},\mathbf{q}) c^\dagger_{\mathbf{k}+\mathbf{q}\sigma} c^\dagger_{\mathbf{k}'-\mathbf{q}\sigma'} c_{\mathbf{k}'\sigma'} c_{\mathbf{k}\sigma},$$

where $W_{\mathbf{kq}}$ is defined by $W_{\mathbf{kq}} = \frac{|M_{\mathbf{q}}|^2 \hbar\omega_{\mathbf{q}}}{[\varepsilon(\mathbf{k}+\mathbf{q})-\varepsilon(\mathbf{k})]^2 - (\hbar\omega_{\mathbf{q}})^2}$. Here, $M_{\mathbf{q}}$ is the electron–phonon matrix element.

A pair of electrons interacting in the presence of a Fermi sea of "spectator electrons" is described by $H = \sum_{\boldsymbol{\ell},\sigma} \varepsilon_{\boldsymbol{\ell}} c^\dagger_{\boldsymbol{\ell}\sigma} c_{\boldsymbol{\ell}\sigma} - \frac{1}{2} V \sum_{\boldsymbol{\ell}\boldsymbol{\ell}'\sigma} c^\dagger_{\boldsymbol{\ell}'\sigma} c^\dagger_{-\boldsymbol{\ell}'\bar{\sigma}} c_{-\boldsymbol{\ell}\bar{\sigma}} c_{\boldsymbol{\ell}\sigma}$, where $\varepsilon_{\boldsymbol{\ell}} = \frac{\hbar^2 \ell^2}{2m}$ and the strength of the interaction, V, is taken as a constant for a small region of $\mathbf{k}$-space close to the Fermi surface. A variational trial function $\Psi = \sum_{\mathbf{k}} a_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} c^\dagger_{-\mathbf{k}\bar{\sigma}} \mid G >$ gives us $\frac{1}{V} = \sum_{\mathbf{k}} \frac{1}{2\varepsilon_{\mathbf{k}} - E}$. Here, $\mid G\rangle$ is the Fermi sea of spectator electrons, $\mid G\rangle = \Pi_{|\mathbf{k}|}^{k<k_F} c^\dagger_{\mathbf{k}\sigma} c^\dagger_{-\mathbf{k}\bar{\sigma}} \mid 0\rangle$. Approximating the sum by an integral over the energy ε, we have $E \simeq 2E_F - 2\hbar\omega_D e^{-2/gV}$. The quantity $2\hbar\omega_D e^{-2/gV}$ is the *binding energy* of the Cooper pair.

In the BCS theory, H is rewritten as $H = H_0 + H_1$, where

$$H_0 = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} \left(c^\dagger_{\mathbf{k}\uparrow} c_{\mathbf{k}\uparrow} + c^\dagger_{-\mathbf{k}\downarrow} c_{-\mathbf{k}\downarrow} \right) \text{ and } H_1 = -V \sum_{\mathbf{k}\mathbf{k}'}{}' c^\dagger_{\mathbf{k}'\uparrow} c^\dagger_{-\mathbf{k}'\downarrow} c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow}.$$

Introducing $\alpha^\dagger_{\mathbf{k}\sigma} = u_{\mathbf{k}} c^\dagger_{\mathbf{k}\sigma} + v_{-\mathbf{k}} c_{-\mathbf{k}\bar{\sigma}}$ and $\alpha_{\mathbf{k}\sigma} = u_{\mathbf{k}} c_{\mathbf{k}\sigma} + v_{-\mathbf{k}} c^\dagger_{-\mathbf{k}\bar{\sigma}}$ the noninteracting Hamiltonian becomes $H_0 = E_0 + \sum_{\mathbf{k},\sigma} \mid \tilde{\varepsilon}_{\mathbf{k}} \mid \alpha^\dagger_{\mathbf{k}\sigma} \alpha_{\mathbf{k}\sigma}$. The ground state of the noninteracting electron gas (filled Fermi sphere) is given by $\mid \mathrm{GS}\rangle = \prod_{\mathbf{k}\sigma} \alpha_{\mathbf{k}\sigma} \alpha_{-\mathbf{k}\bar{\sigma}} \mid \mathrm{VAC}\rangle$, where $\mid \mathrm{VAC}\rangle$ is the true vacuum state.

The Bogoliubov–Valatin transformation defined by

$$\alpha_{\mathbf{k}} = u_{\mathbf{k}} c_{k\uparrow} - v_{\mathbf{k}} c^\dagger_{-\mathbf{k}\downarrow} \; ; \; \beta^\dagger_{\mathbf{k}} = u_{\mathbf{k}} c^\dagger_{-\mathbf{k}\downarrow} + v_{\mathbf{k}} c_{\mathbf{k}\uparrow}$$
$$\alpha^\dagger_{\mathbf{k}} = u_{\mathbf{k}} c^\dagger_{\mathbf{k}\uparrow} + v_{\mathbf{k}} c_{-\mathbf{k}\bar{\downarrow}} \; ; \; \beta_{\mathbf{k}} = u_{\mathbf{k}} c_{-\mathbf{k}\downarrow} + v_{\mathbf{k}} c^\dagger_{\mathbf{k}\uparrow}$$

gives

$$H = H(0) + H(2) + H(4),$$

where

$$H(0) = 2\sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} v^2_{\mathbf{k}} - V \sum_{\mathbf{k}\mathbf{k}'} u_{\mathbf{k}} v_{\mathbf{k}} u_{\mathbf{k}'} v_{\mathbf{k}'} \; ; \; H(2) = \sum_{\mathbf{k}} E_{\mathbf{k}} (\alpha^\dagger_{\mathbf{k}} \alpha_{\mathbf{k}} + \beta^\dagger_{\mathbf{k}} \beta_{\mathbf{k}}).$$

Here $E_{\mathbf{k}} = \tilde{\varepsilon}_{\mathbf{k}} (u^2_{\mathbf{k}} - v^2_{\mathbf{k}}) + 2\Delta u_{\mathbf{k}} v_{\mathbf{k}}$ and $H(4)$ contains interactions between the elementary excitations. The equation for the energy gap Δ is given by

$$1 = \frac{V}{2} \sum_{\mathbf{k}} \frac{1}{\sqrt{\tilde{\varepsilon}^2_{\mathbf{k}} + \Delta^2}} \text{ and } \Delta = 2\hbar\omega_q \mathrm{e}^{-\frac{2}{g(E_\mathrm{F})V}}.$$

The ground-state wave function Ψ_0 of the superconducting system is

$$\mid \Psi_0\rangle = \prod_{\mathbf{k}} (u_{\mathbf{k}} + v_{\mathbf{k}} c^\dagger_{\mathbf{k}} c^\dagger_{-\mathbf{k}}) \mid \mathrm{VAC}\rangle.$$

The energy of a quasiparticle is $E_{\mathbf{k}} = \sqrt{\tilde{\varepsilon}^2_{\mathbf{k}} + \Delta^2}$, where $\tilde{\varepsilon}_{\mathbf{k}} = \frac{\hbar^2 \mathbf{k}^2}{2m} - E_\mathrm{F}$. The density of quasiparticle states in the superconductor is given by

$$g_\mathrm{S}(E) = \frac{m k_\mathrm{F}}{\pi^2 \hbar^2} \frac{E}{\sqrt{E^2 - \Delta^2}}.$$

The *type I* and *type II* superconductors are distinguished by whether the ratio of the coherence length ξ_0 to the London penetration depth Λ_L is large or small compared to unity. The local relation $\mathbf{j}(\mathbf{r}) = -\frac{n_\mathrm{S} e^2}{mc} \mathbf{A}(\mathbf{r})$ is valid only for type II materials where $\Lambda_\mathrm{L} \gg \xi_0$ and $\mathbf{A}(\mathbf{r})$ varies slowly on the scale of ξ_0. For the case $\Lambda_\mathrm{eff} \ll \xi_0$, the vector potential $\mathbf{A}(\mathbf{r})$ is finite only in a surface layer and we have

$$\mathbf{j}(\mathbf{r}) = -\frac{n_\mathrm{S} e^2}{mc} \frac{\Lambda_\mathrm{eff}}{\xi_0} \mathbf{A}(\mathbf{r})$$

leading to $\Lambda_\mathrm{eff} \approx \Lambda_\mathrm{L} (\frac{\xi_0}{\Lambda_\mathrm{L}})^{1/3}$. The *intermediate state* is a mixture of normal and superconducting regions due to geometric factors and it has a domain structure. In the extreme dirty limit of $\xi_0 \gg l$, we have

$$\mathbf{j}(\mathbf{r}) = -\frac{n_\mathrm{S} e^2}{mc} \frac{l}{\xi_0} \mathbf{A}(\mathbf{r}),$$

and the effective penetration depth $\Lambda_\mathrm{eff} = \Lambda_\mathrm{L} \sqrt{\xi_0 / l}$ is increased greatly.

16

The Fractional Quantum Hall Effect: The Paradigm for Strongly Interacting Systems

16.1 Electrons Confined to a Two-Dimensional Surface in a Perpendicular Magnetic Field

The study of the electronic properties of quasi two-dimensional systems has been a very exciting area of condensed matter physics during the last quarter of the twentieth century. Among the most interesting discoveries in this area are the incompressible states showing integral and fractional quantum Hall effects. Incompressible quantum liquid states of the integral quantum Hall effect result from an energy gap in the single particle spectrum. The incompressibility of the fractional quantum Hall effect is completely the result of electron–electron interactions in a highly degenerate fractionally filled Landau level. Since the quantum Hall effect involves electrons moving on a two-dimensional surface in the presence of a perpendicular magnetic field, we begin with a description of this problem.

The application of a large dc magnetic field perpendicular to the two-dimensional layer results in some notable novel physics. The Hamiltonian describing the motion of a single electron (of mass μ) confined to the x–y plane, in the presence of a dc magnetic field $\boldsymbol{B} = B\hat{z}$, is simply

$$H = (2\mu)^{-1}\left[\boldsymbol{p} + \frac{e}{c}\boldsymbol{A}(\boldsymbol{r})\right]^2 . \tag{16.1}$$

The vector potential $\boldsymbol{A}(\boldsymbol{r})$ is given by $\boldsymbol{A}(\boldsymbol{r}) = \frac{1}{2}B(-y\hat{x} + x\hat{y})$ in a *symmetric* gauge. We use $\hat{x}$, $\hat{y}$, and $\hat{z}$ as unit vectors along the Cartesian axes. The Schrödinger equation $(H - E)\Psi(\boldsymbol{r}) = 0$ has eigenstates[1]

$$\Psi_{nm}(r, \phi) = e^{\mathrm{i}m\phi}u_{nm}(r), \tag{16.2}$$

$$E_{nm} = \tfrac{1}{2}\hbar\omega_c(2n + 1 + m + |m|), \tag{16.3}$$

[1] See, for example, L.D. Landau, E.M. Lifshitz, *Q*uantum Mechanics (Pergamon, Oxford, 1977), p. 458; S. Gasiorowicz *Q*uantum Mechanics (Wiley, New York, 1996), ch. 13.

where n and m are principal and angular momentum quantum numbers, respectively, and $\omega_c (= eB/\mu c)$ is the cyclotron angular frequency. The radial function $u(r)$ in (16.2) satisfies the differential equation

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\eta^2} + \eta^{-1}\frac{\mathrm{d}u}{\mathrm{d}\eta} - (m^2\eta^{-1} + \eta^2 - \epsilon)u = 0, \tag{16.4}$$

where η and ϵ are, respectively, defined by $\eta = \sqrt{eB/2\hbar c}r = \frac{r}{\sqrt{2}l_0}$ and $\epsilon = 4E/\hbar\omega_c - 2m$. Here, $l_0 = \sqrt{\hbar c/eB}$ is the magnetic length. The radial wavefunctions $u_{nm}(r)$ can be expressed in terms of associated Laguerre polynomials L_n^m as

$$u_{nm}(\eta) = \eta^{|m|}\mathrm{e}^{-\eta^2/2}L_n^{|m|}(\eta^2). \tag{16.5}$$

Here, $L_0^{|m|}(\eta^2)$ is independent of η and $L_1^{|m|}(\eta^2) \propto (|m| + 1 - \eta^2)$. The lowest energy level has $n = 0$ and $m = 0, -1, -2, \ldots$. The first excited level has $n = 1$ and $m = 0, -1, -2, \ldots$, or $n = 0$ and $m = 1$, etc. These highly degenerate levels are separated from neighboring levels by $\hbar\omega_c$. These quantized energy levels are called *Landau levels*; the lowest Landau level wavefunction can be written as

$$\Psi_{0m} = \mathcal{N}_m z^{|m|}\mathrm{e}^{-|z|^2/4{l_0}^2}, \tag{16.6}$$

where $\mathcal{N}_m$ is the normalization constant and z stands for $z(=x - \mathrm{i}y) = re^{-\mathrm{i}\phi}$. The maximum value of $|\Psi_{0m}(z)|^2$ occurs at $r_m \propto m^{1/2}$.

For a finite size sample of area $\mathcal{S} = \pi R^2$, the number of single particle states in the lowest Landau level is given by $N_\phi = B\mathcal{S}/\phi_0$, where $\phi_0 = hc/e$ is the quantum of magnetic flux. The filling factor ν of a given Landau level is defined by N/N_ϕ, so that ν^{-1} is simply equal to the number of flux quanta of the dc magnetic field per electron. For a sample of radius R, the number of allowed values of $|m|$ is simply N_ϕ. For the lowest Landau level, degeneracy of the level is N_ϕ because the allowed values of $|m|$ are given by $|m| = 0, 1, 2, \ldots, N_\phi - 1$.

16.2 Integral Quantum Hall Effect

The integral quantum Hall effect occurs when N electrons exactly fill an integral number of Landau levels resulting in an integral value of the filling factor ν. When ν is equal to an integer, there is an energy gap (equal to $\hbar\omega_c$) between the filled states and the empty states. This makes the noninteracting electron system incompressible, because an infinitesimal decrease in the area $\mathcal{A}$, which decreases N_ϕ, requires a finite energy $\hbar\omega_c$ to promote an electron across the energy gap into the first unoccupied Landau level. This incompressibility is responsible for the *integral quantum Hall effect*.[2] To understand the minima

[2] K. von Klitzing, G. Dorda, M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).

in the diagonal resistivity ρ_{xx} and the plateaus in the Hall resistivity ρ_{xy}, it is necessary to notice that each Landau level, broadened by collisions with defects and phonons, must contain both extended states and localized states. The extended states lie in the central portion of the broadened Landau level, and the localized states in the wings. As the chemical potential ζ sweeps through the Landau level (by varying either B or the particle number N), zeros of ρ_{xx} (at $T = 0$K) and flat plateaus of ρ_{xy} occur when ζ lies within the localized states.

A many particle wavefunction of N electrons at filling factor $\nu = 1$ can be constructed by antisymmetrizing the product function which places one electron in each of the N states with $0 \leq |m| \leq N_\phi - 1$. Here, the product function should be antisymmetric under exchange of any two electrons, and the many particle wavefunction is written, for $\nu = 1$, as

$$\Psi_1(z_1, \ldots, z_N) = \mathcal{A}\{u_0(z_1)u_1(z_2)\cdots u_{N-1}(z_N)\} \tag{16.7}$$

where $\mathcal{A}$ denotes the antisymmetrizing operator. Since $u_{|m|}(z) \propto z^{|m|}e^{-|z^2|/4l_0^2}$, as given by (16.6), (16.7) can be written out as follows:

$$\Psi_1(z_1, \ldots, z_N) \propto \begin{vmatrix} 1 & 1 & \ldots & 1 \\ z_1 & z_2 & \ldots & z_N \\ z_1^2 & z_2^2 & \ldots & z_N^2 \\ \vdots & \vdots & \ldots & \vdots \\ z_1^{N-1} & z_2^{N-1} & \ldots & z_N^{N-1} \end{vmatrix} e^{-\frac{1}{4l_0^2}\sum_{i=1,N}|z_i|^2}. \tag{16.8}$$

The determinant in (16.8) is the well-known Van der Monde determinant written, simply, as $\Pi_{N\geq i>j\geq 1}(z_i - z_j)$. This is easily demonstrated by subtracting column j from column i and noting $z_{ij} = z_i - z_j$ is a common factor. Since it is true for every $i \neq j$, the result is apparent. Then the N-particle wavefunction corresponding to a filled Landau level becomes

$$\Psi_1(z_1, \ldots, z_N) \propto \Pi_{N\geq i>j\geq 1} z_{ij} \mathrm{e}^{-\frac{1}{4l_0^2}\sum_{k=1,N}|z_k|^2}. \tag{16.9}$$

In (16.9), the highest power of z_j is $N-1$. This means that the allowed values of $|m|$ are equal to $0, 1, 2, \ldots, N-1$ or that the Landau level degeneracy is given by $N_\phi = N$ and hence $\nu = N/N_\phi = 1$. We could obtain (16.9) by the requirement of antisymmetry imposed on the product of single particle eigenfunctions.

