<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Name</th>
<th>Symbol</th>
<th>Name</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actinium</td>
<td>Ac</td>
<td>Hafnium</td>
<td>Hf</td>
<td>Praseodymium</td>
<td>Pr</td>
</tr>
<tr>
<td>Aluminum</td>
<td>Al</td>
<td>Helium</td>
<td>He</td>
<td>Promethium</td>
<td>Pm</td>
</tr>
<tr>
<td>Americium</td>
<td>Am</td>
<td>Holmium</td>
<td>Ho</td>
<td>Protactinium</td>
<td>Pa</td>
</tr>
<tr>
<td>Antimony</td>
<td>Sb</td>
<td>Hydrogen</td>
<td>H</td>
<td>Radium</td>
<td>Ra</td>
</tr>
<tr>
<td>Argon</td>
<td>Ar</td>
<td>Indium</td>
<td>In</td>
<td>Radon</td>
<td>Rn</td>
</tr>
<tr>
<td>Arsenic</td>
<td>As</td>
<td>Iodine</td>
<td>I</td>
<td>Rhenium</td>
<td>Re</td>
</tr>
<tr>
<td>Astatine</td>
<td>At</td>
<td>Iridium</td>
<td>Ir</td>
<td>Rhodium</td>
<td>Rh</td>
</tr>
<tr>
<td>Barium</td>
<td>Ba</td>
<td>Iron</td>
<td>Fe</td>
<td>Rubidium</td>
<td>Rb</td>
</tr>
<tr>
<td>Berkelium</td>
<td>Bk</td>
<td>Krypton</td>
<td>Kr</td>
<td>Ruthenium</td>
<td>Ru</td>
</tr>
<tr>
<td>Beryllium</td>
<td>Be</td>
<td>Lanthanum</td>
<td>La</td>
<td>Samarium</td>
<td>Sm</td>
</tr>
<tr>
<td>Bismuth</td>
<td>Bi</td>
<td>Lawrencium</td>
<td>Lr</td>
<td>Scandium</td>
<td>Sc</td>
</tr>
<tr>
<td>Boron</td>
<td>B</td>
<td>Lead</td>
<td>Pb</td>
<td>Selenium</td>
<td>Se</td>
</tr>
<tr>
<td>Bromine</td>
<td>Br</td>
<td>Lithium</td>
<td>Li</td>
<td>Silicon</td>
<td>Si</td>
</tr>
<tr>
<td>Cadmium</td>
<td>Cd</td>
<td>Lutetium</td>
<td>Lu</td>
<td>Silver</td>
<td>Ag</td>
</tr>
<tr>
<td>Calcium</td>
<td>Ca</td>
<td>Magnesium</td>
<td>Mg</td>
<td>Sodium</td>
<td>Na</td>
</tr>
<tr>
<td>Californium</td>
<td>Cf</td>
<td>Manganese</td>
<td>Mn</td>
<td>Strontium</td>
<td>Sr</td>
</tr>
<tr>
<td>Carbon</td>
<td>C</td>
<td>Mendelevium</td>
<td>Md</td>
<td>Sulfur</td>
<td>S</td>
</tr>
<tr>
<td>Cerium</td>
<td>Ce</td>
<td>Mercury</td>
<td>Hg</td>
<td>Tantalum</td>
<td>Ta</td>
</tr>
<tr>
<td>Cesium</td>
<td>Cs</td>
<td>Molybdenum</td>
<td>Mo</td>
<td>Technetium</td>
<td>Tc</td>
</tr>
<tr>
<td>Chlorine</td>
<td>Cl</td>
<td>Neodymium</td>
<td>Nd</td>
<td>Tellurium</td>
<td>Te</td>
</tr>
<tr>
<td>Chromium</td>
<td>Cr</td>
<td>Neon</td>
<td>Ne</td>
<td>Terbium</td>
<td>Tb</td>
</tr>
<tr>
<td>Cobalt</td>
<td>Co</td>
<td>Neptunium</td>
<td>Np</td>
<td>Thallium</td>
<td>Tl</td>
</tr>
<tr>
<td>Copper</td>
<td>Cu</td>
<td>Nickel</td>
<td>Ni</td>
<td>Thorium</td>
<td>Th</td>
</tr>
<tr>
<td>Curium</td>
<td>Cm</td>
<td>Niobium</td>
<td>Nb</td>
<td>Thulium</td>
<td>Tm</td>
</tr>
<tr>
<td>Dysprosium</td>
<td>Dy</td>
<td>Nitrogen</td>
<td>N</td>
<td>Tin</td>
<td>Sn</td>
</tr>
<tr>
<td>Einsteinium</td>
<td>Es</td>
<td>Nobelium</td>
<td>No</td>
<td>Titanium</td>
<td>Ti</td>
</tr>
<tr>
<td>Erbium</td>
<td>Er</td>
<td>Osmium</td>
<td>Os</td>
<td>Tungsten</td>
<td>W</td>
</tr>
<tr>
<td>Europium</td>
<td>Eu</td>
<td>Oxygen</td>
<td>O</td>
<td>Uranium</td>
<td>U</td>
</tr>
<tr>
<td>Fermium</td>
<td>Fm</td>
<td>Palladium</td>
<td>Pd</td>
<td>Vanadium</td>
<td>V</td>
</tr>
<tr>
<td>Fluorine</td>
<td>F</td>
<td>Phosphorus</td>
<td>P</td>
<td>Xenon</td>
<td>Xe</td>
</tr>
<tr>
<td>Francium</td>
<td>Fr</td>
<td>Platinum</td>
<td>Pt</td>
<td>Ytterbium</td>
<td>Yb</td>
</tr>
<tr>
<td>Gadolinium</td>
<td>Gd</td>
<td>Plutonium</td>
<td>Pu</td>
<td>Yttrium</td>
<td>Y</td>
</tr>
<tr>
<td>Gallium</td>
<td>Ga</td>
<td>Polonium</td>
<td>Po</td>
<td>Zinc</td>
<td>Zn</td>
</tr>
<tr>
<td>Germanium</td>
<td>Ge</td>
<td>Potassium</td>
<td>K</td>
<td>Zirconium</td>
<td>Zr</td>
</tr>
<tr>
<td>Gold</td>
<td>Au</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The notation used to describe the electronic configuration of atoms and ions is discussed in all textbooks of introductory atomic physics. The letters \( s, p, d, \ldots \) signify electrons having orbital angular momentum 0, 1, 2, \ldots in units \( \hbar \); the number to the left of the letter denotes the principal quantum number of one orbit, and the superscript to the right denotes the number of electrons in the orbit.
CHARLES KITTEL

Introduction to Solid State Physics

SEVENTH EDITION

John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore
About the Author

Charles Kittel taught solid state physics at Berkeley from 1951 to 1978; earlier he was a member of the solid state group at the Bell Laboratories. His undergraduate work was at M.I.T. and at Cambridge University, followed by graduate work at the University of Wisconsin. He is a member of the National Academy of Science and of the American Academy of Arts and Sciences.

His research in solids began with studies of ferromagnetic, antiferromagnetic, and paramagnetic resonance, along with work on magnetic domains, spin waves, and domain boundaries in ferromagnets and ferroelectrics. His work on the single domain structure of fine particles has had broad application in magnetic recording, geomagnetism, and biomagnetism. Along with collaborators at Berkeley he did the first work on cyclotron resonance in semiconductors, which led to the understanding of the band structure of silicon, germanium, and indium antimonide, together with the theory of their impurity states. He also worked on the interpretation of magnetoplasma resonance in semiconductors and of Alfvén resonance in electron-hole drops in germanium.

The first edition of ISSP integrated the elementary aspects of solid state physics for study by seniors and beginning graduate students. Now in its seventh edition, ISSP plays the same part for the current generation of students.
Preface

This book is the seventh edition of an elementary text on solid state physics for senior and beginning graduate students of physical science and engineering. The book is an update of the sixth edition of 1986 and includes additions, improvements, and corrections made in that edition in 13 successive printings—which it was time to pull together—and a number of new topics besides. Significant advances in the field have been added or discussed more fully; thus high temperature superconductors are treated, and results of scanning tunneling microscopy are displayed; the treatment of fiber optics is expanded. There are discussions, among other topics, of nanostructures, superlattices, Bloch/Wannier levels, Zener tunneling, light-emitting diodes, and new magnetic materials. The additions have been made within a boundary condition intended to keep the text within one volume and at a reasonable price.

The theoretical level of the text itself has not been changed. There is more discussion of useful materials. The treatment of elastic constants and elastic waves which was dropped after the fourth edition has now been returned because, as many have pointed out, the matter is useful and not easily accessible elsewhere. The treatment of superconductors is much more extensive than is usual in a text at this level: either you do it or you don’t.

Solid state physics is concerned with the properties, often astonishing and often of great utility, that result from the distribution of electrons in metals, semiconductors, and insulators. The book also tells how the excitations and imperfections of real solids can be understood with simple models whose power and scope are now firmly established. The subject matter supports a profitable interplay of experiment, application, and theory. The book, in English and in many translations, has helped give several generations of students a picture of the process. Students also find the field attractive because of the frequent possibility of working in small groups.

Instructors will use the book as the foundation of a course in their own way, yet there are two general patterns to the introduction, selection and order of the basic material. If students have a significant preparation in elementary quantum mechanics, they will like to begin with the quantum theory of elec-
trons in one-dimensional solids, starting with the free electron gas in Chapter 6 and energy bands in Chapter 7. One will need to treat the reciprocal lattice in three dimensions (Chapter 2) before plunging into semiconductors (Chapter 8) and Fermi surfaces (Chapter 9). Crystal structures, crystal binding, and phonons could be considered as recreational reading. In a more gradual approach, the first eight chapters through the physics of semiconductors are read consecutively as a one-semester introduction to the field.

What about the necessary statistical mechanics? A vague discomfort at the thought of the chemical potential is still characteristic of a physics education. This intellectual gap is due to the obscurity of the writings of J. Willard Gibbs, who discovered and understood the matter 100 years ago. Herbert Kroemer and I have clarified the physics of the chemical potential in the early chapters of our book on thermal physics.

Review series give excellent extended treatments of all the subjects treated in this book and many more besides; thus with good conscience I give few references to original papers. In these omissions no lack of honor is intended to those who first set sail on these seas.

The crystallographic notation conforms with current usage in physics. Important equations are repeated in SI and CGS-Gaussian units, where these differ. Exceptions are figure captions, chapter summaries, some problems, and any long section of text where a single indicated substitution will translate from CGS to SI. Chapter Contents pages discuss conventions adopted to make parallel usage simple. The dual usage in this book has been found useful and acceptable.

Tables are in conventional units. The symbol $e$ denotes the charge on the proton and is positive. The notation (18) refers to Equation (18) of the current chapter, but (3.18) refers to Equation 18 of Chapter 3. A caret $\hat{\cdot}$ over a vector refers to a unit vector. Few of the problems are exactly easy; most were devised to carry forward the subject of the chapter. With a few exceptions, the problems are those of the original sixth edition.

Corrections and suggestions will be gratefully received and may be addressed to the author at the Department of Physics, University of California, Berkeley, CA 94720-7300; by email to kittel@uclink4.Berkeley.edu; and by fax to (510) 643-9473.

C. Kittel

An Instructor's Manual is available for this revision; several problems have been added (to Chapter 3 and Chapter 6); one dropped (from Chapter 4), and several corrections made. Instructors who have adopted the text for classroom use should direct a request on departmental letterhead to John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012. Limited requests for permission to copy figures or other material should be addressed to the Permissions Editor at this address.
# Contents

**Guide to Tables**  
ix

**General References**  
xi

1 CRYSTAL STRUCTURE  
2 RECIPROCAL LATTICE  
3 CRYSTAL BINDING AND ELASTIC CONSTANTS  
4 PHONONS I. CRYSTAL VIBRATIONS  
5 PHONONS II. THERMAL PROPERTIES  
6 FREE ELECTRON FERMI GAS  
7 ENERGY BANDS  
8 SEMICONDUCTOR CRYSTALS  
9 FERMI SURFACES AND METALS  
10 PLASMONS, POLARITONS, AND POLARONS  
11 OPTICAL PROCESSES AND EXCITONS  
12 SUPERCONDUCTIVITY  
13 DIELECTRICS AND FERROELECTRICS  
14 DIAMAGNETISM AND PARAMAGNETISM  
15 FERROMAGNETISM AND ANTFERROMAGNETISM  
16 MAGNETIC RESONANCE  
17 NONCRYSTALLINE SOLIDS  
18 POINT DEFECTS  
141  
173  
197  
233  
269  
305  
333  
379  
415  
441  
485  
519  
539
19 SURFACE AND INTERFACE PHYSICS 553
20 DISLOCATIONS 585
21 ALLOYS 609

Appendix

A TEMPERATURE DEPENDENCE OF THE REFLECTION LINES 631
B EWALD CALCULATION OF LATTICE SUMS 634
C QUANTIZATION OF ELASTIC WAVES: PHONONS 638
D FERMI-DIRAC DISTRIBUTION FUNCTION 643
E DERIVATION OF $dk/dt$ EQUATION 646
F BOLTZMANN TRANSPORT EQUATION 647
G VECTOR POTENTIAL, FIELD MOMENTUM, AND GAUGE TRANSFORMATIONS 651
H COOPER PAIRS 656
I Ginsburg-Landau Equation 658
J ELECTRON-PHONON COLLISIONS 662

Subject Index 667
Table of SI Prefixes 673
### Guide to Tables

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Fourteen lattice types in three dimensions</td>
<td>10</td>
</tr>
<tr>
<td>1.2</td>
<td>Characteristics of cubic lattices</td>
<td>12</td>
</tr>
<tr>
<td>1.3</td>
<td>Crystal structures of the elements</td>
<td>23</td>
</tr>
<tr>
<td>1.4</td>
<td>Density and atomic concentration of the elements</td>
<td>24</td>
</tr>
<tr>
<td>3.1</td>
<td>Cohesive energies of the elements</td>
<td>57</td>
</tr>
<tr>
<td>3.2</td>
<td>Melting temperatures of the elements</td>
<td>58</td>
</tr>
<tr>
<td>3.3</td>
<td>Bulk modulii and compressibilities of the elements</td>
<td>59</td>
</tr>
<tr>
<td>3.4</td>
<td>Properties of inert gas crystals</td>
<td>60</td>
</tr>
<tr>
<td>3.5</td>
<td>Ionization energies</td>
<td>61</td>
</tr>
<tr>
<td>3.6</td>
<td>Electron affinities of negative ions</td>
<td>68</td>
</tr>
<tr>
<td>3.7</td>
<td>Properties of alkali halide crystals</td>
<td>73</td>
</tr>
<tr>
<td>3.8</td>
<td>Fractional ionic character of bonds</td>
<td>76</td>
</tr>
<tr>
<td>3.9</td>
<td>Atomic and ionic radii</td>
<td>78</td>
</tr>
<tr>
<td>3.10</td>
<td>Standard radii of ions</td>
<td>79</td>
</tr>
<tr>
<td>3.11</td>
<td>Elastic stiffness constants of cubic crystals at 0 K and 300 K</td>
<td>91</td>
</tr>
<tr>
<td>3.12</td>
<td>Elastic stiffness constants of cubic crystals at 300 K</td>
<td>92</td>
</tr>
<tr>
<td>5.1</td>
<td>Debye temperature and thermal conductivity</td>
<td>126</td>
</tr>
<tr>
<td>5.2</td>
<td>Phonon mean free paths</td>
<td>133</td>
</tr>
<tr>
<td>6.1</td>
<td>Free electron Fermi surface parameters for metals</td>
<td>150</td>
</tr>
<tr>
<td>6.2</td>
<td>Electron heat capacity of metals</td>
<td>157</td>
</tr>
<tr>
<td>6.3</td>
<td>Electrical conductivity and resistivity of metals</td>
<td>160</td>
</tr>
<tr>
<td>6.4</td>
<td>Hall coefficients</td>
<td>167</td>
</tr>
<tr>
<td>6.5</td>
<td>Lorenz numbers</td>
<td>168</td>
</tr>
<tr>
<td>8.1</td>
<td>Energy gaps in semiconductors</td>
<td>201</td>
</tr>
<tr>
<td>8.2</td>
<td>Effective masses of holes and electrons</td>
<td>214</td>
</tr>
<tr>
<td>8.3</td>
<td>Carrier mobilities at room temperature</td>
<td>221</td>
</tr>
<tr>
<td>8.4</td>
<td>Static dielectric constants of semiconductors</td>
<td>223</td>
</tr>
<tr>
<td>8.5</td>
<td>Donor ionization energies</td>
<td>224</td>
</tr>
<tr>
<td>8.6</td>
<td>Acceptor ionization energies</td>
<td>225</td>
</tr>
<tr>
<td>8.7</td>
<td>Electron and hole concentrations in semimetals</td>
<td>229</td>
</tr>
</tbody>
</table>
10.1 Ultraviolet transmission limits of alkali metals
10.2 Volume plasmon energies
10.3 Lattice frequencies
10.4 Polaron masses and coupling constants

11.1 Exciton binding energies
11.2 Electron-hole liquid parameters
11.3 Acronyms of current experimental methods for band structure studies

12.1 Superconductivity parameters of the elements
12.2 Superconductivity of selected compounds
12.3 Energy gaps in superconductors
12.4 Isotope effect in superconductors
12.5 Coherence length and penetration depth

13.1 Electronic polarizabilities of ions
13.2 Ferroelectric crystals
13.3 Antiferroelectric crystals

14.1 Magneton numbers of lanthanide group ions
14.2 Magneton numbers of iron group ions

15.1 Critical point exponents of ferromagnets
15.2 Ferromagnetic crystals
15.3 Antiferromagnetic crystals

16.1 Nuclear magnetic resonance data
16.2 Knight shifts in NMR

18.1 Diffusion constants and activation energies
18.2 Activation energy for a positive ion vacancy
18.3 Experimental F center absorption energies

19.1 Top layer atomic relaxation
19.2 Electron work functions

20.1 Elastic limit and shear modulus
20.2 Dislocation densities

21.1 Electron/atom ratios of electron compounds
Selected General References

Statistical physics background
C. Kittel and H. Kroemer, Thermal physics, 2nd ed., Freeman, 1980. Has a full, clear discussion of the chemical potential and of semiconductor statistics; cited as TP.

Intermediate text

Advanced texts
C. Kittel, Quantum theory of solids, 2nd revised printing, Wiley, 1987, with solutions appendix by C. Y. Fong; cited as QTS.

Applied solid state

Review series
F. Seitz and others, Solid state physics, advances in research and applications, Vols. 1–(48), plus supplements. This valuable continuing series is often cataloged as a serial, as if it were a journal, and is cited here as Solid state physics.

Literature guides
There are many good databases, library and organizational; this is the way to go for monograph and journal searches.
1

Crystal Structure

PERIODIC ARRAYS OF ATOMS 3
Lattice translation vectors 4
Basis and the crystal structure 5
Primitive lattice cell 6

FUNDAMENTAL TYPES OF LATTICES 8
Two-dimensional lattice types 8
Three-dimensional lattice types 10

INDEX SYSTEM FOR CRYSTAL PLANES 12

SIMPLE CRYSTAL STRUCTURES 15
Sodium chloride structure 15
Cesium chloride structure 17
Hexagonal close-packed structure 17
Diamond structure 19
Cubic zinc sulfide structure 20

DIRECT IMAGING OF ATOMIC STRUCTURE 20

NONIDEAL CRYSTAL STRUCTURES 21
Random stacking and polytypism 22

CRYSTAL STRUCTURE DATA 22

SUMMARY 25

PROBLEMS 25

1. Tetrahedral angles 25
2. Indices of planes 25
3. Hcp structure 25

REFERENCES 25

UNITS: 1 Å = 1 angstrom = 10^{-8} cm = 0.1 nm = 10^{-10} m.
Figure 1  Relation of the external form of crystals to the form of the elementary building blocks. The building blocks are identical in (a) and (b), but different crystal faces are developed. (c) Cleaving a crystal of rocksalt.
CHAPTER 1: CRYSTAL STRUCTURE

Solid state physics is largely concerned with crystals and electrons in crystals. The study of solid state physics began in the early years of this century following the discovery of x-ray diffraction by crystals and the publication of a series of simple calculations and successful predictions of the properties of crystals.

When a crystal grows in a constant environment, the form develops as if identical building blocks were added continuously (Fig. 1). The building blocks are atoms or groups of atoms, so that a crystal is a three-dimensional periodic array of atoms.

This was known in the 18th century when mineralogists discovered that the index numbers of the directions of all faces of a crystal are exact integers. Only the arrangement of identical particles in a periodic array can account for the law of integral indices,¹ as discussed below.

In 1912 a paper entitled “Interference effects with Röntgen rays” was presented to the Bavarian Academy of Sciences in Munich. In the first part of the paper, Laue developed an elementary theory of the diffraction of x-rays by a periodic array. In the second part, Friedrich and Knipping reported the first experimental observations of x-ray diffraction by crystals.²

The work proved decisively that crystals are composed of a periodic array of atoms. With an established atomic model of a crystal, physicists now could think much further. The studies have been extended to include amorphous or noncrystalline solids, glasses, and liquids. The wider field is known as condensed matter physics, and it is now the largest and probably the most vigorous area of physics.

PERIODIC ARRAYS OF ATOMS

An ideal crystal is constructed by the infinite repetition of identical structural units in space. In the simplest crystals the structural unit is a single atom, as in copper, silver, gold, iron, aluminum, and the alkali metals. But the smallest structural unit may comprise many atoms or molecules.

The structure of all crystals can be described in terms of a lattice, with a group of atoms attached to every lattice point. The group of atoms is called the basis; when repeated in space it forms the crystal structure.

¹R. J. Haüy, Essai d’une théorie sur la structure des cristaux, Paris, 1784; Traité de cristallographie, Paris, 1801.
Lattice Translation Vectors

The lattice is defined by three fundamental translation vectors \( \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \) such that the atomic arrangement looks the same in every respect when viewed from the point \( r \) as when viewed from the point

\[
\mathbf{r}' = \mathbf{r} + u_1 \mathbf{a}_1 + u_2 \mathbf{a}_2 + u_3 \mathbf{a}_3 ,
\]

where \( u_1, u_2, u_3 \) are arbitrary integers. The set of points \( \mathbf{r}' \) defined by (1) for all \( u_1, u_2, u_3 \) defines a lattice.

A lattice is a regular periodic array of points in space. (The analog in two dimensions is called a net, as in Chapter 18.) A lattice is a mathematical abstraction; the crystal structure is formed when a basis of atoms is attached identically to every lattice point. The logical relation is

\[
\text{lattice} + \text{basis} = \text{crystal structure} .
\]

The lattice and the translation vectors \( \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \) are said to be primitive if any two points \( \mathbf{r}, \mathbf{r}' \) from which the atomic arrangement looks the same always satisfy (1) with a suitable choice of the integers \( u_1, u_2, u_3 \). With this definition of the primitive translation vectors, there is no cell of smaller volume that can serve as a building block for the crystal structure.

We often use primitive translation vectors to define the crystal axes. However, nonprimitive crystal axes are often used when they have a simpler relation to the symmetry of the structure. The crystal axes \( \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \) form three adjacent edges of a parallelepiped. If there are lattice points only at the corners, then it is a primitive parallelepiped.

A lattice translation operation is defined as the displacement of a crystal by a crystal translation vector

\[
\mathbf{T} = u_1 \mathbf{a}_1 + u_2 \mathbf{a}_2 + u_3 \mathbf{a}_3 .
\]

Any two lattice points are connected by a vector of this form.

To describe a crystal structure, there are three important questions to answer: What is the lattice? What choice of \( \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \) do we wish to make? What is the basis?

More than one lattice is always possible for a given structure, and more than one set of axes is always possible for a given lattice. The basis is identified once these choices have been made. Everything (including the x-ray diffraction pattern) works out correctly in the end provided that (3) has been satisfied.

The symmetry operations of a crystal carry the crystal structure into itself. These include the lattice translation operations. Further, there are rotation and reflection operations, called point operations. About lattice points or certain special points within an elementary parallelepiped it may be possible to apply rotations and reflections that carry the crystal into itself.

Finally, there may exist compound operations made up of combined translation and point operations. Textbooks on crystallography are largely devoted to
Figure 2  Portion of a crystal of an imaginary protein molecule, in a two-dimensional world. (We picked a protein molecule because it is not likely to have a special symmetry of its own.) The atomic arrangement in the crystal looks exactly the same to an observer at \( r' \) as to an observer at \( r \), provided that the vector \( T \) which connects \( r' \) and \( r \) may be expressed as an integral multiple of the vectors \( a_1 \) and \( a_2 \). In this illustration, \( T = -a_1 + 3a_2 \). The vectors \( a_1 \) and \( a_2 \) are primitive translation vectors of the two-dimensional lattice.

Figure 3  Similar to Fig. 2, but with protein molecules associated in pairs. The crystal translation vectors are \( a_1 \) and \( a_2 \). A rotation of \( \pi \) radians about any point marked \( \times \) will carry the crystal into itself. This occurs also for equivalent points in other cells, but we have marked the points \( \times \) only within one cell.

The description of symmetry operations. The crystal structure of Fig. 2 is drawn to have only translational symmetry operations. The crystal structure of Fig. 3 allows both translational and point symmetry operations.

Basis and the Crystal Structure

A basis of atoms is attached to every lattice point, with every basis identical in composition, arrangement, and orientation. Figure 4 shows how a crystal structure is formed by adding a basis to every lattice point. The lattice is indicated by dots in Figs. 2 and 3, but in Fig. 4c the dots are omitted.
The number of atoms in the basis may be one, or it may be more than one. The position of the center of an atom $j$ of the basis relative to the associated lattice point is

$$r_j = x_j a_1 + y_j a_2 + z_j a_3 .$$

We may arrange the origin, which we have called the associated lattice point, so that $0 \leq x_j, y_j, z_j \leq 1$.

**Primitive Lattice Cell**

The parallelepiped defined by primitive axes $a_1$, $a_2$, $a_3$ is called a **primitive cell** (Fig. 5b). A primitive cell is a type of cell or unit cell. (The adjective unit is superfluous and not needed.) A cell will fill all space by the repetition of suitable crystal translation operations. A primitive cell is a minimum-volume cell.

There are many ways of choosing the primitive axes and primitive cell for a given lattice. The number of atoms in a primitive cell or primitive basis is always the same for a given crystal structure.
Figure 5a  Lattice points of a space lattice in two dimensions. All pairs of vectors $a_1$, $a_2$ are translation vectors of the lattice. But $a_1''$, $a_2'''$ are not primitive translation vectors because we cannot form the lattice translation $T$ from integral combinations of $a_1''$ and $a_2'''$. All other pairs shown of $a_1$ and $a_2$ may be taken as the primitive translation vectors of the lattice. The parallelograms 1, 2, 3 are equal in area and any of them could be taken as the primitive cell. The parallelogram 4 has twice the area of a primitive cell.

Figure 5b  Primitive cell of a space lattice in three dimensions.

Figure 5c  Suppose these points are identical atoms; sketch in on the figure a set of lattice points, a choice of primitive axes, a primitive cell, and the basis of atoms associated with a lattice point.

There is always one lattice point per primitive cell. If the primitive cell is a parallelepiped with lattice points at each of the eight corners, each lattice point is shared among eight cells, so that the total number of lattice points in the cell is one: $8 \times \frac{1}{8} = 1$.

The volume of a parallelepiped with axes $a_1$, $a_2$, $a_3$ is

$$ V_c = |a_1 \cdot (a_2 \times a_3) |, $$

by elementary vector analysis. The basis associated with a primitive cell is called a primitive basis. No basis contains fewer atoms than a primitive basis contains.
Figure 6 A primitive cell may also be chosen following this procedure: (1) draw lines to connect a given lattice point to all nearby lattice points; (2) at the midpoint and normal to these lines, draw new lines or planes. The smallest volume enclosed in this way is the Wigner-Seitz primitive cell. All space may be filled by these cells, just as by the cells of Fig. 5.

Another way of choosing a primitive cell is shown in Fig. 6. This is known to physicists as a Wigner-Seitz cell.

**FUNDAMENTAL TYPES OF LATTICES**

Crystal lattices can be carried or mapped into themselves by the lattice translations \( T \) and by various other symmetry operations. A typical symmetry operation is that of rotation about an axis that passes through a lattice point. Lattices can be found such that one-, two-, three-, four-, and sixfold rotation axes carry the lattice into itself, corresponding to rotations by \( 2\pi, 2\pi/2, 2\pi/3, 2\pi/4, \) and \( 2\pi/6 \) radians and by integral multiples of these rotations. The rotation axes are denoted by the symbols 1, 2, 3, 4, and 6.

We cannot find a lattice that goes into itself under other rotations, such as by \( 2\pi/7 \) radians or \( 2\pi/5 \) radians. A single molecule properly designed can have any degree of rotational symmetry, but an infinite periodic lattice cannot. We can make a crystal from molecules that individually have a fivefold rotation axis, but we should not expect the lattice to have a fivefold rotation axis. In Fig. 7 we show what happens if we try to construct a periodic lattice having fivefold symmetry: the pentagons do not fit together to fill all space, showing that we cannot combine fivefold point symmetry with the required translational periodicity.

By lattice point group we mean the collection of symmetry operations which, applied about a lattice point, carry the lattice into itself. The possible rotations have been listed. We can have mirror reflections \( m \) about a plane through a lattice point. The inversion operation is composed of a rotation of \( \pi \) followed by reflection in a plane normal to the rotation axis; the total effect is to replace \( r \) by \( -r \). The symmetry axes and symmetry planes of a cube are shown in Fig. 8.

**Two-Dimensional Lattice Types**

There is an unlimited number of possible lattices because there is no natural restriction on the lengths of the lattice translation vectors or on the angle \( \varphi \) between them. The lattice in Fig. 5a was drawn for arbitrary \( a_1 \) and \( a_2 \). A general lattice such as this is known as an oblique lattice and is invariant only under rotation of \( \pi \) and \( 2\pi \) about any lattice point.
Figure 7 A fivefold axis of symmetry cannot exist in a periodic lattice because it is not possible to fill the area of a plane with a connected array of pentagons. We can, however, fill all the area of a plane with just two distinct designs of "tiles" or elementary polygons. A quasicrystal is a quasiperiodic nonrandom assembly of two types of figures. Quasicrystals are discussed at the end of Chapter 2.

Figure 8 (a) A plane of symmetry parallel to the faces of a cube. (b) A diagonal plane of symmetry in a cube. (c) The three tetrad axes of a cube. (d) The four triad axes of a cube. (e) The six diad axes of a cube.
But special lattices of the oblique type can be invariant under rotation of \(2\pi/3\), \(2\pi/4\), or \(2\pi/6\), or under mirror reflection. We must impose restrictive conditions on \(a_1\) and \(a_2\) if we want to construct a lattice that will be invariant under one or more of these new operations. There are four distinct types of restriction, and each leads to what we may call a special lattice type. Thus there are five distinct lattice types in two dimensions, the oblique lattice and the four special lattices shown in Fig. 9. Bravais lattice is the common phrase for a distinct lattice type; we say that there are five Bravais lattices or nets in two dimensions.

**Three-Dimensional Lattice Types**

The point symmetry groups in three dimensions require the 14 different lattice types listed in Table 1. The general lattice is triclinic, and there are 13 special lattices. These are grouped for convenience into systems classified according to seven types of cells, which are triclinic, monoclinic, orthorhombic, tetragonal, cubic, trigonal, and hexagonal. The division into systems is expressed in the table in terms of the axial relations that describe the cells.

The cells in Fig. 10 are conventional cells; of these only the sc is a primitive cell. Often a nonprimitive cell has a more obvious relation with the point symmetry operations than has a primitive cell.

<table>
<thead>
<tr>
<th>System</th>
<th>Number of lattices</th>
<th>Restrictions on conventional cell axes and angles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triclinic</td>
<td>1</td>
<td>(a_1 \neq a_2 \neq a_3) (\alpha \neq \beta \neq \gamma)</td>
</tr>
<tr>
<td>Monoclinic</td>
<td>2</td>
<td>(a_1 \neq a_2 \neq a_3) (\alpha = \gamma = 90^\circ \neq \beta)</td>
</tr>
<tr>
<td>Orthorhombic</td>
<td>4</td>
<td>(a_1 \neq a_2 \neq a_3) (\alpha = \beta = \gamma = 90^\circ)</td>
</tr>
<tr>
<td>Tetragonal</td>
<td>2</td>
<td>(a_1 = a_2 \neq a_3) (\alpha = \beta = \gamma = 90^\circ)</td>
</tr>
<tr>
<td>Cubic</td>
<td>3</td>
<td>(a_1 = a_2 = a_3) (\alpha = \beta = \gamma = 90^\circ)</td>
</tr>
<tr>
<td>Trigonal</td>
<td>1</td>
<td>(a_1 = a_2 = a_3) (\alpha = \beta = \gamma &lt; 120^\circ, \neq 90^\circ)</td>
</tr>
<tr>
<td>Hexagonal</td>
<td>1</td>
<td>(a_1 = a_2 \neq a_3) (\alpha = \beta = 90^\circ) (\gamma = 120^\circ)</td>
</tr>
</tbody>
</table>
Figure 9

(a) Square lattice
\[ |a_1| = |a_2|; \varphi = 90^\circ \]

(b) Hexagonal lattice
\[ |a_1| = |a_2|; \varphi = 120^\circ \]

(c) Rectangular lattice
\[ |a_1| \neq |a_2|; \varphi = 90^\circ \]

(d) Centered rectangular lattice;
axes are shown for both the primitive cell and for the rectangular unit cell, for which \[ |a_1| \neq |a_2|; \varphi = 90^\circ \].

Figure 10

The cubic space lattices. The cells shown are the conventional cells.
### Table 2 Characteristics of cubic lattices

<table>
<thead>
<tr>
<th></th>
<th>Simple</th>
<th>Body-centered</th>
<th>Face-centered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume, conventional cell</td>
<td>$a^3$</td>
<td>$a^3$</td>
<td>$a^3$</td>
</tr>
<tr>
<td>Lattice points per cell</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Volume, primitive cell</td>
<td>$a^3$</td>
<td>$\frac{1}{2} a^3$</td>
<td>$\frac{1}{4} a^3$</td>
</tr>
<tr>
<td>Lattice points per unit volume</td>
<td>$1/a^3$</td>
<td>$2/a^3$</td>
<td>$4/a^3$</td>
</tr>
<tr>
<td>Number of nearest neighbors$^a$</td>
<td>6</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Nearest-neighbor distance</td>
<td>$a$</td>
<td>$3^{1/2} a/2 = 0.866a$</td>
<td>$a/2^{1/2} = 0.707a$</td>
</tr>
<tr>
<td>Number of second neighbors</td>
<td>12</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Second neighbor distance</td>
<td>$2^{1/2} a$</td>
<td>$a$</td>
<td>$a$</td>
</tr>
<tr>
<td>Packing fraction$^b$</td>
<td>$\frac{1}{6} \pi$</td>
<td>$\frac{1}{4} \pi \sqrt{3}$</td>
<td>$\frac{1}{6} \pi \sqrt{2}$</td>
</tr>
<tr>
<td></td>
<td>= 0.524</td>
<td>= 0.680</td>
<td>= 0.740</td>
</tr>
</tbody>
</table>


$^b$The packing fraction is the maximum proportion of the available volume that can be filled with hard spheres.

There are three lattices in the cubic system: the simple cubic (sc) lattice, the body-centered cubic (bcc) lattice, and the face-centered cubic (fcc) lattice. The characteristics of the three cubic lattices are summarized in Table 2.

A primitive cell of the bcc lattice is shown in Fig. 11, and the primitive translation vectors are shown in Fig. 12. The primitive translation vectors of the fcc lattice are shown in Fig. 13. Primitive cells by definition contain only one lattice point, but the conventional bcc cell contains two lattice points, and the fcc cell contains four lattice points.

The position of a point in a cell is specified by (4) in terms of the atomic coordinates $x$, $y$, $z$. Here each coordinate is a fraction of the axial length $a_1$, $a_2$, $a_3$ in the direction of the coordinate axis, with the origin taken at one corner of the cell. Thus the coordinates of the body center of a cell are $\frac{1}{2} \frac{1}{2} \frac{1}{2}$, and the face centers include $\frac{1}{2} 0 \frac{1}{2}$, $0 \frac{1}{2} \frac{1}{2}$.

In the hexagonal system the primitive cell is a right prism based on a rhombus with an included angle of $120^\circ$. Figure 14 shows the relationship of the rhombic cell to a hexagonal prism.

### INDEX SYSTEM FOR CRYSTAL PLANES

The orientation of a crystal plane is determined by three points in the plane, provided they are not collinear. If each point lay on a different crystal axis, the plane could be specified by giving the coordinates of the points in terms of the lattice constants $a_1$, $a_2$, $a_3$. 
Figure 11 Body-centered cubic lattice, showing a primitive cell. The primitive cell shown is a rhombohedron of edge $\frac{1}{2} \sqrt{3} a$, and the angle between adjacent edges is $109^\circ28'$.

Figure 12 Primitive translation vectors of the body-centered cubic lattice; these vectors connect the lattice point at the origin to lattice points at the body centers. The primitive cell is obtained on completing the rhombohedron. In terms of the cube edge $a$ the primitive translation vectors are:

$$a_1 = \frac{1}{2}a(\hat{x} + \hat{y} - \hat{z}) ; \quad a_2 = \frac{1}{2}a(-\hat{x} + \hat{y} + \hat{z}) ;$$
$$a_3 = \frac{1}{2}a(\hat{x} - \hat{y} + \hat{z}) .$$

Figure 13 The rhombohedral primitive cell of the face-centered cubic crystal. The primitive translation vectors $a_1, a_2, a_3$ connect the lattice point at the origin with lattice points at the face centers. As drawn, the primitive vectors are:

$$a_1 = \frac{1}{4}a(\hat{x} + \hat{y}) ; \quad a_2 = \frac{1}{4}a(\hat{y} + \hat{z}) ; \quad a_3 = \frac{1}{4}a(\hat{z} + \hat{x}) .$$

The angles between the axes are $60^\circ$. Here $\hat{x}, \hat{y}, \hat{z}$ are the Cartesian unit vectors.

Figure 14 Relation of the primitive cell in the hexagonal system (heavy lines) to a prism of hexagonal symmetry. Here:

$$a_1 = a_2 \neq a_3.$$
Figure 15  This plane intercepts the $a_1$, $a_2$, $a_3$ axes at $3a_1$, $2a_2$, $2a_3$. The reciprocals of these numbers are $\frac{1}{3}$, $\frac{1}{2}$, $\frac{1}{2}$. The smallest three integers having the same ratio are 2, 3, 3, and thus the indices of the plane are (233).

However, it turns out to be more useful for structure analysis to specify the orientation of a plane by the indices determined by the following rules (Fig. 15).

- Find the intercepts on the axes in terms of the lattice constants $a_1$, $a_2$, $a_3$. The axes may be those of a primitive or nonprimitive cell.
- Take the reciprocals of these numbers and then reduce to three integers having the same ratio, usually the smallest three integers. The result, enclosed in parentheses ($hkl$), is called the index of the plane.

For the plane whose intercepts are 4, 1, 2, the reciprocals are $\frac{1}{4}$, 1, and $\frac{1}{2}$; the smallest three integers having the same ratio are (142). For an intercept at infinity, the corresponding index is zero. The indices of some important planes in a cubic crystal are illustrated by Fig. 16.

The indices ($hkl$) may denote a single plane or a set of parallel planes. If a plane cuts an axis on the negative side of the origin, the corresponding index is negative, indicated by placing a minus sign above the index: ($hkl$). The cube faces of a cubic crystal are (100), (010), (001), (100), (010), and (001). Planes equivalent by symmetry may be denoted by curly brackets (braces) around indices; the set of cube faces is {100}. When we speak of the (200) plane we mean a plane parallel to (100) but cutting the $a_1$ axis at $\frac{1}{2}a$.

The indices $[uvw]$ of a direction in a crystal are the set of the smallest integers that have the ratio of the components of a vector in the desired direction, referred to the axes. The $a_1$ axis is the [100] direction; the $-a_2$ axis is the
[010] direction. In cubic crystals the direction \([hkl]\) is perpendicular to a plane \((hkl)\) having the same indices, but this is not generally true in other crystal systems.

**SIMPLE CRYSTAL STRUCTURES**

We discuss simple crystal structures of general interest: the sodium chloride, cesium chloride, hexagonal close-packed, diamond, and cubic zinc sulfide structures.

**Sodium Chloride Structure**

The sodium chloride, NaCl, structure is shown in Figs. 17 and 18. The lattice is face-centered cubic; the basis consists of one Na atom and one Cl atom separated by one-half the body diagonal of a unit cube. There are four units of NaCl in each unit cube, with atoms in the positions

<table>
<thead>
<tr>
<th>Cl:</th>
<th>000 ;</th>
<th>110 ;</th>
<th>101 ;</th>
<th>011 ;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na:</td>
<td>111 ;</td>
<td>001 ;</td>
<td>010 ;</td>
<td>100 ;</td>
</tr>
</tbody>
</table>
Figure 17  We may construct the sodium chloride crystal structure by arranging $\text{Na}^+$ and $\text{Cl}^-$ ions alternately at the lattice points of a simple cubic lattice. In the crystal each ion is surrounded by six nearest neighbors of the opposite charge. The space lattice is fcc, and the basis has one $\text{Cl}^-$ ion at 000 and one $\text{Na}^+$ ion at $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}$. The figure shows one conventional cubic cell. The ionic diameters here are reduced in relation to the cell in order to clarify the spatial arrangement.

Figure 18  Model of sodium chloride. The sodium ions are smaller than the chlorine ions. (Courtesy of A. N. Holden and P. Singer.)

Figure 19  Natural crystals of lead sulfide, PbS, which has the NaCl crystal structure. (Photograph by B. Burleson.)

Figure 20  The cesium chloride crystal structure. The space lattice is simple cubic, and the basis has one Cs$^+$ ion at 000 and one Cl$^-$ ion at $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}$. 
Each atom has as nearest neighbors six atoms of the opposite kind. Representative crystals having the NaCl arrangement include those in the following table. The cube edge $a$ is given in angstroms; $1 \text{Å} = 10^{-8} \text{cm} = 10^{-10} \text{m} = 0.1 \text{nm}$.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>$a$</th>
<th>Crystal</th>
<th>$a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiH</td>
<td>4.08 Å</td>
<td>AgBr</td>
<td>5.77 Å</td>
</tr>
<tr>
<td>MgO</td>
<td>4.20</td>
<td>PbS</td>
<td>5.92</td>
</tr>
<tr>
<td>MnO</td>
<td>4.43</td>
<td>KCl</td>
<td>6.29</td>
</tr>
<tr>
<td>NaCl</td>
<td>5.63</td>
<td>KBr</td>
<td>6.59</td>
</tr>
</tbody>
</table>

Figure 19 is a photograph of crystals of lead sulfide (PbS) from Joplin, Missouri. The Joplin specimens form in beautiful cubes.

**Cesium Chloride Structure**

The cesium chloride structure is shown in Fig. 20. There is one molecule per primitive cell, with atoms at the corners 000 and body-centered positions $\frac{1}{8}$ of the simple cubic space lattice. Each atom may be viewed as at the center of a cube of atoms of the opposite kind, so that the number of nearest neighbors or coordination number is eight.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>$a$</th>
<th>Crystal</th>
<th>$a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BeCu</td>
<td>2.70 Å</td>
<td>LiHg</td>
<td>3.29 Å</td>
</tr>
<tr>
<td>AlNi</td>
<td>2.88</td>
<td>NH₄Cl</td>
<td>3.87</td>
</tr>
<tr>
<td>CuZn ($\beta$-brass)</td>
<td>2.94</td>
<td>TlBr</td>
<td>3.97</td>
</tr>
<tr>
<td>CuPd</td>
<td>2.99</td>
<td>CsCl</td>
<td>4.11</td>
</tr>
<tr>
<td>AgMg</td>
<td>3.28</td>
<td>TlI</td>
<td>4.20</td>
</tr>
</tbody>
</table>

**Hexagonal Close-packed Structure (hcp)**

There are an infinite number of ways of arranging identical spheres in a regular array that maximizes the packing fraction (Fig. 21). One is the face-centered cubic structure; another is the hexagonal close-packed structure (Fig. 22). The fraction of the total volume occupied by the spheres is 0.74 for both structures. No structure, regular or not, has denser packing.

Spheres are arranged in a single closest-packed layer $A$ by placing each sphere in contact with six others. This layer may serve as either the basal plane of an hcp structure or the (111) plane of the fcc structure. A second similar layer $B$ may be added by placing each sphere of $B$ in contact with three spheres of the bottom layer, as in Fig. 21. A third layer $C$ may be added in two ways. We obtain the fcc structure if the spheres of the third layer are added over the holes in the first layer that are not occupied by $B$. We obtain the hcp structure when the spheres in the third layer are placed directly over the centers of the spheres in the first layer.
Figure 21 A close-packed layer of spheres is shown, with centers at points marked A. A second and identical layer of spheres can be placed on top of this, above and parallel to the plane of the drawing, with centers over the points marked B. There are two choices for a third layer. It can go in over A or over C. If it goes in over A the sequence is ABABAB... and the structure is hexagonal close-packed. If the third layer goes in over C the sequence is ABCABCABC... and the structure is face-centered cubic.

Figure 22 The hexagonal close-packed structure. The atom positions in this structure do not constitute a space lattice. The space lattice is simple hexagonal with a basis of two identical atoms associated with each lattice point. The lattice parameters $a$ and $c$ are indicated, where $a$ is in the basal plane and $c$ is the magnitude of the axis $a_3$ of Fig. 14.

Figure 23 The primitive cell has $a_1 = a_2$, with an included angle of 120°. The $c$ axis (or $a_3$) is normal to the plane of $a_1$ and $a_2$. The ideal hcp structure has $c = 1.633\ a$. The two atoms of one basis are shown as solid circles. One atom of the basis is at the origin; the other atom is at $\frac{3}{2}a_1 + \frac{1}{2}a_2 + \frac{1}{2}a_3$.

The hcp structure has the primitive cell of the hexagonal lattice, but with a basis of two atoms (Fig. 23). The fcc primitive cell has a basis of one atom (Fig. 13).

The ratio $c/a$ (or $a_3/a_1$) for hexagonal closest-packing of spheres has the value $(\frac{\sqrt{2}}{2})^{1/2} = 1.633$, as in Problem 3. It is usual to refer to crystals as hcp even if the actual $c/a$ ratio departs somewhat from this theoretical value.

The number of nearest-neighbor atoms is 12 for both hcp and fcc structures. If the binding energy (or free energy) depends only on the number of
Figure 24  Atomic positions in the cubic cell of the diamond structure projected on a cube face; fractions denote height above the base in units of a cube edge. The points at 0 and 1\(\frac{1}{2}\) are on the fcc lattice; those at 1\(\frac{1}{4}\) and 1\(\frac{3}{4}\) are on a similar lattice displaced along the body diagonal by one-fourth of its length. With a fcc space lattice, the basis consists of two identical atoms at 000; 111.

nearest-neighbor bonds per atom, there would be no difference in energy between the fcc and hcp structures.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>c/a</th>
<th>Crystal</th>
<th>c/a</th>
<th>Crystal</th>
<th>c/a</th>
</tr>
</thead>
<tbody>
<tr>
<td>He</td>
<td>1.633</td>
<td>Zn</td>
<td>1.861</td>
<td>Zr</td>
<td>1.594</td>
</tr>
<tr>
<td>Be</td>
<td>1.581</td>
<td>Cd</td>
<td>1.886</td>
<td>Gd</td>
<td>1.592</td>
</tr>
<tr>
<td>Mg</td>
<td>1.623</td>
<td>Co</td>
<td>1.622</td>
<td>Lu</td>
<td>1.586</td>
</tr>
<tr>
<td>Ti</td>
<td>1.586</td>
<td>Y</td>
<td>1.570</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Diamond Structure

The space lattice of diamond is fcc. The primitive basis has two identical atoms at 000; 111 associated with each point of the fcc lattice, as in Fig. 24. Thus the conventional unit cube contains eight atoms. There is no way to choose the primitive cell such that the basis of diamond contains only one atom.

The tetrahedral bonding characteristic of the diamond structure is shown in Fig. 25. Each atom has 4 nearest neighbors and 12 next nearest neighbors. The diamond structure is relatively empty: the maximum proportion of the available volume which may be filled by hard spheres is only 0.34, which is 46 percent of the filling factor for a closest-packed structure such as fcc or hcp. The
diamond structure is an example of the directional covalent bonding found in column IV of the periodic table of elements.

Carbon, silicon, germanium, and tin can crystallize in the diamond structure, with lattice constants \( a = 3.56, 5.43, 5.65, \) and \( 6.46 \) Å, respectively. Here \( a \) is the edge of the conventional cubic cell.

**Cubic Zinc Sulfide Structure**

The diamond structure may be viewed as two fcc structures displaced from each other by one-quarter of a body diagonal. The cubic zinc sulfide (zinc blende) structure results when Zn atoms are placed on one fcc lattice and S atoms on the other fcc lattice, as in Fig. 26. The conventional cell is a cube. The coordinates of the Zn atoms are \( 000; 0\frac{1}{2}; \frac{1}{2}0; \frac{1}{2}\frac{1}{2} \); the coordinates of the S atoms are \( \frac{1}{4}\frac{1}{4}; \frac{1}{4}\frac{1}{4}; \frac{1}{4}\frac{1}{4}; \frac{1}{4}\frac{1}{4} \). The lattice is fcc. There are four molecules of ZnS per conventional cell. About each atom there are four equally distant atoms of the opposite kind arranged at the corners of a regular tetrahedron.

The diamond structure allows a center-of-inversion symmetry operation at the midpoint of every line between nearest-neighbor atoms. The inversion operation carries an atom at \( r \) into an atom at \( -r \). The cubic ZnS structure does not have inversion symmetry. Examples of the cubic zinc sulfide structure are

<table>
<thead>
<tr>
<th>Crystal</th>
<th>( a )</th>
<th>Crystal</th>
<th>( a )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CuF</td>
<td>4.26 Å</td>
<td>ZnSe</td>
<td>5.65 Å</td>
</tr>
<tr>
<td>SiC</td>
<td>4.35</td>
<td>GaAs</td>
<td>5.65</td>
</tr>
<tr>
<td>CuCl</td>
<td>5.41</td>
<td>AlAs</td>
<td>5.66</td>
</tr>
<tr>
<td>ZnS</td>
<td>5.41</td>
<td>CdS</td>
<td>5.82</td>
</tr>
<tr>
<td>AlP</td>
<td>5.45</td>
<td>InSb</td>
<td>6.46</td>
</tr>
<tr>
<td>GaP</td>
<td>5.45</td>
<td>AgI</td>
<td>6.47</td>
</tr>
</tbody>
</table>

The close equality of several pairs, notably (Al,Ga)P and (Al,Ga)As, makes possible the construction of semiconductor heterojunctions (Chapter 19).

**DIRECT IMAGING OF ATOMIC STRUCTURE**

Direct images of crystal structure have been produced by transmission electron microscopy. Perhaps the most beautiful images are produced by scanning tunneling microscopy; in STM (Chapter 19) one exploits the large variations in quantum tunneling as a function of the height of a fine metal tip above the surface of a crystal. The image of Figure 27 was produced in this way; see also Figures 12.19 and 19.21. An STM method has been developed that will assemble single atoms into an organized layer nanometer structure on a crystal substrate: see the electron corral in Figure 19.21.
NONIDEAL CRYSTAL STRUCTURES

The ideal crystal of classical crystallographers is formed by the periodic repetition of identical units in space. But no general proof has been given that the ideal crystal is the state of minimum energy of identical atoms at absolute zero. At finite temperatures this is not likely to be true—see the discussion of lattice defects in Chapter 18. Further, it is not always possible for a structure to attain the equilibrium state in a reasonable time—see the discussion of glasses in Chapter 17. Many structures that occur in nature are not entirely periodic; see the quasicrystals treated at the end of Chapter 2. We give some examples here that supplement those in the chapters just cited.
Random Stacking and Polytypism

The fcc and hcp structures are made up of close-packed planes of atoms. The structures differ in the stacking sequence of the planes, fcc having the sequence ABCABC . . . and hcp having the sequence ABABAB . . . . Structures are known in which the stacking sequence of close-packed planes is random. This is known as random stacking and may be thought of as crystalline in two dimensions and noncrystalline or glasslike in the third.

Polytypism is characterized by a stacking sequence with a long repeat unit along the stacking axis. The best known example is zinc sulfide, ZnS, in which more than 150 polytypes have been identified, with the longest periodicity being 360 layers. Another example is silicon carbide, SiC, which occurs with more than 45 stacking sequences of the close-packed layers. The polytype of SiC known as 393R has a primitive cell with \( a = 3.079 \, \text{Å} \) and \( c = 989.6 \, \text{Å} \). The longest primitive cell observed for SiC has a repeat distance of 594 layers. A given sequence is repeated many times within a single crystal. The mechanism that induces such long-range crystallographic order is not a long-range force as such, but is associated with the presence of spiral steps due to dislocations in the growth nucleus (Chapter 20).

CRYSTAL STRUCTURE DATA

In Table 3 we list the more common crystal structures and lattice structures of the elements. Values of the atomic concentration and the density are given in Table 4.

Many elements occur in several crystal structures and transform from one to the other as the temperature or pressure is varied. Sometimes two structures coexist at the same temperature and pressure, although one may be slightly more stable.

The reader who wishes to look up the crystal structure of a substance may consult the excellent compilation by Wyckoff listed in the references at the end of the chapter. Structure Reports and the journals Acta Crystallographica and Zeitschrift für Kristallographie are valuable aids.
Table 3  Crystal structures of the elements

The data given are at room temperature for the most common form, or at the stated temperature in deg K. For further descriptions of the elements see Wyckoff, Vol. 1, Chap. 2. Structures labeled complex are described there.