16.3 Fractional Quantum Hall Effect

When the filling factor ν is smaller than unity, the standard approach of placing N particles in the lowest energy single particle states is not applicable, because more degenerate states than the number of particles are present

in the lowest Landau level. For example, for the case of $\nu = 1/3$, it is not apparent how to construct antisymmetric product function for N electrons in $3N$ states to describe *fractional quantum Hall states*. In this case, no gap occurs in the absence of electron–electron interaction, and it is not easy to understand why fractional quantum Hall states are incompressible. At very high values of the applied magnetic field, there is only one relevant energy scale in the problem, the Coulomb scale e^2/ℓ_0. In that case, standard many body perturbation theory is not applicable. Laughlin used remarkable physical insight to propose a ground state wavefunction, for filling factor $\nu = 1/n$,[3]

$$\Psi_{1/n}(1, 2, \cdots, N) = \prod_{i>j} z_{ij}^n \, \mathrm{e}^{-\sum_l |z_l|^2/4l_0^2}, \tag{16.10}$$

where n is an odd integer.

The Laughlin wavefunction has the properties that (1) it is antisymmetric under interchange of any pair of particles as long as n is odd, (2) particles stay farther apart and have lower Coulomb repulsion for $n > 1$, and (3) because the wavefunction contains terms with z_i^m for $0 \leq m \leq n(N-1)$, $N_\phi - 1$, the largest value of m in the Landau level, is equal to $n(N-1)$ giving $\nu = N/N_\phi \longrightarrow 1/n$ for large systems in agreement with experiment.[4]

16.4 Numerical Studies

Remarkable confirmation of Laughlin's hypothesis was obtained by exact diagonalization carried out for relatively small systems. Exact diagonalization of the interaction Hamiltonian within the Hilbert subspace of the lowest Landau level is a very good approximation at large values of B, where $\hbar\omega_c \gg e^2/l_0$. Although real experiments are performed on a two-dimensional plane, it is more convenient to use a spherical two dimensional surface for numerical diagonalization studies. Haldane introduced the idea of putting a small number of electrons on a spherical surface at the center of which is located a magnetic monopole. We consider the case that the N electrons are confined to a Haldane surface of radius R. At the center of the sphere, a magnetic monopole of strength $2Q\phi_0$, where $2Q$ is an integer, is located, as illustrated in Fig. 16.1. The radial magnetic field is written as

$$\boldsymbol{B} = \frac{2Q\phi_0}{4\pi R^2}\hat{R}, \tag{16.11}$$

where $\hat{R}$ is a unit vector in the radial direction. The single particle Hamiltonian can be expressed as

$$H_0 = \frac{1}{2mR^2}\left(\boldsymbol{l} - \hbar Q\hat{R}\right)^2. \tag{16.12}$$

[3] R.B.Laughlin, Phys. Rev. Lett. **50**, 1395 (1983).
[4] D.C. Tsui, H.L. Stormer, A.C. Gossard, Phys. Rev. Lett. **48**, 1559 (1982).

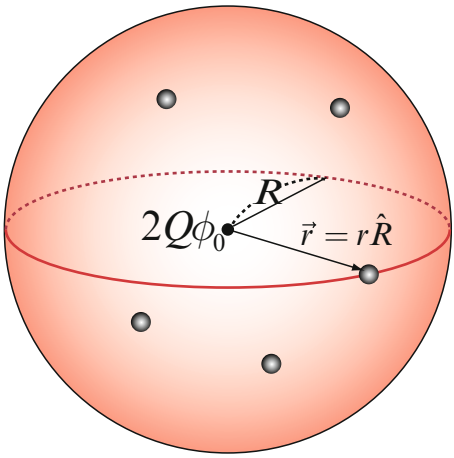


Fig. 16.1. Haldane sphere of radius R with magnetic monopoles of strength $2Q$ located at the center of the sphere

Here, $\boldsymbol{l}$ is the orbital angular momentum operator. The components of $\boldsymbol{l}$ satisfy the usual commutation rules $[l_\alpha, l_\beta] = \mathrm{i}\hbar\epsilon_{\alpha\beta\gamma}l_\gamma$, where the eigenvalues of l^2 and l_z are, respectively, $\hbar^2 l(l+1)$ and $\hbar m$.[5] The single particle eigenstates of (16.12) denoted by $|Q, l, m\rangle$ are called *monopole harmonics*. The states $|Q, l, m\rangle$ are eigenfunctions of l^2 and l_z as well as of H_0, the single particle Hamiltonian, with eigenvalues

$$\varepsilon(Q, l, m) = \frac{\hbar\omega_\mathrm{c}}{2Q}[l(l+1) - Q^2]. \tag{16.13}$$

In writing (16.13), we noted that $\boldsymbol{\Lambda}\cdot\hat{R} = \hat{R}\cdot\boldsymbol{\Lambda} = 0$ and, hence, $\boldsymbol{l}\cdot\hat{R} = \hat{R}\cdot\boldsymbol{l} = \hbar Q$. Because this energy must be positive, the allowed values of l are given by $l_n = Q + n$, where $n = 0, 1, 2, \ldots$. The lowest Landau level (or angular momentum shell) occurs for $l_0 = Q$ and has the energy $\varepsilon_0 = \hbar\omega_\mathrm{c}/2$, which is independent of m as long as m is a nonpositive integer. Therefore, the lowest Landau level has $(2Q+1)$-fold degeneracy. The nth excited Landau level occurs for $l_n = Q + n$ with energy

$$\varepsilon_n = \frac{\hbar\omega_\mathrm{c}}{2Q}[(Q+n)(Q+n+1) - Q^2]. \tag{16.14}$$

An N-particle eigenfunction of the lowest Landau level can be written, in general, as

$$|m_1, m_2, \cdots, m_N\rangle = c^\dagger_{m_N} \cdots c^\dagger_{m_2} c^\dagger_{m_1}|0\rangle. \tag{16.15}$$

[5] We note that, in the presence of the magnetic field, the total angular momentum is given by $\boldsymbol{\Lambda} = \boldsymbol{r} \times [-\mathrm{i}\hbar\nabla + e\boldsymbol{A}(\boldsymbol{r})]$ and that the eigenvalues of Λ^2 are not equal to $l(l+1)\hbar^2$. Here, $\boldsymbol{A}$ is the vector potential and $[\Lambda_i, \hat{R}_j] = \mathrm{i}\hbar\varepsilon_{ijk}\hat{R}_k$.

Here, $|m_i| \leq Q$ and $c^{\dagger}_{m_i}$ creates an electron in state $|l_0, m_i\rangle$. Since we are concentrating on a partially filled lowest Landau level we have only $2Q+1$ degenerate single particle states. The number of possible way of constructing N-electron antisymmetric states from $2Q+1$ single particle states or choosing N distinct values of m out of the $2Q+1$ allowed values is given by

$$G_{NQ} = \begin{pmatrix} 2Q+1 \\ N \end{pmatrix} = \frac{(2Q+1)!}{N!(2Q+1-N)!}. \tag{16.16}$$

Then, there are G_{NQ} N-electron states in the Hilbert subspace of the lowest Landau level. For the Laughlin $\nu = 1/m$ state, we have $2Q_{\nu=1/m} = m(N-1)$. For example, for the case of $2Q = 9$ and $N = 4$ ($\nu = 1/3$ state for 4 electrons in the lowest Landau level of degeneracy $2Q+1 = 10$), we have $l = 4.5$ and there are $G_{NQ} = 10!/[4!(10-4)! = 210$ of four-electron states in the Hilbert space of the lowest Landau level.

Table 16.1 lists the values of the electron angular momentum $\ell_{\rm e}$, $2Q+1$ (the Landau level degeneracy), G_{NQ} (the number of antisymmetric N-electron states), $L_{\rm MAX}$ (the largest possible angular momentum of the system), and

Table 16.1. The angular momentum $\ell_{\rm e}$ of an electron in the lowest Landau level; $2Q+1$ the Landau level degeneracy; G_{NQ} the dimension of N-electron Hilbert space; $L_{\rm Max}$ the maximum value of the total angular momentum L; the allowed L-values for a system consisting of N electrons on the surface of a Haldane sphere. The exponent of the allowed L-values indicates the number of times an L-multiplet appears and the number in parenthesis denotes the total number of L-multiplets

N	$\ell_{\rm e}$	$2Q+1$	G_{NQ}	$L_{\rm Max}$	Allowed L-values
3	3.0	7	35	6	$6 \oplus 4 \oplus 3 \oplus 2 \oplus 0$ (5)
4	4.5	10	210	12	$12 \oplus 10 \oplus 9 \oplus 8^2 \oplus 7 \oplus 6^3 \oplus 5 \oplus 4^3 \oplus 3 \oplus 2^2 \oplus 0^2$ (18)
5	6.0	13	1,287	20	$20 \oplus 18 \oplus 17 \oplus 16^2 \oplus 15^2 \oplus 14^3 \oplus 13^3 \oplus 12^5 \oplus 11^4 \oplus 10^6 \oplus 9^5 \oplus 8^7 \oplus 6^7 \oplus 5^5 \oplus 4^6 \oplus 3^3 \oplus 2^4 \oplus 1 \oplus 0^2$ (73)
6	7.5	16	8,008	30	$30 \oplus 28 \oplus \cdots$ (338)
7	9.0	19	50,382	42	$42 \oplus 40 \oplus \cdots$ (1,656)
8	10.5	22	319,770	56	$56 \oplus 54 \oplus \cdots$ (8,512)
9	12.0	25	2,042,975	72	$72 \oplus 70 \oplus \cdots$ (45,207)
10	13.5	28	13,123,110	90	$90 \oplus 88 \oplus \cdots$ (246,448)

the allowed values of L (the total angular momentum) with a superscript indicating how many times they appear. The number in parenthesis in the allowed L-value column is the total number of different L-multiplets that appear. For three electrons there are five such states, all with different L values. For four electrons there are 18 states; $L = 12$, 10, 9, 7, 5, and 3 each appearing once, $L = 8$, 2, and 0 each appearing twice, and $L = 6$ and 4 each three times. For $N = 10$ and $Q = 13.5$ ($\nu = 1/3$ state of 10 electrons) $G_{NQ} = 13,123,110$ and there are 246,448 distinct L multiplets with $0 \leq L \leq 90$.

The numerical problem is to diagonalize the interaction Hamiltonian

$$H_{\rm int} = \sum_{i<j} V(|\boldsymbol{r}_i - \boldsymbol{r}_j|) \tag{16.17}$$

in the G_{NQ} dimensional space. The problem is facilitated by first determining the eigenfunctions $|LM\alpha\rangle$ of the total angular momentum. Here, $\hat{L} = \sum_i \hat{l}_i$, $M = \sum_i m_i$, and α is an additional label that accounts for distinct multiplets with the same total angular momentum L (for example, for the five electron system the seven $L = 6$ states correspond to seven different values of α). The 210 four-electron states of four electrons give us 18×18 matrix that is block diagonal with two 3×3 blocks, three 2×2 blocks, and six 1×1 blocks. For small numbers of electrons these finite matrices can easily be diagonalized to obtain the many-body eigenvalues and eigenfunctions.[6]

In a plane geometry, the allowed values of m, the z-component of the single particle angular momentum, are 0, 1, $\cdots$, $N_\phi - 1$. $M = \sum_i m_i$ is the total $z-$component of angular momentum, where the sum is over all occupied states. It can be divided into the center-of-mass (CM) and relative (R) contributions $M_{\rm CM} + M_{\rm R}$. The connection between the planar and spherical geometries is as follows:

$$M = Nl + L_z, \quad M_{\rm R} = Nl - L, \quad M_{\rm CM} = L + L_z \tag{16.18}$$

The interactions depend only on $M_{\rm R}$, so $|M_{\rm R}, M_{\rm CM}\rangle$ acts just like $|L, L_z\rangle$. The absence of boundary conditions and the complete rotational symmetry make the spherical geometry attractive to theorists. Many experimentalists prefer using the $|M_{\rm R}, M_{\rm CM}\rangle$ states of the planar geometry. The calculations give the eigenenergies E as a function of the total angular momentum L.

[6] Because $H_{\rm int}$ is a scalar, the Wigner–Eckart theorem

$$\langle L'M'\alpha'|H_{\rm int}|LM\alpha\rangle = \delta_{LL'}\delta_{MM'}\langle L'\alpha'|H_{\rm int}|L\alpha\rangle$$

tells us that matrix elements of $H_{\rm int}$ are independent of M and vanish unless $L' = L$. This reduces the size of the matrix to be diagonalized enormously. For example, for $N = 10$ and $Q = 27/2$ ($\nu = 1/3$ state of ten electrons) $G_{NQ} = 13,123,110$ and there are 246,448 distinct L multiplets with $0 \leq L \leq 90$. However, the largest matrix diagonalized is only 7069 by 7069.

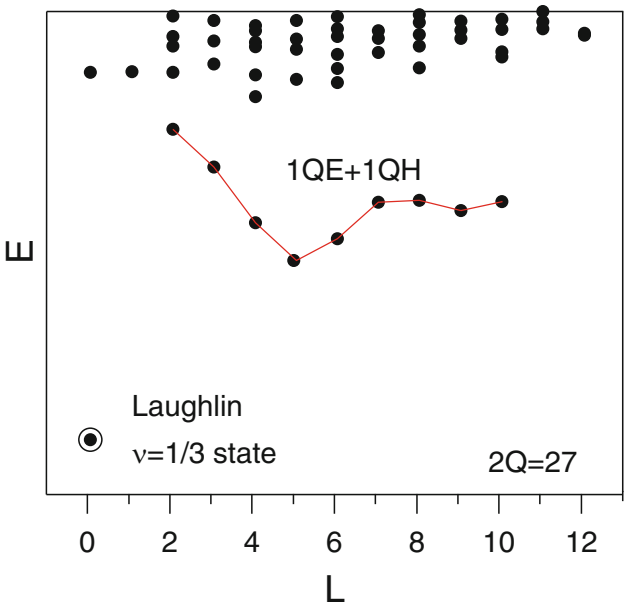


Fig. 16.2. The energy spectrum of 10 electrons in the lowest Landau level calculated on a Haldane sphere with $2Q = 27$. The *open circle* denotes the $L = 0$ ground state

The numerical results for the lowest Landau level always show one or more L multiplets forming a low energy band. As an example, the numerical results (E vs L) are shown in Figs. 16.2 and 16.3 for a system of 10 electrons with values of $2Q$ between 25 and 29. It is clear that the states fall into a well defined low energy sector and slightly less well defined excited sectors. The Laughlin $\nu = 1/3$ state occurs at $2Q = 3(N-1) = 27$ and the low energy sector consists of a singlet $L = 0$ state as illustrated in Fig. 16.2. States with larger values of Q contain one, two, or three quasiholes ($2Q = 28, 29, 30$), and states with smaller values of Q, such as $2Q = 25$ or 26, contain quasielectrons in the ground states. For $2Q = 26$ the system is one single-particle state shy of having the Laughlin $\nu = 1/3$ filling. In this case the low energy sector corresponds to having a single Laughlin quasielectron of angular momentum $L = 5$.