<table>
<thead>
<tr>
<th>Element</th>
<th>Hcp</th>
<th>Bcc</th>
<th>Fcc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li 78K</td>
<td>3.75</td>
<td>3.491</td>
<td>6.12</td>
</tr>
<tr>
<td>Na 5K</td>
<td>4.225</td>
<td>5.225</td>
<td>5.585</td>
</tr>
<tr>
<td>K 5K</td>
<td>5.225</td>
<td>5.585</td>
<td>6.045</td>
</tr>
<tr>
<td>Rb 5K</td>
<td>5.585</td>
<td>6.08</td>
<td>6.045</td>
</tr>
<tr>
<td>Cs 5K</td>
<td>6.045</td>
<td>6.08</td>
<td>6.045</td>
</tr>
</tbody>
</table>

Crystal structure, a lattice parameter, in Å

<table>
<thead>
<tr>
<th>Element</th>
<th>Hcp</th>
<th>Bcc</th>
<th>Fcc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hf</td>
<td>5.05</td>
<td>5.05</td>
<td>5.05</td>
</tr>
<tr>
<td>Ta</td>
<td>5.27</td>
<td>5.27</td>
<td>5.27</td>
</tr>
<tr>
<td>W</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
<tr>
<td>Re</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
<tr>
<td>Os</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
<tr>
<td>Ir</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
<tr>
<td>Pt</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
<tr>
<td>Au</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
<tr>
<td>Hg</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
<tr>
<td>Tl</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
<tr>
<td>Pb</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
<tr>
<td>Bi</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
<tr>
<td>Po</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
<tr>
<td>At</td>
<td>5.15</td>
<td>5.15</td>
<td>5.15</td>
</tr>
</tbody>
</table>

I. Crystal Structure

<table>
<thead>
<tr>
<th>Element</th>
<th>Hcp</th>
<th>Bcc</th>
<th>Fcc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ce</td>
<td>5.16</td>
<td>5.16</td>
<td>5.16</td>
</tr>
<tr>
<td>Pr</td>
<td>3.67</td>
<td>3.67</td>
<td>3.67</td>
</tr>
<tr>
<td>Nd</td>
<td>3.66</td>
<td>3.66</td>
<td>3.66</td>
</tr>
<tr>
<td>Pm</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Sm</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Eu</td>
<td>4.58</td>
<td>4.58</td>
<td>4.58</td>
</tr>
<tr>
<td>Gd</td>
<td>5.78</td>
<td>5.78</td>
<td>5.78</td>
</tr>
<tr>
<td>Tb</td>
<td>5.70</td>
<td>5.70</td>
<td>5.70</td>
</tr>
<tr>
<td>Dy</td>
<td>5.65</td>
<td>5.65</td>
<td>5.65</td>
</tr>
<tr>
<td>Ho</td>
<td>5.62</td>
<td>5.62</td>
<td>5.62</td>
</tr>
<tr>
<td>Er</td>
<td>5.59</td>
<td>5.59</td>
<td>5.59</td>
</tr>
<tr>
<td>Tm</td>
<td>5.56</td>
<td>5.56</td>
<td>5.56</td>
</tr>
<tr>
<td>Yb</td>
<td>5.54</td>
<td>5.54</td>
<td>5.54</td>
</tr>
<tr>
<td>Lu</td>
<td>5.55</td>
<td>5.55</td>
<td>5.55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Element</th>
<th>Hcp</th>
<th>Bcc</th>
<th>Fcc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Th</td>
<td>5.08</td>
<td>5.08</td>
<td>5.08</td>
</tr>
<tr>
<td>Pa</td>
<td>3.92</td>
<td>3.92</td>
<td>3.92</td>
</tr>
<tr>
<td>U</td>
<td>3.24</td>
<td>3.24</td>
<td>3.24</td>
</tr>
<tr>
<td>Np</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Pu</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Am</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cm</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Bk</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cf</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Es</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Fm</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Md</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lr</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
Table 4  Density and atomic concentration

The data are given at atmospheric pressure and room temperature, or at the stated temperature in deg K. (Crystal modifications as for Table 3.)
SUMMARY

- A lattice is an array of points related by the lattice translation operator \( T = u_1a_1 + u_2a_2 + u_3a_3 \), where \( u_1, u_2, u_3 \) are integers and \( a_1, a_2, a_3 \) are the crystal axes.

- To form a crystal we attach to every lattice point an identical basis composed of \( s \) atoms at the positions \( r_j = x_ja_1 + y_ja_2 + z_ja_3 \), with \( j = 1, 2, \ldots, s \). Here \( x, y, z \) may be selected to have values between 0 and 1.

- The axes \( a_1, a_2, a_3 \) are primitive for the minimum cell volume \( |a_1 \cdot a_2 \times a_3| \) for which the crystal can be constructed from a lattice translation operator \( T \) and a basis at every lattice point.

Problems

1. Tetrahedral angles. The angles between the tetrahedral bonds of diamond are the same as the angles between the body diagonals of a cube, as in Fig. 12. Use elementary vector analysis to find the value of the angle.

2. Indices of planes. Consider the planes with indices (100) and (001); the lattice is fcc, and the indices refer to the conventional cubic cell. What are the indices of these planes when referred to the primitive axes of Fig. 13?

3. Hcp structure. Show that the \( c/a \) ratio for an ideal hexagonal close-packed structure is \( (\frac{12}{7})^{1/2} = 1.633 \). If \( c/a \) is significantly larger than this value, the crystal structure may be thought of as composed of planes of closely packed atoms, the planes being loosely stacked.

References

ELEMENTARY

CRYSTALLOGRAPHY

CRYSTAL GROWTH
Series: Journal of Crystal Growth, includes proceedings of the International Conferences on Crystal Growth.

CLASSICAL TABLES AND HANDBOOKS


Reciprocal Lattice

DIFFRACTION OF WAVES BY CRYSTALS 29
Bragg law 29

SCATTERED WAVE AMPLITUDE 30
Fourier analysis 30
Reciprocal lattice vectors 33
Diffraction conditions 34
Laue equations 36

BRILLOUIN ZONES 37
Reciprocal lattice to sc lattice 40
Reciprocal lattice to bcc lattice 40
Reciprocal lattice to fcc lattice 41

FOURIER ANALYSIS OF THE BASIS 42
Structure factor of the bcc lattice 44
Structure factor of the fcc lattice 45
Atomic form factor 45

QUASICRYSTALS 48

SUMMARY 49

PROBLEMS 51
1. Interplanar separation 51
2. Hexagonal space lattice 51
3. Volume of Brillouin zone 51
4. Width of diffraction maximum 51
5. Structure factor of diamond 51
6. Form factor of atomic hydrogen 52
7. Diatomic line 52

REFERENCES 52
Figure 1  Wavelength versus particle energy, for photons, neutrons, and electrons.

Figure 2  Derivation of the Bragg equation $2d \sin \theta = n\lambda$, here $d$ is the spacing of parallel atomic planes and $2\pi n$ is the difference in phase between reflections from successive planes. The reflecting planes have nothing to do with the surface planes bounding the particular specimen.
CHAPTER 2: RECIPROCAL LATTICE

DIFFRACTION OF WAVES BY CRYSTALS

Bragg Law

We study crystal structure through the diffraction of photons, neutrons, and electrons (Fig. 1). The diffraction depends on the crystal structure and on the wavelength. At optical wavelengths such as 5000 Å the superposition of the waves scattered elastically by the individual atoms of a crystal results in ordinary optical refraction. When the wavelength of the radiation is comparable with or smaller than the lattice constant, we may find diffracted beams in directions quite different from the incident direction.

W. L. Bragg presented a simple explanation of the diffracted beams from a crystal. The Bragg derivation is simple but is convincing only because it reproduces the correct result. Suppose that the incident waves are reflected specularly from parallel planes of atoms in the crystal, with each plane reflecting only a very small fraction of the radiation, like a lightly silvered mirror. In specular (mirrorlike) reflection the angle of incidence is equal to the angle of reflection. The diffracted beams are found when the reflections from parallel planes of atoms interfere constructively, as in Fig. 2. We treat elastic scattering, in which the energy of the x-ray is not changed on reflection. Inelastic scattering, with excitation of elastic waves, is discussed in Appendix A.

Consider parallel lattice planes spaced \( d \) apart. The radiation is incident in the plane of the paper. The path difference for rays reflected from adjacent planes is \( 2d \sin \theta \), where \( \theta \) is measured from the plane. Constructive interference of the radiation from successive planes occurs when the path difference is an integral number \( n \) of wavelengths \( \lambda \), so that

\[
2d \sin \theta = n\lambda
\]  

This is the Bragg law. Bragg reflection can occur only for wavelength \( \lambda \leq 2d \). This is why we cannot use visible light.

Although the reflection from each plane is specular, for only certain values of \( \theta \) will the reflections from all parallel planes add up in phase to give a strong reflected beam. If each plane were perfectly reflecting, only the first plane of a parallel set would see the radiation, and any wavelength would be reflected. But each plane reflects \( 10^{-3} \) to \( 10^{-5} \) of the incident radiation, so that \( 10^3 \) to \( 10^5 \) planes may contribute to the formation of the Bragg-reflected beam in a perfect crystal. Reflection by a single plane of atoms is treated in Chapter 19 on surface physics.

The Bragg law is a consequence of the periodicity of the lattice. Notice that the law does not refer to the composition of the basis of atoms associated with every lattice point. We shall see, however, that the composition of the basis
determines the relative intensity of the various orders of diffraction (denoted by $n$ above) from a given set of parallel planes. Experimental results for Bragg reflection from single crystals are shown in Figs. 3 and 4, for rotation about a fixed axis.

SCATTERED WAVE AMPLITUDE

The Bragg derivation of the diffraction condition (1) gives a neat statement of the condition for the constructive interference of waves scattered from the lattice points. We need a deeper analysis to determine the scattering intensity from the basis of atoms, which means from the spatial distribution of electrons within each cell.

From (1.3), a crystal is invariant under any translation of the form $T = u_1a_1 + u_2a_2 + u_3a_3$, where $u_1$, $u_2$, $u_3$ are integers and $a_1$, $a_2$, $a_3$ are the crystal axes. Any local physical property of the crystal is invariant under $T$, such as the charge concentration, electron number density, or magnetic moment density.

Fourier Analysis

What is most important to us here is that the electron number density $n(r)$ is a periodic function of $r$, with periods $a_1$, $a_2$, $a_3$ in the directions of the three crystal axes. Thus

$$n(r + T) = n(r) \ . \ (2)$$

Such periodicity creates an ideal situation for Fourier analysis. The most interesting properties of crystals are directly related to the Fourier components of the electron density.

We consider first a function $n(x)$ with period $a$ in the direction $x$, in one dimension. We expand $n(x)$ in a Fourier series of sines and cosines:

$$n(x) = n_0 + \sum_{p>0} [C_p \cos(2\pi px/a) + S_p \sin(2\pi px/a)] \ , \ (3)$$

where the $p$'s are positive integers and $C_p$, $S_p$ are real constants, called the Fourier coefficients of the expansion. The factor $2\pi/a$ in the arguments ensures that $n(x)$ has the period $a$:

$$n(x + a) = n_0 + \Sigma [C_p \cos(2\pi px/a + 2\pi p) + S_p \sin(2\pi px/a + 2\pi p)]$$

$$= n_0 + \Sigma [C_p \cos(2\pi px/a) + S_p \sin(2\pi px/a)] = n(x) \ . \ (4)$$

We say that $2\pi p/a$ is a point in the reciprocal lattice or Fourier space of the crystal. In one dimension these points lie on a line. The reciprocal lattice points tell us the allowed terms in the Fourier series (4) or (5). A term is allowed if it is consistent with the periodicity of the crystal, as in Fig. 5; other points in the reciprocal space are not allowed in the Fourier expansion of a periodic function.
Figure 3 Sketch of a monochromator which by Bragg reflection selects a narrow spectrum of x-ray or neutron wavelengths from a broad spectrum incident beam. The upper part of the figure shows the analysis (obtained by reflection from a second crystal) of the purity of a 1.16 Å beam of neutrons from a calcium fluoride crystal monochromator. The main beam is that not reflected from the second crystal. (After G. Bacon.)

Figure 4 X-ray diffractometer recording of powdered silicon, showing a counter recording of the diffracted beams. (Courtesy of W. Parrish.)

Figure 5 A periodic function $n(x)$ of period $a$, and the terms $2\pi p/a$ that may appear in the Fourier transform $n(x) = \sum n_p \exp(i2\pi px/a)$. The magnitudes of the individual terms $n_p$ are not plotted.
It is a great convenience to write the series (4) in the compact form

\[ n(x) = \sum_{p} n_{p} \exp(i2\pi px/a) \ , \tag{5} \]

where the sum is over all integers \( p \): positive, negative, and zero. The coefficients \( n_{p} \) now are complex numbers. To ensure that \( n(x) \) is a real function, we require

\[ n^{*}_{-p} = n_{p} \ , \tag{6} \]

for then the sum of the terms in \( p \) and \( -p \) is real. The asterisk on \( n^{*}_{-p} \) denotes the complex conjugate of \( n_{-p} \).

With \( \varphi = 2\pi px/a \), the sum of the terms in \( p \) and \( -p \) in (5) can be shown to be real if (6) is satisfied. The sum is

\[ n_{p}(\cos \varphi + i \sin \varphi) + n_{-p}(\cos \varphi - i \sin \varphi) \]
\[ = (n_{p} + n_{-p})\cos \varphi + i(n_{p} - n_{-p})\sin \varphi \ , \tag{7} \]

which in turn is equal to the real function

\[ 2\text{Re}\{n_{p}\} \cos \varphi - 2\text{Im}\{n_{p}\} \sin \varphi \ , \tag{8} \]

if (6) is satisfied. Here \( \text{Re}\{n_{p}\} \) and \( \text{Im}\{n_{p}\} \) denote the real and imaginary parts of \( n_{p} \). Thus the number density \( n(x) \) is a real function, as desired.

The extension of the Fourier analysis to periodic functions \( n(r) \) in three dimensions is straightforward. We must find a set of vectors \( \mathbf{G} \) such that

\[ n(r) = \sum_{\mathbf{G}} n_{\mathbf{G}} \exp(i\mathbf{G} \cdot r) \tag{9} \]

is invariant under all crystal translations \( T \) that leave the crystal invariant. It will be shown below that the set of Fourier coefficients \( n_{\mathbf{G}} \) determines the x-ray scattering amplitude.

**Inversion of Fourier Series.** We now show that the Fourier coefficient \( n_{p} \) in the series (5) is given by

\[ n_{p} = a^{-1} \int_{0}^{a} dx \ n(x) \exp(-i2\pi px/a) \ . \tag{10} \]

Substitute (5) in (10) to obtain

\[ n_{p} = a^{-1} \sum_{p'} n_{p'} \int_{0}^{a} dx \ \exp[i2\pi(p' - p)x/a] \ . \tag{11} \]

If \( p' \neq p \) the value of the integral is

\[ \frac{a}{i2\pi(p' - p)}(e^{i2\pi(p' - p)} - 1) = 0 \ , \]
because \( p' - p \) is an integer and \( \exp(i2\pi(p\text{integer})) = 1 \). For the term \( p' = p \) the integrand is \( \exp(i0) = 1 \), and the value of the integral is \( a \), so that \( n_p = a^{-1}n_p' = n_p \), which is an identity, so that (10) is an identity.

Similarly, the inversion of (9) gives

\[
n_C = V_c^{-1} \int_{\text{cell}} dV \ n(r) \exp(-i\mathbf{G} \cdot \mathbf{r}) .
\]

(12)

Here \( V_c \) is the volume of a cell of the crystal.

**Reciprocal Lattice Vectors**

To proceed further with the Fourier analysis of the electron concentration we must find the vectors \( \mathbf{G} \) of the Fourier sum \( \Sigma n_\mathbf{G} \exp(i\mathbf{G} \cdot \mathbf{r}) \) as in (9). There is a powerful, somewhat abstract procedure for doing this. The procedure forms the theoretical basis for much of solid state physics, where Fourier analysis is the order of the day.

We construct the axis vectors \( \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3 \) of the reciprocal lattice:

\[
\mathbf{b}_1 = 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3} \quad \mathbf{b}_2 = 2\pi \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3} \quad \mathbf{b}_3 = 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3} .
\]

(13)

The factors \( 2\pi \) are not used by crystallographers but are convenient in solid state physics.

If \( \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \) are primitive vectors of the crystal lattice, then \( \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3 \) are primitive vectors of the reciprocal lattice. Each vector defined by (13) is orthogonal to two axis vectors of the crystal lattice. Thus \( \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3 \) have the property

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

(14)

where \( \delta_{ij} = 1 \) if \( i = j \) and \( \delta_{ij} = 0 \) if \( i \neq j \).

Points in the reciprocal lattice are mapped by the set of vectors

\[
\mathbf{G} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + v_3\mathbf{b}_3 ,
\]

(15)

where \( v_1, v_2, v_3 \) are integers. A vector \( \mathbf{G} \) of this form is a reciprocal lattice vector.

Every crystal structure has two lattices associated with it, the crystal lattice and the reciprocal lattice. A diffraction pattern of a crystal is, as we shall show, a map of the reciprocal lattice of the crystal. A microscope image, if it could be resolved on a fine enough scale, is a map of the crystal structure in real space. The two lattices are related by the definitions (13). Thus when we rotate a crystal in a holder, we rotate both the direct lattice and the reciprocal lattice.

Vectors in the direct lattice have the dimensions of [length]; vectors in the reciprocal lattice have the dimensions of [1/length]. The reciprocal lattice is a lattice in the Fourier space associated with the crystal. The term is motivated
below. Wavevectors are always drawn in Fourier space, so that every position in Fourier space may have a meaning as a description of a wave, but there is a special significance to the points defined by the set of \( G \)'s associated with a crystal structure.

The vectors \( G \) in the Fourier series (9) are just the reciprocal lattice vectors (15), for then the Fourier series representation of the electron density has the desired invariance under any crystal translation \( T = u_1a_1 + u_2a_2 + u_3a_3 \) as defined by (1.3). From (9),

\[
n(r + T) = \sum_G n_G \exp(iG \cdot r) \exp(iG \cdot T) .
\]

But \( \exp(iG \cdot T) = 1 \), because

\[
\exp(iG \cdot T) = \exp[i(v_1b_1 + v_2b_2 + v_3b_3) \cdot (u_1a_1 + u_2a_2 + u_3a_3)]
= \exp[i2\pi(v_1u_1 + v_2u_2 + v_3u_3)] .
\]

The argument of the exponential has the form \( 2\pi i \) times an integer, because \( v_1u_1 + v_2u_2 + v_3u_3 \) is an integer, being the sum of products of integers. Thus by (9) we have the desired invariance, \( n(r + T) = n(r) \).

This result proves that the Fourier representation of a function periodic in the crystal lattice can contain components \( n_G \exp(iG \cdot r) \) only at the reciprocal lattice vectors \( G \) as defined by (15).

**Diffraction Conditions**

**Theorem.** The set of reciprocal lattice vectors \( G \) determines the possible x-ray reflections.

We see in Fig. 6 that the difference in phase factors is \( \exp[i(k - k') \cdot r] \) between beams scattered from volume elements \( r \) apart. The wavevectors of the incoming and outgoing beams are \( k \) and \( k' \). The amplitude of the wave scattered from a volume element is proportional to the local electron concentration \( n(r) \). The total amplitude of the scattered wave in the direction of \( k' \) is proportional to the integral over the crystal of \( n(r) \, dV \) times the phase factor \( \exp[i(k - k') \cdot r] \).

In other words, the amplitude of the electric or magnetic field vectors in the scattered electromagnetic wave is proportional to the following integral which defines the quantity \( F \) that we call the scattering amplitude:

\[
F = \int dV \, n(r) \exp[i(k - k') \cdot r] = \int dV \, n(r) \exp(-i\Delta k \cdot r) ,
\]

where

\[
k + \Delta k = k' .
\]

Here \( \Delta k \) measures the change in wavevector and is called the scattering vector (Fig. 7). We add \( \Delta k \) to \( k \) to obtain \( k' \), the wavevector of the scattered beam.
Figure 6  The difference in path length of the incident wave \( k \) at the points \( O, r \) is \( r \sin \varphi \), and the difference in phase angle is \( (2\pi r \sin \varphi)/\lambda \), which is equal to \( k \cdot r \). For the diffracted wave the difference in phase angle is \( -k' \cdot r \). The total difference in phase angle is \( (k - k') \cdot r \), and the wave scattered from \( dV \) at \( r \) has the phase factor \( \exp[i(k - k') \cdot r] \) relative to the wave scattered from a volume element at the origin \( O \).

Figure 7  Definition of the scattering vector \( \Delta k \) such that \( k + \Delta k = k' \). In elastic scattering the magnitudes satisfy \( k' = k \). Further, in Bragg scattering from a periodic lattice any allowed \( \Delta k \) must equal some reciprocal lattice vector \( G \).

We introduce into (18) the Fourier components (9) of \( n(r) \) to obtain for the scattering amplitude

\[
F = \sum_{G} \int dV \, n_G \, \exp[i(G - \Delta k) \cdot r].
\]

When the scattering vector \( \Delta k \) is equal to a particular reciprocal lattice vector,

\[
\Delta k = G,
\]

the argument of the exponential vanishes and \( F = V n_G \). It is a simple exercise (Problem 4) to show that \( F \) is negligibly small when \( \Delta k \) differs significantly from any reciprocal lattice vector.

In elastic scattering of a photon its energy \( \hbar \omega \) is conserved, so that the frequency \( \omega' = c k' \) of the emergent beam is equal to the frequency of the incident beam. Thus the magnitudes \( k \) and \( k' \) are equal, and \( k^2 = k'^2 \), a result that holds also for electron and neutron beams. From (21) we found \( \Delta k = G \) or
\( k + G = k' \), so that the diffraction condition is written as \((k + G)^2 = k'^2\), or

\[
2k \cdot G + G^2 = 0.
\]

(22)

This is the central result of the theory of elastic scattering of waves in a periodic lattice. If \( G \) is a reciprocal lattice vector, so is \(-G\), and with this substitution we can write (22) as

\[
2k \cdot G = G^2.
\]

(23)

This particular expression is often used as the condition for diffraction.

Equation (23) is another statement of the Bragg condition (1). The result of Problem 1 is that the spacing \( d(hkl) \) between parallel lattice planes that are normal to the direction \( G = h b_1 + k b_2 + l b_3 \) is \( d(hkl) = 2\pi/|G| \). Thus the result \( 2k \cdot G = G^2 \) may be written as

\[
2(2\pi/\lambda) \sin \theta = 2\pi/d(hkl),
\]

or \( 2d(hkl) \sin \theta = \lambda \). Here \( \theta \) is the angle between the incident beam and the crystal plane.

The integers \( hkl \) that define \( G \) are not necessarily identical with the indices of an actual crystal plane, because the \( hkl \) may contain a common factor \( n \), whereas in the definition of the indices in Chapter 1 the common factor has been eliminated. We thus obtain the Bragg result:

\[
2d \sin \theta = n\lambda,
\]

(24)

where \( d \) is the spacing between adjacent parallel planes with indices \( h/n, k/n, l/n \).

**Laue Equations**

The original result (21) of diffraction theory, namely that \( \Delta k = G \), may be expressed in another way to give what are called the Laue equations. These are valuable because of their geometrical representation (see Chapter 19).

Take the scalar product of both \( \Delta k \) and \( G \) successively with \( a_1, a_2, a_3 \). From (14) and (15) we get

\[
a_1 \cdot \Delta k = 2\pi v_1; \quad a_2 \cdot \Delta k = 2\pi v_2; \quad a_3 \cdot \Delta k = 2\pi v_3.
\]

(25)

These equations have a simple geometrical interpretation. The first equation \( a_1 \cdot \Delta k = 2\pi v_1 \) tells us that \( \Delta k \) lies on a certain cone about the direction of \( a_1 \). The second equation tells us that \( \Delta k \) lies on a cone about \( a_2 \) as well, and the third equation requires that \( \Delta k \) lies on a cone about \( a_3 \).

Thus, at a reflection \( \Delta k \) must satisfy all three equations; it must lie at the common line of intersection of three cones, which is a severe condition that can be satisfied only by systematic sweeping or searching in wavelength or crystal orientation—or else by sheer accident.
A beautiful construction, the Ewald construction, is exhibited in Fig. 8. This helps us visualize the nature of the accident that must occur in order to satisfy the diffraction condition in three dimensions. The condition in two dimensions (diffraction from a surface layer) is treated in Chapter 19.

Reflection from a single plane of atoms takes place in the directions of the lines of intersection of two cones, for example the cones defined by the first two of the Laue equations (25). Now two cones will in general intercept each other provided the wavevector of the particles in the incident beam exceeds some threshold value determined by the first two Laue equations. No accidental coincidence is required, unlike the problem of diffraction in 3D. This matter is of prime importance in the diffraction of low energy electrons from the surface of a crystal.

**BRILLOUIN ZONES**

Brillouin gave the statement of the diffraction condition that is most widely used in solid state physics, which means in the description of electron energy band theory and of the elementary excitations of other kinds.
A Brillouin zone is defined as a Wigner-Seitz primitive cell in the reciprocal lattice. (The construction in the direct lattice was shown in Fig. 1.6.) The value of the Brillouin zone is that it gives a vivid geometrical interpretation of the diffraction condition $2k \cdot G = G^2$ of Eq. (23). We divide both sides by 4 to obtain

$$k \cdot (\frac{1}{2}G) = (\frac{1}{2}G)^2. \tag{26}$$

We work in reciprocal space, the space of the $k$'s and $G$'s. Select a vector $G$ from the origin to a reciprocal lattice point. Construct a plane normal to this vector $G$ at its midpoint. This plane forms a part of the zone boundary (Fig. 9a). An x-ray beam in the crystal will be diffracted if its wavevector $k$ has the magnitude and direction required by (26). The diffracted beam will then be in the direction $k - G$, as we see from (19) with $\Delta k = -G$. Thus the Brillouin construction exhibits all the wavevectors $k$ which can be Bragg-reflected by the crystal.

The set of planes that are the perpendicular bisectors of the reciprocal lattice vectors is of general importance in the theory of wave propagation in crystals. A wave whose wavevector drawn from the origin terminates on any of these planes will satisfy the condition for diffraction.

These planes divide the Fourier space of the crystal into fragments, as shown in Fig. 9b for a square lattice. The central square is a primitive cell of the reciprocal lattice. It is a Wigner-Seitz cell of the reciprocal lattice.
Figure 10 Construction of the first Brillouin zone for an oblique lattice in two dimensions. We first draw a number of vectors from $O$ to nearby points in the reciprocal lattice. Next we construct lines perpendicular to these vectors at their midpoints. The smallest enclosed area is the first Brillouin zone.

Figure 11 Crystal and reciprocal lattices in one dimension. The basis vector in the reciprocal lattice is $b$, of length equal to $2\pi/a$. The shortest reciprocal lattice vectors from the origin are $b$ and $-b$. The perpendicular bisectors of these vectors form the boundaries of the first Brillouin zone. The boundaries are at $k = \pm \pi/a$.

The central cell in the reciprocal lattice is of special importance in the theory of solids, and we call it the first Brillouin zone. The first Brillouin zone is the smallest volume entirely enclosed by planes that are the perpendicular bisectors of the reciprocal lattice vectors drawn from the origin.

The first Brillouin zone of an oblique lattice in two dimensions is constructed in Fig. 10 and of a linear lattice in one dimension in Fig. 11. The zone boundaries of the linear lattice are at $k = \pm \pi/a$, where $a$ is the primitive axis of the crystal lattice.
Historically, Brillouin zones are not part of the language of x-ray diffraction analysis of crystal structures, but the zones are an essential part of the analysis of the electronic energy-band structure of crystals. The special utility of the first Brillouin zone is developed in Chapter 9.

**Reciprocal Lattice to sc Lattice**

The primitive translation vectors of a simple cubic lattice may be taken as the set
\[
a_1 = a\hat{x} \quad a_2 = a\hat{y} \quad a_3 = a\hat{z}.
\]  
(27a)

Here \(\hat{x}, \hat{y}, \hat{z}\) are orthogonal vectors of unit length. The volume of the cell is \(a_1 \cdot a_2 \times a_3 = a^3\). The primitive translation vectors of the reciprocal lattice are found from the standard prescription (13):
\[
b_1 = (2\pi/a)\hat{x} \quad b_2 = (2\pi/a)\hat{y} \quad b_3 = (2\pi/a)\hat{z}.
\]  
(27b)

Here the reciprocal lattice is itself a simple cubic lattice, now of lattice constant \(2\pi/a\).

The boundaries of the first Brillouin zones are the planes normal to the six reciprocal lattice vectors \(\pm b_1, \pm b_2, \pm b_3\) at their midpoints:
\[
\pm \frac{1}{2}b_1 = \pm (\pi/a)\hat{x} \quad \pm \frac{1}{2}b_2 = \pm (\pi/a)\hat{y} \quad \pm \frac{1}{2}b_3 = \pm (\pi/a)\hat{z}.
\]

The six planes bound a cube of edge \(2\pi/a\) and of volume \((2\pi/a)^3\); this cube is the first Brillouin zone of the sc crystal lattice.

**Reciprocal Lattice to bcc Lattice**

The primitive translation vectors of the bcc lattice (Fig. 12) are
\[
a_1 = \frac{1}{2}a(-\hat{x} + \hat{y} + \hat{z}) \quad a_2 = \frac{1}{2}a(\hat{x} - \hat{y} + \hat{z}) \quad a_3 = \frac{1}{2}a(\hat{x} + \hat{y} - \hat{z}),
\]  
(28)

where \(a\) is the side of the conventional cube and \(\hat{x}, \hat{y}, \hat{z}\) are orthogonal unit vectors parallel to the cube edges. The volume of the primitive cell is
\[
V = |a_1 \cdot a_2 \times a_3| = \frac{1}{2}a^3.
\]  
(29)

The primitive translations of the reciprocal lattice are defined by (13). We have, using (28),
\[
b_1 = (2\pi/a)(\hat{y} + \hat{z}) \quad b_2 = (2\pi/a)(\hat{x} + \hat{z}) \quad b_3 = (2\pi/a)(\hat{x} + \hat{y}).
\]  
(30)

Note by comparison with Fig. 14 (p. 42) that these are just the primitive vectors of an fcc lattice, so that an fcc lattice is the reciprocal lattice of the bcc lattice.

The general reciprocal lattice vector is, for integral \(v_1, v_2, v_3\),
\[
G = v_1b_1 + v_2b_2 + v_3b_3 = (2\pi/a)[(v_2 + v_3)\hat{x} + (v_1 + v_3)\hat{y} + (v_1 + v_2)\hat{z}].
\]  
(31)

The shortest \(G\)'s are the following 12 vectors, where all choices of sign are independent:
\[
(2\pi/a)(\pm \hat{y} \pm \hat{z}) \quad (2\pi/a)(\pm \hat{x} \pm \hat{z}) \quad (2\pi/a)(\pm \hat{x} \pm \hat{y}).
\]  
(32)
The primitive cell of the reciprocal lattice is the parallelepiped described by the $b_1$, $b_2$, $b_3$ defined by (30). The volume of this cell in reciprocal space is $b_1 \cdot b_2 \times b_3 = 2(2\pi/a)^3$. The cell contains one reciprocal lattice point, because each of the eight corner points is shared among eight parallelepipeds. Each parallelepiped contains one-eighth of each of eight corner points.

In solid state physics we take the central (Wigner-Seitz) cell of the reciprocal lattice as the first Brillouin zone. Each such cell contains one lattice point at the central point of the cell. This zone (for the bcc lattice) is bounded by the planes normal to the 12 vectors of Eq. (32) at their midpoints. The zone is a regular 12-faced solid, a rhombic dodecahedron, as shown in Fig. 13. The vectors from the origin to the center of each face are

$$(\pi/a)(\pm \hat{x} \pm \hat{z}); \quad (\pi/a)(\pm \hat{y} \pm \hat{z}); \quad (\pi/a)(\pm \hat{x} \pm \hat{y}).$$  \hspace{1cm} (33)

All choices of sign are independent, giving 12 vectors.

**Reciprocal Lattice to fcc Lattice**

The primitive translation vectors of the fcc lattice of Fig. 14 are

$$a_1 = \frac{1}{2} a(\hat{y} + \hat{z}); \quad a_2 = \frac{1}{2} a(\hat{x} + \hat{z}); \quad a_3 = \frac{1}{2} a(\hat{x} + \hat{y}).$$  \hspace{1cm} (34)

The volume of the primitive cell is

$$V = |a_1 \cdot a_2 \times a_3| = \frac{1}{4} a^3$$  \hspace{1cm} (35)

The primitive translation vectors of the lattice reciprocal to the fcc lattice are

$$b_1 = (2\pi/a)(-\hat{x} + \hat{y} + \hat{z}); \quad b_2 = (2\pi/a)(\hat{x} - \hat{y} + \hat{z}); \quad b_3 = (2\pi/a)(\hat{x} + \hat{y} - \hat{z}).$$  \hspace{1cm} (36)
These are primitive translation vectors of a bcc lattice, so that the bcc lattice is reciprocal to the fcc lattice. The volume of the primitive cell of the reciprocal lattice is $4(2\pi/a)^3$.

The shortest $G$’s are the eight vectors:

$$(2\pi/a)(\pm \hat{x} \pm \hat{y} \pm \hat{z}) \quad .$$  

(37)

The boundaries of the central cell in the reciprocal lattice are determined for the most part by the eight planes normal to these vectors at their midpoints. But the corners of the octahedron thus formed are cut by the planes that are the perpendicular bisectors of six other reciprocal lattice vectors:

$$(2\pi/a)(\pm 2\hat{x}) ; \quad (2\pi/a)(\pm 2\hat{y}) ; \quad (2\pi/a)(\pm 2\hat{z}) \quad .$$  

(38)

Note that $(2\pi/a)(2\hat{x})$ is a reciprocal lattice vector because it is equal to $b_2 + b_3$. The first Brillouin zone is the smallest bounded volume about the origin, the truncated octahedron shown in Fig. 15. The six planes bound a cube of edge $4\pi/a$ and (before truncation) of volume $(4\pi/a)^3$.

**FOURIER ANALYSIS OF THE BASIS**

When the diffraction condition $\Delta k = G$ of Eq. (21) is satisfied, the scattering amplitude is determined by (18), which for a crystal of $N$ cells may be written as

$$F_G = N \int dV n(r) \exp(-iG \cdot r) = NS_G \quad (39)$$

The quantity $S_G$ is called the structure factor and is defined as an integral over a single cell, with $r = 0$ at one corner.
Often it is useful to write the electron concentration $n(r)$ as the superposition of electron concentration functions $n_j$ associated with each atom $j$ of the cell. If $\mathbf{r}_j$ is the vector to the center of atom $j$, then the function $n_j(\mathbf{r} - \mathbf{r}_j)$ defines the contribution of that atom to the electron concentration at $\mathbf{r}$. The total electron concentration at $\mathbf{r}$ due to all atoms in the cell is the sum

$$n(r) = \sum_{j=1}^{s} n_j(\mathbf{r} - \mathbf{r}_j) \tag{40}$$

over the $s$ atoms of the basis. The decomposition of $n(r)$ is not unique, for we cannot always say how much charge is associated with each atom. This is not an important difficulty.

The structure factor defined by (39) may now be written as integrals over the $s$ atoms of a cell:

$$S_G = \sum_j \int dV \ n_j(\mathbf{r} - \mathbf{r}_j) \exp(-i\mathbf{G} \cdot \mathbf{r}) =$$

$$\sum_j \exp(-i\mathbf{G} \cdot \mathbf{r}_j) \int dV \ n_j(\rho) \exp(-i\mathbf{G} \cdot \rho), \tag{41}$$

where $\rho = \mathbf{r} - \mathbf{r}_j$.

We now define the atomic form factor as

$$f_j = \int dV \ n_j(\rho) \exp(-i\mathbf{G} \cdot \rho). \tag{42}$$

integrated over all space. If $n_j(\rho)$ is an atomic property, $f_j$ is an atomic property.
We combine (41) and (42) to obtain the structure factor of the basis in the form

$$S_C = \sum_j f_j \exp(-iG \cdot r_j) .$$  \hfill (43)

The usual form of this result follows on writing for atom $j$:

$$r_j = x_ja_1 + y_ja_2 + z_ja_3 ,$$  \hfill (44)

as in (1.4). Then, for the reflection labelled by $v_1$, $v_2$, $v_3$ we have

$$G \cdot r_j = (v_1b_1 + v_2b_2 + v_3b_3) \cdot (x_ja_1 + y_ja_2 + z_ja_3)$$

$$= 2\pi(v_1x_j + v_2y_j + v_3z_j) ,$$  \hfill (45)

so that (43) becomes

$$S_C(v_1v_2v_3) = \sum_j f_j \exp[-i2\pi(v_1x_j + v_2y_j + v_3z_j)] .$$  \hfill (46)

The structure factor $S$ need not be real because the scattered intensity will involve $S^*S$, where $S^*$ is the complex conjugate of $S$ so that $S^*S$ is real.

At a zero of $S_C$ the scattered intensity will be zero, even though $G$ is a perfectly good reciprocal lattice vector. What happens if we choose the cell in another way, as a conventional cell instead of a primitive cell, for example? The basis is changed, but in such a way that the physical scattering is unchanged. Thus for two choices, 1 and 2, it is not hard to satisfy yourself from (39) that

$$N_1(\text{cell}) \times S_1(\text{basis}) = N_2(\text{cell}) \times S_2(\text{basis}) .$$

**Structure Factor of the bcc Lattice**

The bcc basis referred to the cubic cell has identical atoms at $x_1 = y_1 = z_1 = 0$ and at $x_2 = y_2 = z_2 = \frac{1}{2}$. Thus (46) becomes

$$S(v_1v_2v_3) = f\{1 + \exp[-i\pi(v_1 + v_2 + v_3)]\} ,$$  \hfill (47)

where $f$ is the form factor of an atom. The value of $S$ is zero whenever the exponential has the value $-1$, which is whenever the argument is $-i\pi \times$ (odd integer). Thus we have

$$S = 0 \quad \text{when } v_1 + v_2 + v_3 = \text{odd integer} ;$$

$$S = 2f \quad \text{when } v_1 + v_2 + v_3 = \text{even integer} .$$

Metallic sodium has a bcc structure. The diffraction pattern does not contain lines such as (100), (300), (111), or (221), but lines such as (200), (110), and (222) will be present; here the indices $(v_1v_2v_3)$ are referred to a cubic cell. What is the physical interpretation of the result that the (100) reflection vanishes?

The (100) reflection normally occurs when reflections from the planes that bound the cubic cell differ in phase by $2\pi$. In the bcc lattice there is an intervening plane (Fig. 16) of atoms, labeled the second plane in the figure, which is
equal in scattering power to the other planes. Situated midway between them, it gives a reflection retarded in phase by \( \pi \) with respect to the first plane, thereby canceling the contribution from that plane. The cancellation of the (100) reflection occurs in the bcc lattice because the planes are identical in composition. A similar cancellation can easily be found in the hcp structure.

**Structure Factor of the fcc Lattice**

The basis of the fcc structure referred to the cubic cell has identical atoms at \( 000; 0\frac{1}{2}\frac{1}{2}; \frac{1}{2}0\frac{1}{2}; \frac{1}{2}\frac{1}{2}0 \). Thus (46) becomes

\[
S(v_1v_2v_3) = f[1 + \exp[-i\pi(v_2 + v_3)] + \exp[-i\pi(v_1 + v_3)]
+ \exp[-i\pi(v_1 + v_2)]]. \tag{48}
\]

If all indices are even integers, \( S = 4f \); similarly if all indices are odd integers. But if only one of the integers is even, two of the exponents will be odd multiples of \(-i\pi\) and \(S\) will vanish. If only one of the integers is odd, the same argument applies and \(S\) will also vanish.

Thus in the fcc lattice no reflections can occur for which the indices are partly even and partly odd. The point is beautifully illustrated by Fig. 17: both KCl and KBr have an fcc lattice, but KCl simulates an sc lattice because the \( K^+ \) and \( Cl^- \) ions have equal numbers of electrons.

**Atomic Form Factor**

In the expression (46) for the structure factor, there occurs the quantity \( f_j \), which is a measure of the scattering power of the \( j \)th atom in the unit cell. The value of \( f \) involves the number and distribution of atomic electrons, and the wavelength and angle of scattering of the radiation. We now give a classical calculation of the scattering factor.
The scattered radiation from a single atom takes account of interference effects within the atom. We defined the form factor in (42):

$$f_j = \int dV n_j(r) \exp(-i \mathbf{G} \cdot \mathbf{r}),$$

with the integral extended over the electron concentration associated with a single atom. Let $\mathbf{r}$ make an angle $\alpha$ with $\mathbf{G}$; then $\mathbf{G} \cdot \mathbf{r} = Gr \cos \alpha$. If the electron distribution is spherically symmetric about the origin, then

$$f_j = 2\pi \int dr \ r^2 \ d(\cos \alpha) n_j(r) \exp(-i Gr \cos \alpha)$$

$$= 2\pi \int dr \ r^2 n_j(r) \cdot \frac{e^{iGr} - e^{-iGr}}{iGr},$$

after integration over $d(\cos \alpha)$ between $-1$ and $1$. Thus the form factor is given by

$$f_j = 4\pi \int dr \ n_j(r) r^2 \frac{\sin Gr}{Gr}$$

(50)

If the same total electron density were concentrated at $r = 0$, only $Gr = 0$ would contribute to the integrand. In this limit $(\sin Gr)/Gr = 1$, and

$$f_j = 4\pi \int dr \ n_j(r) r^2 = Z,$$

(51)
the number of atomic electrons. Therefore $f$ is the ratio of the radiation amplitude scattered by the actual electron distribution in an atom to that scattered by one electron localized at a point.

In the forward direction $G = 0$, and $f$ reduces again to the value $Z$. Values of the atomic form factor $f$ for atoms may be found in the *International tables for x-ray crystallography*, Vol. 3.

The overall electron distribution in a solid as seen in x-ray diffraction is fairly close to that of the appropriate free atoms. This statement does not mean that the outermost or valence electrons are not redistributed in forming the solid; it means only that the x-ray reflection intensities are represented well by the free atom values of the form factors and are not very sensitive to small redistributions of the electrons.

As an example, Batterman and co-workers find agreement within 1 percent in a comparison of the x-ray intensities of Bragg reflections of metallic iron, copper, and aluminum with the theoretical free atom values from wavefunction calculations. The results for aluminum are shown in Fig. 18.

There have been many attempts to obtain direct x-ray evidence about the actual electron distribution in a covalent chemical bond, particularly in crystals having the diamond structure. The question now lies within the limits of what can be explored by x-ray diffraction methods. In silicon at a point midway
between two nearest-neighbor atoms, there is an appreciable increase in electron concentration over what is expected from the overlap of the electron densities calculated for two free atoms.

Scattering from crystal surfaces is treated in Chapter 19. It is shown in Appendix A that thermal motion does not broaden a diffraction line, but only reduces the intensity. The lost intensity reappears as long, low wings about the position of the diffraction line.

QUASICRYSTALS

In 1984 quasicrystals were first observed;¹ these are structures which cannot be indexed to any Bravais lattice and "which have symmetries intermediate between a crystal and a liquid." They were first observed in grains of size 2 \( \mu \text{m} \) in an alloy of Al with 14 at. pct Mn. The smaller Mn atoms are each surrounded by 12 Al atoms arranged at the corners of an icosahedron. The structure is made up of parallel icosahedra attached at their edges. Crystals cannot exhibit the fivefold symmetry of an icosahedron, but a crystal can be constructed by nucleation at a center cell, followed by outward growth from there. All of the space of a node can be filled by repeating the basic unit (see Figures 19 and 1.7 for the picture in two dimensions), although the "parallel" part of the specification does give a long-range orientational order to the structure. It is perhaps surprising that the x-ray diffraction pattern of such a structure can have fivefold symmetry; that is how they were first observed.

The known quasicrystals are intermetallic alloys and are very poor electrical conductors; they are nearly insulators with a somewhat well-defined band gap (Chapter 7) at the Fermi level. They are of great interest intellectually in expanding the definition of crystal lattice.

A distinctly different crystal diffraction pattern results from an almost periodic structure, one that is neither rigorously periodic nor simply amorphous (as for a glass, Chapter 17). An almost periodic structure can be expressed in one dimension if we are given the electron charge density wave:

\[
\rho(x) = \sum C_n \cos[2\pi n (1 + \tau)x/a],
\]

where \( \tau \) is an irrational fraction. The terms in \( 2\pi n/a \) by themselves give the usual lattice with translational periodicity \( a \). When the terms in \( \tau \) are added, the charge density is almost periodic; that is, the period \( (1 + \tau)a \) is not an integral multiple of the period \( a \), because \( \tau \) is irrational. The period gives a long-range nonrandom order to the structure, and the long-range order gives a diffraction pattern, which appears split off from the pattern defined by the short-range order. This is dominated by the reciprocal lattice points in \( n \), but

will appear to be clustered and spread out (broadened). The diffraction pattern of a three-dimensional quasicrystal is quite different, however; the pattern is well defined and can have the fivefold symmetry by which quasicrystals were first discovered. A computer-generated diffraction pattern with fivefold symmetry is shown in Figure 20.

**SUMMARY**

- Various statements of the Bragg condition:
  \[ 2d \sin \theta = n\lambda ; \quad \Delta k = \mathbf{G} ; \quad 2k \cdot \mathbf{G} = \mathbf{G}^2 . \]

- Laue conditions:
  \[ a_1 \cdot \Delta k = 2\pi v_1 ; \quad a_2 \cdot \Delta k = 2\pi v_2 ; \quad a_3 \cdot \Delta k = 2\pi v_3 . \]

- The primitive translation vectors of the reciprocal lattice are
  \[ b_1 = 2\pi \frac{a_2 \times a_3}{a_1 \cdot a_2 \times a_3} ; \quad b_2 = 2\pi \frac{a_3 \times a_1}{a_1 \cdot a_2 \times a_3} ; \quad b_3 = 2\pi \frac{a_1 \times a_2}{a_1 \cdot a_2 \times a_3} . \]

Here \( a_1, a_2, a_3 \) are the primitive translation vectors of the crystal lattice.

- A reciprocal lattice vector has the form
  \[ \mathbf{G} = v_1 b_1 + v_2 b_2 + v_3 b_3 , \]
  where \( v_1, v_2, v_3 \) are integers or zero.

- The scattered amplitude in the direction \( \mathbf{k}' = \mathbf{k} + \Delta \mathbf{k} = \mathbf{k} + \mathbf{G} \) is proportional to the geometrical structure factor:
\[ S_G = \sum_j f_j \exp(-i\mathbf{r}_j \cdot \mathbf{G}) = \sum_j f_j \exp[-i2\pi(x_j v_1 + y_j v_2 + z_j v_3)] , \]

where \( j \) runs over the \( s \) atoms of the basis, and \( f_j \) is the atomic form factor \((49)\) of the \( j \)th atom of the basis. The expression on the right-hand side is written for a reflection \((v_1 v_2 v_3)\), for which \( \mathbf{G} = v_1 b_1 + v_2 b_2 + v_3 b_3 \).

- Any function invariant under a lattice translation \( T \) may be expanded in a Fourier series of the form

\[ n(r) = \sum_G n_G \exp(i\mathbf{G} \cdot \mathbf{r}) . \]

- The first Brillouin zone is the Wigner-Seitz primitive cell of the reciprocal lattice. Only waves whose wavevector \( \mathbf{k} \) drawn from the origin terminates on a surface of the Brillouin zone can be diffracted by the crystal.

- **Crystal lattice**
  - Simple cubic
  - Body-centered cubic
  - Face-centered cubic

- **First Brillouin zone**
  - Cube
  - Rhombic dodecahedron (Fig. 13)
  - Truncated octahedron (Fig. 15)

---

**Figure 20** Photograph of the calculated Fourier transform (diffraction pattern) of an icosahedral quasicrystal along one of the fivefold axes, illustrating the fivefold symmetry. The transform is calculated from a theoretical computer-generated model, by hael Jacob.
Problems

1. **Interplanar separation.** Consider a plane $hkl$ in a crystal lattice. (a) Prove that the reciprocal lattice vector $G = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3$ is perpendicular to this plane. (b) Prove that the distance between two adjacent parallel planes of the lattice is $d(hkl) = 2\pi|G|$. (c) Show for a simple cubic lattice that $d^2 = a^2(h^2 + k^2 + l^2)$.

2. **Hexagonal space lattice.** The primitive translation vectors of the hexagonal space lattice may be taken as

$$a_1 = (3^{1/2}a/2)\hat{x} + (a/2)\hat{y}; \quad a_2 = -(3^{1/2}a/2)\hat{x} + (a/2)\hat{y}; \quad a_3 = c\hat{z}.$$  

(a) Show that the volume of the primitive cell is $(3^{1/2}/2)a^2c$.

(b) Show that the primitive translations of the reciprocal lattice are

$$b_1 = (2\pi/3^{1/2}a)\hat{x} + (2\pi/a)\hat{y}; \quad b_2 = -(2\pi/3^{1/2}a)\hat{x} + (2\pi/a)\hat{y}; \quad b_3 = (2\pi/c)\hat{z},$$  

so that the lattice is its own reciprocal, but with a rotation of axes.

(c) Describe and sketch the first Brillouin zone of the hexagonal space lattice.

3. **Volume of Brillouin zone.** Show that the volume of the first Brillouin zone is $(2\pi)^3/V_c$, where $V_c$ is the volume of a crystal primitive cell. Hint: The volume of a Brillouin zone is equal to the volume of the primitive parallelepiped in Fourier space. Recall the vector identity $(c \times a) \times (a \times b) = (c \cdot a \times b)a$.

4. **Width of diffraction maximum.** We suppose that in a linear crystal there are identical point scattering centers at every lattice point $\rho_m = ma$, where $m$ is an integer. By analogy with (20) the total scattered radiation amplitude will be proportional to $F = \Sigma \exp[-ima \cdot \Delta k]$. The sum over $M$ lattice points is

$$F = \frac{1 - \exp[-iM(a \cdot \Delta k)]}{1 - \exp[-i(a \cdot \Delta k)],}$$  

by the use of the series

$$\sum_{m=0}^{M-1} x^m = \frac{1 - x^M}{1 - x}.$$  

(a) The scattered intensity is proportional to $|F|^2$. Show that

$$|F|^2 = F^*F = \frac{\sin^2\frac{1}{4}M(a \cdot \Delta k)}{\sin^2\frac{1}{4}(a \cdot \Delta k)}.$$  

(b) We know that a diffraction maximum appears when $a \cdot \Delta k = 2\pi h$, where $h$ is an integer. We change $\Delta k$ slightly and define $\epsilon$ in $a \cdot \Delta k = 2\pi h + \epsilon$ such that $\epsilon$ gives the position of the first zero in $\sin\frac{1}{4}M(a \cdot \Delta k)$. Show that $\epsilon = 2\pi/M$, so that the width of the diffraction maximum is proportional to $1/M$ and can be extremely narrow for macroscopic values of $M$. The same result holds true for a three-dimensional crystal.

5. **Structure factor of diamond.** The crystal structure of diamond is described in Chapter 1. The basis consists of eight atoms if the cell is taken as the conventional cube.
Figure 21 Neutron diffraction pattern for powdered diamond. (After G. Bacon.)

(a) Find the structure factor $S$ of this basis. (b) Find the zeros of $S$ and show that the allowed reflections of the diamond structure satisfy $h + v_2 + v_3 = 4n$, where all indices are even and $n$ is any integer, or else all indices are odd (Fig. 21). (Notice that $h, k, l$ may be written for $v_1, v_2, v_3$ and this is often done.)

6. Form factor of atomic hydrogen. For the hydrogen atom in its ground state, the number density is $n(r) = (\pi a_0^2)^{-1} \exp(-2r/a_0)$, where $a_0$ is the Bohr radius. Show that the form factor is $f_C = 16/(4 + C^2 a_0^2)^2$.

7. Diatomic line. Consider a line of atoms $ABAB \ldots AB$, with an $A-B$ bond length of $\frac{1}{2}a$. The form factors are $f_A, f_B$ for atoms $A, B$, respectively. The incident beam of x-rays is perpendicular to the line of atoms. (a) Show that the interference condition is $n\lambda = a \cos \theta$, where $\theta$ is the angle between the diffracted beam and the line of atoms. (b) Show that the intensity of the diffracted beam is proportional to $|f_A - f_B|^2$ for $n$ odd, and to $|f_A + f_B|^2$ for $n$ even. (c) Explain what happens if $f_A = f_B$.

References

X-RAY DIFFRACTION

NEUTRON DIFFRACTION
Crystal Binding and Elastic Constants

CRYS TALS OF INERT GASES 55
IONIC CRYSTALS 66
COVALENT CRYSTALS 72
METALS 75
HYDROGEN BONDS 76
ATOMIC RADII 77
ANALYSIS OF ELASTIC STRAINS 80
ELASTIC COMPLIANCE AND STIFFNESS CONSTANTS 83
ELASTIC WAVES IN CUBIC CRYSTALS 87
SUMMARY 90
PROBLEMS 92

1. Quantum solid 92
2. Cohesive energy of bcc and fcc neon 92
3. Solid molecular hydrogen 92
4. Possibility of ionic crystals $R^+R^-$ 93
5. Linear ionic crystal 93
6. Cubic ZnS structure 93
7. Divalent ionic crystals 93
8. Young’s modulus and Poisson’s ratio 93
9. Longitudinal wave velocity 93
10. Transverse wave velocity 94
11. Effective shear constant 94
12. Determinantal approach 94
13. General propagation direction 94
14. Stability criteria 95

REFERENCES 95
Figure 1  The principal types of crystalline binding. In (a) neutral atoms with closed electron shells are bound together weakly by the van der Waals forces associated with fluctuations in the charge distributions. In (b) electrons are transferred from the alkali atoms to the halogen atoms, and the resulting ions are held together by attractive electrostatic forces between the positive and negative ions. In (c) the valence electrons are taken away from each alkali atom to form a communal electron sea in which the positive ions are dispersed. In (d) the neutral atoms appear to be bound together by the overlapping parts of their electron distributions.
CHAPTER 3: CRYSTAL BINDING

In this chapter we are concerned with the question: What holds a crystal together? The attractive electrostatic interaction between the negative charges of the electrons and the positive charges of the nuclei is entirely responsible for the cohesion of solids. Magnetic forces have only a weak effect on cohesion, and gravitational forces are negligible. Specialized terms categorize distinctive situations: exchange energy, van der Waals forces, and covalent bonds. The observed differences between the forms of condensed matter are caused in the final analysis by differences in the distribution of the outermost electrons and the ion cores (Fig. 1).