16.5 Statistics of Identical Particles in Two Dimension

Let us consider a system consisting of two particles each of charge $-e$ and mass μ, confined to a plane, in the presence of a perpendicular dc magnetic field $\boldsymbol{B} = (0, 0, B) = \nabla \times \boldsymbol{A}(\boldsymbol{r})$. Since $\boldsymbol{A}(\boldsymbol{r})$ is linear in the coordinate $\boldsymbol{r} = (x, y)$, (for example, $\boldsymbol{A}(\boldsymbol{r}) = \frac{1}{2}B(-y, x)$ in a symmetric gauge), the Hamiltonian separates into the center-of-mass and relative coordinate pieces $\boldsymbol{R} = \frac{1}{2}(\boldsymbol{r}_1 + \boldsymbol{r}_2)$

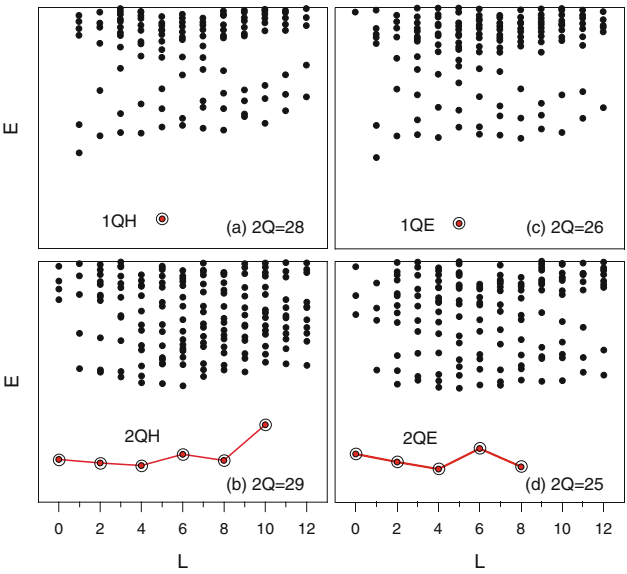


Fig. 16.3. The energy spectra of 10 electrons in the lowest Landau level calculated on a Haldane sphere with $2Q = 28, 29, 26, 25$. The *open circles* and *solid lines* mark the lowest energy bands with the fewest composite fermion quasiparticles of $n_{\rm QH} = 1$ for $2Q = 28$ in (**a**), $n_{\rm QH} = 2$ for $2Q = 29$ in (**b**), $n_{\rm QE} = 1$ for $2Q = 26$ in (**c**), and $n_{\rm QE} = 2$ for $2Q = 25$ in (**d**)

and $\boldsymbol{r} = \boldsymbol{r}_2 - \boldsymbol{r}_1$ being the center-of-mass and relative coordinates, respectively. The energy spectra for the center-of-mass and relative motion of the particles are identical to that of a single particle of mass μ and charge $-e$. We have seen that, as given by (16.6), for the lowest Landau level, the single particle wavefunction is

$$\Psi_{0m}(\boldsymbol{r}_1) = \mathcal{N}_m r_1^m \mathrm{e}^{-\mathrm{i}m\phi} \mathrm{e}^{-r_1^2/4{l_0}^2}.$$

For the relative motion ϕ is equal to $\phi_1 - \phi_2$, and the interchange of the pair, denoted by $P\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = \Psi(\boldsymbol{r}_2, \boldsymbol{r}_1)$, is accomplished by replacing ϕ by $\phi + \pi$.

For two identical particles initially at positions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ in a three dimensional space, the amplitude for the path that takes the system from the initial state $(\boldsymbol{r}_1, \boldsymbol{r}_2)$ to the same final state $(\boldsymbol{r}_1, \boldsymbol{r}_2)$ depends on the angle of rotation ϕ of the vector $\boldsymbol{r}_{12} (= \boldsymbol{r}_1 - \boldsymbol{r}_2)$. The end points represented by $\phi = \pi$ or 0 correspond to exchange or non-exchange processes, and the angle ϕ is only defined modulo 2π. The angle of rotation ϕ is not a well-defined quantity in three dimension, but the statistics can not be arbitrary. Under the exchange of two particles, the wavefunction picks up either a plus sign named bosonic statistics or a minus sign named fermionic statistics with no other possibilities. Since two consecutive interchanges must result in the original wavefunction, $\mathrm{e}^{im\pi}$ must be equal to either $+1$ (m is even; bosons) or -1 (m is odd; fermions).

In two dimensions the angle ϕ is perfectly well-defined for a given trajectory. It is possible to keep track of how many times the angle ϕ winds around the origin. Any two trajectories can not be deformed continuously into one another since any two particles can not go through each other. The space of particle trajectories falls into disconnected pieces that cannot be deformed into one another if $|\boldsymbol{r}_{ij}|$ is not allowed to vanish. Each piece has a definite winding number. Therefore, it is not enough to specify the initial and final configurations to characterize a given system completely. In constructing path integrals, the weighting of trajectories can depend on a new parameter θ (defined modulo 2π) through a factor $\mathrm{e}^{\mathrm{i}\theta\phi/\pi}$. For $\theta = 0$ or $\theta = \pi$ we have the conventional boson or fermion statistics. For the most general case, we have

$$P_{12}\Psi(1,2) = \mathrm{e}^{\mathrm{i}\theta}\Psi(1,2). \tag{16.19}$$

For arbitrary values of θ the particles are called *anyons* and satisfy a new form of quantum statistics.[7]

Let us consider a simple Lagrangian describing the relative motion of two interacting particles, the relative position and reduced mass of which are denoted by $\boldsymbol{r}[= (r,\phi)]$ and μ, respectively. A simple way to realize anyon statistics is to add a term $\hbar\beta\dot{\phi}$ called Chern–Simons term to the Lagrangian, where $\beta(\equiv \theta/\pi) = q\Phi/hc$ is the anyon parameter with $0 \le \beta \le 1$. While q and Φ are a fictitious charge and flux, θ is the numerical parameter of $0 \le \theta \le 1$. For example, if

$$\mathcal{L} = \frac{1}{2}\mu(\dot{r}^2 + r^2\dot{\phi}^2) - V(r) + \hbar\beta\dot{\phi} \tag{16.20}$$

the added (fictitious charge–flux) Chern–Simons term does not affect the classical equations of motions because q and Φ are time independent. However, the canonical angular momentum is given by $p_\phi(= \frac{\partial\mathcal{L}}{\partial\dot{\phi}}) = \mu r^2\dot{\phi} + \hbar\beta$. Because $\mathrm{e}^{2\pi\mathrm{i}p_\phi/\hbar}$ generates rotations of 2π, $\hbar^{-1}p_\phi$ must have integral eigenvalues ℓ. However, the gauge invariant kinetic angular momentum, given by $p_\phi - \hbar\beta$, can take on fractional values, which will result in fractional quantum statistics for the particles.

16.6 Chern–Simons Gauge Field

Let us consider a two-dimensional system of particles satisfying some particular statistics and described by a Hamiltonian

$$H = \frac{1}{2\mu}\sum_i \left[\boldsymbol{p}_i + \frac{e}{c}\boldsymbol{A}(\boldsymbol{r}_i)\right]^2 + \sum_{i>j} V(r_{ij}). \tag{16.21}$$

[7] A. Lerda, *Anyons*: Quantum Mechanics of Particles with Fractional Statistics, Lecture Notes in Physics (Springer, Berlin, 1992) and F. Wilczek, Fractional Statistics and Anyon Superconductivity (World Scientific, Singapore, 1990).

Then, we can change the statistics by attaching to each particle a fictitious charge q and flux tube carrying magnetic flux Φ. The fictitious vector potential $\boldsymbol{a}(\boldsymbol{r}_i)$ at the position of the ith particle caused by flux tubes, each carrying flux of Φ, on all the other particles at $\boldsymbol{r}_j(\neq \boldsymbol{r}_i)$ is written as

$$\boldsymbol{a}(\boldsymbol{r}_i) = \Phi \sum_{j \neq i} \frac{\hat{z} \times \boldsymbol{r}_{ij}}{r_{ij}^2}. \tag{16.22}$$

The *Chern–Simons gauge field* due to the gauge potential $\boldsymbol{a}(\boldsymbol{r}_i)$ becomes

$$\boldsymbol{b}(\boldsymbol{r}) = \Phi \sum_i \delta(\boldsymbol{r} - \boldsymbol{r}_i)\hat{z}, \tag{16.23}$$

where $\boldsymbol{r}_i$ is the position of the ith particle carrying gauge potential $\boldsymbol{a}(\boldsymbol{r}_i)$. Since no two electrons can occupy the same position and a given electron can never sense the δ-function type gauge field due to other electrons, $\boldsymbol{b}(\boldsymbol{r})$ has no effect on the classical equations of motion.

In a quantum mechanical system, we rewrite the vector potential $\boldsymbol{a}(\boldsymbol{r})$ as follows:

$$\boldsymbol{a}(\boldsymbol{r}_i) = \Phi \int \mathrm{d}^2 r_1 \frac{\hat{z} \times (\boldsymbol{r} - \boldsymbol{r}_1)}{|\boldsymbol{r} - \boldsymbol{r}_1|^2} \psi^\dagger(\boldsymbol{r}_1)\psi(\boldsymbol{r}_1). \tag{16.24}$$

Here, $\psi^\dagger(\boldsymbol{r}_1)\psi(\boldsymbol{r}_1)$ denotes the density operator $\rho(r_1)$ for the electron liquid and the gauge potential $\boldsymbol{a}(\boldsymbol{r})$ introduces a phase factor into the quantum mechanical wavefunction.

Chern–Simons transformation is a singular gauge transformation which transforms an electron creation operator $\psi_e^\dagger(\boldsymbol{r})$ into a composite particle creation operator $\psi^\dagger(\boldsymbol{r})$ as follows:

$$\psi^\dagger(\boldsymbol{r}) = \psi_e^\dagger(\boldsymbol{r}) \mathrm{e}^{\mathrm{i}\alpha \int \mathrm{d}^2 r' arg(\boldsymbol{r} - r')\psi_e^\dagger(\boldsymbol{r}')\psi_e(\boldsymbol{r}')}. \tag{16.25}$$

Here, $arg(\boldsymbol{r} - \boldsymbol{r}')$ denotes the angle that the vector $\boldsymbol{r} - \boldsymbol{r}'$ makes with the x-axis and α is a gauge parameter. Then, the kinetic energy operator K_e of an electron is transformed into

$$K_{\mathrm{CS}} = \frac{1}{2\mu} \int \mathrm{d}^2 r \psi^\dagger(\boldsymbol{r}) \left[-\mathrm{i}\hbar\nabla + \frac{e}{c}\boldsymbol{A}(\boldsymbol{r}) + \frac{e}{c}\boldsymbol{a}(\boldsymbol{r}) \right]^2 \psi(\boldsymbol{r}), \tag{16.26}$$

where $\boldsymbol{a}(\boldsymbol{r})$ is the total gauge potential formed at the position $\boldsymbol{r}$ due to the Chern–Simons flux attached to other particles.

$$\boldsymbol{a}(\boldsymbol{r}) = \alpha\phi_0 \int \mathrm{d}^2 r' \frac{\hat{z} \times (\boldsymbol{r} - \boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|^2} \psi^\dagger(\boldsymbol{r}')\psi(\boldsymbol{r}').$$

Hence, the Chern–Simons transformation corresponds to a transformation attaching to each particle a flux tube of fictitious magnetic flux $\Phi(=\alpha\phi_0)$

and a fictitious charge $-e$ so that the particle could couple to the flux tube attached to other particles.

The new Hamiltonian, through Chern–Simons gauge transformation, is obtained by simply replacing $\frac{e}{c}\boldsymbol{A}(\boldsymbol{r}_i)$ in (16.21) by $\frac{e}{c}\boldsymbol{A}(\boldsymbol{r}_i) + \frac{e}{c}\boldsymbol{a}(\boldsymbol{r}_i)$.

$$H_{\mathrm{CS}} = \frac{1}{2\mu}\int \mathrm{d}^2 r \psi^\dagger(\boldsymbol{r})\left[\boldsymbol{p} + \frac{e}{c}\boldsymbol{A}(\boldsymbol{r}) + \frac{e}{c}\boldsymbol{a}(\boldsymbol{r})\right]^2 \psi(\boldsymbol{r}) + \sum_{i>j} V(r_{ij}). \quad (16.27)$$

The *composite fermions* obtained in this way carry both electric charge and magnetic flux. The Chern–Simons transformation is a gauge transformation and hence the composite fermion energy spectrum is identical with the original electron spectrum. Since attached fluxes are localized on electrons and the magnetic field acting on each electron is unchanged, the classical Hamiltonian of the system is also unchanged. However, the quantum mechanical Hamiltonian includes additional terms describing an additional charge–flux interaction, which arises from the Aharanov–Bohm phase attained when one electron's path encircles the flux tube attached to another electron.

The net effect of the additional Chern–Simons term is to replace the statistics parameter θ describing the particle statistics in (16.19) with $\theta + \pi\Phi\frac{q}{hc}$. If $\Phi = p\frac{hc}{e}$ when p is an integer, then $\theta \to \theta + \pi pq/e$. For the case of $q = e$ and $p = 1$, $\theta = 0 \to \theta = \pi$ converting bosons to fermions and $\theta = \pi \to \theta = 2\pi$ converting fermions to bosons. For $p = 2$, the statistics would be unchanged by the Chern–Simons terms.

The Hamiltonian H_{CS} contains terms proportional to $\boldsymbol{a}^n(\boldsymbol{r})$ $(n = 0, 1, 2)$. The $\boldsymbol{a}^1(\boldsymbol{r})$ term gives rise to a standard two-body interaction. The $\boldsymbol{a}^2(\boldsymbol{r})$ term gives three-body interactions containing the operator

$$\Psi^\dagger(\boldsymbol{r})\Psi(\boldsymbol{r})\Psi^\dagger(\boldsymbol{r}_1)\Psi(\boldsymbol{r}_1)\Psi^\dagger(\boldsymbol{r}_2)\Psi(\boldsymbol{r}_2).$$

The three-body terms are complicated, and they are frequently neglected. The Chern–Simons Hamiltonian introduced via a gauge transformation is considerably more complicated than the original Hamiltonian given by (16.21). Simplification results only when the mean-field approximation is made. This is accomplished by replacing the operator $\rho(\boldsymbol{r})$ in the Chern–Simons vector potential (16.24) by its mean-field value n_S, the uniform equilibrium electron density. The resulting mean-field Hamiltonian is a sum of single particle Hamiltonians in which, instead of the external field B, an *effective* magnetic field $B^* = B + \alpha\phi_0 n_S$ appears.

16.7 Composite Fermion Picture

The difficulty in trying to understand the fractionally filled Landau level in two dimensional systems comes from the enormous degeneracy that is present in the noninteracting many body states. The lowest Landau level contains N_ϕ

states and $N_\phi = B\mathcal{S}/\phi_0$, the number of flux quanta threading the sample of area $\mathcal{S}$. Therefore, $N_\phi/N = \nu^{-1}$ is equal to the number of flux quanta per electron. Let us think of the $\nu = 1/3$ state as an example; it has three flux quanta per electron. If we attach to each electron a fictitious charge $q(= -e$, the electron charge) and a fictitious flux tube (carrying flux $\Phi = 2p\phi_0$ directed opposite to B, where p is an integer and ϕ_0 the flux quantum), the net effect is to give us the Hamiltonian described by Eqs.(16.21) and (16.22) and to leave the statistical parameter θ unchanged. The electrons are converted into composite fermions which interact through the gauge field term as well as through the Coulomb interaction.

Why does one want to make this transformation, which results in a much more complicated Hamiltonian? The answer is simple if the gauge field $\boldsymbol{a}(\boldsymbol{r}_i)$ is replaced by its mean value, which simply introduces an effective magnetic field $B^* = B + \langle b\rangle$. Here, $\langle b\rangle$ is the average magnetic field associated with the fictitious flux. In the mean field approach, the magnetic field due to attached flux tubes is evenly spread over the occupied area $\mathcal{S}$. The mean field composite fermions obtained in this way move in an effective magnetic field B^*. Since, for $\nu = 1/3$ state, B corresponds to three flux quanta per electron and $\langle b\rangle$ corresponds to two flux quanta per electron directed opposite to the original magnetic field B, we see that $B^* = \frac{1}{3}B$. The effective magnetic field B^* acting on the composite fermions gives a composite fermion Landau level containing $\frac{1}{3}N_\phi$ states, or exactly enough states to accommodate our N particles. Therefore, the $\nu = 1/3$ electron Landau level is converted, by the composite fermion transformation, to a $\nu^* = 1$ composite fermion Landau level. Now, the ground state is the antisymmetric product of single particle states containing N composite fermions in exactly N states. The properties of a filled (composite fermion) Landau level is well investigated in two dimension. The fluctuations about the mean field can be treated by standard many body perturbation theory. The vector potential associated with fluctuation beyond the mean field level is given by $\delta\boldsymbol{a}(\boldsymbol{r}) = \boldsymbol{a}(\boldsymbol{r}) - \langle\boldsymbol{a}(\boldsymbol{r})\rangle$. The perturbation to the mean field Hamiltonian contains both linear and quadratic terms in $\delta\boldsymbol{a}(\boldsymbol{r})$, resulting in both two body and three body interaction terms.