The cohesive energy of a crystal is defined as the energy that must be added to the crystal to separate its components into neutral free atoms at rest, at infinite separation, with the same electronic configuration. The term lattice energy is used in the discussion of ionic crystals and is defined as the energy that must be added to the crystal to separate its component ions into free ions at rest at infinite separation.

Values of the cohesive energy of the crystalline elements are given in Table 1. Notice the wide variation in cohesive energy between different columns of the periodic table. The inert gas crystals are weakly bound, with cohesive energies less than a few percent of the cohesive energies of the elements in the C, Si, Ge . . . column. The alkali metal crystals have intermediate values of the cohesive energy. The transition element metals (in the middle columns) are quite strongly bound. The melting temperatures (Table 2) and bulk modulii (Table 3) vary roughly as the cohesive energies.

CRYSTALS OF INERT GASES

The inert gases form the simplest crystals. The electron distribution is very close to that of the free atoms. Their properties at absolute zero are summarized in Table 4. The crystals are transparent insulators, weakly bound, with low melting temperatures. The atoms have very high ionization energies (see Table 5). The outermost electron shells of the atoms are completely filled, and the distribution of electron charge in the free atom is spherically symmetric. In the crystal the inert gas atoms pack together as closely as possible: the crystal structures (Fig. 2) are all cubic close-packed (fcc), except He³ and He⁴.

1Zero-point motion of the atoms (kinetic energy at absolute zero) is a quantum effect that plays a dominant role in He³ and He⁴. They do not solidify at zero pressure even at absolute zero. The average fluctuation at 0 K of a He atom from its equilibrium position is of the order of 30 to 40 percent of the nearest-neighbor distance. The heavier the atom, the less important are the zero-point effects. If we omit zero-point motion, we calculate a molar volume of 9 cm³ mol⁻¹ for solid helium, as compared with the observed values of 27.5 and 36.8 cm³ mol⁻¹ for liquid He⁴ and liquid He³, respectively. In the ground state of helium we must take account of the zero-point motion of the atoms.
What holds an inert gas crystal together? The electron distribution in the crystal cannot be significantly distorted from the electron distribution around the free atoms, because the cohesive energy of an atom in the crystal is only 1 percent or less of the ionization energy of an atomic electron. Thus not much energy is available to distort the free atom charge distributions. Part of this distortion gives the van der Waals interaction.

*Van der Waals-London Interaction*

Consider two identical inert gas atoms at a separation $R$ large in comparison with the radii of the atoms. What interactions exist between the two neutral atoms? If the charge distributions on the atoms were rigid, the interaction between atoms would be zero, because the electrostatic potential of a spherical distribution of electronic charge is canceled outside a neutral atom by the electrostatic potential of the charge on the nucleus. Then the inert gas atoms could show no cohesion and could not condense. But the atoms induce dipole moments in each other, and the induced moments cause an attractive interaction between the atoms.
Energy required to form separated neutral atoms in their ground electronic state from the solid at 0 K at 1 atm. The data were supplied by Prof. Leo Brewer in units kcal per mole, revised to May 4, 1977, after LBL Report 3720 Rev.

<table>
<thead>
<tr>
<th>Li</th>
<th>Be</th>
<th>Na</th>
<th>Mg</th>
<th>Al</th>
<th>Si</th>
<th>P</th>
<th>S</th>
<th>Cl</th>
<th>Ar</th>
</tr>
</thead>
<tbody>
<tr>
<td>158.</td>
<td>320.</td>
<td>107.</td>
<td>145.</td>
<td>327.</td>
<td>446.</td>
<td>331.</td>
<td>275.</td>
<td>135.</td>
<td>7.74</td>
</tr>
<tr>
<td>1.63</td>
<td>3.32</td>
<td>1.113</td>
<td>1.51</td>
<td>3.39</td>
<td>4.63</td>
<td>3.42</td>
<td>2.85</td>
<td>1.40</td>
<td>0.080</td>
</tr>
<tr>
<td>37.7</td>
<td>76.5</td>
<td>25.67</td>
<td>34.7</td>
<td>78.1</td>
<td>106.7</td>
<td>79.16</td>
<td>65.75</td>
<td>32.2</td>
<td>1.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>K</th>
<th>Ca</th>
<th>Sc</th>
<th>Ti</th>
<th>V</th>
<th>Cr</th>
<th>Mn</th>
<th>Fe</th>
<th>Co</th>
<th>Ni</th>
<th>Cu</th>
<th>Zn</th>
<th>Ga</th>
<th>Ge</th>
<th>As</th>
<th>Se</th>
<th>Br</th>
<th>Kr</th>
</tr>
</thead>
<tbody>
<tr>
<td>90.1</td>
<td>178.</td>
<td>376.</td>
<td>468.</td>
<td>512.</td>
<td>395.</td>
<td>282.</td>
<td>413.</td>
<td>492.</td>
<td>336.</td>
<td>130.</td>
<td>271.</td>
<td>372.</td>
<td>285.3</td>
<td>237.</td>
<td>118.</td>
<td>11.2</td>
<td></td>
</tr>
<tr>
<td>0.934</td>
<td>1.48</td>
<td>3.90</td>
<td>4.85</td>
<td>5.31</td>
<td>4.10</td>
<td>2.92</td>
<td>4.28</td>
<td>4.39</td>
<td>4.44</td>
<td>3.49</td>
<td>1.85</td>
<td>2.81</td>
<td>3.85</td>
<td>2.96</td>
<td>2.46</td>
<td>1.22</td>
<td>0.116</td>
</tr>
<tr>
<td>21.54</td>
<td>42.5</td>
<td>89.9</td>
<td>111.8</td>
<td>122.4</td>
<td>94.5</td>
<td>67.4</td>
<td>98.7</td>
<td>101.3</td>
<td>102.4</td>
<td>80.4</td>
<td>31.04</td>
<td>64.8</td>
<td>88.8</td>
<td>68.2</td>
<td>56.7</td>
<td>28.18</td>
<td>2.68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rb</th>
<th>Sr</th>
<th>Y</th>
<th>Zr</th>
<th>Nb</th>
<th>Mo</th>
<th>Tc</th>
<th>Ru</th>
<th>Rh</th>
<th>Pd</th>
<th>Ag</th>
<th>Cd</th>
<th>In</th>
<th>Sn</th>
<th>Sb</th>
<th>Te</th>
<th>I</th>
<th>Xe</th>
</tr>
</thead>
<tbody>
<tr>
<td>82.2</td>
<td>166.</td>
<td>422.</td>
<td>603.</td>
<td>659.</td>
<td>661.</td>
<td>650.</td>
<td>554.</td>
<td>376.</td>
<td>284.</td>
<td>112.</td>
<td>243.</td>
<td>303.</td>
<td>265.</td>
<td>211.</td>
<td>107.</td>
<td>15.9</td>
<td></td>
</tr>
<tr>
<td>0.852</td>
<td>1.72</td>
<td>4.37</td>
<td>6.25</td>
<td>7.57</td>
<td>6.82</td>
<td>6.85</td>
<td>6.74</td>
<td>5.75</td>
<td>3.89</td>
<td>2.95</td>
<td>1.16</td>
<td>2.52</td>
<td>3.44</td>
<td>2.75</td>
<td>2.19</td>
<td>1.11</td>
<td>0.16</td>
</tr>
<tr>
<td>19.64</td>
<td>39.7</td>
<td>100.8</td>
<td>144.2</td>
<td>174.5</td>
<td>157.2</td>
<td>158.</td>
<td>155.4</td>
<td>132.5</td>
<td>89.8</td>
<td>68.0</td>
<td>26.73</td>
<td>58.1</td>
<td>72.4</td>
<td>63.4</td>
<td>50.34</td>
<td>25.62</td>
<td>3.80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cs</th>
<th>Ba</th>
<th>La</th>
<th>Hf</th>
<th>Ta</th>
<th>W</th>
<th>Re</th>
<th>Os</th>
<th>Ir</th>
<th>Pt</th>
<th>Au</th>
<th>Hg</th>
<th>Tl</th>
<th>Pb</th>
<th>Bi</th>
<th>Po</th>
<th>At</th>
<th>Rn</th>
</tr>
</thead>
<tbody>
<tr>
<td>77.6</td>
<td>183.</td>
<td>431.</td>
<td>621.</td>
<td>782.</td>
<td>859.</td>
<td>785.</td>
<td>788.</td>
<td>670.</td>
<td>564.</td>
<td>368.</td>
<td>182.</td>
<td>196.</td>
<td>210.</td>
<td>144.</td>
<td>19.5</td>
<td>19.5</td>
<td></td>
</tr>
<tr>
<td>0.804</td>
<td>1.90</td>
<td>4.47</td>
<td>6.44</td>
<td>8.10</td>
<td>8.90</td>
<td>8.03</td>
<td>8.17</td>
<td>6.94</td>
<td>5.84</td>
<td>3.81</td>
<td>3.81</td>
<td>1.88</td>
<td>2.03</td>
<td>2.16</td>
<td>1.50</td>
<td>0.202</td>
<td>4.66</td>
</tr>
<tr>
<td>18.54</td>
<td>43.7</td>
<td>103.1</td>
<td>148.4</td>
<td>186.9</td>
<td>205.2</td>
<td>185.2</td>
<td>188.4</td>
<td>160.1</td>
<td>134.7</td>
<td>87.96</td>
<td>15.5</td>
<td>43.4</td>
<td>46.78</td>
<td>50.2</td>
<td>34.5</td>
<td>4.66</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fr</th>
<th>Ra</th>
<th>Ac</th>
<th>Ce</th>
<th>Pr</th>
<th>Nd</th>
<th>Pm</th>
<th>Sm</th>
<th>Eu</th>
<th>Gd</th>
<th>Tb</th>
<th>Dy</th>
<th>Ho</th>
<th>Er</th>
<th>Tm</th>
<th>Yb</th>
<th>Lu</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.2</td>
<td>98.9</td>
<td>6.20</td>
<td>536.</td>
<td>456.</td>
<td>347.</td>
<td>264.</td>
<td>385.</td>
<td>288.</td>
<td>399.</td>
<td>70.2</td>
<td>72.3</td>
<td>75.8</td>
<td>55.8</td>
<td>37.1</td>
<td>102.2</td>
<td>38.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Th</th>
<th>Pa</th>
<th>U</th>
<th>Np</th>
<th>Pu</th>
<th>Am</th>
<th>Cm</th>
<th>Bk</th>
<th>Cf</th>
<th>Es</th>
<th>Fm</th>
<th>Md</th>
<th>No</th>
<th>Lr</th>
</tr>
</thead>
<tbody>
<tr>
<td>598.</td>
<td>620.</td>
<td>128.</td>
<td>109.</td>
<td>83.0</td>
<td>63.0</td>
<td>92.1</td>
<td>39.9</td>
<td>7.74</td>
<td>0.080</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 Crystal Binding
Table 2 Melting points, in K.

(After R. H. Lamoreaux, LBL Report 4995)
Table 3  Isothermal bulk modulii and compressibilities at room temperature

After K. Gschneidner, Jr., Solid state physics 16, 275-426 (1964); several data are from F. Birch, in Handbook of physical constants, Geological Society of America Memoir 97, 107-173 (1966). Original references should be consulted when values are needed for research purposes. Values in parentheses are estimates. Letters in parentheses refer to the crystal form. Letters in brackets refer to the temperature:

\[ [a] = 77 \text{ K}; \ [b] = 273 \text{ K}; \ [c] = 1 \text{ K}; \ [d] = 4 \text{ K}; \ [e] = 81 \text{ K}. \]

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>C</th>
<th>N</th>
<th>O</th>
<th>F</th>
<th>Ne</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.78</td>
<td>4.43</td>
<td>0.012</td>
<td>0.562</td>
<td>0.226</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>He</td>
<td>1.168</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bulk modulus in units \(10^{12} \text{ dyn/cm}^2\) or \(10^{11} \text{ N/m}^2\)
Compressibility in units \(10^{-12} \text{ cm}^3/\text{dyn} \times 10^{-11} \text{ m}^3/\text{N}\)

Al  | Si  | P (b) | S (r) | Cl  | Ar (a) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.722</td>
<td>0.988</td>
<td>0.304</td>
<td>0.178</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>1.385</td>
<td>1.012</td>
<td>3.29</td>
<td>5.62</td>
<td>7.9</td>
</tr>
</tbody>
</table>

Li  | Be  | Na  | Mg  |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.116</td>
<td>1.003</td>
<td>0.068</td>
<td>0.354</td>
</tr>
<tr>
<td>8.62</td>
<td>0.997</td>
<td>14.7</td>
<td>2.82</td>
</tr>
<tr>
<td>K</td>
<td>Ca</td>
<td>Sc</td>
<td>Ti</td>
</tr>
<tr>
<td>0.032</td>
<td>0.152</td>
<td>0.435</td>
<td>1.051</td>
</tr>
<tr>
<td>31</td>
<td>6.58</td>
<td>2.30</td>
<td>0.951</td>
</tr>
</tbody>
</table>

Rb  | Sr  | Y   | Zr  | Nb  | Mo  | Tc  | Ru  | Rh  | Pd  | Ag  | Cd  | In  | Sn (g) | Sb  | Te  | I   | Xe   |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.031</td>
<td>0.116</td>
<td>0.366</td>
<td>0.833</td>
<td>1.702</td>
<td>2.725</td>
<td>2.97</td>
<td>3.208</td>
<td>2.704</td>
<td>1.808</td>
<td>1.007</td>
<td>0.467</td>
<td>0.411</td>
<td>1.11</td>
<td>0.383</td>
<td>0.230</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>8.62</td>
<td>2.73</td>
<td>1.20</td>
<td>0.587</td>
<td>0.366</td>
<td>0.34</td>
<td>0.311</td>
<td>0.369</td>
<td>0.553</td>
<td>0.993</td>
<td>2.14</td>
<td>2.43</td>
<td>0.901</td>
<td>2.61</td>
<td>4.35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cs  | Ba  | La  | Hf  | Ta  | W   | Re  | Os  | Ir  | Pt  | Au  | Hg (c) | Tl  | Pb  | Bi  | Po  | At  | Rn |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.020</td>
<td>0.103</td>
<td>0.243</td>
<td>1.09</td>
<td>2.00</td>
<td>3.232</td>
<td>3.72</td>
<td>4.18</td>
<td>3.55</td>
<td>2.783</td>
<td>1.732</td>
<td>0.382</td>
<td>0.359</td>
<td>0.430</td>
<td>0.315</td>
<td>(0.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>9.97</td>
<td>4.12</td>
<td>0.92</td>
<td>0.50</td>
<td>0.309</td>
<td>0.269</td>
<td>(0.24)</td>
<td>0.282</td>
<td>0.359</td>
<td>0.577</td>
<td>2.60</td>
<td>2.79</td>
<td>2.33</td>
<td>3.17</td>
<td>(3.8)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fr  | Ra  | Ac  | Ce (y) | Pr  | Nd  | Pm  | Sm  | Eu  | Gd  | Tb  | Dy  | Ho  | Er  | Tm  | Yb  | Lu  |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.020</td>
<td>0.132</td>
<td>0.132</td>
<td>0.239</td>
<td>0.306</td>
<td>0.327</td>
<td>0.35</td>
<td>0.294</td>
<td>0.147</td>
<td>0.383</td>
<td>0.399</td>
<td>0.364</td>
<td>0.397</td>
<td>0.411</td>
<td>0.397</td>
<td>0.133</td>
<td>0.411</td>
</tr>
<tr>
<td>50</td>
<td>(7.6)</td>
<td>(4.3)</td>
<td>4.18</td>
<td>3.27</td>
<td>3.06</td>
<td>(2.85)</td>
<td>3.40</td>
<td>6.80</td>
<td>2.61</td>
<td>2.51</td>
<td>2.60</td>
<td>2.52</td>
<td>2.43</td>
<td>2.52</td>
<td>7.52</td>
<td>2.43</td>
</tr>
</tbody>
</table>

Th | Pa  | U   | Np  | Pu  | Am  | Cm  | Bk  | Cf  | Es  | Fm  | Md  | No  | Lr  |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.543</td>
<td>0.76</td>
<td>0.987</td>
<td>0.68</td>
<td>0.54</td>
<td>1.84</td>
<td>0.3</td>
<td>1.01</td>
<td>1.5</td>
<td>1.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As a model we consider two identical linear harmonic oscillators 1 and 2 separated by $R$. Each oscillator bears charges $\pm e$ with separations $x_1$ and $x_2$, as in Fig. 3. The particles oscillate along the $x$ axis. Let $p_1$ and $p_2$ denote the momenta. The force constant is $C$. Then the Hamiltonian of the unperturbed system is

$$\mathcal{H}_0 = \frac{1}{2m} p_1^2 + \frac{1}{2} C x_1^2 + \frac{1}{2m} p_2^2 + \frac{1}{2} C x_2^2.$$  

(1)

Each uncoupled oscillator is assumed to have the frequency $\omega_0$ of the strongest optical absorption line of the atom. Thus $C = m\omega_0^2$.

Let $\mathcal{H}_1$ be the coulomb interaction energy of the two oscillators. The geometry is shown in the figure. The internuclear coordinate is $R$. Then

$$\mathcal{H}_1 = \frac{e^2}{R} + \frac{e^2}{R + x_1 - x_2} - \frac{e^2}{R + x_1} - \frac{e^2}{R - x_2};$$  

(CGS)  

(2)

in the approximation $|x_1|, |x_2| \ll R$ we expand (2) to obtain in lowest order:

$$\mathcal{H}_1 \approx -\frac{2e^2 x_1 x_2}{R^3}.$$  

(3)

Table 4  Properties of inert gas crystals

(Extrapolated to 0 K and zero pressure)

<table>
<thead>
<tr>
<th>Nearest-neighbor distance, in Å</th>
<th>Experimental cohesive energy, kJ/mol</th>
<th>Experimental cohesive energy, eV/atom</th>
<th>Melting point, K</th>
<th>Ionization potential of free atom, eV</th>
<th>Parameters in Lennard-Jones potential, Eq. 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>He (liquid at zero pressure)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ne</td>
<td>3.13</td>
<td>1.88</td>
<td>0.02</td>
<td>24</td>
<td>14 in $10^{-16}$ erg</td>
</tr>
<tr>
<td>Ar</td>
<td>3.76</td>
<td>7.74</td>
<td>0.080</td>
<td>84</td>
<td>50 in $10^{-16}$ erg</td>
</tr>
<tr>
<td>Kr</td>
<td>4.01</td>
<td>11.2</td>
<td>0.116</td>
<td>117</td>
<td>167 in $10^{-16}$ erg</td>
</tr>
<tr>
<td>Xe</td>
<td>4.35</td>
<td>16.0</td>
<td>0.17</td>
<td>161</td>
<td>225 in $10^{-16}$ erg</td>
</tr>
</tbody>
</table>


Table 5 Ionization energies

The total energy required to remove the first two electrons is the sum of the first and second ionization potentials. Source: National Bureau of Standards Circular 467.

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>Li</th>
<th>Be</th>
<th>Na</th>
<th>Mg</th>
<th>Al</th>
<th>Si</th>
<th>P</th>
<th>S</th>
<th>Cl</th>
<th>Ar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>13.595</td>
<td>5.39</td>
<td>9.32</td>
<td>5.14</td>
<td>7.64</td>
<td>5.40</td>
<td>5.56</td>
<td>6.85</td>
<td>5.34</td>
<td>8.15</td>
<td>15.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>81.01</td>
<td>27.53</td>
<td>52.43</td>
<td>22.67</td>
<td>30.99</td>
<td>20.46</td>
<td>22.98</td>
<td>30.20</td>
<td>36.81</td>
<td>43.38</td>
</tr>
</tbody>
</table>

---

- Energy to remove one electron, in eV
- Energy to remove two electrons, in eV

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>C</th>
<th>N</th>
<th>O</th>
<th>F</th>
<th>Ne</th>
</tr>
</thead>
<tbody>
<tr>
<td>71</td>
<td>8.30</td>
<td>11.26</td>
<td>14.54</td>
<td>13.61</td>
<td>17.42</td>
<td>21.56</td>
</tr>
<tr>
<td>112</td>
<td>33.45</td>
<td>35.64</td>
<td>44.14</td>
<td>48.76</td>
<td>52.40</td>
<td>62.63</td>
</tr>
</tbody>
</table>

---

3 Crystal Binding
The total hamiltonian with the approximate form (3) for \( \mathcal{H}_1 \) can be diagonalized by the normal mode transformation

\[
x_s = \frac{1}{\sqrt{2}}(x_1 + x_2) \quad ; \quad x_a = \frac{1}{\sqrt{2}}(x_1 - x_2) ,
\]

(4)
or, on solving for \( x_1 \) and \( x_2 \),

\[
x_1 = \frac{1}{\sqrt{2}}(x_s + x_a) \quad ; \quad x_2 = \frac{1}{\sqrt{2}}(x_s - x_a) .
\]

(5)
The subscripts \( s \) and \( a \) denote symmetric and antisymmetric modes of motion. Further, we have the momenta \( p_s, p_a \) associated with the two modes:

\[
p_1 = \frac{1}{\sqrt{2}}(p_s + p_a) \quad ; \quad p_2 = \frac{1}{\sqrt{2}}(p_s - p_a) .
\]

(6)
The total hamiltonian \( \mathcal{H}_0 + \mathcal{H}_1 \) after the transformations (5) and (6) is

\[
\mathcal{H} = \left[ \frac{1}{2m} p_s^2 + \frac{1}{2} \left( C - \frac{2\epsilon^2}{R^3} \right) x_s^2 \right] + \left[ \frac{1}{2m} p_a^2 + \frac{1}{2} \left( C + \frac{2\epsilon^2}{R^3} \right) x_a^2 \right] .
\]

(7)
The two frequencies of the coupled oscillators are found by inspection of (7) to be

\[
\omega = \left[ \left( C \pm \frac{2\epsilon^2}{R^3} \right) / m \right]^{1/2} = \omega_0 \left[ 1 \pm \frac{1}{2} \left( \frac{2\epsilon^2}{CR^3} \right) - \frac{1}{8} \left( \frac{2\epsilon^2}{CR^3} \right)^2 + \cdots \right] ,
\]

(8)
with \( \omega_0 \) given by \( (C/m)^{1/2} \). In (8) we have expanded the square root.

The zero point energy of the system is \( \frac{1}{2}\hbar(\omega_s + \omega_a) \), because of the interaction the sum is lowered from the uncoupled value \( 2 \cdot \frac{1}{2}\hbar \omega_0 \) by

\[
\Delta U = \frac{1}{2}\hbar(\Delta \omega_s + \Delta \omega_a) = -\hbar \omega_0 \cdot \frac{1}{8} \left( \frac{2\epsilon^2}{CR^3} \right)^2 = -\frac{A}{R^6} .
\]

(9)
This attractive interaction varies as the minus sixth power of the separation of the two oscillators.

This is called the van der Waals interaction, known also as the London interaction or the induced dipole-dipole interaction. It is the principal attractive interaction in crystals of inert gases and also in crystals of many organic molecules. The interaction is a quantum effect, in the sense that \( \Delta U \to 0 \) as \( \hbar \to 0 \). Thus the zero point energy of the system is lowered by the dipole-dipole coupling of Eq. (3). The van der Waals interaction does not depend for its existence on any overlap of the change densities of the two atoms.

An approximate value of \( A \) for identical atoms is given by \( \hbar \omega_0 \alpha^2 \), where \( \hbar \omega_0 \) is the energy of the strongest optical absorption line and \( \alpha \) is the electronic polarizability, Chapter 13.
**Repulsive Interaction**

As the two atoms are brought together their charge distributions gradually overlap (Fig. 4), thereby changing the electrostatic energy of the system. At sufficiently close separations the overlap energy is repulsive, in large part because of the Pauli exclusion principle. The elementary statement of the principle is that two electrons cannot have all their quantum numbers equal. When the charge distributions of two atoms overlap, there is a tendency for electrons from atom $B$ to occupy in part states of atom $A$ already occupied by electrons of atom $A$, and vice versa.

The Pauli principle prevents multiple occupancy, and electron distributions of atoms with closed shells can overlap only if accompanied by the partial promotion of electrons to unoccupied high energy states of the atoms. Thus the electron overlap increases the total energy of the system and gives a repulsive contribution to the interaction. An extreme example in which the overlap is complete is shown in Fig. 5.

We make no attempt here to evaluate the repulsive interaction from first principles. Experimental data on the inert gases can be fitted well by an empirical repulsive potential of the form $B/R^{12}$, where $B$ is a positive constant, when used together with a long-range attractive potential of the form of (9). The constants $A$ and $B$ are empirical parameters determined from independent measurements made in the gas phase; the data used include the virial coefficients and the viscosity. It is usual to write the total potential energy of two atoms at separation $R$ as

$$U(R) = 4\varepsilon \left[ \left( \frac{\sigma}{R} \right)^{12} - \left( \frac{\sigma}{R} \right)^{6} \right],$$

where $\varepsilon$ and $\sigma$ are the new parameters, with $4\varepsilon\sigma^6 = A$ and $4\varepsilon\sigma^{12} = B$. The potential (10) is known as the Lennard-Jones potential, Fig. 6. The force between the two atoms is given by $-dU/dr$. Values of $\varepsilon$ and $\sigma$ given in Table 4 can

---

5 The overlap energy naturally depends on the radial distribution of charge about each atom. The mathematical calculation is always complicated even if the charge distribution is known.
Figure 5 The effect of Pauli principle on the repulsive energy: in an extreme example, two hydrogen atoms are pushed together until the protons are almost in contact. The energy of the electron system alone can be taken from observations on atomic He, which has two electrons. In (a) the electrons have antiparallel spins and the Pauli principle has no effect: the electrons are bound by \(-78.98\) eV. In (b) the spins are parallel: the Pauli principle forces the promotion of an electron from a 1s \(\uparrow\) orbital of H to a 2s \(\uparrow\) orbital of He. The electrons now are bound by \(-59.38\) eV, less than (a) by \(19.60\) eV. This is the amount by which the Pauli principle has increased the repulsion. We have omitted the repulsive coulomb energy of the two protons, which is the same in both (a) and (b).

Figure 6 Form of the Lennard-Jones potential (10) which describes the interaction of two inert gas atoms. The minimum occurs at \(R/\sigma = 2^{1/6} \approx 1.12\). Notice how steep the curve is inside the minimum, and how flat it is outside the minimum. The value of \(U\) at the minimum is \(-\epsilon\); and \(U = 0\) at \(R = \sigma\).
be obtained from gas-phase data, so that calculations on properties of the solid do not involve disposable parameters.

Other empirical forms for the repulsive interaction are widely used, in particular the exponential form \( \lambda \exp(-R/\rho) \), where \( \rho \) is a measure of the range of the interaction. This is generally as easy to handle analytically as the inverse power law form.