The idea of a composite fermion was introduced initially to represent an electron with an attached flux tube which carries an even number α $(= 2p)$ of flux quanta. In the mean field approximation the composite fermion filling factor ν^* is given by the number of flux quanta per electron of the dc field less the composite fermion flux per electron, i.e.

$$\nu^{*-1} = \nu^{-1} - \alpha. \tag{16.28}$$

We remember that ν^{-1} is equal to the number of flux quanta of the applied magnetic field B per electron, and α is the (even) number of Chern–Simons flux quanta (oriented oppositely to the applied magnetic field B) attached to each electron in the Chern–Simons transformation. Negative ν^* means the effective magnetic field B^* seen by the composite fermions is oriented opposite to the original magnetic field B. Equation (16.28) implies that when

$\nu^* = \pm 1, \pm 2, \ldots$ and a nondegenerate mean field composite fermion ground state occurs, then

$$\nu = \frac{\nu^*}{1 + \alpha\nu^*} \tag{16.29}$$

generates, for $\alpha = 2$, condensed states at $\nu = 1/3, 2/5, 3/7, \ldots$ and $\nu = 1, 2/3, 3/5, \ldots$. These are the most pronounced fractional quantum Hall states observed in experiment. The $\nu^* = 1$ states correspond to Laughlin $\nu = \frac{1}{1+\alpha}$ states. If ν^* is not an integer, the low lying states contain a number of quasiparticles ($N_{\mathrm{QP}} \leq N$) in the neighboring incompressible state with integral ν^*. The mean field Hamiltonian of noninteracting composite fermions is known to give a good description of the low lying states of interacting electrons in the lowest Landau level.

It is quite remarkable to note that the mean field picture predicts not only the *Jain sequence* of incompressible ground states, given by $\nu = \frac{\nu^*}{1+2p\nu^*}$ (with integer p), but also the correct band of low energy states for any value of the applied magnetic field. This is illustrated very nicely for the case of N electrons on a Haldane sphere. In the spherical geometry, one can introduce an effective monopole strength $2Q^*$ seen by one composite fermion. When the monopole strength seen by an electron has the value $2Q$, $2Q^*$ is given, since the α flux quanta attached to every other composite fermion must be subtracted from the original monopole strength $2Q$, by

$$2Q^* = 2Q - \alpha(N - 1). \tag{16.30}$$

This equation reflects the fact that a given composite fermion senses the vector potential produced by the Chern–Simons flux on all other particles, but not its own Chern–Simons flux.

Now, $|Q^*| = l_0^*$ plays the role of the angular momentum of the lowest composite fermion shell just as $Q = l_0$ was the angular momentum of the lowest electron shell. When $2Q$ is equal to an odd integer $(1+\alpha)$ times $(N-1)$, the composite fermion shell l_0^* is completely filled ($\nu^* = 1$), and an $L = 0$ incompressible Laughlin state at filling factor $\nu = (1 + \alpha)^{-1}$ results. When $2|Q^*| + 1$ is smaller than N, quasielectrons appear in the shell $l_{\mathrm{QE}} = l_0^* + 1$. Similarly, when $2|Q^*| + 1$ is larger than N, quasiholes appear in the shell $l_{\mathrm{QH}} = l_0^*$. The low-energy sector of the energy spectrum consists of the states with the minimum number of quasiparticle excitations required by the value of $2Q^*$ and N. The first excited band of states will contain one additional quasielectron – quasihole pair. The total angular momentum of these states in the lowest energy sector can be predicted by addition of the angular momenta (l_{QH} or l_{QE}) of the n_{QH} or n_{QE} quasiparticles treated as identical fermions. In Table 16.2 we demonstrate how these allowed L values are found for a ten electron system with $2Q$ in the range $29 \geq 2Q \geq 15$. By comparing with numerical results presented in Fig. 16.1, one can readily observe that the total angular momentum multiplets appearing in the lowest energy sector are correctly predicted by this simple mean field Chern–Simons picture.

Table 16.2. The effective CF monopole strength $2Q^*$, the number of CF quasiparticles (quasiholes $n_{\rm QH}$ and quasielectrons $n_{\rm QE}$), the quasiparticle angular momenta $l_{\rm QE}$ and $l_{\rm QH}$, the composite fermion and electron filling factors ν^* and ν, and the angular momenta L of the lowest lying band of multiplets for a ten electron system at $2Q$ between 29 and 15

$2Q$	29	28	27	26	25	24	23	22	21	15
$2Q^*$	11	10	9	8	7	6	5	4	3	-3
$n_{\rm QH}$	2	1	0	0	0	0	0	0	0	0
$n_{\rm QE}$	0	0	0	1	2	3	4	5	6	6
$l_{\rm QH}$	5.5	5	4.5	4	3.5	3	2.5	2	1.5	1.5
$l_{\rm QE}$	6.5	6	5.5	5	4.5	4	3.5	3	2.5	2.5
ν^*			1						2	-2
ν			1/3						2/5	2/3
L	10,8,6, 4,2,0	5	0	5	8,6,4, 2,0	9,7,6,5, $4,3^2,1$	8,6,5, $4^2,2^2,0$	5,3,1	0	0

For example, the Laughlin $L = 0$ ground state at $\nu = 1/3$ occurs when $2l_0^* = N-1$, so that the N composite fermions fill the lowest shell with angular momentum $l_0^* (= \frac{N-1}{2})$. The composite fermion quasielectron and quasihole states occur at $2l_0^* = N - 1 \pm 1$ and have one too many (for quasielectron) or one too few (for quasihole) quasiparticles to give integral filling. The single quasiparticle states ($n_{\rm QP} = 1$) occur at angular momentum $N/2$, for example, at $l_{\rm QE} = 5$ with $2Q^* = 8$ and $l_{\rm QH} = 5$ with $2Q^* = 10$ for $N = 10$ as indicated in Table 16.2. The two quasielectron or two quasihole states ($n_{\rm QP} = 2$) occur at $2l_0^* = N - 1 \mp 2$, and they have $2l_{\rm QE} = N - 1$ and $2l_{\rm QH} = N + 1$. For example, we expect that, for $N = 10$, $l_{\rm QE} = 4.5$ with $2Q^* = 7$ and $l_{\rm QH} = 5.5$ with $2Q^* = 11$ as indicated in Table 16.2, leading to low energy bands with $L = 0 \oplus +2 \oplus +4 \oplus +6 \oplus +8$ for two quasielectrons and $L = 0 \oplus +2 \oplus +4 \oplus +6 \oplus +8 \oplus 10$ for two quasiholes. In the mean field picture, which neglects quasiparticle-quasiparticle interactions, these bands are degenerate.

We emphasize that the low lying excitations can be described in terms of the number of quasiparticles $n_{\rm QE}$ and $n_{\rm QH}$. The total angular momentum can be obtained by addition of the individual quasiparticle angular momenta, being careful to treat the quasielectron excitations as a set of fermions and quasihole excitations as a set of fermions distinguishable from the quasielectron excitations. The energy of the excited state would simply be the sum of the individual quasiparticle energies if interactions between quasiparticles were neglected. However, interactions partially remove the degeneracy of different states having the same values of $n_{\rm QE}$ and $n_{\rm QH}$. Numerical results in Fig. 16.3b and d illustrate that two quasiparticles with different L have different energies. From this numerical data one can obtain the residual interaction $V_{\rm QP}(L')$ of a quasiparticle pair as a function of the pair angular momentum L'. In Fig. 16.2, in addition to the lowest energy band of multiplets,

the first excited band containing one additional quasielectron-quasihole pair can be observed. The "magnetoroton" band can be observed lying between the $L = 0$ Laughlin ground state of incompressible quantum liquid and a continuum of higher energy states. The band contains one quasihole with $l_{\mathrm{QH}} = 9/2$ and one quasielectron with $l_{\mathrm{QE}} = 11/2$. By adding the angular momenta of these two distinguishable particles, a band comprising L of $1(=l_{\mathrm{QE}} - l_{\mathrm{QH}}) \leq L \leq 10(=l_{\mathrm{QE}} + l_{\mathrm{QH}})$ would be predicted. But, from Fig. 16.2 we conjecture that the state with $L = 1$ is either forbidden or pushed up by interactions into the higher energy continuum above the magnetoroton band. Furthermore, the states in the band are not degenerate indicating residual interactions that depend on the angular momentum of the pair L'. Other bands that are not quite so clearly defined can also be observed in Fig. 16.3.

Although fluctuations beyond the mean field interact via both Coulomb and Chern–Simons gauge interactions, the mean field composite fermion picture is remarkably successful in predicting the low-energy multiplets in the spectrum of N electrons on a Haldane sphere. It was suggested originally that this success of the mean field picture results from the cancellation of the Coulomb and Chern–Simons gauge interactions among fluctuations beyond the mean field level. It was conjectured that the composite fermion transformation converts a system of strongly interaction electrons into one of weakly interacting composite fermions. The mean field Chern–Simons picture introduces a new energy scale $\hbar\omega_{\mathrm{c}}^*$ proportional to the effective magnetic field B^*, in addition to the energy scale e^2/l_0 ($\propto \sqrt{B}$) associated with the electron–electron Coulomb interaction. The Chern–Simons gauge interactions convert the electron system to the composite fermion system. The Coulomb interaction lifts the degeneracy of the noninteracting electron bands. The low lying multiplets of interacting electrons will be contained in a band of width e^2/l_0 about the lowest electron Landau level. The noninteracting composite fermion spectrum contains a number of bands separated by $\hbar\omega_{\mathrm{c}}^*$. However, for large values of the applied magnetic field B, the Coulomb energy can be made arbitrarily small compared to the Chern–Simons energy $\hbar\omega_{\mathrm{c}}^*$, resulting in the former being too small to reproduce the separation of levels present in the mean field composite fermion spectrum. The new energy scale is very large compared with the Coulomb scale, and it is totally irrelevant to the determination of the low energy spectrum. Despite the satisfactory description of the allowed angular momentum multiplets, the magnitude of the mean field composite fermion energies is completely wrong. The structure of the low-energy states is quite similar to that of the fully interacting electron system but completely different from that of the noninteracting system. The magnetoroton energy does not occur at the effective cyclotron energy $\hbar\omega_{\mathrm{c}}^*$. What is clear is that the success of the composite fermion picture does not result from a cancellation between Chern–Simons gauge interactions and Coulomb interactions.

16.8 Fermi Liquid Picture

The numerical result of the type displayed in Fig. 16.2 could be understood in a very simple way within the composite fermion picture. For the 10 particle system, the Laughlin $\nu = 1/3$ incompressible ground state at $L = 0$ occurs for $2Q = 3(N-1) = 27$. The low lying excited states consist of a single quasiparticle pair, with the quasielectron and quasihole having angular momentum $l_{\rm QE} = 11/2$ and $l_{\rm QH} = 9/2$. The mean field composite fermion picture does not account for quasiparticle interactions and would give a magnetoroton band of degenerate states with $1 \le L \le 10$ at $2Q = 27$. It also predicts the degeneracies of the bands of two identical quasielectron states at $2Q = 25$ and of two identical quasihole states at $2Q = 29$.

The energy spectra of states containing more than one composite fermion quasiparticle can be described in the following phenomenological *Fermi liquid model*. The creation of an elementary excitation, quasielectron or quasihole, in a Laughlin incompressible ground state requires a finite energy, $\varepsilon_{\rm QE}$ or $\varepsilon_{\rm QH}$, respectively. In a state containing more than one Laughlin quasiparticles, quasiparticles interact with one another through appropriate quasiparticle-quasiparticle *pseudopotentials*, $V_{\rm QP-QP'}$. Here, $V_{\rm QP-QP'}(L')$ is defined as the interaction energy of a pair of electrons as a function of the total angular momentum L' of the pair.

An estimate of the quasiparticle energies can be obtained by comparing the energy of a single quasielectron (for example, for the 10 electron system, the energy of the ground state at $L = N/2 = 5$ for $2Q = 27 - 1 = 26$) or a single quasihole (the $L = N/2 = 5$ ground state at $2Q = 27 + 1 = 28$ for the 10 electron system) with the Laughlin $L = 0$ ground state at $2Q = 27$. There can be finite size effects, because the quasiparticle states occur at different values of $2Q$ from that of the ground state. But estimation of reliable $\varepsilon_{\rm QE}$ and $\varepsilon_{\rm QH}$ should be possible for a macroscopic system by using the correct magnetic length $l_0 = R/\sqrt{Q}$ (R is the radius of the Haldane sphere) in the units of energy e^2/l_0 at each value of $2Q$ and by extrapolating the results as a function of N^{-1} to an infinite system.[8]

The quasiparticle pseudopotentials $V_{\rm QP-QP'}$ can be obtained by subtracting from the energies of the two quasiparticle states obtained numerically, for example, for the ten particle system at $2Q = 25$(2QE state), $2Q = 27$(QE − QH state), and $2Q = 29$(2QH state) the energy of the Laughlin ground state at $2Q = 27$ and two energies of appropriate noninteracting quasiparticles. As for the single quasiparticle, the energies calculated at different $2Q$ must be taken in correct units of $e^2/l_0 = \sqrt{Q}e^2/R$ to avoid finite size effects.

[8] P. Sitko, S.-N. Yi, K.-S. Yi, J.J. Quinn, Phys. Rev. Lett. **76**, 3396 (1996).

16.9 Pseudopotentials

Electron pair states in the spherical geometry are characterized by a pair angular momentum $L'(=L_{12})$. The Wigner–Eckart theorem tells us that the interaction energy $V_n(L')$ depends only on L' and the Landau level index n. The reason of the success of the mean field Chern–Simons picture can be seen by examining the behavior of the pseudopotential $V_{\mathrm{QP}-\mathrm{QP}'}(L')$ of a pair of particles. In the mean field approximation the energy necessary to create a quasielectron–quasihole pair is $\hbar\omega_{\mathrm{c}}^*$. However, the quasiparticles will interact with the Laughlin condensed state through the fluctuation Hamiltonian. The renormalized quasiparticle energy will include this self-energy, which is difficult to calculate. We can determine the quasiparticle energies phenomenologically using exact numerical results as input data. The picture we are using is very reminiscent of Fermi liquid theory. The ground state is the Laughlin condensed state; it plays the role of a vacuum state. The elementary excitations are quasielectrons and quasiholes. The total energy can be expressed as

$$E = E_0 + \sum_{\mathrm{QP}} \varepsilon_{\mathrm{QP}} n_{\mathrm{QP}} + \frac{1}{2} \sum_{\mathrm{QP},\mathrm{QP}'} V_{\mathrm{QP}-\mathrm{QP}'}(L) n_{\mathrm{QP}} n_{\mathrm{QP}'}. \tag{16.31}$$

The last term represents the interactions between pair of quasiparticles in a state of angular momentum L. One can take the energy spectra of finite systems, and compare the two quasiparticle states, such as $|2\mathrm{QE}\rangle$, $|2\mathrm{QH}\rangle$, or $|1\mathrm{QE}+1\mathrm{QH}\rangle$, with the composite fermion picture. The values of $V_{\mathrm{QP}-\mathrm{QP}'}(L)$ are obtained by subtracting the energies of the noninteracting quasiparticles from the numerical values of $E(L)$ for the $|1\mathrm{QP}+1\mathrm{QP}'\rangle$ states after the appropriate positive background energy correction. It is worth noting that the interaction energy for unlike quasiparticles depends on the total angular momentum L while for like quasiparticles it depends on the relative angular momentum $\mathcal{R}$, which is defined by $\mathcal{R} = L_{\mathrm{Max}} - L$. One can understand it by considering the motions in the two dimensional plane. Oppositely charged quasiparticles form bound states, in which both charges drift in the direction perpendicular to the line connecting them, and their spatial separation is related to the total angular momentum L. Like charges repel one another orbiting around one another due to the effect of the dc magnetic field. Their separation is related to their relative angular momentum $\mathcal{R}$.[9]

If $V_{\mathrm{QP}-\mathrm{QP}'}(L')$ is a "harmonic" pseudopotential of the form

$$V_{\mathrm{H}}(L') = A + BL'(L'+1)$$

[9] The angular momentum L_{12} of a pair of identical fermions in an angular momentum shell or a Landau level is quantized, and the convenient quantum number to label the pair states is the relative angular momentum $\mathcal{R} = 2l_{\mathrm{QP}} - L_{12}$ (on a sphere) or relative angular momentum m (on a plane).

every angular momentum multiplet having the same value of the total angular momentum L has the same energy. Here, A and B are constants and we mean that the interactions, which couple only states with the same total angular momentum, introduce no correlations. Any linear combination of eigenstates with the same total angular momentum has the same energy. We define $V_{\mathrm{QP}-\mathrm{QP}'}(L')$ to be "superharmonic" ("subharmonic") at $L' = 2l - \mathcal{R}$ if it increases approaching this value more quickly (slowly) than the harmonic pseudopotential appropriate at $L' - 2$. For harmonic and subharmonic pseudopotentials, Laughlin correlations do not occur. In Figs. 16.3b and d, it is clear that residual quasiparticle–quasiparticle interactions are present. If they were not present, then all of the 2QE states in frame (b) would be degenerate, as would all of the 2QH states in frame (d). In fact, these frames give us the pseudopotentials $V_{\mathrm{QE}}(\mathcal{R})$ and $V_{\mathrm{QH}}(\mathcal{R})$, up to an overall constant, describing the interaction energy of pairs with angular momentum $L' = 2l - \mathcal{R}$.

Figure 16.4 gives a plot of $V_n(L')$ vs $L'(L'+1)$ for the $n = 0$ and $n = 1$ Landau levels. For electrons in the lowest Landau level ($n = 0$), $V_0(L')$ is superharmonic at every value of L'. For excited Landau levels ($n \geq 1$), $V_n(L')$ is not superharmonic at all allowed values of L'. The allowed values of L' for a pair of fermions each of angular momentum l are given by $L' = 2l - \mathcal{R}$, where the relative angular momentum $\mathcal{R}$ must be an odd integer. We often write the pseudopotential as $V(\mathcal{R})$ since $L' = 2l - \mathcal{R}$. For the lowest Landau level $V_0(\mathcal{R})$ is superharmonic everywhere. This is apparent for the largest values of L' in

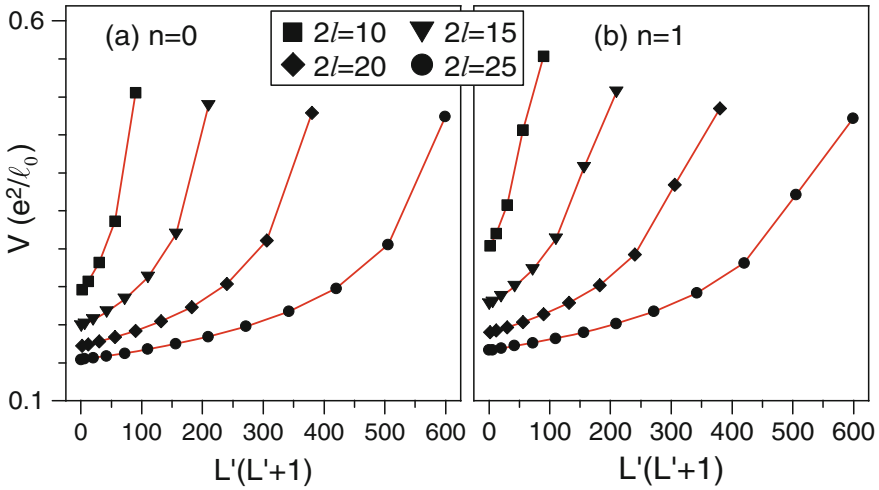


Fig. 16.4. Pseudopotential $V_n(L')$ of the Coulomb interaction in the lowest (**a**) and the first excited Landau level (**b**) as a function of the eigenvalue of the squared pair angular momentum $L'(L'+1)$. Here, n indicates the Landau level index. *Squares* ($l = 5$), *triangles* ($l = 15/2$), *diamonds* ($l = 10$), and *circles* ($l = 25/2$) indicate data for different values of $Q = l + n$

Fig. 16.4. For the first excited Landau level V_1 increases between $L' = 2l - 3$ and $L' = 2l - 1$, but it increases either harmonically or more slowly, and hence $V_1(\mathcal{R})$ is superharmonic only for $\mathcal{R} > 1$. Generally, for higher Landau levels (for example, $n = 2, 3, 4, \cdots$) $V_n(L')$ increases more slowly or even decreases at the largest values of L'. The reason for this is that the wavefunctions of the higher Landau levels have one or more nodes giving structure to the electron charge density. When the separation between the particles becomes comparable to the scale of the structure, the repulsion is weaker than for structureless particles.[10]

When plotted as a function of $\mathcal{R}$, the pseudopotentials calculated for small systems containing different number of electrons (hence for different values of quasiparticle angular momenta l_{QP}) behave similarly and, for $N \to \infty$, i.e., $2Q \to \infty$, they seem to converge to the limiting pseudopotentials $V_{\mathrm{QP-QP'}}(\mathcal{R} = m)$ describing an infinite planar system.

The number of electrons required to have a system of quasiparticle pairs of reasonable size is, in general, too large for exact diagonalization in terms of electron states and the Coulomb pseudopotential. However, by restricting our consideration to the quasiparticles in the partially field composite fermion shell and by using $V_{\mathrm{QP}}(\mathcal{R})$ obtained from numerical studies of small systems of electrons, the numerical diagonalization can be reduced to manageable size.[11] Furthermore, because the important correlations and the nature of the ground state are primarily determined by the short range part of the pseudopotential, such as at small values of $\mathcal{R}$ or small quasiparticle–quasiparticle separations, the numerical results for small systems should describe the essential correlations quite well for systems of any size.

In Fig. 16.5 we display $V_{\mathrm{QE}}(\mathcal{R})$ and $V_{\mathrm{QH}}(\mathcal{R})$ obtained from numerical diagonalization of N $(6 \leq N \leq 11)$ electron systems appropriate to quasiparticles of the $\nu = 1/3$ and $\nu = 1/5$ Laughlin incompressible quantum liquid states. We note that the behavior of quasielectrons is similar for $\nu = 1/3$ and $\nu = 1/5$ states, and the same is true for quasiholes of the $\nu = 1/3$ and $\nu = 1/5$ Laughlin states. Because $V_{\mathrm{QE}}(\mathcal{R} = 1) < V_{\mathrm{QE}}(\mathcal{R} = 3)$ and $V_{\mathrm{QE}}(\mathcal{R} = 5) < V_{\mathrm{QE}}(\mathcal{R} = 7)$, we can readily ascertain that $V_{\mathrm{QE}}(\mathcal{R})$ is subharmonic at $\mathcal{R} = 1$ and $\mathcal{R} = 5$. Similarly, $V_{\mathrm{QH}}(\mathcal{R})$ is subharmonic at $\mathcal{R} = 3$ and possibly at $\mathcal{R} = 7$.

There are clearly finite size effects since $V_{\mathrm{QP}}(\mathcal{R})$ is different for different values of the electron number N. However, $V_{\mathrm{QP}}(\mathcal{R})$ converges to a rather well defined limit when plotted as a function of N^{-1}. The results are quite accurate up to an overall constant, which is of no significance when we are interested

[10] As for a conduction electron and a valence hole pair in a semiconductor, the motion of a quasielectron–quasihole pair, which does not carry a net electric charge is not quantized in a magnetic field. The appropriate quantum number to label the states is the continuous wavevector k, which is given by $k = L/R = L/l_0\sqrt{Q}$ on a sphere.

[11] The quasiparticle pseudopotentials determined in this way are quite accurate up to an overall constant which has no effect on the correlations.

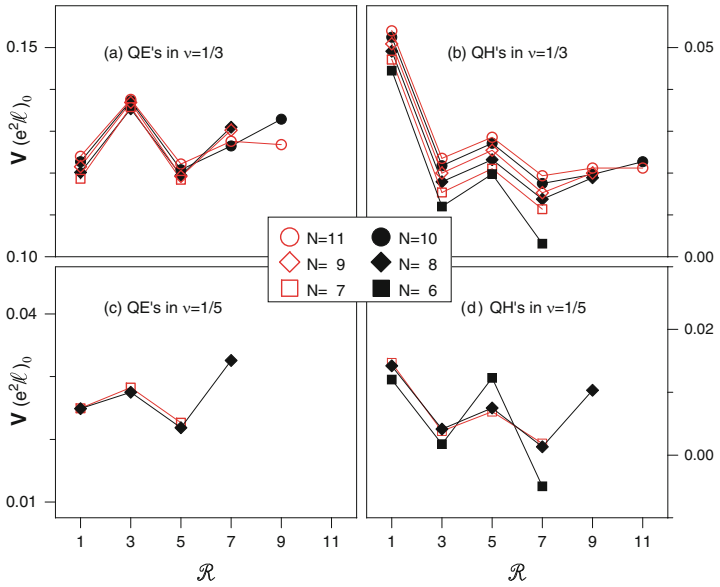


Fig. 16.5. Pseudopotentials $V(\mathcal{R})$ of a pair of quasielectrons and quasiholes in Laughlin $\nu = 1/3$ and $\nu = 1/5$ states, as a function of relative pair angular momentum $\mathcal{R}$. Different *symbols* denote data obtained in the diagonalization of between six and eleven electrons

only in the behavior of $V_{\mathrm{QP-QP}}$ as a function of $\mathcal{R}$. Once the quasiparticle–quasiparticle pseudopotentials and the bare quasiparticle energies are known, one can evaluate the energies of states containing three or more quasiparticles. Figure 16.6 illustrates typical energy spectra of three quasiparticles in the Laughlin incompressible ground states. The spectrum in frame (a) shows the energy spectrum of three quasielectrons in the Laughlin $\nu = 1/3$ state of eleven electrons. In frame (b) we show the energy spectrum of three quasiholes in the nine electron system at the same filling. The results from exact numerical diagonalization of the eleven and nine electron systems are represented by the crosses and the Fermi liquid results are indicated by solid circles. The exact energies above the dashed lines correspond to higher energy states containing additional quasielectron–quasihole pairs. It should be noted that in the mean field composite fermion model, which neglects the quasiparticle–quasiparticle interactions, all of the three quasiparticle states would be degenerate and the energy gap separating the three quasiparticle states from higher energy states would be equal to $\hbar\omega_{\mathrm{c}}^{*} = \hbar\omega_{\mathrm{c}}/3$. Although the fit is not perfect in Fig. 16.6, the agreement is quite good and it justifies the use of the Fermi liquid picture to describe non-Laughlin type compressible states at filling factor $\nu \neq \frac{1}{2p+1}$.

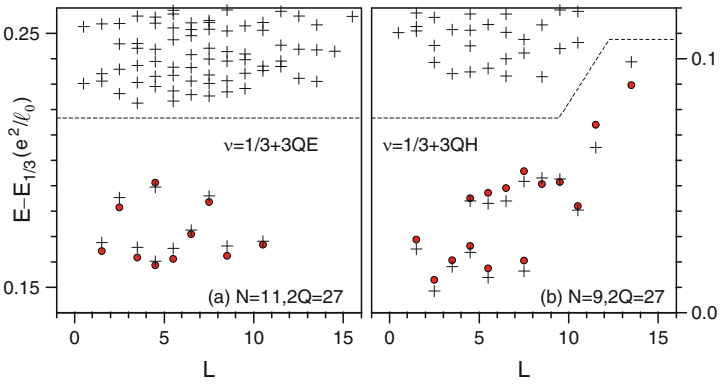


Fig. 16.6. The energy spectra of three quasielectrons (**a**) and three quasiholes (**b**) in the Laughlin $\nu = 1/3$ state. The solid circles denote the Fermi liquid calculation using pseudopotentials illustrated in Figs. 16.5a and b. The crosses correspond to exact spectra obtained in full diagonalization of the Coulomb interaction for electrons of $N = 11$ (**a**) and $N = 9$ (**b**) at $2Q = 27$

16.10 Angular Momentum Eigenstates

We have already seen that a spin polarized shell containing N fermions each with angular momentum l can be described by eigenfunctions of the total angular momentum $\hat{L} = \sum_i \hat{l}_i$ and its z-component $M = \sum_i m_i$. We define $f_L(N, l)$ as the number of multiplets of total angular momentum L that can be formed from N fermions each with angular momentum l. We usually label these multiplets as $|l^N; L\alpha\rangle$, where it is understood that each multiplet contains $2L+1$ states having $-L \le M \le L$, and α is the label that distinguishes different multiplets with the same value of L. We define $\hat{L}_{ij} = \hat{l}_i + \hat{l}_j$, the angular momentum of the pair i, j each with angular momentum l.

We can write[12]

$$|l^N; L\alpha\rangle = \sum_{L'\alpha'} \sum_{L_{12}} G_{L\alpha,L'\alpha'}(L_{12}) |l^2, L_{12}; l^{N-2}, L'\alpha'; L\rangle, \tag{16.32}$$

where $|l^{N-2}, L'\alpha'\rangle$ is the α' multiplet of total angular momentum L' of $N-2$ fermions each with angular momentum l. From $|l^{N-2}, L'\alpha'\rangle$ and $|l^2, L_{12}\rangle$ one

[12] The following theorems are quite useful:
Theorem 1: $\hat{L}^2 + N(N-2)\hat{l}^2 = \sum_{\langle i,j\rangle} \hat{L}_{ij}^2$, where the sum is over all pairs.
Theorem 2: $f_L(N, l) \ge f_L(N, l^*)$, where $l^* = l - (N-1)$ and $2l \ge N-1$.
Theorem 3: If $b_L(N, l)$ is the boson equivalent of $f_L(N, l)$, then $b_L(N, l_B) = f_L(N, l_F)$, if $l_B = l_F - \frac{1}{2}(N-1)$.

The first theorem can be proven using the definitions of $\hat{L}^2$ and $\sum_{\langle i,j\rangle} \hat{L}_{ij}^2$ and eliminating $\hat{l}_i \cdot \hat{l}_j$ from the pair of equations. The other two theorems are almost obvious conjectures to a physicist, but there exist rigorous mathematical proofs of their validity.

can construct an eigenfunction of total angular momentum L. Here, we mean that the angular momentum multiplet $|l^N; L\alpha\rangle$ obtained from N fermions, each with angular momentum l, can be formed from $|l^{N-2}, L'\alpha'\rangle$ and a pair wavefunction $|l^2, L_{12}\rangle$ by using $G_{L\alpha,L'\alpha'}(L_{12})$ the *coefficient of fractional grandparentage*. The $G_{L\alpha,L'\alpha'}(L_{12})$ produces a totally antisymmetric eigenfunction $|l^N; L\alpha\rangle$, even though $|l^2, L_{12}; l^{N-2}, L'\alpha'; L\rangle$ is not antisymmetric under exchange of particle 1 or 2 with any of the other particles.

Equation (16.32) is useful because of a very simple identity involving N fermions:

$$\hat{L}^2 + N(N-2)\hat{l}^2 - \sum_{\langle i,j\rangle} \hat{L}_{ij}^2 = 0, \tag{16.33}$$

where $\hat{L}$ is the total angular momentum operator, $\hat{L}_{ij} = \hat{l}_i + \hat{l}_j$, and the sum is over all pairs. Taking the expectation value of this identity in the state $|l^N; L\alpha\rangle$ gives the following useful result:

$$L(L+1) + N(N-2)l(l+1) = \langle l^N; L\alpha| \sum_{\langle i,j\rangle} \hat{L}_{ij}^2 |l^N; L\alpha\rangle. \tag{16.34}$$

Because (16.32) expresses the totally antisymmetric eigenfunction $|l^N; L\alpha\rangle$ as a linear combination of states of well defined pair angular momentum $\hat{L}_{ij}$, the right hand side of Eq.(16.34) can be written as

$$\frac{1}{2}N(N-1)\sum_{\alpha} \mathcal{P}_{L\alpha}(L_{12})L_{12}(L_{12}+1). \tag{16.35}$$

In this expression $\mathcal{P}_{L\alpha}(L_{12})$ is defined by

$$\mathcal{P}_{L\alpha}(L_{12}) = \sum_{L'\alpha'} |G_{L\alpha,L'\alpha'}(L_{12})|^2, \tag{16.36}$$

and is the probability that $|l^N; L\alpha\rangle$ contains pairs with pair angular momentum L_{12}. It is interesting to note that the expectation value of square of the pair angular momentum summed over all pairs is totally independent of the multiplet α. It depends only on the total angular momentum L. Because the eigenfunctions $|l^N; L\alpha\rangle$ are orthonormal, one can show that

$$\sum_{L_{12}} \sum_{L'\alpha'} G_{L\alpha,L'\alpha'}(L_{12}) G_{L\beta,L'\alpha'}(L_{12}) = \delta_{\alpha\beta}. \tag{16.37}$$

From (16.34)–(16.37) we have two useful *sum rules* involving $\mathcal{P}_{L\alpha}(L_{12})$. They are

$$\sum_{L_{12}} \mathcal{P}_{L\alpha}(L_{12}) = 1 \tag{16.38}$$

and

$$\frac{1}{2}N(N-1)\sum_{L_{12}} L_{12}(L_{12}+1)\mathcal{P}_{L\alpha}(L_{12}) = L(L+1)+N(N-2)l(l+1). \quad (16.39)$$

The energy of the multiplet $|l^N; L\alpha\rangle$ is given by

$$E_\alpha(L) = \frac{1}{2}N(N-1)\sum_{L_{12}} \mathcal{P}_{L\alpha}(L_{12})V(L_{12}), \quad (16.40)$$

where $V(L_{12})$ is the pseudopotential describing the interaction energy of a pair of fermions with pair angular momentum L_{12}.