**Equilibrium Lattice Constants**

If we neglect the kinetic energy of the inert gas atoms, the cohesive energy of an inert gas crystal is given by summing the Lennard-Jones potential (10) over all pairs of atoms in the crystal. If there are \( N \) atoms in the crystal, the total potential energy is

\[
U_{\text{tot}} = \frac{1}{2}N(4\epsilon) \left[ \sum_j' \left( \frac{\sigma}{p_j R} \right)^{12} - \sum_j' \left( \frac{\sigma}{p_j R} \right)^{6} \right],
\]

where \( p_j R \) is the distance between reference atom \( i \) and any other atom \( j \), expressed in terms of the nearest neighbor distance \( R \). The factor \( \frac{1}{2} \) occurs with the \( N \) to compensate for counting twice each pair of atoms.

The summations in (11) have been evaluated, and for the fcc structure

\[
\sum_j' p_j^{-12} = 12.13188 \quad \text{and} \quad \sum_j' p_j^{-6} = 14.45392.
\]

There are 12 nearest-neighbor sites in the fcc structure; we see that the series are rapidly converging and have values not far from 12. The nearest neighbors contribute most of the interaction energy of inert gas crystals. The corresponding sums for the hcp structure are 12.13229 and 14.45489.

If we take \( U_{\text{tot}} \) in (11) as the total energy of the crystal, the equilibrium value \( R_0 \) is given by requiring that \( U_{\text{tot}} \) be a minimum with respect to variations in the nearest neighbor distance \( R \):

\[
\frac{dU_{\text{tot}}}{dR} = 0 = -2N\epsilon \left[ (12)(12.13) \frac{\sigma^{12}}{R^{13}} - (6)(14.45) \frac{\sigma^{6}}{R^{7}} \right],
\]

whence

\[
\frac{R_0}{\sigma} = 1.09,
\]

the same for all elements with an fcc structure. The observed values of \( R_0/\sigma \), using the independently determined values of \( \sigma \) given in Table 4, are:

<table>
<thead>
<tr>
<th>Ne</th>
<th>Ar</th>
<th>Kr</th>
<th>Xe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.14</td>
<td>1.11</td>
<td>1.10</td>
<td>1.09</td>
</tr>
</tbody>
</table>

The agreement with (14) is remarkable. The slight departure of \( R_0/\sigma \) for the lighter atoms from the universal value 1.09 predicted for inert gases can be
explained by zero-point quantum effects. From measurements on the gas phase we have predicted the lattice constant of the crystal.

**Cohesive Energy**

The cohesive energy of inert gas crystals at absolute zero and at zero pressure is obtained by substituting (12) and (14) in (11):

$$U_{\text{tot}}(R) = 2N\varepsilon \left[ (12.13) \left( \frac{\sigma}{R} \right)^{12} - (14.45) \left( \frac{\sigma}{R} \right)^{6} \right],$$

and, at $R = R_0$,

$$U_{\text{tot}}(R_0) = -(2.15)(4N\varepsilon),$$

the same for all inert gases. This is the calculated cohesive energy when the atoms are at rest. Quantum-mechanical corrections act to reduce the binding by 28, 10, 6, and 4 percent of Eq. (16) for Ne, Ar, Kr, and Xe, respectively.

The heavier the atom, the smaller the quantum correction. We can understand the origin of the quantum correction by considering a model in which an atom is confined by fixed boundaries. If the particle has the quantum wavelength $\lambda$, where $\lambda$ is determined by the boundaries, then the particle has kinetic energy $p^2/2M = (h/\lambda)^2/2M$ with the de Broglie relation $p = h/\lambda$ for the connection between the momentum and the wavelength of a particle. On this model the quantum zero-point correction to the energy is inversely proportional to the mass. The final calculated cohesive energies agree with the experimental values of Table 4 within 7 to 1 percent.

One consequence of the quantum kinetic energy is that a crystal of the isotope Ne$^{20}$ is observed to have a larger lattice constant than a crystal of Ne$^{22}$. The higher quantum kinetic energy of the lighter isotope expands the lattice, because the kinetic energy is reduced by expansion. The observed lattice constants (extrapolated to absolute zero from 2.5 K) are Ne$^{20}$, 4.4644 Å; Ne$^{22}$, 4.4559 Å.

**IONIC CRYSTALS**

Ionic crystals are made up of positive and negative ions. The ionic bond results from the electrostatic interaction of oppositely charged ions. Two common crystal structures found for ionic crystals, the sodium chloride and the cesium chloride structures, were shown in Chapter 1.

The electronic configurations of all ions of a simple ionic crystal correspond to closed electronic shells, as in the inert gas atoms. In lithium fluoride the configuration of the neutral atoms are, according to the periodic table in the front endpapers of this book, Li:1s$^2$2s$^2$, F:1s$^2$2s$^2$2p$^5$. The singly charged ions have the configurations Li$^+$:1s$^2$, F$^-$:1s$^2$2s$^2$2p$^6$, as for helium and neon, respectively. Inert gas atoms have closed shells, and the charge distributions are spherically symmetric. We expect that the charge distributions on each ion in
an ionic crystal will have approximately spherical symmetry, with some distortion near the region of contact with neighboring atoms. This picture is confirmed by x-ray studies of electron distributions (Fig. 7).

A quick estimate suggests that we are not misguided in looking to electrostatic interactions for a large part of the binding energy of an ionic crystal. The distance between a positive ion and the nearest negative ion in crystalline sodium chloride is $2.81 \times 10^{-8}$ cm, and the attractive coulomb part of the potential energy of the two ions by themselves is 5.1 eV. This value may be compared (Fig. 8) with the experimental value of 7.9 eV per molecular unit for the lattice energy of crystalline NaCl with respect to separated $\text{Na}^+$ and $\text{Cl}^-$ ions. We now calculate the energy more closely.
Table 6  Electron affinities of negative ions


<table>
<thead>
<tr>
<th>Atom</th>
<th>Electron affinity energy</th>
<th>Atom</th>
<th>Electron affinity energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>0.7542 eV</td>
<td>Si</td>
<td>1.39 eV</td>
</tr>
<tr>
<td>Li</td>
<td>0.62</td>
<td>P</td>
<td>0.74</td>
</tr>
<tr>
<td>C</td>
<td>1.27</td>
<td>S</td>
<td>2.08</td>
</tr>
<tr>
<td>O</td>
<td>1.46</td>
<td>Cl</td>
<td>3.61</td>
</tr>
<tr>
<td>F</td>
<td>3.40</td>
<td>Br</td>
<td>3.36</td>
</tr>
<tr>
<td>Na</td>
<td>0.55</td>
<td>I</td>
<td>3.06</td>
</tr>
<tr>
<td>Al</td>
<td>0.46</td>
<td>K</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Electrostatic or Madelung Energy

The long-range interaction between ions with charge \( \pm q \) is the electrostatic interaction \( \pm q^2/\ell \), attractive between ions of opposite charge and repulsive between ions of the same charge. The ions arrange themselves in whatever crystal structure gives the strongest attractive interaction compatible with the repulsive interaction at short distances between ion cores. The repulsive interactions between ions with inert gas configurations are similar to those between inert gas atoms. The van der Waals part of the attractive interaction in ionic crystals makes a relatively small contribution to the cohesive energy in ionic crystals, of the order of 1 or 2 percent. The main contribution to the binding energy of ionic crystals is electrostatic and is called the Madelung energy.

If \( U_{ij} \) is the interaction energy between ions \( i \) and \( j \), we define a sum \( U_i \) which includes all interactions involving the ion \( i \):

\[
U_i = \sum_j' U_{ij} ,
\]

(17)

where the summation includes all ions except \( j = i \). We suppose that \( U_{ij} \) may be written as the sum of a central field repulsive potential of the form \( \lambda \exp(-r/\ell) \), where \( \lambda \) and \( \ell \) are empirical parameters, and a coulomb potential \( \pm q^2/\ell \). Thus (CGS)

\[
U_{ij} = \lambda \exp(-r_{ij}/\ell) \pm q^2/r_{ij} ,
\]

(18)

where the + sign is taken for the like charges and the − sign for unlike charges. In SI units the coulomb interaction is \( \pm q^2/4\pi \varepsilon_0 r \); we write this section in CGS units in which the coulomb interaction is \( \pm q^2/r \).
The repulsive term describes the fact that each ion resists overlap with the electron distributions of neighboring ions. We treat the strength $\lambda$ and range $\rho$ as constants to be determined from observed values of the lattice constant and compressibility; we have used the exponential form of the empirical repulsive potential rather than the $R^{-12}$ form used for the inert gases. The change is made because it may give a better representation of the repulsive interaction. For the ions, we do not have gas-phase data available to permit the independent determination of $\lambda$ and $\rho$. We note that $\rho$ is a measure of the range of the repulsive interaction: when $r = \rho$, the repulsive interaction is reduced to $e^{-1}$ of the value at $r = 0$.

In the NaCl structure the value of $U_i$ does not depend on whether the reference ion $i$ is a positive or a negative ion. The sum in (17) can be arranged to converge rapidly, so that its value will not depend on the site of the reference ion in the crystal, as long as it is not near the surface. We neglect surface effects and write the total lattice energy $U_{\text{tot}}$ of a crystal composed of $N$ molecules or 2N ions as $U_{\text{tot}} = NU_i$. Here $N$, rather than 2N, occurs because we must count each pair of interactions only once or each bond only once. The total lattice energy is defined as the energy required to separate the crystal into individual ions at an infinite distance apart.

It is convenient again to introduce quantities $p_{ij}$ such that $r_{ij} = p_{ij}R$, where $R$ is the nearest-neighbor separation in the crystal. If we include the repulsive interaction only among nearest neighbors, we have

$$U_{ij} = \begin{cases} \lambda \exp(-R/\rho) - \frac{q^2}{R} & \text{(nearest neighbors)} \\ \pm \frac{1}{p_{ij}} \frac{q^2}{R} & \text{(otherwise)} \end{cases}$$

Thus

$$U_{\text{tot}} = NU_i = N\left(z\lambda e^{-R/\rho} - \frac{\alpha q^2}{R}\right),$$

where $z$ is the number of nearest neighbors of any ion and

$$\alpha = \sum_j \frac{(\pm)}{p_{ij}} = \text{Madelung constant}.$$

The sum should include the nearest-neighbor contribution, which is just $z$. The $(\pm)$ sign is discussed just before (25). The value of the Madelung constant is of central importance in the theory of an ionic crystal. Methods for its calculation are discussed below.
At the equilibrium separation \( dU_{\text{tot}}/dR = 0 \), so that

(CGS) \[
N \frac{dU_i}{dR} = -\frac{NZ\lambda}{\rho} \exp(-R/\rho) + \frac{N\alpha q^2}{R^2} = 0 ,
\]

or

(CGS) \[
R_0^2 \exp(-R_0/\rho) = \rho \alpha q^2 / 2\lambda .
\]

This determines the equilibrium separation \( R_0 \) if the parameters \( \rho, \lambda \) of the repulsive interaction are known. For SI, replace \( q^2 \) by \( q^2/4\pi\varepsilon_0 \).

The total lattice energy of the crystal of \( 2N \) ions at their equilibrium separation \( R_0 \) may be written, using (20) and (23), as

(CGS) \[
U_{\text{tot}} = -\frac{N\alpha q^2}{R_0} \left( 1 - \frac{\rho}{R_0} \right) .
\]

The term \(-N\alpha q^2/R_0\) is the Madelung energy. We shall find that \( \rho \) is of the order of \( 0.1R_0 \), so that the repulsive interaction has a very short range.

**Evaluation of the Madelung Constant**

The first calculation of the coulomb energy constant \( \alpha \) was made by Madelung. A powerful general method for lattice sum calculations was developed by Ewald and is developed in Appendix B. Computers are now used for the calculations.

The definition of the Madelung constant \( \alpha \) is, by (21),

\[
\alpha = \sum_j' \frac{(\pm)}{p_j} .
\]

For (20) to give a stable crystal it is necessary that \( \alpha \) be positive. If we take the reference ion as a negative charge the plus sign will apply to positive ions and the minus sign to negative ions.

An equivalent definition is

\[
\frac{\alpha}{R} = \sum_j' \frac{(\pm)}{r_j} ,
\]

where \( r_j \) is the distance of the \( j \)th ion from the reference ion and \( R \) is the nearest-neighbor distance. It must be emphasized that the value given for \( \alpha \) will depend on whether it is defined in terms of the nearest-neighbor distance \( R \) or in terms of the lattice parameter \( a \) or in terms of some other relevant length.

As an example, we compute the Madelung constant for the infinite line of ions of alternating sign in Fig. 9. Pick a negative ion as reference ion, and let \( R \)
Figure 9  Line of ions of alternating signs, with distance $R$ between ions.

denote the distance between adjacent ions. Then

$$\frac{\alpha}{R} = 2\left[\frac{1}{R} - \frac{1}{2R} + \frac{1}{3R} - \frac{1}{4R} + \cdots\right],$$

or

$$\alpha = 2\left[\frac{1}{2} - \frac{1}{3} + \frac{1}{4} + \cdots\right];$$

the factor 2 occurs because there are two ions, one to the right and one to the left, at equal distances $r_j$. We sum this series by the expansion

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots.$$ 

Thus the Madelung constant for the one-dimensional chain is $\alpha = 2 \ln 2$.

In three dimensions the series presents greater difficulty. It is not possible to write down the successive terms by a casual inspection. More important, the series will not converge unless the successive terms in the series are arranged so that the contributions from the positive and negative terms nearly cancel.

Typical values of the Madelung constant are listed below, based on unit charges and referred to the nearest-neighbor distance:

<table>
<thead>
<tr>
<th>Structure</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium chloride, NaCl</td>
<td>1.747565</td>
</tr>
<tr>
<td>Cesium chloride, CsCl</td>
<td>1.762675</td>
</tr>
<tr>
<td>Zinc blende, cubic ZnS</td>
<td>1.6381</td>
</tr>
</tbody>
</table>

The Madelung and repulsive contributions to the binding of a KCl crystal are shown in Fig. 10. Properties of alkali halide crystals having the sodium chloride structure are given in Table 7. The calculated values of the lattice energy are in exceedingly good agreement with the observed values.
Figure 10  Energy per molecule of KCl crystal, showing Madelung and repulsive contributions.

COVALENT CRYSTALS

The covalent bond is the classical electron pair or homopolar bond of chemistry, particularly of organic chemistry. It is a strong bond: the bond between two carbon atoms in diamond with respect to separated neutral atoms is comparable with the bond strength in ionic crystals.

The covalent bond has strong directional properties (Fig. 11). Thus carbon, silicon, and germanium have the diamond structure, with atoms joined to four nearest neighbors at tetrahedral angles, even though this arrangement gives a low filling of space, 0.34 of the available space, compared with 0.74 for a close-packed structure. The tetrahedral bond allows only four nearest neighbors, whereas a close-packed structure has 12. We should not overemphasize the similarity of the bonding of carbon and silicon. Carbon gives biology, but silicon gives geology and semiconductor technology.

The covalent bond is usually formed from two electrons, one from each atom participating in the bond. The electrons forming the bond tend to be partly localized in the region between the two atoms joined by the bond. The spins of the two electrons in the bond are antiparallel.
Table 7 Properties of alkali halide crystals with the NaCl structure

All values (except those in brackets) at room temperature and atmospheric pressure, with no correction for changes in $R_0$ and $U$ from absolute zero. Values in brackets at absolute zero temperature and zero pressure, from private communication by L. Brewer.

<table>
<thead>
<tr>
<th></th>
<th>Nearest-neighbor separation $R_h$, in Å</th>
<th>Bulk modulus $B$, in $10^{11}$ dyn/cm$^2$ or $10^{10}$ N/m$^2$</th>
<th>Repulsive energy parameter $z\lambda$, in $10^{-8}$ erg</th>
<th>Repulsive range parameter $\rho$, in Å</th>
<th>Lattice energy compared to free ions, in kcal/mol</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiF</td>
<td>2.014</td>
<td>6.71</td>
<td>0.296</td>
<td>0.291</td>
<td>242.3[246.8]</td>
</tr>
<tr>
<td>LiCl</td>
<td>2.570</td>
<td>2.98</td>
<td>0.490</td>
<td>0.330</td>
<td>198.9[201.8]</td>
</tr>
<tr>
<td>LiBr</td>
<td>2.751</td>
<td>2.38</td>
<td>0.591</td>
<td>0.340</td>
<td>189.8</td>
</tr>
<tr>
<td>LiI</td>
<td>3.000</td>
<td>(1.71)</td>
<td>0.599</td>
<td>0.366</td>
<td>177.7</td>
</tr>
<tr>
<td>NaF</td>
<td>2.317</td>
<td>4.65</td>
<td>0.641</td>
<td>0.290</td>
<td>214.4[217.9]</td>
</tr>
<tr>
<td>NaCl</td>
<td>2.820</td>
<td>2.40</td>
<td>1.05</td>
<td>0.321</td>
<td>182.6[185.3]</td>
</tr>
<tr>
<td>NaBr</td>
<td>2.989</td>
<td>1.99</td>
<td>1.33</td>
<td>0.328</td>
<td>173.6[174.3]</td>
</tr>
<tr>
<td>NaI</td>
<td>3.237</td>
<td>1.51</td>
<td>1.58</td>
<td>0.345</td>
<td>163.2[162.3]</td>
</tr>
<tr>
<td>KF</td>
<td>2.674</td>
<td>3.05</td>
<td>1.31</td>
<td>0.298</td>
<td>189.8[194.5]</td>
</tr>
<tr>
<td>KCl</td>
<td>3.147</td>
<td>1.74</td>
<td>2.05</td>
<td>0.326</td>
<td>165.8[169.5]</td>
</tr>
<tr>
<td>KBr</td>
<td>3.298</td>
<td>1.48</td>
<td>2.30</td>
<td>0.336</td>
<td>158.5[159.3]</td>
</tr>
<tr>
<td>KI</td>
<td>3.533</td>
<td>1.17</td>
<td>2.85</td>
<td>0.348</td>
<td>149.9[151.1]</td>
</tr>
<tr>
<td>RbF</td>
<td>2.815</td>
<td>2.62</td>
<td>1.78</td>
<td>0.301</td>
<td>181.4</td>
</tr>
<tr>
<td>RbCl</td>
<td>3.291</td>
<td>1.56</td>
<td>3.19</td>
<td>0.323</td>
<td>159.3</td>
</tr>
<tr>
<td>RbBr</td>
<td>3.445</td>
<td>1.30</td>
<td>3.03</td>
<td>0.338</td>
<td>152.6</td>
</tr>
<tr>
<td>RbI</td>
<td>3.671</td>
<td>1.06</td>
<td>3.99</td>
<td>0.348</td>
<td>144.9</td>
</tr>
</tbody>
</table>

Data from various tables by M. P. Tosi, Solid state physics 16, 1 (1964).
The binding of molecular hydrogen is a simple example of a covalent bond. The strongest binding (Fig. 12) occurs when the spins of the two electrons are antiparallel. The binding depends on the relative spin orientation not because there are strong magnetic dipole forces between the spins, but because the Pauli principle modifies the distribution of charge according to the spin orientation. This spin-dependent coulomb energy is called the exchange interaction.

The Pauli principle gives a strong repulsive interaction between atoms with filled shells. If the shells are not filled, electron overlap can be accommodated without excitation of electrons to high energy states and the bond will be shorter. Compare the bond length (2 Å) of Cl₂ with the interatomic distance (3.76 Å) of Ar in solid Ar; also compare the cohesive energies given in Table 1. The difference between Cl₂ and Ar₂ is that the Cl atom has five electrons in the 3p shell and the Ar atom has six, filling the shell, so that the repulsive interaction is stronger in Ar than in Cl.

The elements C, Si, and Ge lack four electrons with respect to filled shells, and thus these elements (for example) can have an attractive interaction associated with charge overlap. The electron configuration of carbon is 1s²2s²2p². To form a tetrahedral system of covalent bonds the carbon atom must first be promoted to the electronic configuration 1s²2s²2p³. This promotion from the ground state requires 4 eV, an amount more than regained when the bonds are formed.

There is a continuous range of crystals between the ionic and the covalent limits. It is often important to estimate the extent a given bond is ionic or covalent. A semiempirical theory of the fractional ionic or covalent character of
Figure 12. Energy of molecular hydrogen (H₂) referred to separated neutral atoms. A negative energy corresponds to binding. The curve N refers to a classical calculation with free atom charge densities; A is the result for parallel electron spins, taking the Pauli exclusion principle into account, and S (the stable state) for antiparallel spins. The density of charge is represented by contour lines for the states A and S.

A bond in a dielectric crystal has been developed with considerable success by J. C. Phillips, Table 8, p. 76.

METALS

Metals are characterized by high electrical conductivity, and a large number of electrons in a metal are free to move about, usually one or two per atom. The electrons available to move about are called conduction electrons. The valence electrons of the atom become the conduction electrons of the metal.

In some metals the interaction of the ion cores with the conduction electrons always makes a large contribution to the binding energy, but the characteristic feature of metallic binding is the lowering of the energy of the valence electrons in the metal as compared with the free atom. This lowering is exhibited by several simple models treated in Chapters 7 and 9.
Table 8  Fractional ionic character of bonds in binary crystals

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Fractional ionic character</th>
<th>Crystal</th>
<th>Fractional ionic character</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>0.00</td>
<td>GaAs</td>
<td>0.31</td>
</tr>
<tr>
<td>SiC</td>
<td>0.18</td>
<td>GaSb</td>
<td>0.26</td>
</tr>
<tr>
<td>Ge</td>
<td>0.00</td>
<td>AgCl</td>
<td>0.86</td>
</tr>
<tr>
<td>ZnO</td>
<td>0.62</td>
<td>AgBr</td>
<td>0.85</td>
</tr>
<tr>
<td>ZnS</td>
<td>0.62</td>
<td>AgI</td>
<td>0.77</td>
</tr>
<tr>
<td>ZnSe</td>
<td>0.63</td>
<td>MgO</td>
<td>0.84</td>
</tr>
<tr>
<td>ZnTe</td>
<td>0.61</td>
<td>MgS</td>
<td>0.79</td>
</tr>
<tr>
<td>CdO</td>
<td>0.79</td>
<td>MgSe</td>
<td>0.79</td>
</tr>
<tr>
<td>CdS</td>
<td>0.69</td>
<td>LiF</td>
<td>0.92</td>
</tr>
<tr>
<td>CdSe</td>
<td>0.70</td>
<td>NaCl</td>
<td>0.94</td>
</tr>
<tr>
<td>CdTe</td>
<td>0.67</td>
<td>RbF</td>
<td>0.96</td>
</tr>
<tr>
<td>InP</td>
<td>0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>InAs</td>
<td>0.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>InSb</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


The binding energy of an alkali metal crystal is considerably less than that of an alkali halide crystal: the bond formed by a conduction electron is not very strong. The interatomic distances are relatively large in the alkali metals because the kinetic energy of the conduction electrons is lower at large interatomic distances. This leads to weak binding. Metals tend to crystallize in relatively close packed structures: hcp, fcc, bcc, and some other closely related structures, and not in loosely-packed structures such as diamond.

In the transition metals there is additional binding from inner electron shells. Transition metals and the metals immediately following them in the periodic table have large d-electron shells and are characterized by high binding energy.

**HYDROGEN BONDS**

Because neutral hydrogen has only one electron, it should form a covalent bond with only one other atom. It is known, however, that under certain conditions an atom of hydrogen is attracted by rather strong forces to two atoms, thus forming a hydrogen bond between them, with a bond energy of the order of 0.1 eV. It is believed that the hydrogen bond is largely ionic in character, being formed only between the most electronegative atoms, particularly F, O, and N.
Figure 13 The hydrogen difluoride ion HF$_2$ is stabilized by a hydrogen bond. The sketch is of an extreme model of the bond, extreme in the sense that the proton is shown bare of electrons.

In the extreme ionic form of the hydrogen bond, the hydrogen atom loses its electron to another atom in the molecule; the bare proton forms the hydrogen bond. The atoms adjacent to the proton are so close that more than two of them would get in each other's way; thus the hydrogen bond connects only two atoms (Fig. 13).

The hydrogen bond is an important part of the interaction between H$_2$O molecules and is responsible together with the electrostatic attraction of the electric dipole moments for the striking physical properties of water and ice. It is important in certain ferroelectric crystals (Chapter 13).

**ATOMIC RADII**

Distances between atoms in crystals can be measured very accurately by x-ray diffraction, often to 1 part in $10^5$. Can we say that the observed distance between atoms may be assigned partly to atom A and partly to atom B? Can a definite meaning be assigned to the radius of an atom or an ion, irrespective of the nature and composition of the crystal?

Strictly, the answer is no. The charge distribution around an atom is not limited by a rigid spherical boundary. Nonetheless, the concept of an atomic radius is fruitful in predicting interatomic spacing. The existence and probable lattice constants of phases that have not yet been synthesized can be predicted from the additive properties of the atomic radii. Further, the electronic configuration of the constituent atoms often can be inferred by comparison of measured and predicted values of the lattice constants.

To make predictions of lattice constants it is convenient to assign (Table 9) sets of self-consistent radii to various types of bonds: one set for ionic crystals with the constituent ions 6-coordinated in inert gas closed-shell configurations, another set for the ions in tetrahedrally-coordinated structures, and another set for 12-coordinated metals (close-packed).

The predicted self-consistent radii of the cation Na$^+$ and the anion F$^-$ as given in Table 9 would lead to 0.97 Å + 1.36 Å = 2.33 Å for the interatomic separation in the crystal NaF, as compared with the observed 2.32 Å. This agreement is much better than if we assume atomic (neutral) configurations for
Table 9  Atomic and ionic radii

Values approximate only. Units are 1 Å = 10^{-10} m. For original references see W. B. Pearson, *Crystal chemistry and physics of metals and alloys*, Wiley, 1972.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>B</th>
<th>C</th>
<th>N</th>
<th>O</th>
<th>F</th>
<th>Ne</th>
<th>Al</th>
<th>Si</th>
<th>P</th>
<th>S</th>
<th>Cl</th>
<th>Ar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li</td>
<td>Be</td>
<td>0.68</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.58</td>
</tr>
<tr>
<td>Na</td>
<td>Mg</td>
<td>0.97</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.88</td>
</tr>
</tbody>
</table>

When in tetrahedral covalent bonds:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>K</th>
<th>Ca</th>
<th>Sc</th>
<th>Ti</th>
<th>V</th>
<th>Cr</th>
<th>Mn</th>
<th>Fe</th>
<th>Co</th>
<th>Ni</th>
<th>Cu</th>
<th>Zn</th>
<th>Ga</th>
<th>Ge</th>
<th>Al</th>
<th>Si</th>
<th>P</th>
<th>S</th>
<th>Cl</th>
<th>Ar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.33</td>
<td>0.99</td>
<td></td>
<td></td>
<td>0.81</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.38</td>
<td>1.98</td>
<td>1.64</td>
<td>1.46</td>
<td>1.35</td>
<td>1.28</td>
<td>1.26</td>
<td>1.27</td>
<td>1.25</td>
<td>1.25</td>
<td>1.28</td>
<td>1.25</td>
<td>1.39</td>
<td>1.41</td>
<td>1.37</td>
<td>1.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.55</td>
<td>2.15</td>
<td>1.80</td>
<td>1.60</td>
<td>1.47</td>
<td>1.40</td>
<td>1.36</td>
<td>1.34</td>
<td>1.35</td>
<td>1.38</td>
<td>1.45</td>
<td>1.57</td>
<td>1.66</td>
<td>1.55</td>
<td>1.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cs</td>
<td>Ba</td>
<td>1.67</td>
<td>1.35</td>
<td>1.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.73</td>
<td>2.24</td>
<td>1.88</td>
<td>1.58</td>
<td>1.47</td>
<td>1.41</td>
<td>1.38</td>
<td>1.35</td>
<td>1.39</td>
<td>1.44</td>
<td>1.57</td>
<td>1.72</td>
<td>1.75</td>
<td>1.70</td>
<td>1.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When in 12-coordinated metals:

|     |     |  Hf |  Ta | W  | Re | Os  | Ir  | Pt  | Au  | Hg  | Tl  | Pb  | Bi  | Po  | At  | Rn  | Ce  | Pr  | Nd  | Pm  | Sm  | Eu  | Gd  | Tb  | Dy  | Ho  | Er  | Tm  | Yb  | Lu  |
|-----|-----|-----|-----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1.75| 1.37| 1.11|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

When in inert gas (filled shell) configuration:

|     |     |  Th |  Pa | U  | Np | Pu  | Am  | Cm  | Bk  | Cf  | Es  | Fm  | Md  | No  | Lr  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.99| 0.90| 0.83|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

Radii of ions when in tetrahedral covalent bonds:

|     |     |  Th |  Pa | U  | Np | Pu  | Am  | Cm  | Bk  | Cf  | Es  | Fm  | Md  | No  | Lr  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1.80| 1.63| 1.56|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

Radii of ionic radii in 12-coordinated metals:

|     |     |  Th |  Pa | U  | Np | Pu  | Am  | Cm  | Bk  | Cf  | Es  | Fm  | Md  | No  | Lr  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1.80| 1.63| 1.56|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

Standard radii for ions in inert gas (filled shell) configuration:

|     |     |  Th |  Pa | U  | Np | Pu  | Am  | Cm  | Bk  | Cf  | Es  | Fm  | Md  | No  | Lr  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1.80| 1.63| 1.56|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

Th <sub>2</sub>P <sub>4</sub> <sup>2+</sup> 1.58- 1.64 1.81
Table 10 Use of the standard radii of ions given in Table 9

The interionic distance $D$ is represented by $D_N = R_C + R_A + \Delta_N$, for ionic crystals, where $N$ is the coordination number of the cation (positive ion), $R_C$ and $R_A$ are the standard radii of the cation and anion, and $\Delta_N$ is a correction for coordination number. Room temperature. (After Zachariasen.)

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\Delta_N$ (Å)</th>
<th>$N$</th>
<th>$\Delta_N$ (Å)</th>
<th>$N$</th>
<th>$\Delta_N$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.50</td>
<td>5</td>
<td>-0.05</td>
<td>9</td>
<td>+0.11</td>
</tr>
<tr>
<td>2</td>
<td>-0.31</td>
<td>6</td>
<td>0</td>
<td>10</td>
<td>+0.14</td>
</tr>
<tr>
<td>3</td>
<td>-0.19</td>
<td>7</td>
<td>+0.04</td>
<td>11</td>
<td>+0.17</td>
</tr>
<tr>
<td>4</td>
<td>-0.11</td>
<td>8</td>
<td>+0.08</td>
<td>12</td>
<td>+0.19</td>
</tr>
</tbody>
</table>

Na and F, for this would lead to 2.58 Å for the interatomic separation in the crystal. The latter value is $\frac{1}{2}$ (n.n. distance in metallic Na + interatomic distance in gaseous F₂).

The interatomic distance between C atoms in diamond is 1.54 Å; one-half of this is 0.77 Å. In silicon, which has the same crystal structure, one-half the interatomic distance is 1.17 Å. In SiC each atom is surrounded by four atoms of the opposite kind. If we add the C and Si radii just given, we predict 1.94 Å for the length of the C-Si bond, in fair agreement with the 1.89 Å observed for the bond length. This is the kind of agreement (a few percent) that we shall find in using tables of atomic radii.³

Ionic Crystal Radii

In Table 9 we include a set of ionic crystal radii in inert gas configurations. The ionic radii are to be used in conjunction with Table 10. Let us consider BaTiO₃ in Fig. 13.10, with a lattice constant of 4.004 Å at room temperature. Each Ba⁺⁺ ion has 12 nearest O⁻⁻ ions, so that the coordination number is 12 and the correction $\Delta_{12}$ of Table 10 applies. If we suppose that the structure is determined by the Ba-O contacts, we have $D_{12} = 1.35 + 1.40 + 0.19 = 2.94$ Å, or $a = 4.16$ Å; if the Ti-O contact determines the structure, we have $D_6 = 0.68 + 1.40 = 2.08$ or $a = 4.16$ Å. The actual lattice constant is somewhat smaller than the estimates and may perhaps suggest that the bonding is not purely ionic, but is partly covalent.

ANALYSIS OF ELASTIC STRAINS

We consider the elastic properties of a crystal viewed as a homogeneous continuous medium rather than as a periodic array of atoms. The continuum approximation is usually valid for elastic waves of wavelengths $\lambda$ longer than $10^{-6}$ cm, which means for frequencies below $10^{11}$ or $10^{12}$ Hz. Some of the material below looks complicated because of the unavoidable multiplicity of subscripts on the symbols. The basic physical ideas are simple: we use Hooke’s law and Newton’s second law. Hooke’s law states that in an elastic solid the strain is directly proportional to the stress. The law applies to small strains only. We say that we are in the nonlinear region when the strains are so large that Hooke’s law is no longer satisfied.

We specify the strain in terms of the components $e_{xx}$, $e_{yy}$, $e_{zz}$, $e_{xy}$, $e_{yz}$, $e_{zx}$ which are defined below. We treat infinitesimal strains only. We shall not distinguish in our notation between isothermal (constant temperature) and adiabatic (constant entropy) deformations. The small differences between the isothermal and adiabatic elastic constants are not often of importance at room temperature and below.

We imagine that three orthogonal vectors $\mathbf{\hat{x}}$, $\mathbf{\hat{y}}$, $\mathbf{\hat{z}}$ of unit length are embedded securely in the unstrained solid, as shown in Fig. 14. After a small uniform deformation of the solid has taken place the axes are distorted in orientation and in length. In a uniform deformation each primitive cell of the crystal is deformed in the same way. The new axes $x'$, $y'$, $z'$ may be written in terms of the old axes:

$$
x' = (1 + e_{xx})\mathbf{\hat{x}} + e_{xy}\mathbf{\hat{y}} + e_{xz}\mathbf{\hat{z}} ;
$$
$$
y' = e_{yx}\mathbf{\hat{x}} + (1 + e_{yy})\mathbf{\hat{y}} + e_{yz}\mathbf{\hat{z}} ;
$$
$$
z' = e_{zx}\mathbf{\hat{x}} + e_{zy}\mathbf{\hat{y}} + (1 + e_{zz})\mathbf{\hat{z}} .
$$

(26)
The coefficients $e_{\alpha\beta}$ define the deformation; they are dimensionless and have values $\ll 1$ if the strain is small. The original axes were of unit length, but the new axes will not necessarily be of unit length. For example,

$$
x' \cdot x' = 1 + 2e_{xx} + e_{xx}^2 + e_{xy}^2 + e_{xz}^2 ,
$$
whence $x' \approx 1 + e_{xx} + \cdots$. The fractional changes of length of the $\mathbf{\hat{x}}$, $\mathbf{\hat{y}}$, and $\mathbf{\hat{z}}$ axes are $e_{xx}$, $e_{yy}$, $e_{zz}$, respectively, to the first order.

What is the effect of the deformation (26) on an atom originally at $r = x\mathbf{\hat{x}} + y\mathbf{\hat{y}} + z\mathbf{\hat{z}}$? The origin is taken at some other atom. If the deformation is uniform, then after deformation the point will be at the position $^4 r' = xx' + yy' + zz'$. The displacement $R$ of the deformation is defined by

$$
R = r' - r = x(x' - \mathbf{\hat{x}}) + y(y' - \mathbf{\hat{y}}) + z(z' - \mathbf{\hat{z}}) ,
$$

(27)

This is obviously correct if we choose the $\mathbf{\hat{x}}$ axis such that $r = xx\mathbf{\hat{x}}$, then $r' = xx'$ by definition of $x'$.
or, from (26),
\[
R(r) = (x \varepsilon_{xx} + y \varepsilon_{yx} + z \varepsilon_{zx}) \hat{x} + (x \varepsilon_{xy} + y \varepsilon_{yy} + z \varepsilon_{zy}) \hat{y} \\
+ (x \varepsilon_{xz} + y \varepsilon_{yz} + z \varepsilon_{zz}) \hat{z} .
\] (28)

This may be written in a more general form by introducing \( u, v, w \) such that the displacement is given by
\[
R(r) = u(r) \hat{x} + v(r) \hat{y} + w(r) \hat{z} .
\] (29)

If the deformation is nonuniform we must relate \( u, v, w \) to the local strains. We take the origin of \( r \) close to the region of interest; then comparison of (28) and (29) gives, by Taylor series expansion of \( R \) using \( R(0) = 0 \),
\[
x \varepsilon_{xx} = x \frac{\partial u}{\partial x} ; \quad y \varepsilon_{yx} = y \frac{\partial u}{\partial y} ; \quad \text{etc.}
\] (30)

It is usual to work with coefficients \( e_{\alpha \beta} \) rather than \( \varepsilon_{\alpha \beta} \). We define the strain components \( e_{xx}, e_{yy}, e_{zz} \) by the relations
\[
e_{xx} \equiv \varepsilon_{xx} = \frac{\partial u}{\partial x} ; \quad e_{yy} \equiv \varepsilon_{yy} = \frac{\partial v}{\partial y} ; \quad e_{zz} \equiv \varepsilon_{zz} = \frac{\partial w}{\partial z} ,
\] (31)

using (30). The other strain components \( e_{xy}, e_{yz}, e_{xz} \) are defined in terms of the changes in angle between the axes: using (26) we may define
\[
\begin{align*}
    e_{xy} & \equiv x' \cdot y' \equiv \varepsilon_{yx} + \varepsilon_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} ; \\
    e_{yz} & \equiv y' \cdot z' \equiv \varepsilon_{zy} + \varepsilon_{yz} = \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} ; \\
    e_{zx} & \equiv z' \cdot x' \equiv \varepsilon_{zx} + \varepsilon_{xz} = \frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} . 
\end{align*}
\]  

(32)

We may replace the \( \equiv \) signs by \( = \) signs if we neglect terms of order \( \varepsilon^2 \). The six dimensionless coefficients \( e_{\alpha\beta}(=e_{\beta\alpha}) \) completely define the strain.

**Dilation**

The fractional increase of volume associated with a deformation is called the dilation. The dilation is negative for hydrostatic pressure. The unit cube of edges \( \hat{x}, \hat{y}, \hat{z} \) has a volume after deformation of

\[
    V' = x' \cdot y' \times z' ,
\]

by virtue of a well-known result for the volume of a parallelepiped having edges \( x', y', z' \). From (26) we have

\[
    x' \cdot y' \times z' = \begin{vmatrix}
    1 + \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\
    \varepsilon_{yx} & 1 + \varepsilon_{yy} & \varepsilon_{yz} \\
    \varepsilon_{zx} & \varepsilon_{zy} & 1 + \varepsilon_{zz}
\end{vmatrix} \equiv 1 + \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} .
\]

(34)

Products of two strain components have been neglected. The dilation \( \delta \) is then given by

\[
    \delta = \frac{V' - V}{V} \equiv \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} .
\]

(35)

**Stress Components**

The force acting on a unit area in the solid is defined as the stress. There are nine stress components: \( X_x, X_y, X_z, Y_x, Y_y, Y_z, Z_x, Z_y, Z_z \). The capital letter indicates the direction of the force, and the subscript indicates the normal to the plane to which the force is applied. In Fig. 15 the stress component \( X_x \) represents a force applied in the \( x \) direction to a unit area of a plane whose normal lies in the \( x \) direction; the stress component \( X_y \) represents a force applied in the \( x \) direction to a unit area of a plane whose normal lies in the \( y \) direction. The number of independent stress components is reduced from nine to six by applying to an elementary cube (as in Fig. 16) the condition that the angular acceleration vanish,\(^5\) and hence that the total torque must be zero. It

\(^5\)This does not mean we cannot treat problems in which there is angular acceleration; it just means that we can use the static situation to define the elastic constants.
follows that

\[ Y_x = Z_y ; \quad Z_x = X_z ; \quad X_y = Y_x . \] (36)

The six independent stress components may be taken as \(X_x, Y_y, Z_z, Y_z, Z_x, X_y\).

Stress components have the dimensions of force per unit area or energy per unit volume. The strain components are ratios of lengths and are dimensionless.

**ELASTIC COMPLIANCE AND STIFFNESS CONSTANTS**

Hooke's law states that for sufficiently small deformations the strain is directly proportional to the stress, so that the strain components are linear functions of the stress components:
\[ e_{xx} = S_{11}X_x + S_{12}Y_y + S_{13}Z_z + S_{14}X_y + S_{15}X_z + S_{16}Y_y ; \]
\[ e_{yy} = S_{21}X_x + S_{22}Y_y + S_{23}Z_z + S_{24}X_y + S_{25}X_z + S_{26}Y_y ; \]
\[ e_{zz} = S_{31}X_x + S_{32}Y_y + S_{33}Z_z + S_{34}X_y + S_{35}X_z + S_{36}Y_y ; \]
\[ e_{yz} = S_{41}X_x + S_{42}Y_y + S_{43}Z_z + S_{44}X_y + S_{45}X_z + S_{46}Y_y ; \]
\[ e_{zx} = S_{51}X_x + S_{52}Y_y + S_{53}Z_z + S_{54}X_y + S_{55}X_z + S_{56}Y_y ; \]
\[ e_{xy} = S_{61}X_x + S_{62}Y_y + S_{63}Z_z + S_{64}X_y + S_{65}X_z + S_{66}Y_y . \] (37)

\[ X_x = C_{11}e_{xx} + C_{12}e_{yy} + C_{13}e_{zz} + C_{14}e_{xy} + C_{15}e_{zx} + C_{16}e_{zy} ; \]
\[ Y_y = C_{21}e_{xx} + C_{22}e_{yy} + C_{23}e_{zz} + C_{24}e_{xy} + C_{25}e_{zx} + C_{26}e_{zy} ; \]
\[ Z_z = C_{31}e_{xx} + C_{32}e_{yy} + C_{33}e_{zz} + C_{34}e_{xy} + C_{35}e_{zx} + C_{36}e_{zy} ; \]
\[ Y_z = C_{41}e_{xx} + C_{42}e_{yy} + C_{43}e_{zz} + C_{44}e_{xy} + C_{45}e_{zx} + C_{46}e_{zy} ; \]
\[ Z_x = C_{51}e_{xx} + C_{52}e_{yy} + C_{53}e_{zz} + C_{54}e_{yz} + C_{55}e_{zx} + C_{56}e_{yz} ; \]
\[ X_y = C_{61}e_{xx} + C_{62}e_{yy} + C_{63}e_{zz} + C_{64}e_{yz} + C_{65}e_{zx} + C_{66}e_{xy} . \] (38)

The quantities \( S_{11}, \ S_{12}, \ldots \) are called elastic compliance constants or elastic constants; the quantities \( C_{11}, C_{12}, \ldots \) are called the elastic stiffness constants or moduli of elasticity. The \( S \)'s have the dimensions of \([\text{area}] / [\text{force}]\) or \([\text{volume}] / [\text{energy}]\). The \( C \)'s have the dimensions of \([\text{force}] / [\text{area}]\) or \([\text{energy}] / [\text{volume}]\).

**Elastic Energy Density**

The 36 constants in (37) or in (38) may be reduced in number by several considerations. The elastic energy density \( U \) is a quadratic function of the strains, in the approximation of Hooke's law (recall the expression for the energy of a stretched spring). Thus we may write

\[ U = \frac{1}{2} \sum_{\lambda=1}^{6} \sum_{\mu=1}^{6} \tilde{C}_{\lambda\mu} e_{\lambda} e_{\mu} , \] (39)

where the indices 1 through 6 are defined as:

\[ 1 = xx ; \quad 2 = yy ; \quad 3 = zz ; \quad 4 = yz ; \quad 5 = xy ; \quad 6 = xy . \] (40)

The \( \tilde{C} \)'s are related to the \( C \)'s of (38), as we see in (42) below.

The stress components are found from the derivative of \( U \) with respect to the associated strain component. This result follows from the definition of potential energy. Consider the stress \( X_x \) applied to one face of a unit cube, the opposite face being held at rest:

\[ X_x = \frac{\partial U}{\partial e_{xx}} = \frac{\partial U}{\partial e_1} = C_{11}e_1 + \frac{1}{2} \sum_{\beta=2}^{6} (\tilde{C}_{1\beta} + \tilde{C}_{\beta 1}) e_\beta . \] (41)
Note that only the combination \( \frac{1}{2}(\tilde{C}_{\alpha\beta} + \tilde{C}_{\beta\alpha}) \) enters the stress-strain relations. It follows that the elastic stiffness constants are symmetrical:

\[
C_{\alpha\beta} = \frac{1}{2}(\tilde{C}_{\alpha\beta} + \tilde{C}_{\beta\alpha}) = C_{\beta\alpha} .
\] (42)

Thus the thirty-six elastic stiffness constants are reduced to twenty-one.

**Elastic Stiffness Constants of Cubic Crystals**

The number of independent elastic stiffness constants is reduced further if the crystal possesses symmetry elements. We now show that in cubic crystals there are only three independent stiffness constants.

We assert that the elastic energy density of a cubic crystal is

\[
U = \frac{1}{2}C_{11}(\varepsilon_{xx}^2 + \varepsilon_{yy}^2 + \varepsilon_{zz}^2) + \frac{1}{2}C_{44}(\varepsilon_{yz}^2 + \varepsilon_{zx}^2 + \varepsilon_{xy}^2) + C_{12}(\varepsilon_{yy}\varepsilon_{zz} + \varepsilon_{zz}\varepsilon_{xx} + \varepsilon_{xx}\varepsilon_{yy}) ,
\] (43)

and that no other quadratic terms occur; that is

\[
(\varepsilon_{xx}\varepsilon_{xy} + \cdots) ; \quad (\varepsilon_{yx}\varepsilon_{zx} + \cdots) ; \quad (\varepsilon_{xx}\varepsilon_{yz} + \cdots)
\] (44)
do not occur.

The minimum symmetry requirement for a cubic structure is the existence of four three-fold rotation axes. The axes are in the [111] and equivalent directions (Fig. 17). The effect of a rotation of \(2\pi/3\) about these four axes is to interchange the \(x, y, z\) axes according to the schemes

\[
x \rightarrow y \rightarrow z \rightarrow x ; \quad -x \rightarrow z \rightarrow -y \rightarrow -x ;
\]

\[
x \rightarrow z \rightarrow -y \rightarrow x ; \quad -x \rightarrow y \rightarrow z \rightarrow -x ,
\] (45)

according to the axis chosen. Under the first of these schemes, for example,

\[
\varepsilon_{xx}^2 + \varepsilon_{yy}^2 + \varepsilon_{zz}^2 \rightarrow \varepsilon_{yy}^2 + \varepsilon_{zz}^2 + \varepsilon_{xx}^2 ,
\]

![Figure 17 Rotation by 2π/3 about the axis marked 3 changes x → y, y → z, and z → x.](image)
and similarly for the other terms in parentheses in (43). Thus (43) is invariant under the operations considered. But each of the terms exhibited in (44) is odd in one or more indices. A rotation in the set (45) can be found which will change the sign of the term, because \( e_{xy} = -e_{x(-y)} \), for example. Thus the terms (44) are not invariant under the required operations.

It remains to verify that the numerical factors in (43) are correct. By (41)

\[
\frac{\partial U}{\partial e_{xx}} = X_x = C_{11} e_{xx} + C_{12}(e_{yy} + e_{zz}) .
\]

(46)
The appearance of \( C_{11} e_{xx} \) agrees with (38). On further comparison, we see that

\[
C_{12} = C_{13} ; \quad C_{14} = C_{15} = C_{16} = 0 .
\]

(47)

Further, from (43),

\[
\frac{\partial U}{\partial e_{xy}} = X_y = C_{44} e_{xy} ;
\]

(48)
on comparison with (38) we have

\[
C_{61} = C_{62} = C_{63} = C_{64} = C_{65} = 0 ; \quad C_{66} = C_{44} .
\]

(49)

Thus from (43) we find that the array of values of the elastic stiffness constants is reduced for a cubic crystal to the matrix

\[
\begin{array}{cccccc}
  & e_{xx} & e_{yy} & e_{zz} & e_{yz} & e_{zx} & e_{xy} \\
 X_x & C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\
 Y_y & C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\
 Z_z & C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\
 Y_z & 0 & 0 & 0 & C_{44} & 0 & 0 \\
 Z_x & 0 & 0 & 0 & 0 & C_{44} & 0 \\
 X_y & 0 & 0 & 0 & 0 & 0 & C_{44} \\
\end{array}
\]

(50)

For cubic crystals the stiffness and compliance constants are related by

\[
C_{44} = 1/S_{44} ; \quad C_{11} - C_{12} = (S_{11} - S_{12})^{-1} ;
\]

\[
C_{11} + 2C_{12} = (S_{11} + 2S_{12})^{-1} .
\]

(51)

These relations follow on evaluating the inverse matrix to (50).

**Bulk Modulus and Compressibility**

Consider the uniform dilation \( e_{xx} = e_{yy} = e_{zz} = \frac{1}{2} \delta \). For this deformation the energy density (43) of a cubic crystal is

\[
U = \frac{1}{2}(C_{11} + 2C_{12})\delta^2 .
\]

(52)

We may define the bulk modulus \( B \) by the relation

\[
U = \frac{1}{2}B\delta^2 ,
\]

(53)
which is equivalent to the definition \(-V \, dp/dV\). For a cubic crystal

\[ B = \frac{1}{3}(C_{11} + 2C_{12}) \]  

(54)

The compressibility \(K\) is defined as \(K = 1/B\). Values of \(B\) and \(K\) are given in Table 3.

ELASTIC WAVES IN CUBIC CRYSTALS

By considering as in Figs. 18 and 19 the forces acting on an element of volume in the crystal we obtain the equation of motion in the \(x\) direction

\[ \rho \frac{\partial^2 u}{\partial t^2} = \frac{\partial X_x}{\partial x} + \frac{\partial X_y}{\partial y} + \frac{\partial X_z}{\partial z} ; \]  

(55)

here \(\rho\) is the density and \(u\) is the displacement in the \(x\) direction. There are similar equations for the \(y\) and \(z\) directions. From (38) and (50) it follows that for a cubic crystal

\[ \rho \frac{\partial^2 u}{\partial t^2} = C_{11} \frac{\partial^2 e_{xx}}{\partial x^2} + C_{12} \left( \frac{\partial^2 e_{yy}}{\partial x^2} + \frac{\partial^2 e_{zz}}{\partial x^2} \right) + C_{44} \left( \frac{\partial^2 e_{yy}}{\partial y^2} + \frac{\partial^2 e_{zz}}{\partial y^2} \right) ; \]  

(56)

here the \(x\), \(y\), \(z\) directions are parallel to the cube edges. Using the definitions (31) and (32) of the strain components we have

\[ \rho \frac{\partial^2 u}{\partial t^2} = C_{11} \frac{\partial^2 u}{\partial x^2} + C_{44} \left( \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) + \left( C_{12} + C_{44} \right) \left( \frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 w}{\partial x \partial z} \right) \]  

(57a)

where \(u, v, w\) are components of the displacement \(\mathbf{R}\) as defined by (29).
Figure 19 If springs A and B are stretched equally, the block between them experiences no net force. This illustrates the fact that a uniform stress $X_s$ in a solid does not give a net force on a volume element. If the spring at B is stretched more than the spring at A, the block between them will be accelerated by the force $X_s(B) - X_s(A)$.

The corresponding equations of motion for $\frac{\partial^2 v}{\partial t^2}$ and $\frac{\partial^2 w}{\partial t^2}$ are found directly from (57a) by symmetry:

$$\rho \frac{\partial^2 v}{\partial t^2} = C_{11} \frac{\partial^2 v}{\partial y^2} + C_{44} \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial z^2} \right) + (C_{12} + C_{44}) \left( \frac{\partial^2 u}{\partial x \partial y} + \frac{\partial^2 w}{\partial y \partial z} \right); \\
$$

$$\rho \frac{\partial^2 w}{\partial t^2} = C_{11} \frac{\partial^2 w}{\partial z^2} + C_{44} \left( \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} \right) + (C_{12} + C_{44}) \left( \frac{\partial^2 u}{\partial x \partial z} + \frac{\partial^2 v}{\partial y \partial z} \right).$$

(57b) (57c)

We now look for simple special solutions of these equations.

**Waves in the [100] Direction**

One solution of (57a) is given by a longitudinal wave

$$u = u_0 \exp \left[ i(Kx - \omega t) \right],$$

where $u$ is the $x$ component of the particle displacement. Both the wavevector and the particle motion are along the $x$ cube edge. Here $K = 2\pi/\lambda$ is the wavevector and $\omega = 2\pi v$ is the angular frequency. If we substitute (58) into (57a) we find

$$\omega^2 \rho = C_{11} K^2;$$

thus the velocity $\omega/K$ of a longitudinal wave in the [100] direction is

$$v_s = \nu \lambda = \omega/K = (C_{11}/\rho)^{1/2}. \quad (60)$$

Consider a transverse or shear wave with the wavevector along the $x$ cube edge and with the particle displacement $v$ in the $y$ direction:

$$v = v_0 \exp \left[ i(Kx - \omega t) \right].$$

(61)

On substitution in (57b) this gives the dispersion relation

$$\omega^2 \rho = C_{44} K^2;$$

thus the velocity $\omega/K$ of a transverse wave in the [100] direction is

$$v_s = (C_{44}/\rho)^{1/2}. \quad (63)$$

The identical velocity is obtained if the particle displacement is in the $z$ direc-
tion. Thus for \( \mathbf{K} \) parallel to [100] the two independent shear waves have equal velocities. This is not true for \( \mathbf{K} \) in a general direction in the crystal.

**Waves in the [110] Direction**

There is a special interest in waves that propagate in a face diagonal direction of a cubic crystal, because the three elastic constants can be found simply from the three propagation velocities in this direction.

Consider a shear wave that propagates in the \( xy \) plane with particle displacement \( w \) in the \( z \) direction

\[
w = w_0 \exp \left[ i(K_x x + K_y y - \omega t) \right],
\]

whence (32c) gives

\[
\omega^2 \rho = C_{44}(K_x^2 + K_y^2) = C_{44}K^2,
\]

independent of propagation direction in the plane.