Equation (16.40) together with our sum rules, (16.38) and (16.39) on $\mathcal{P}_{L\alpha}(L_{12})$ gives the remarkable result that, for a harmonic pseudopotential $V_{\rm H}(L_{12})$, the energy $E_\alpha(L)$ is totally independent of α and every multiplet that has the same total angular momentum L has the same energy. This proves that the harmonic pseudopotential introduces no correlations and the degeneracy of multiplets of a given L remains under the interactions. Any linear combination of the eigenstates of the total angular momentum having the same eigenvalue L is an eigenstate of the harmonic pseudopotential. Only the anharmonic part of the pseudopotential $\Delta V(\mathcal{R}) = V(\mathcal{R}) - V_{\rm H}(\mathcal{R})$ causes correlations.

16.11 Correlations in Quantum Hall States

Since the harmonic pseudopotential introduces no correlations, only the anharmonic part of the pseudopotential $\Delta V(\mathcal{R}) = V(\mathcal{R}) - V_H(\mathcal{R})$ lifts the degeneracy of the multiplets with a given L. The simplest anharmonic pseudopotential is the case in which $\Delta V(\mathcal{R}) = U\delta_{\mathcal{R},1}$. If U is positive, the lowest energy multiplet for each value of L is the one with the minimum value of the probability $\mathcal{P}_L(\mathcal{R}=1)$ for pairs with $\mathcal{R}=1$. That is, the state of the lowest energy will tend to avoid pair states with $\mathcal{R}=1$ to the maximum possible extent.[13] This is exactly what we mean by Laughlin correlations. In fact, if $U \to \infty$, the only states with finite energy are those for which $\mathcal{P}_L(\mathcal{R}=1)$, the probability for pairs with $\mathcal{R}=1$, vanishes. This cannot occur for $2Q < 3(N-1)$, the value of $2Q$ at which the Laughlin $L=0$ incompressible quantum liquid state occurs. If U is negative, the lowest energy state will have a maximum value of $\mathcal{P}_L(\mathcal{R}=1)$. This would certainly lead to the formation of clusters. However, this simple pseudopotential appears to be unrealistic, with nothing to prevent compact droplet formation and charge separation.

[13] Avoiding $\mathcal{R}=1$ is equivalent to avoiding pair states with $m=1$ in the planar geometry.

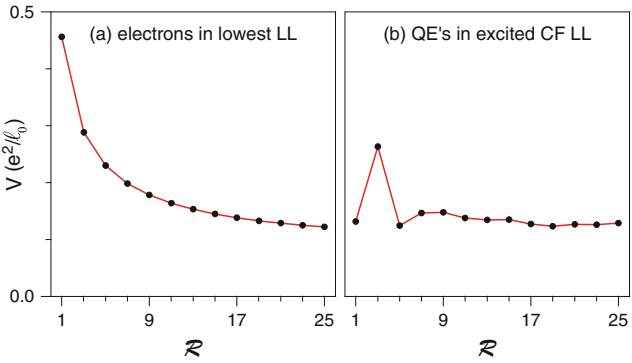


Fig. 16.7. Pseudopotentials as a function of relative pair angular momentum $\mathcal{R}$ for electrons in the lowest Landau level (**a**) and for quasielectrons in the first excited composite fermion Landau level (**b**). l_0 is the magnetic length

$\Delta V(\mathcal{R}) = U\delta_{\mathcal{R},3}$ gives the lowest energy states when $\mathcal{P}_L(\mathcal{R} = 3)$ is either a minimum for positive U or a maximum for negative U.

The pseudopotential describing the interaction energy of a pair of electrons in the first excited Landau level was shown in Fig. 16.4b. It rises faster than a harmonic potential $V_{\rm H}(\mathcal{R})$ for $\mathcal{R} = 3$ and 5, but not for $\mathcal{R} = 1$, where it appears to be harmonic or slightly subharmonic. In this situation, a Laughlin correlated state with a minimum value of $\mathcal{P}_L(\mathcal{R} = 1)$ will have higher energy than a state in which some increase of $\mathcal{P}_L(\mathcal{R} = 1)$ (from its Laughlin value) is made at the expense of $\mathcal{P}_L(\mathcal{R} = 3)$, while the sum rules (16.38) and (16.39) are still satisfied. Figure 16.7 illustrates the pseudopotentials $V_0(\mathcal{R})$ for electrons in the lowest Landau level and $V_{\rm QE}(\mathcal{R})$ for composite fermion quasielectrons in the first excited composite fermion Landau level.[14] The pseudopotential $V_{\rm QE}(\mathcal{R})$ describing the interaction of Laughlin quasiparticles is not superharmonic at all allowed values of $\mathcal{R}$ or L'.

In Fig. 16.8 the low energy spectrum of $N = 12$ electrons at $2l = 29$ and the corresponding spectrum for $N_{\rm QE} = 4$ quasielectrons at $2l_{\rm QE} = 9$, which are obtained in numerical experiments, are shown.[15] The calculation for $2l_{\rm QE} = 9$ and $N_{\rm QE} = 4$ is almost trivial in comparison to that of $N = 12$ at $2l = 29$, but the low-energy spectra are in reasonably good agreement, giving us confidence in using $V_{\rm QP}(\mathcal{R})$ to describe the composite fermion quasiparticles.

In Fig. 16.9 probability functions of pair states $\mathcal{P}(\mathcal{R})$ for the $L = 0$ ground states of the 12 electron system and the four quasielectron system illustrated

[14] S.-Y. Lee, V.W. Scarola, J.K. Jain, Phys. Rev. Lett. **87**, 256803 (2001).

[15] The composite fermion transformation applied to the electrons gives an effective composite fermion angular momentum $l^* = l - (N - 1) = 7/2$. The lowest composite fermion Landau level can accommodate $2l^* + 1 = 8$ of the particles, so that the first excited composite fermion Landau level contains the remaining four quasiparticles each of angular momentum $l_{\rm QE} = 9/2$.

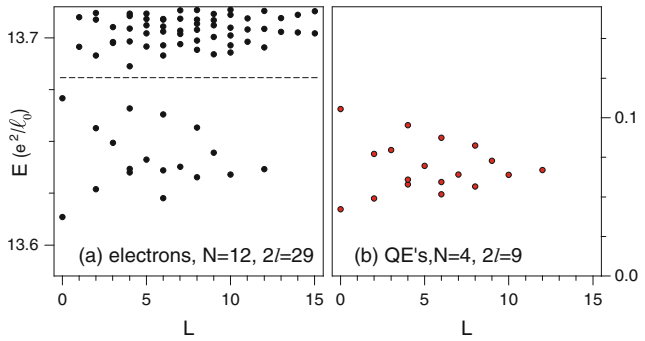


Fig. 16.8. Energy spectra for $N = 12$ electrons in the lowest Landau level with $2l = 29$ (**a**) and for $N = 4$ quasielectrons in the first excited composite fermion Landau level with $2l = 9$ (**b**). The energy scales are the same, but the quasielectron spectrum obtained using $V_{\rm QE}(\mathcal{R})$ is determined only up to an arbitrary constant

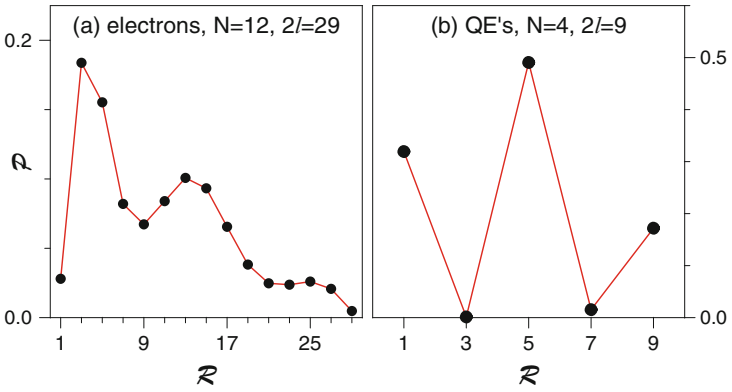


Fig. 16.9. Pair probability functions $\mathcal{P}(\mathcal{R})$ for the $L = 0$ ground states of the 12 electron system (**a**) and the four quasielectron system (**b**) shown in Fig. 16.8

in Fig. 16.8. The electrons are clearly Laughlin correlated avoiding $\mathcal{R} = 1$ pair states, but the quasielectrons are not Laughlin correlated because they avoid $\mathcal{R} = 3$ and $\mathcal{R} = 7$ pair states but not $\mathcal{R} = 1$ state.

It is known that, when $V_{\rm QP}(L')$ is not superharmonic, the interacting particles form pairs or larger clusters in order to lower the total energy.[16] These pairing correlations can lead to a nondegenerate incompressible ground state. Standard numerical calculations for N electrons are not very useful for studying novel incompressible quantum liquid states at $\nu = 3/8$, $3/10$, $4/11$, and $4/13$. Convincing numerical results would require values of N too large to diagonalize directly. However, one can look at states containing a small number

[16] A. Wojs, J.J. Quinn, Phys. Rev. B **69**, 205322 (2004).

Table 16.3. Some novel family of incompressible quantum liquid states resulting from pairing of composite fermion quasiparticles in the lowest Landau level

ν_{QE}	2/3	1/2	1/3	ν_{QH}	2/7	1/4	1/5
ν	5/13	3/8	4/11	ν	5/17	3/10	4/13

of quasiparticles of an incompressible quantum liquid state, and use $V_{\mathrm{QP}}(L')$ discussed here to obtain the spectrum of quasiparticle states. As an example, Table 16.3 shows a novel family of incompressible states resulting from the scheme of quasiparticle pairing.

Problems

16.1. The many particle wavefunction is written, for $\nu = 1$, by

$$\Psi_1(z_1, \ldots, z_N) = \mathcal{A}\{u_0(z_1)u_1(z_2)\cdots u_{N-1}(z_N)\},$$

where $\mathcal{A}$ denotes the antisymmetrizing operator. Demonstrate explicitly that $\Psi_1(z_1, \ldots, z_N)$ can be written as follows:

$$\Psi_1(z_1, \ldots, z_N) \propto \begin{vmatrix} 1 & 1 & \cdots 1 \\ z_1 & z_2 & \cdots z_N \\ z_1^2 & z_2^2 & \cdots z_N^2 \\ \vdots & \vdots & \cdots \vdots \\ z_1^{N-1} & z_2^{N-1} & \cdots z_N^{N-1} \end{vmatrix} e^{-\frac{1}{4l_0^2}\sum_{i=1,N}|z_i|^2}.$$

16.2. Consider a system of N electrons confined to a Haldane surface of radius R. There is a magnetic monopole of strength $2Q\phi_0$ at the center of the sphere.

(a) Demonstrate that, in the presence of a radial magnetic field $\boldsymbol{B} = \frac{2Q\phi_0}{4\pi R^2}\hat{R}$, the single particle Hamiltonian is given by

$$H_0 = \frac{1}{2mR^2}\left(\boldsymbol{l} - \hbar Q\hat{R}\right)^2.$$

Here, $\hat{R}$ and $\boldsymbol{l}$ are, respectively, a unit vector in the radial direction and the angular momentum operator.

(b) Show that the single particle eigenvalues of H_0 are written as

$$\varepsilon(Q, l, m) = \frac{\hbar\omega_c}{2Q}[l(l+1) - Q^2].$$

16.3. Figure 16.4 displays $V_{\mathrm{QE}}(\mathcal{R})$ and $V_{\mathrm{QH}}(\mathcal{R})$ obtained from numerical diagonalization of N ($6 \le N \le 11$) electron systems appropriate to quasiparticles of the $\nu = 1/3$ and $\nu = 1/5$ Laughlin incompressible quantum liquid states. Demonstrate that $V_{\mathrm{QP}}(\mathcal{R})$ converges to a rather well defined limit by plotting $V_{\mathrm{QP}}(\mathcal{R})$ as a function of N^{-1} at $\mathcal{R} = 1$, 3, and 5.

16.4. Consider a system of N fermions and prove an identity given by

$$\hat{L}^2 + N(N-2)\hat{l}^2 - \sum_{\langle i,j\rangle}\hat{L}_{ij}^2 = 0.$$

Here, $\hat{L}$ is the total angular momentum operator, $\hat{L}_{ij} = \hat{l}_i + \hat{l}_j$, and the sum is over all pairs. *Hint:* One can write out the definitions of $\hat{L}^2$ and $\sum_{\langle i,j\rangle}\hat{L}_{ij}^2$ and eliminate $\hat{l}_i \cdot \hat{l}_j$ from the pair of equations.

16.5. Demonstrate that the expectation value of square of the pair angular momentum L_{ij} summed over all pairs is totally independent of the multiplet α and depends only on the total angular momentum L.

16.6. Derive the two sum rules involving $\mathcal{P}_{L\alpha}(L_{12})$, i.e., the probability that the multiplet $|l^N; N\alpha\rangle$ contains pairs having pair angular momentum L_{12}:

$$\frac{1}{2}N(N-1)\sum_{L_{12}} L_{12}(L_{12}+1)\mathcal{P}_{L\alpha}(L_{12}) = L(L+1) + N(N-2)l(l+1)$$

and

$$\sum_{L_{12}} \mathcal{P}_{L\alpha}(L_{12}) = 1.$$

16.7. Show that, for harmonic pseudopotential $V_{\mathrm{H}}(L_{12})$, the energy of the multiplet $|l^N; L\alpha\rangle$ is given by

$$E_\alpha(L) = N\left[\frac{1}{2}(N-1)A + B(N-2)l(l+1)\right] + BL(L+1).$$

Summary

In this chapter, we introduce basic concepts commonly used to interpret experimental data on the quantum Hall effect. We begin with a description of two dimensional electrons in the presence of a perpendicular magnetic field. The occurrence of incompressible quantum fluid states of a two-dimensional system is reviewed as a result of electron–electron interactions in a highly degenerate fractionally filled Landau level. The idea of harmonic pseudopotential is introduced and residual interactions among the quasiparticles are analyzed. For electrons in the lowest Landau level the interaction energy of a pair of particles is shown to be superharmonic at every value of pair angular momenta.

The Hamiltonian of an electron (of mass μ) confined to the x-y plane, in the presence of a dc magnetic field $\boldsymbol{B} = B\hat{z}$, is simply $H = (2\mu)^{-1}\left[\boldsymbol{p} + \frac{e}{c}\boldsymbol{A}(\boldsymbol{r})\right]^2$, where $\boldsymbol{A}(\boldsymbol{r})$ is given by $\boldsymbol{A}(\boldsymbol{r}) = \frac{1}{2}B(-y\hat{x} + x\hat{y})$ in a $s$ymmetric gauge. The Schrödinger equation $(H - E)\Psi(\boldsymbol{r}) = 0$ has eigenstates described by

$$\Psi_{nm}(r, \phi) = e^{\mathrm{i}m\phi}u_{nm}(r) \text{ and } E_{nm} = \frac{1}{2}\hbar\omega_c(2n + 1 + m + |m|),$$

where n and m are principal and angular momentum quantum numbers, respectively, and $\omega_c(= eB/\mu c)$ is the cyclotron angular frequency. The lowest Landau level wavefunction can be written as $\Psi_{0m} = \mathcal{N}_m z^{|m|}\mathrm{e}^{-|z|^2/4{l_0}^2}$ where $\mathcal{N}_m$ is the normalization constant and z stands for $z(= x - \mathrm{i}y) = re^{-\mathrm{i}\phi}$. The filling factor ν of a given Landau level is defined by N/N_ϕ, so that ν^{-1} is simply equal to the number of flux quanta of the dc magnetic field per electron. The integral quantum Hall effect occurs when N electrons exactly fill an integral number of Landau levels resulting in an integral value of the filling factor ν. The energy gap (equal to $\hbar\omega_c$) between the filled states and the empty states makes the noninteracting electron system incompressible. A many particle wavefunction of N electrons at filling factor $\nu = 1$ becomes

$$\Psi_1(z_1, \ldots, z_N) \propto \mathit{\Pi}_{N \geq i > j \geq 1} z_{ij}\mathrm{e}^{-\frac{1}{4l_0^2}\sum_{k=1,N}|z_k|^2}.$$

For filling factor $\nu = 1/n$, Laughlin ground state wavefunction is written as

$$\Psi_{1/n}(1, 2, \cdots, N) = \prod_{i>j} z_{ij}^n \ \mathrm{e}^{-\sum_l |z_l|^2/4l_0^2},$$

where n is an odd integer.

It is convenient to introduce a Haldane sphere at the center of which is located a magnetic monopole and a small number of electrons are confined on its surface. The numerical problem is to diagonalize the interaction Hamiltonian $H_{\mathrm{int}} = \sum_{i<j} V(|\boldsymbol{r}_i - \boldsymbol{r}_j|)$. The calculations give the eigenenergies E as a function of the total angular momentum L.

Considering a two dimensional system of particles described by a Hamiltonian

$$H = \frac{1}{2\mu}\sum_i \left[\boldsymbol{p}_i + \frac{e}{c}\boldsymbol{A}(\boldsymbol{r}_i)\right]^2 + \sum_{i>j} V(r_{ij}),$$

we can change the statistics by attaching to each particle a fictitious charge q and flux tube carrying magnetic flux Φ. The fictitious vector potential $\boldsymbol{a}(\boldsymbol{r}_i)$ at the position of the ith particle caused by flux tubes, each carrying flux of Φ, on all the other particles at $\boldsymbol{r}_j(\neq \boldsymbol{r}_i)$ is written as $\boldsymbol{a}(\boldsymbol{r}_i) = \Phi\sum_{j\neq i}\frac{\hat{z}\times\boldsymbol{r}_{ij}}{r_{ij}^2}$. The *Chern–Simons gauge field* due to the gauge potential $\boldsymbol{a}(\boldsymbol{r}_i)$ becomes $\boldsymbol{b}(\boldsymbol{r}) = \Phi\sum_i \delta(\boldsymbol{r}-\boldsymbol{r}_i)\hat{z}$, where $\boldsymbol{r}_i$ is the position of the ith particle carrying gauge potential $\boldsymbol{a}(\boldsymbol{r}_i)$. The new Hamiltonian, through Chern–Simons gauge transformation, is

$$H_{\mathrm{CS}} = \frac{1}{2\mu}\int \mathrm{d}^2 r\psi^\dagger(\boldsymbol{r})\left[\boldsymbol{p} + \frac{e}{c}\boldsymbol{A}(\boldsymbol{r}) + \frac{e}{c}\boldsymbol{a}(\boldsymbol{r})\right]^2\psi(\boldsymbol{r}) + \sum_{i>j} V(r_{ij}).$$

The net effect of the additional Chern–Simons term is to replace the statistics parameter θ with $\theta + \pi\Phi\frac{q}{hc}$. If $\Phi = p\frac{hc}{e}$ when p is an integer, then $\theta \to \theta + \pi pq/e$. For the case of $q = e$ and $p = 2$, the statistics would be unchanged by the Chern–Simons terms, and the gauge interactions convert the electrons system to the composite fermions which interact through the gauge field term as well as through the Coulomb interaction. In the mean field approach, the composite fermions move in an effective magnetic field B^*. The composite fermion filling factor ν^* is given by ${\nu^*}^{-1} = \nu^{-1} - \alpha$. The mean field picture predicts not only the sequence of incompressible ground states, given by $\nu = \frac{\nu^*}{1+2p\nu^*}$ (with integer p), but also the correct band of low energy states for any value of the applied magnetic field. The low lying excitations can be described in terms of the number of quasiparticles n_{QE} and n_{QH}.

In a state containing more than one Laughlin quasiparticles, quasiparticles interact with one another through appropriate quasiparticle-quasiparticle *pseudopotentials*, $V_{\mathrm{QP}-\mathrm{QP}'}$. The total energy can be expressed as

$$E = E_0 + \sum_{\mathrm{QP}}\varepsilon_{\mathrm{QP}}n_{\mathrm{QP}} + \frac{1}{2}\sum_{\mathrm{QP},\mathrm{QP}'} V_{\mathrm{QP}-\mathrm{QP}'}(L)n_{\mathrm{QP}}n_{\mathrm{QP}'}.$$

If $V_{\mathrm{QP}-\mathrm{QP}'}(L')$ is a "harmonic" pseudopotential of the form $V_{\mathrm{H}}(L') = A + BL'(L'+1)$ every angular momentum multiplet having the same value of the total angular momentum L has the same energy. We define $V_{\mathrm{QP}-\mathrm{QP}'}(L')$ to be "superharmonic" ("subharmonic") at $L' = 2l - \mathcal{R}$ if it increases approaching this value more quickly (slowly) than the harmonic pseudopotential appropriate at $L' - 2$. For harmonic and subharmonic pseudopotentials, Laughlin correlations do not occur. Since the harmonic pseudopotential introduces no correlations, only the anharmonic part of the pseudopotential $\Delta V(\mathcal{R}) = V(\mathcal{R}) - V_H(\mathcal{R})$ lifts the degeneracy of the multiplets with a given L.

A

Operator Method for the Harmonic Oscillator Problem

Hamiltonian

The Hamiltonian of a particle of mass m moving in a one-dimensional harmonic potential is

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2. \tag{A.1}$$

The quantum mechanical operators p and x satisfy the commutation relation $[p,\ x]_- = -\imath\hbar$ where $\imath = \sqrt{-1}$. The Hamiltonian can be written

$$H = \frac{1}{2m}\left(m\omega x - \imath p\right)\left(m\omega + \imath p\right) + \frac{1}{2}\hbar\omega. \tag{A.2}$$

To see the equivalence of (A.1) and (A.2) one need only multiply out the product in (A.2) remembering that p and x are operators which do not commute. Equation (A.2) can be rewritten by

$$H = \hbar\omega\left\{\frac{\left(m\omega x - \imath p\right)}{\sqrt{2m\hbar\omega}}\ \frac{\left(m\omega x + \imath p\right)}{\sqrt{2m\hbar\omega}} + \frac{1}{2}\right\}. \tag{A.3}$$

We now define the operator a and its adjoint $a^\dagger$ by the relations

$$a\ = \tfrac{m\omega x + \imath p}{\sqrt{2m\hbar\omega}} \tag{A.4}$$

$$a^\dagger = \tfrac{m\omega x - \imath p}{\sqrt{2m\hbar\omega}}. \tag{A.5}$$

These two equations can be solved for the operators x and p to give

$$x\ = \left(\tfrac{\hbar}{2m\omega}\right)^{1/2}\left(a^\dagger + a\right), \tag{A.6}$$

$$p = \imath\left(\tfrac{m\hbar\omega}{2}\right)^{1/2}\left(a^\dagger - a\right). \tag{A.7}$$

It follows from the commutation relation satisfied by x and p that

$$\left[a,\, a^{\dagger}\right]_{-} = 1, \tag{A.8}$$

$$[a,\, a]_{-} = \left[a^{\dagger},\, a^{\dagger}\right]_{-} = 0. \tag{A.9}$$

By using the relation

$$[A, BC]_{-} = B\,[A, C]_{-} + [A, B]_{-}\,C, \tag{A.10}$$

it is not difficult to prove that

$$\begin{aligned}
\left[a,\, a^{\dagger 2}\right]_{-} &= 2a^{\dagger}, \\
\left[a,\, a^{\dagger 3}\right]_{-} &= 3a^{\dagger 2}, \\
&\vdots \\
\left[a,\, a^{\dagger n}\right]_{-} &= n a^{\dagger n-1}.
\end{aligned} \tag{A.11}$$

Here, $a^{\dagger}$ and a are called as *raising* and *lowering operators*, respectively.

From (A.3)–(A.5) it can be seen that

$$H = \hbar\omega\left(a^{\dagger}a + \frac{1}{2}\right). \tag{A.12}$$

Now, assume that $|n>$ is an eigenvector of H with an eigenvalue ε_n. Operate on $|n>$ with $a^{\dagger}$, and consider the energy of the resulting state. We can certainly write

$$H\left(a^{\dagger}|n>\right) = a^{\dagger}H|n> + \left[H, a^{\dagger}\right]|n>. \tag{A.13}$$

But we have assumed that $H|n>= \varepsilon_n|n>$, and we can evaluate the commutator $[H, a^{\dagger}]$.

$$\begin{aligned}
\left[H, a^{\dagger}\right] &= \hbar\omega\left[a^{\dagger}a, a^{\dagger}\right] = \hbar\omega a^{\dagger}\left[a, a^{\dagger}\right] \\
&= \hbar\omega a^{\dagger}.
\end{aligned} \tag{A.14}$$

Therefore, (A.13) gives

$$Ha^{\dagger}|n>= \left(\varepsilon_n + \hbar\omega\right)a^{\dagger}|n>. \tag{A.15}$$

Equation (A.15) tells us that if $|n>$ is an eigenvector of H with eigenvalue ε_n, then $a^{\dagger}|n>$ is also an eigenvector of H with eigenvalue $\varepsilon_n + \hbar\omega$. Exactly the same technique can be used to show that

$$Ha|n>= \left(\varepsilon_n - \hbar\omega\right)a|n>. \tag{A.16}$$

Thus, $a^{\dagger}$ and a act like raising and lowering operators, raising the energy by $\hbar\omega$ or lowering it by $\hbar\omega$.

Ground State

Since $V(x) \geq 0$ everywhere, the energy must be greater than or equal to zero. Suppose the ground state of the system is denoted by $|0>$. Then, by applying the operator a to $|0>$ we generate a state whose energy is lower by $\hbar\omega$, i.e.,

$$Ha|0>=(\varepsilon_0-\hbar\omega)\,a|0>. \tag{A.17}$$

The only possible way for (A.17) to be consistent with the assumption that $|0>$ was the ground state is to have $a|0>$ give zero. Thus, we have

$$a|0>=0. \tag{A.18}$$

If we use the position representation where $\Psi_0(x)$ is the ground state wavefunction and p can be represented by $p=-\imath\hbar\partial/\partial x$, (A.18) becomes a simple first-order differential equation

$$\left(\frac{\partial}{\partial x}+\frac{m\omega}{\hbar}x\right)\Psi_0(x)=0. \tag{A.19}$$

One can see immediately see that the solution of (A.19) is

$$\Psi_0(x)=N_0\mathrm{e}^{-\frac{1}{2}\alpha^2x^2}, \tag{A.20}$$

where N_0 is a normalization constant, and $\alpha^2=\frac{m\omega}{\hbar}$. The normalization constant is given by $N_0=\alpha^{1/2}\pi^{-1/4}$. The energy is given by $\varepsilon_0=\frac{\hbar\omega}{2}$, since $a^\dagger a|0>=0$.

Excited States

We can generate all the excited states by using the operator $a^\dagger$ to raise the system to the next higher energy level, i.e., if we label the nth excited state by $|n>$,

$$\begin{aligned}
|1>&\propto a^\dagger|0>, \qquad \varepsilon_1=\hbar\omega\left(1+\frac{1}{2}\right),\\
|2>&\propto {a^\dagger}^2|0>, \qquad \varepsilon_2=\hbar\omega\left(2+\frac{1}{2}\right),\\
&\vdots\\
|n>&\propto {a^\dagger}^n|0>, \qquad \varepsilon_n=\hbar\omega\left(n+\frac{1}{2}\right).
\end{aligned} \tag{A.21}$$

Because $a^\dagger$ creates one quantum of excitation and a annihilates one, $a^\dagger$ and a are often called *creation* and *annihilation operators*, respectively.

If we wish to normalize the eigenfunctions $|n>$ we can write

$$|n>=C_n{a^\dagger}^n|0>\,. \tag{A.22}$$

Assume that $|0>$ is normalized (see (A.20)). Then, we can write

$$\langle n|n\rangle = |C_n|^2 \left\langle 0 \left| a^n a^{\dagger n} \right| 0 \right\rangle . \tag{A.23}$$

Using the relations given by (A.12) allows one to show that

$$a^n a^{\dagger n}|0>= n!|0> . \tag{A.24}$$

So that

$$|n>= \frac{1}{\sqrt{n!}} a^{\dagger n}|0> \tag{A.25}$$

is the normalized eigenfunction for the nth excited state.

One can use $\Psi_0(x) = \alpha^{1/2}\pi^{-1/4}\mathrm{e}^{-\frac{1}{2}\alpha^2 x^2}$ and express $a^{\dagger n}$ in terms of p and x to obtain

$$\Psi_n(x) = \frac{1}{\sqrt{n!}} \left[\frac{-\imath\left(-\imath\hbar\partial/\partial x\right) + m\omega x}{\sqrt{2m\hbar\omega}} \right]^n \frac{\alpha^{1/2}}{\pi^{1/4}} \mathrm{e}^{-\frac{\alpha^2 x^2}{2}}, \tag{A.26}$$

This can be simplified a little to the form

$$\Psi_n(x) = \frac{\left(\alpha/\sqrt{\pi}\right)^{1/2}(-)^n}{\alpha^n \left(2^n n!\right)^{1/2}} \left(\frac{\partial}{\partial x} - \alpha^2 x \right)^n \mathrm{e}^{-\frac{\alpha^2 x^2}{2}}. \tag{A.27}$$

Summary

The Hamiltonian of the simple harmonic oscillator can be written

$$H = \hbar\omega \left(a^\dagger a + \frac{1}{2} \right). \tag{A.28}$$

and $H|n>= \hbar\omega(n+\frac{1}{2})|n>$. The excited eigenkets can be written

$$|n>= \frac{1}{\sqrt{n!}} a^{\dagger n}|0> . \tag{A.29}$$

The eigenfunctions (A.29) form a complete orthonormal set, i.e.,

$$\langle n|m\rangle = \delta_{nm}, \tag{A.30}$$

and

$$\sum_n |n\rangle \langle n| = 1. \tag{A.31}$$

The creation and annihilation operators satisfy the commutation relation

$$[a,\ a^\dagger] = 1.$$

Problems

A.1. Prove that $[\hat{A}, \hat{B}\hat{C}]_- = \hat{B}[\hat{A}, \hat{C}]_- + [\hat{A}, \hat{B}]_-\hat{C}$, where $\hat{A}$, $\hat{B}$, and $\hat{C}$ are quantum mechanical operators.

A.2. Prove that $[\hat{a}, (\hat{a}^\dagger)^n]_- = n(\hat{a}^\dagger)^{n-1}$.

B

Neutron Scattering

A beam of neutrons interacts with a crystal through a potential

$$V(\mathbf{r}) = \sum_{\mathbf{R}_i} v(\mathbf{r} - \mathbf{R}_i), \tag{B.1}$$

where $\mathbf{r}$ is the position operator of the neutron, and $\mathbf{R}_i$ is the position operator of the i^{th} atom in the crystal. It is common to write $v(\mathbf{r} - \mathbf{R}_i)$ in terms of its Fourier transform $v(\mathbf{r}) = \sum_{\mathbf{k}} v_{\mathbf{k}} \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}$. Then, (B.1) can be rewritten

$$V(\mathbf{r}) = \sum_{\mathbf{k},\mathbf{R}_i} v_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{r}-\mathbf{R}_i)}. \tag{B.2}$$

The potential $v(\mathbf{r})$ is very short-range, and $v_{\mathbf{k}}$ is almost independent of $\mathbf{k}$. The $\mathbf{k}$-independent coefficient $v_{\mathbf{k}}$ is usually expressed as $v = \frac{2\pi\hbar^2 a}{M_n}$, where a is defined as the *scattering length* and M_n is the mass of the neutron.

The initial state of the system can be expressed as

$$\Psi_i(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{r}) = V^{-1/2} \mathrm{e}^{\mathrm{i}\frac{\mathbf{p}}{\hbar}\cdot\mathbf{r}} |n_1, n_2, \ldots, n_N\rangle . \tag{B.3}$$

Here, $V^{-1/2}\mathrm{e}^{\mathrm{i}\frac{\mathbf{p}}{\hbar}\cdot\mathbf{r}}$ is the initial state of a neutron of momentum $\mathbf{p}$. The ket $|n_1, n_2, \ldots, n_N\rangle$ represents the initial state of the crystal, with n_i phonons in mode i. The final state, after the neutron is scattered, is

$$\Psi_f(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{r}) = V^{-1/2} \mathrm{e}^{\mathrm{i}\frac{\mathbf{p}'}{\hbar}\cdot\mathbf{r}} |m_1, m_2, \cdots, m_N\rangle . \tag{B.4}$$

The transition rate for going from Ψ_i to Ψ_f can be calculated from Fermi's golden rule.

$$R_{i\to f} = \frac{2\pi}{\hbar} |\langle \Psi_f | V | \Psi_i \rangle|^2 \delta(E_f - E_i) . \tag{B.5}$$

Here, E_i and E_f are the initial and final energies of the entire system. Let us write $\varepsilon_i = E_i - \frac{p^2}{2M_n}$ and $\varepsilon_f = E_f - \frac{p'^2}{2M_n}$. The total rate of scattering out of

initial state Ψ_i is given by

$$R_{\text{outof } i} = \frac{2\pi}{\hbar} \sum_f \delta\left(\varepsilon_f - \varepsilon_i - \hbar\omega\right) \left|\langle \Psi_f \left| V \right| \Psi_i \rangle\right|^2, \tag{B.6}$$

where $\hbar\omega = \frac{p'^2 - p^2}{2M_n}$ is the change in energy of the neutron. If we write $\mathbf{p}' = \mathbf{p} + \hbar\mathbf{k}$, where $\hbar\mathbf{k}$ is the momentum transfer, the matrix element becomes

$$\sum_{i,k} \left\langle m_1, m_2, \ldots, m_N \left| v_{\mathbf{k}} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_i} \right| n_1, n_2, \ldots, n_N \right\rangle. \tag{B.7}$$

But we can take $v_{\mathbf{k}} = v$ outside the sum since it is a constant. In addition, we can write $\mathbf{R}_j = \mathbf{R}_j^0 + \mathbf{u}_j$ and

$$\mathbf{u}_j = \sum_{\mathbf{q}\lambda} \left(\frac{\hbar}{2MN\omega_{\mathbf{q}\lambda}} \right)^{1/2} \mathrm{e}^{i\mathbf{q}\cdot\mathbf{R}_j^0} \hat{\varepsilon}_{\mathbf{q}\lambda} \left(a_{\mathbf{q}\lambda} - a^\dagger_{-\mathbf{q}\lambda} \right). \tag{B.8}$$

The matrix element of $\mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{u}_j}$ between harmonic oscillator states $|n_1, n_2, \ldots, n_N\rangle$ and $|m_1, m_2, \ldots, m_N\rangle$ is exactly what we evaluated earlier in studying the Mössbauer effect. By using our earlier results and then summing over the atoms in the crystal, one can obtain the transition rate. The cross-section is related to the transition rate divided by the incident flux.

One can find the following result for the cross-section:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega \mathrm{d}\omega} = \frac{p'}{p} N \frac{a^2}{\hbar} S(q, \omega), \tag{B.9}$$

where $\mathrm{d}\Omega$ is solid angle, $\mathrm{d}\omega$ is energy transfer, N is the number of atoms in the crystal, a is the scattering length, and $S(q, \omega)$ is called the *dynamic structure factor*. It is given by

$$S(q,\omega) = N^{-1} \sum_f \left| \sum_j \left\langle m_1, \ldots, m_N \left| \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\mathbf{u}_j} \right| n_1, \ldots, n_N \right\rangle \right|^2 \delta\left(\varepsilon_f - \varepsilon_i - \hbar\omega\right). \tag{B.10}$$

Again, there is an elastic scattering part of $S(q, \omega)$, corresponding to no-phonon emission or absorption in the scattering process. For that case $S(q, \omega)$ is given by

$$S_0(q,\omega) = \mathrm{e}^{-2W} \delta(\omega) N \sum_K \delta_{q,K}. \tag{B.11}$$

Here, e^{-2W} is the *Debye–Waller factor*. W is proportional to

$$\left[\left\langle n_1, \ldots, n_N \left| \left[\mathbf{q}\cdot\mathbf{u}_0\right]^2 \right| n_1, \ldots, n_N \right\rangle \right].$$

From (B.11) we see that there are Bragg peaks. In the harmonic approximation the peaks are δ-functions [because of $\delta(\omega)$] due to energy conservation. The peaks occur at momentum transfer $\mathbf{p}' - \mathbf{p} = \mathbf{K}$, a reciprocal lattice vector.

In the early days of X-ray scattering there was some concern over whether the motion of the atoms (both zero point and thermal motion) would broaden the δ-function peaks and make X-ray diffraction unobservable. The result, in the harmonic approximation, is that the δ-function peaks are still there, but their amplitude is reduced by the Debye–Waller factor e^{-2W}.

For the one-phonon contribution to the cross-section, we obtain

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega \mathrm{d}\omega} = N\mathrm{e}^{-2W}\frac{p'}{p}a^2\sum_{\lambda}\frac{\left(\mathbf{q}\cdot\hat{\varepsilon}_{\mathbf{q}\lambda}\right)^2}{2M\omega_{\mathbf{q}\lambda}}\left\{\left(1+n_{\mathbf{q}\lambda}\right)\delta\left(\omega+\omega_{\mathbf{q}\lambda}\right)+n_{\mathbf{q}\lambda}\delta\left(\omega-\omega_{\mathbf{q}\lambda}\right)\right\}. \tag{B.12}$$

There are still unbroadened δ-function peaks at $\varepsilon_f \pm \hbar\omega_{\mathbf{q}\lambda} = \varepsilon_i$, corresponding to the emission or absorption of a phonon. The peaks occur at a scattering angle determined from $\mathbf{p}' - \mathbf{p} = \mathbf{q} + \mathbf{K}$ where $\mathbf{K}$ is a reciprocal lattice vector. The amplitude again contains the Debye–Waller factor e^{-2W}. Inelastic neutron scattering allows a experimentalist to determine the phonon frequencies $\omega_{\mathbf{q}\lambda}$ as a function of $\mathbf{q}$ and of λ.

The broadening of the δ-function peaks occurs only when anharmonic terms are included in the calculation. Anharmonic forces lead to phonon–phonon scattering and to finite phonon lifetimes.

Bibliography

General Reading

N.W. Ashcroft, N.D. Mermin, *Solid State Physics* (Thomson Learning, 1976)
O. Madelung, *Introduction to Solid-State Theory* (Springer, Berlin, 1978)
C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 2005)
G. Grosso, G.P. Parravicini, *Solid State Physics* (Academic, London, 2000)
H. Ibach, H. Lüth, *Solid-State Physics An Introduction to Principles of Material Science* (Springer, Heidelberg, 2003)

Chapter 1 Crystal Structures

R.W.G. Wyckoff, *Crystal Structures*, vol. 1–5 (Wiley, New York, 1963–1968)
G. Burns, *Solid State Physics* (Academic, New York, 1985)

Chapter 2 Lattice Vibrations

L. Brillouin, *Wave Propagation in Periodic Structures* (Dover, New York, 1953)
R.A. Smith, *Wave Mechanics of Crystalline Solids*, 2nd edn. (Chapman & Hall, London, 1969)
J.M. Ziman, *Electrons and Phonons* (Oxford University Press, Oxford, 1972)
P.F. Choquard, *The Anharmonic Crystal* (Benjamin, New York, 1967)
G.K. Wertheim, *Mössbauer Effect: Principles and Applications* (Academic, New York, 1964)

Chapter 3 Free Electron Theory of Metals

K. Huang, *Statistical Mechanics*, 2nd edn. (Wiley, New York, 1987)
C. Kittel, *Elementary Statistical Physics* (Wiley, New York, 1958)
R. Kubo, *Statistical Mechanics*, 7th edn. (North-Holland, Amsterdam, 1988)
R.K. Pathria, *Statistical Mechanics*, 7th edn. (Pergamon, Oxford, 1972)
A.H. Wilson *The Theory of Metals*, 2nd edn. (Cambridge University Press, Cambridge, 1965)

Chapter 4 Elements of Band Theory

J. Callaway, *Energy Band Theory* (Academic, New York, 1964)
J.M. Ziman, *Electrons and Phonons* (Oxford University Press, Oxford, 1972)
W.A. Harrision, *Electronic Structure and the Properties of Solids* (Freeman, San Francisco, 1980)
J. Callaway, *Quantum Theory of the Solid State*, 2nd edn. (Academic, New York, 1991)

Chapter 5 Use of Elementary Group Theory in Calculating Band Structure

V. Heine *Group Theory in Quantum Mechanics – An Introduction to Its Present Usage* (Pergamon, Oxford, 1977)
S.L. Altmann, *Band Theory of Solids: An Introduction from the Point of View of Symmetry* (Clarendon, Oxford, 1994)
H. Jones, *The Theory of Brillouin Zones and Electronic States in Crystals* (North-Holland, Amsterdam, 1975)
L.M. Falicov, *group theory and Its applications* (University of Chicago Press, Chicago, 1966)
M. Tinkham, *Group Theory and Quantum Mechanics* (McGraw-Hill, New York, 1964)
P.Y. Yu, M. Cardona, *Fundamentals of Semiconductors* (Springer, Berlin, 1996)

Chapter 6 More Band Theory and the Semiclassical Approximation

J.C. Slater, *Symmetry and Energy Band in Crystals* (Dover, New York, 1972)
D. Long, *Energy Bands in Semiconductors* (Wiley, New York, 1968)
J. Callaway, *Energy Bands Theory* (Academic, New York, 1964)

Chapter 7 Semiconductors

J.C. Phillips, *Bonds and Bands in Semiconductors* (Academic, New York, 1973)
A. Anselm, *Introduction to Semiconductor Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1981)
P.S. Kireev, *Semiconductor Physics* (Mir, Moscow, 1978)
P.Y. Yu, M. Cardona, *Fundamentals of Semiconductors* (Springer, Berlin, 1996)
K.F. Brennan, *THE PHYSICS OF SEMICONDUCTORS with applications to optoelectronic devices* (Cambridge University Press, Cambridge, 1999)
M. Balkanski, R.F. Wallis, *Semiconductor Physics and Applications* (Oxford University Press, Oxford, 2000)
K. Seeger, *Semiconductor Physics – An Introduction*, 8th edn. (Springer, Berlin, 2002)
S.M. Sze, *Semiconductor Devices: Physics and Technology*, 2nd edn. (Wiley, New York, 2001)
G. Bastard, *wave mechanics applied to semiconductor heterostructures* (Halsted, New York, 1986)
T. Ando, A. Fowler, F. Stern, *Rev. Mod. Phys.* **54**, 437 (1982)
T. Chakraborty, P. Pietiläinen, *The Quantum Hall Effects, Integral and Fractional*, 2nd edn. (Springer, Berlin, 1995)
R.E. Prange, S.M. Girvin (eds.), *The Quantum Hall Effect*, 2nd ed., (Springer, New York, 1990)

Chapter 8 Dielectric Properties of Solids

J.M. Ziman, *Principles of the Theory of Solids* (Cambridge University Press, Cambridge, 1972)
G. Grosso, G.P. Parravicini, *Solid State Physics* (Academic, London, 2000)
J. Callaway, *Quantum Theory of the Solid State*, 2nd edn. (Academic, New York, 1991)
P.Y. Yu, M. Cardona, *Fundamentals of Semiconductors* (Springer, Berlin, 1996)
A.A. Abrikosov, *Fundamentals of the Theory of Metals* (North-Holland, Amsterdam, 1988)
M. Balkanski, *Optical Properties of Semiconductors* (North-Holland, Amsterdam, 1994)

Chapter 9 Magnetism in Solids

D.C. Mattis, *The Theory of Magnetism* vols. I and II (Springer, Berlin, 1981)
R. Kubo, T. Nagamiya (eds.), *Solid Sate Physics* (McGraw-Hill, New York, 1969)
W. Jones, N.H. March, *Theoretical Solid State Physics* (Wiley-Interscience, New York, 1973)

Chapter 10 Magnetic Ordering and Spin Waves

C. Herring, *Direct exchange between well separated atoms* in *Magnetism*, vol. 2B, G. Rado, H. Suhl (eds.) (Academic, New York, 1965)
R.M. White, *Quantum Theory of Magnetism* (Springer, Berlin, 1995)
D.C. Mattis, *The Theory of Magnetism* vols. I and II (Springer, Berlin, 1981)
C. Kittel, *Quantum Theory of Solids*, 2nd edn. (Springer, Berlin, 1995)

Chapter 11 Many Body Interactions – Introduction

D. Pines, *Elementary Excitations in Solids*, 2nd edn. (Benjamin, New York, 1963)
K.S. Singwi, M.P. Tosi, *Correlations in Electron Liquids* in *Solid State Physics* **36**, 177 (1981), ed. by H. Ehrenreich, F. Seitz, D. Turnbull (Academic, New York, 1981)
D. Pines, P. Noizières, *The Theory of Quantum Liquids* (Perseus Books, Cambridge, 1999)
P.L. Taylor, O. Heinonen, *A Quantum Approach to Condensed Matter Physics* (Cambridge University Press, Cambridge, 2002)

Chapter 12 Many Body Interactions – Green's Function Method

A.A. Abrikosov, L.P. Gorkov, I.E. Dzyaloshinski, *Method of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, NJ, 1963)
G. Mahan, *Many-Particle Physics* (Plenum, New York, 2000)
P.L. Taylor, O. Heinonen, *A Quantum Approach to Condensed Matter Physics* (Cambridge University Press, Cambridge, 2002)

Chapter 13 Semiclassical Theory of Electrons

D. Shoenberg, *Phil. Roy. Soc.* (London), Ser **A 255**, 85 (1962)
J. Condon, *Phys. Rev.* **145**, 526 (1966)
J.J. Quinn, *Nature* **317**, 389 (1985)
M.H. Cohen, M.J. Harrison, W.A. Harrison, *Phys. Rev.* **117**, 937 (1960)
M.P. Greene, H.J. Lee, J.J. Quinn, S. Rodriguez, *Phys. Rev.* **177**, 1019 (1969)

Chapter 14 Electrodynamics of Metals

For a review of magnetoplasma surface waves see, for example, J.J. Quinn, K.W. Chiu, *Magnetoplasma Surface Waves in Metals and Semiconductors*, Polaritons, ed. by E. Burstein, F. DeMartini (Pergamon, New York, 1971), p. 259
P.M. Platzman, P.J. Wolff, *Waves and Interactions in Solid State Plasmas*, Solid State Physics-Supplement **13**, ed. by H. Ehrenreich, F. Seitz, D. Turnbull (Academic, New York, 1973)
E.D. Palik, B.G. Wright, *Free-Carrier Magnetooptic Effects*, Semiconductors and Semimetals **3**, ed. by R.K. Willardson, A.C. Beer (Academic, New York, 1967), p. 421
I. Bernstein, *Phys. Rev.* **109**, 10 (1958)

Chapter 15 Superconductivity

F. London, *Superfluids* vols. 1 and 2, (Wiley, New York, 1954)
D. Shoenberg, *Superconductivity*, (Cambridge University Press, Cambridge, 1962)
J.R. Schrieffer, *Superconductivity*, (W.A. Benjamin, New York, 1964)
A.A. Abrikosov, *Fundamentals of the Theory of Metals* (North-Holland, Amsterdam, 1988)
M. Tinkham, *Introduction to Superconductivity* 2nd ed. (McGraw-Hill, New York, 1996)
G. Rickayzen, *Theory of Superconductivity* (Wiley-Interscience, New York, 1965)
R.D. Parks, ed., *Superconductivity* vols. I and II (Dekker, New York, 1969)

Chapter 16 The Fractional Quantum Hall Effect

R.E. Prange, S.M. Girvin (eds.), *The Quantum Hall Effect*, 2nd edn. (Springer, Berlin, 1990)
T. Chakraborty, P. Pietiläinen, *The Quantum Hall Effects* Fractional and Integral, 2nd edn. (Springer, Berlin, 1995)
A. Shapere, F. Wilczek (eds.), *Geometric Phases in Physics* (World Scientific, Singapore, 1989)
F. Wilczek, *Fractional Statistics and Anyon Superconductivity* (World Scientific, Singapore, 1990)
Z.F. Ezawa, *Quantum Hall Effects*: Field Theoretical Approach and Related Topics (World Scientific, Singapore, 2000)
O. Heinonen, *Composite Fermions: A Unified View of the Quantum Hall Regime* (Singapore, World Scientific, 1998)

Index