Consider other waves that propagate in the \( xy \) plane with particle motion in the \( xy \) plane: let

\[
u = v_0 \exp \left[ i(K_x x + K_y y - \omega t) \right]; \quad v = v_0 \exp \left[ i(K_x x + K_y y - \omega t) \right]
\]

From (57a) and (57b),

\[
\omega^2 \rho u = (C_{11}K_x^2 + C_{44}K_y^2)u + (C_{12} + C_{44})K_xK_yv; \\
\omega^2 \rho v = (C_{11}K_y^2 + C_{44}K_x^2)v + (C_{12} + C_{44})K_xK_yu.
\]

This pair of equations has a particularly simple solution for a wave in the [110] direction, for which \( K_x = K_y = K/\sqrt{2} \). The condition for a solution is that the determinant of the coefficients of \( u \) and \( v \) in (67) should equal zero:

\[
\begin{vmatrix}
-\omega^2 \rho + \frac{1}{2}(C_{11} + C_{44})K^2 & \frac{1}{2}(C_{12} + C_{44})K^2 \\
\frac{1}{2}(C_{12} + C_{44})K^2 & -\omega^2 \rho + \frac{1}{2}(C_{11} + C_{44})K^2
\end{vmatrix} = 0.
\]

This equation has the roots

\[
\omega^2 \rho = \frac{1}{2}(C_{11} + C_{12} + 2C_{44})K^2; \quad \omega^2 \rho = \frac{1}{2}(C_{11} - C_{12})K^2.
\]

The first root describes a longitudinal wave; the second root describes a shear wave. How do we determine the direction of particle displacement? The first root when substituted into the upper equation of (67) gives

\[
\frac{1}{2}(C_{11} + C_{12} + 2C_{44})K^2u = \frac{1}{2}(C_{11} + C_{44})K^2u + \frac{1}{2}(C_{12} + C_{44})K^2v,
\]

whence the displacement components satisfy \( u = v \). Thus the particle displacement is along [110] and parallel to the \( \mathbf{K} \) vector (Fig. 20). The second root of (44) when substituted into the upper equation of (67) gives

\[
\frac{1}{2}(C_{11} - C_{12})K^2u = \frac{1}{2}(C_{11} + C_{44})K^2u + \frac{1}{2}(C_{12} + C_{44})K^2v,
\]
Figure 20  Effective elastic constants for the three modes of elastic waves in the principal propagation directions in cubic crystals. The two transverse modes are degenerate for propagation in the [100] and [111] directions.

whence \( u = -v \). The particle displacement is along \([\bar{1}10]\) and perpendicular to the \( K \) vector.

Selected values of the adiabatic elastic stiffness constants of cubic crystals at low temperatures and at room temperature are given in Table 11. Notice the general tendency for the elastic constants to decrease as the temperature is increased. Further values at room temperature alone are given in Table 12.

There are three normal modes of wave motion in a crystal for a given magnitude and direction of the wavevector \( K \). In general, the polarizations (directions of particle displacement) of these modes are not exactly parallel or perpendicular to \( K \). In the special propagation directions \([100]\), \([111]\), and \([110]\) of a cubic crystal two of the three modes for a given \( K \) are such that the particle motion is exactly transverse to \( K \) and in the third mode the motion is exactly longitudinal (parallel to \( K \)). The analysis is much simpler in these special directions than in general directions.

SUMMARY

- Crystals of inert gas atoms are bound by the van der Waals interaction (induced dipole-dipole interaction), and this varies with distance as \( 1/R^6 \).

- The repulsive interaction between atoms arises generally from the electrostatic repulsion of overlapping charge distributions and the Pauli principle, which compels overlapping electrons of parallel spin to enter orbitals of higher energy.

- Ionic crystals are bound by the electrostatic attraction of charged ions of
Table 11  Adiabatic elastic stiffness constants of cubic crystals at low temperature and at room temperature

The values given at 0 K were obtained by extrapolation of measurements carried out down to 4 K. The table was compiled with the assistance of Professor Charles S. Smith.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>$C_{11}$ (in $10^{12}$ dyne/cm$^2$)</th>
<th>$C_{12}$</th>
<th>$C_{44}$</th>
<th>Temperature, K</th>
<th>Density, g/cm$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>5.326</td>
<td>2.049</td>
<td>1.631</td>
<td>0</td>
<td>19.317</td>
</tr>
<tr>
<td></td>
<td>5.233</td>
<td>2.045</td>
<td>1.607</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Ta</td>
<td>2.663</td>
<td>1.582</td>
<td>0.874</td>
<td>0</td>
<td>16.696</td>
</tr>
<tr>
<td></td>
<td>2.609</td>
<td>1.574</td>
<td>0.818</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>1.762</td>
<td>1.249</td>
<td>0.818</td>
<td>0</td>
<td>9.018</td>
</tr>
<tr>
<td></td>
<td>1.654</td>
<td>1.214</td>
<td>0.754</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Ag</td>
<td>1.315</td>
<td>0.973</td>
<td>0.511</td>
<td>0</td>
<td>10.635</td>
</tr>
<tr>
<td></td>
<td>1.240</td>
<td>0.937</td>
<td>0.461</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Au</td>
<td>2.016</td>
<td>1.697</td>
<td>0.454</td>
<td>0</td>
<td>19.488</td>
</tr>
<tr>
<td></td>
<td>1.923</td>
<td>1.631</td>
<td>0.420</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Al</td>
<td>1.143</td>
<td>0.619</td>
<td>0.316</td>
<td>0</td>
<td>2.733</td>
</tr>
<tr>
<td></td>
<td>1.068</td>
<td>0.607</td>
<td>0.282</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>0.0416</td>
<td>0.0341</td>
<td>0.0286</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0370</td>
<td>0.0314</td>
<td>0.0188</td>
<td>295</td>
<td></td>
</tr>
<tr>
<td>Pb</td>
<td>0.555</td>
<td>0.454</td>
<td>0.194</td>
<td>0</td>
<td>11.599</td>
</tr>
<tr>
<td></td>
<td>0.495</td>
<td>0.423</td>
<td>0.149</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Ni</td>
<td>2.612</td>
<td>1.508</td>
<td>1.317</td>
<td>0</td>
<td>8.968</td>
</tr>
<tr>
<td></td>
<td>2.506</td>
<td>1.500</td>
<td>1.235</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Pd</td>
<td>2.341</td>
<td>1.761</td>
<td>0.712</td>
<td>0</td>
<td>12.132</td>
</tr>
<tr>
<td></td>
<td>2.271</td>
<td>1.761</td>
<td>0.717</td>
<td>300</td>
<td></td>
</tr>
</tbody>
</table>

The electrostatic energy of a structure of $2N$ ions of charge $\pm q$ is

(CGs) \[ U = -N \alpha \frac{q^2}{R} = -N \sum \frac{(\pm)q^2}{r_y} \],

where $\alpha$ is the Madelung constant and $R$ is the distance between nearest neighbors.

- Metals are bound by the reduction in the kinetic energy of the valence electrons in the metal as compared with the free atom.
Table 12  Adiabatic elastic stiffness constants of several cubic crystals at room temperature or 300 K.

<table>
<thead>
<tr>
<th></th>
<th>$C_{11}$</th>
<th>$C_{12}$</th>
<th>$C_{44}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond</td>
<td>10.76</td>
<td>1.25</td>
<td>5.76</td>
</tr>
<tr>
<td>Na</td>
<td>0.073</td>
<td>0.062</td>
<td>0.042</td>
</tr>
<tr>
<td>Li</td>
<td>0.135</td>
<td>0.114</td>
<td>0.088</td>
</tr>
<tr>
<td>Ge</td>
<td>1.285</td>
<td>0.483</td>
<td>0.680</td>
</tr>
<tr>
<td>Si</td>
<td>1.66</td>
<td>0.639</td>
<td>0.796</td>
</tr>
<tr>
<td>GaSb</td>
<td>0.885</td>
<td>0.404</td>
<td>0.433</td>
</tr>
<tr>
<td>InSb</td>
<td>0.672</td>
<td>0.367</td>
<td>0.302</td>
</tr>
<tr>
<td>MgO</td>
<td>2.86</td>
<td>0.87</td>
<td>1.48</td>
</tr>
<tr>
<td>NaCl</td>
<td>0.487</td>
<td>0.124</td>
<td>0.126</td>
</tr>
</tbody>
</table>

- A covalent bond is characterized by the overlap of charge distributions of antiparallel electron spin. The Pauli contribution to the repulsion is reduced for antiparallel spins, and this makes possible a greater degree of overlap. The overlapping electrons bind their associated ion cores by electrostatic attraction.

Problems

1. Quantum solid. In a quantum solid the dominant repulsive energy is the zero-point energy of the atoms. Consider a crude one-dimensional model of crystalline He$^4$ with each He atom confined to a line segment of length $L$. In the ground state the wave function within each segment is taken as a half wavelength of a free particle. Find the zero-point kinetic energy per particle.

2. Cohesive energy of bcc and fcc neon. Using the Lennard-Jones potential, calculate the ratio of the cohesive energies of neon in the bcc and fcc structures (Ans. 0.958). The lattice sums for the bcc structures are:

$$\sum_j' p_{ij}^{12} = 9.11418; \quad \sum_j' p_{ij}^{-6} = 12.2533.$$  

3. Solid molecular hydrogen. For H$_2$ one finds from measurements on the gas that the Lennard-Jones parameters are $\epsilon = 50 \times 10^{-16}$ erg and $\sigma = 2.96$ Å. Find the cohesive energy in kJ per mole of H$_2$; do the calculation for an fcc structure. Treat each H$_2$ molecule as a sphere. The observed value of the cohesive energy is 0.751 kJ/mol, much less than we calculated, so that quantum corrections must be very important.
4. **Possibility of ionic crystals** \(R^+R^-\). Imagine a crystal that exploits for binding the coulomb attraction of the positive and negative ions of the same atom or molecule \(R\). This is believed to occur with certain organic molecules, but it is not found when \(R\) is a single atom. Use the data in Tables 5 and 6 to evaluate the stability of such a form of Na in the NaCl structure relative to normal metallic sodium. Evaluate the energy at the observed interatomic distance in metallic sodium, and use 0.78 eV as the electron affinity of Na.

5. **Linear ionic crystal.** Consider a line of \(2N\) ions of alternating charge \(\pm q\) with a repulsive potential energy \(A/R^n\) between nearest neighbors. (a) Show that at the equilibrium separation

\[
(U(R_0) = -\frac{2Nq^2 \ln 2}{R_0} \left(1 - \frac{1}{n}\right).
\]

(b) Let the crystal be compressed so that \(R_0 \rightarrow R_0(1 - \delta)\). Show that the work done in compressing a unit length of the crystal has the leading term \(\frac{1}{2}C\delta^2\), where

\[
C = \frac{(n - 1)q^2 \ln 2}{R_0}.
\]

To obtain the results in SI, replace \(q^2\) by \(q^2/4\pi\epsilon_0\). Note: We should not expect to obtain this result from the expression for \(U(R_0)\), but we must use the complete expression for \(U(R)\).

6. **Cubic ZnS structure.** Using \(\lambda\) and \(\rho\) from Table 7 and the Madelung constants given in the text, calculate the cohesive energy of KCl in the cubic ZnS structure described in Chapter 1. Compare with the value calculated for KCl in the NaCl structure.

7. **Divalent ionic crystals.** Barium oxide has the NaCl structure. Estimate the cohesive energies per molecule of the hypothetical crystals Ba\(^{2+}\)O\(^{-}\) and Ba\(^{2+}\)O\(^{-}\) referred to separated neutral atoms. The observed nearest-neighbor internuclear distance is \(R_0 = 2.76\ \text{Å}\); the first and second ionization potentials of Ba are 5.19 and 9.96 eV; and the electron affinities of the first and second electrons added to the neutral oxygen atom are 1.5 and \(-9.0\) eV. The first electron affinity of the neutral oxygen atom is the energy released in the reaction \(O + e \rightarrow O^-\). The second electron affinity is the energy released in the reaction \(O^- + e \rightarrow O^{2-}\). Which valence state do you predict will occur? Assume \(R_0\) is the same for both forms, and neglect the repulsive energy.

8. **Young's modulus and Poisson's ratio.** A cubic crystal is subject to tension in the [100] direction. Find expressions in terms of the elastic stiffnesses for Young's modulus and Poisson's ratio as defined in Fig. 21.

9. **Longitudinal wave velocity.** Show that the velocity of a longitudinal wave in the [111] direction of a cubic crystal is given by

\[
u_z = \left[H(C_{11} + 2C_{12} + 4C_{44})/\rho\right]^{1/2}.
\]

*Hint:* For such a wave \(u = v = w\). Let \(u = u_0 e^{ik(x+y+z)}/\sqrt{3} e^{-i\omega t}\), and use Eq. (57a).
10. **Transverse wave velocity.** Show that the velocity of transverse waves in the [111] direction of a cubic crystal is given by \( v = \left[ \frac{1}{2}(C_{11} - C_{12} + C_{44})/\rho \right]^{1/2} \). *Hint:* See Problem 9.

11. **Effective shear constant.** Show that the shear constant \( \frac{1}{2}(C_{11} - C_{12}) \) in a cubic crystal is defined by setting \( \varepsilon_{xx} = -\varepsilon_{yy} = \frac{1}{2} \varepsilon \) and all other strains equal to zero, as in Fig. 22. *Hint:* Consider the energy density (43); look for a \( C' \) such that \( U = \frac{1}{2} C' \varepsilon^2 \).

12. **Determinantal approach.** It is known that an \( R \)-dimensional square matrix with all elements equal to unity has roots \( R \) and 0, with the \( R \) occurring once and the zero occurring \( R - 1 \) times. If all elements have the value \( p \), then the roots are \( Rp \) and 0. (a) Show that if the diagonal elements are \( q \) and all other elements are \( p \), then there is one root equal to \( (R - 1)p + q \) and \( R - 1 \) roots equal to \( q - p \). (b) Show from the elastic equation (57) for a wave in the [111] direction of a cubic crystal that the determinantal equation which gives \( \omega^2 \) as a function of \( K \) is

\[
\begin{vmatrix}
q - \omega^2 p & p & p \\
p & q - \omega^2 p & p \\
p & p & q - \omega^2 p
\end{vmatrix} = 0,
\]

where \( q = \frac{1}{2}K^2(C_{11} + 2C_{44}) \) and \( p = \frac{1}{2}K^2(C_{12} + C_{44}) \). This expresses the condition that three linear homogeneous algebraic equations for the three displacement components \( u, v, w \) have a solution. Use the result of part (a) to find the three roots of \( \omega^2 \); check with the results given for Problems 9 and 10.

13. **General propagation direction.** (a) By substitution in (57) find the determinantal equation which expresses the condition that the displacement

\[
R(x) = [u_0 \delta + v_0 \delta + w_0 \delta] \exp \left[ i(K \cdot r - \omega t) \right]
\]

be a solution of the elastic wave equations in a cubic crystal. (b) The sum of the roots of a determinantal equation is equal to the sum of the diagonal elements \( a_{ii} \). Show
from part (a) that the sum of the squares of the three elastic wave velocities in any
direction in a cubic crystal is equal to \((C_{11} + 2C_{44})/\rho\). Recall that \(v_s^2 = \omega^2/K^2\).

14. **Stability criteria.** The criterion that a cubic crystal with one atom in the primitive
cell be stable against small homogeneous deformations is that the energy density
(43) be positive for all combinations of strain components. What restrictions are
thereby imposed on the elastic stiffness constants? (In mathematical language the
problem is to find the conditions that a real symmetric quadratic form should be
positive definite. The solution is given in books on algebra; see also Korn and Korn,
*Mathematical Handbook*, McGraw-Hill, 1961, Sec. 13.5–6.) Ans. \(C_{44} > 0, C_{11} > 0,
C_{11} - C_{12} > 0, \) and \(C_{11} + 2C_{12} > 0\). For an example of the instability which results

**References**

M. Born and K. Huang, *Dynamical theory of crystal lattices*, Oxford, 1954. The classic work on
ionic crystals.


properties of crystals of the inert gases.

(1975).


J. F. Nye, *Physical properties of crystals: their representation by tensors and matrices*, Oxford,
1957.


Chapter 5 treats the thermal properties of phonons. Chapter 10 treats the interaction of phonons and photons, and the interaction of phonons and electrons.
Figure 1  Important elementary excitations in solids. The origins of the concepts and the names of the excitations are discussed by C. T. Walker and G. A. Slack, Am. J. Phys. 38, 1380 (1970).

Figure 2  (Dashed lines) Planes of atoms when in equilibrium. (Solid lines) Planes of atoms when displaced as for a longitudinal wave. The coordinate $u$ measures the displacement of the planes.

Figure 3  Planes of atoms as displaced during passage of a transverse wave.
CHAPTER 4: PHONONS I. CRYSTAL VIBRATIONS

VIBRATIONS OF CRYSTALS WITH MONATOMIC BASIS

Consider the elastic vibrations of a crystal with one atom in the primitive cell. We want to find the frequency of an elastic wave in terms of the wavevector that describes the wave and in terms of the elastic constants.

The mathematical solution is simplest in the [100], [110], and [111] propagation directions in cubic crystals. These are the directions of the cube edge, face diagonal, and body diagonal. When a wave propagates along one of these directions, entire planes of atoms move in phase with displacements either parallel or perpendicular to the direction of the wavevector. We can describe with a single coordinate \( u_s \) the displacement of the plane \( s \) from its equilibrium position. The problem is then one dimensional. For each wavevector there are three modes, one of longitudinal polarization (Fig. 2) and two of transverse polarization (Fig. 3).

We assume that the elastic response of the crystal is a linear function of the forces. That is equivalent to the assumption that the elastic energy is a quadratic function of the relative displacement of any two points in the crystal. Terms in the energy that are linear will vanish in equilibrium—see the minimum in Fig. 3.6. Cubic and higher order terms may be neglected for sufficiently small elastic deformations, but play some role at high temperatures, as we see in Chapter 5.

We assume accordingly that the force on the plane \( s \) caused by the displacement of the plane \( s + p \) is proportional to the difference \( u_{s+p} - u_s \) of their displacements. For brevity we consider only nearest-neighbor interactions, so that \( p = \pm 1 \). The total force on \( s \) comes from planes \( s \pm 1 \):

\[
F_s = C(u_{s+1} - u_s) + C(u_{s-1} - u_s) \,.
\]

This expression is linear in the displacements and is of the form of Hooke’s law.

The constant \( C \) is the force constant between nearest-neighbor planes and will differ for longitudinal and transverse waves. It is convenient hereafter to regard \( C \) as defined for one atom of the plane, so that \( F_s \) is the force on one atom in the plane \( s \).

The equation of motion of the plane \( s \) is

\[
M \frac{d^2 u_s}{dt^2} = C(u_{s+1} + u_{s-1} - 2u_s) \,.
\]

where \( M \) is the mass of an atom. We look for solutions with all displacements
having the time dependence $\exp(-i\omega t)$. Then $d^2u_s/dt^2 = -\omega^2u_s$, and (2) becomes

$$-M\omega^2u_s = C(u_{s+1} + u_{s-1} - 2u_s) .$$

(3)

This is a difference equation in the displacements $u$ and has traveling wave solutions of the form:

$$u_{s\pm 1} = u \exp(isKa) \exp(\pm iKa) ,$$

(4)

where $a$ is the spacing between planes and $K$ is the wavevector. The value to use for $a$ will depend on the direction of $K$.

With (4), we have from (3):

$$-\omega^2Mu \exp(isKa) = Cu\{\exp[i(s+1)Ka] + \exp[i(s-1)Ka] - 2 \exp(isKa)\} .$$

(5)

We cancel $u \exp(isKa)$ from both sides, to leave

$$\omega^2M = -C[\exp(iKa) + \exp(-iKa) - 2] .$$

(6)

With the identity $2 \cos Ka = \exp(iKa) + \exp(-iKa)$, we have the dispersion relation connecting $\omega$ and $K$:

$$\omega^2 = (2C/M)(1 - \cos Ka) .$$

(7)

The boundary of the first Brillouin zone lies at $K = \pm \pi/a$. We show from (7) that the slope of $\omega$ versus $K$ is zero at the zone boundary:

$$d\omega^2/dK = (2Ca/M) \sin Ka = 0$$

(8)

at $K = \pm \pi/a$, for here $\sin Ka = \sin(\pm \pi) = 0$. The special significance of phonon wavevectors that lie on the zone boundary is developed in (12) below.

By a trigonometric identity (7) may be written as

$$\omega^2 = (4C/M) \sin^2 \frac{1}{2}Ka ; \quad \omega = (4C/M)^{1/2}|\sin \frac{1}{2}Ka| .$$

(9)

A plot of $\omega$ versus $K$ is given in Fig. 4.

**First Brillouin Zone**

What range of $K$ is physically significant for elastic waves? Only those in the first Brillouin zone. From (4) the ratio of the displacements of two successive planes is given by

$$\frac{u_{s+1}}{u_s} = \frac{u \exp[i(s+1)Ka]}{u \exp(isKa)} = \exp(iKa) .$$

(10)

The range $-\pi$ to $+\pi$ for the phase $Ka$ covers all independent values of the exponential.

There is absolutely no point in saying that two adjacent atoms are out of phase by more than $\pi$: a relative phase of $1.2\pi$ is physically identical with a
Figure 4  Plot of $\omega$ versus $K$. The region of $K \ll 1/a$ or $\lambda \gg a$ corresponds to the continuum approximation; here $\omega$ is directly proportional to $K$.

Figure 5  The wave represented by the solid curve conveys no information not given by the dashed curve. Only wavelengths longer than $2a$ are needed to represent the motion.

relative phase of $-0.8\pi$, and a relative phase of $4.2\pi$ is identical with $0.2\pi$. We need both positive and negative values of $K$ because waves can propagate to the right or to the left.

The range of independent values of $K$ is specified by

$$-\pi < Ka \leq \pi, \quad \text{or} \quad -\frac{\pi}{a} < K \leq \frac{\pi}{a}.$$ 

This range is the first Brillouin zone of the linear lattice, as defined in Chapter 2. The extreme values are $K_{\text{max}} = \pm \pi/a$.

There is a real difference here from an elastic continuum: in the continuum limit $a \to 0$ and $K_{\text{max}} \to \pm \infty$. Values of $K$ outside of the first Brillouin zone (Fig. 5) merely reproduce lattice motions described by values within the limits $\pm \pi/a$.

We may treat a value of $K$ outside these limits by subtracting the integral multiple of $2\pi/a$ that will give a wavevector inside these limits. Suppose $K$ lies
outside the first zone, but a related wavevector \( K' \) defined by \( K' = K - 2\pi n/a \) lies within the first zone, where \( n \) is an integer. Then the displacement ratio (10) becomes

\[
\frac{u_{s+1}}{u_s} = \exp(iKa) = \exp(i2\pi n) \exp[i(Ka - 2\pi n)] = \exp(iK'a), \tag{11}
\]

because \( \exp(i2\pi n) = 1 \). Thus the displacement can always be described by a wavevector within the first zone. We note that \( 2\pi n/a \) is a reciprocal lattice vector because \( 2\pi/a \) is a reciprocal lattice vector. Thus by subtraction of an appropriate reciprocal lattice vector from \( K \), we always obtain an equivalent wavevector in the first zone.

At the boundaries \( K_{\text{max}} = \pm \pi/a \) of the Brillouin zone the solution \( u_s = u \exp(isKa) \) does not represent a traveling wave, but a standing wave. At the zone boundaries \( sK_{\text{max}} = \pm s\pi \), whence

\[
u_s = u \exp(\pm is\pi) = u (-1)^s . \tag{12} \]

This is a standing wave: alternate atoms oscillate in opposite phases, because \( u_s = \pm 1 \) according to whether \( s \) is an even or an odd integer. The wave moves neither to the right nor to the left.

This situation is equivalent to Bragg reflection of x-rays: when the Bragg condition is satisfied a traveling wave cannot propagate in a lattice, but through successive reflections back and forth, a standing wave is set up.

The critical value \( K_{\text{max}} = \pm \pi/a \) found here satisfies the Bragg condition \( 2d \sin \theta = n\lambda \): we have \( \theta = \frac{1}{2} \pi, \ d = a, \ K = 2\pi/\lambda, \ n = 1 \), so that \( \lambda = 2a \). With x-rays it is possible to have \( n \) equal to other integers besides unity because the amplitude of the electromagnetic wave has a meaning in the space between atoms, but the displacement amplitude of an elastic wave usually has a meaning only at the atoms themselves.

**Group Velocity**

The transmission velocity of a wave packet is the group velocity, given as

\[
\nu_g = \frac{d\omega}{dK},
\]
or

\[
\nu_g = \text{grad}_K \omega(K), \tag{13}
\]

the gradient of the frequency with respect to \( K \). This is the velocity of energy propagation in the medium.

With the particular dispersion relation (9), the group velocity (Fig. 6) is

\[
\nu_g = (Ca^2/M)^{1/2} \cos \frac{1}{2} Ka . \tag{14}
\]

This is zero at the edge of the zone where \( K = \pi/a \). Here the wave is a standing wave, as in (12), and we expect zero net transmission velocity for a standing wave.
Long Wavelength Limit

When \( Ka \ll 1 \) we expand \( \cos Ka \equiv 1 - \frac{1}{2}(Ka)^2 \), so that the dispersion relation (7) becomes

\[
\omega^2 = \left(\frac{C}{M}\right)K^2a^2 .
\]

The result that the frequency is directly proportional to the wavevector in the long wavelength limit is equivalent to the statement that the velocity of sound is independent of frequency in this limit. Thus \( v = \omega/K \), exactly as in the continuum theory of elastic waves—in the continuum limit \( a = 0 \) and thus \( Ka = 0 \).

Derivation of Force Constants from Experiment

In metals the effective forces may be of quite long range, carried from ion to ion through the conduction electron sea (Chapter 10). Interactions have been found between planes of atoms separated by as many as 20 planes. We can make a statement about the range of the forces from the observed dispersion relation for \( \omega \). The generalization of the dispersion relation (7) to \( p \) nearest planes is easily found to be

\[
\omega^2 = \left(\frac{2}{M}\right) \sum_{p>0} C_p(1 - \cos pKa) .
\]

We solve for the interplanar force constants \( C_p \) by multiplying both sides by \( \cos rKa \), where \( r \) is an integer, and integrating over the range of independent values of \( K \):

\[
M \int_{-\pi/a}^{\pi/a} dK \omega_K^2 \cos rKa = 2 \sum_{p>0} C_p \int_{-\pi/a}^{\pi/a} dK (1 - \cos pKa) \cos rKa
= -2\pi C_r/a .
\]

The integral vanishes except for \( p = r \). Thus

\[
C_p = -\frac{Ma}{2\pi} \int_{-\pi/a}^{\pi/a} dK \omega_K^2 \cos pKa
\]

gives the force constant at range \( pa \), for a structure with a monatomic basis.
TWO ATOMS PER PRIMITIVE BASIS

The phonon dispersion relation shows new features in crystals with two or more atoms per primitive basis. Consider, for example, the NaCl or diamond structures, with two atoms in the primitive cell. For each polarization mode in a given propagation direction the dispersion relation \( \omega \) versus \( K \) develops two branches, known as the acoustical and optical branches. We have longitudinal LA and transverse acoustical TA modes, and longitudinal LO and transverse optical TO modes, as in Fig. 7.

If there are \( p \) atoms in the primitive cell, there are \( 3p \) branches to the dispersion relation: 3 acoustical branches and \( 3p - 3 \) optical branches. Thus germanium (Fig. 8a) and KBr (Fig. 8b), each with two atoms in a primitive cell, have six branches: one LA, one LO, two TA, and two TO.

The numerology of the branches follows from the number of degrees of freedom of the atoms. With \( p \) atoms in the primitive cell and \( N \) primitive cells, there are \( pN \) atoms. Each atom has three degrees of freedom, one for each of the \( x, y, z \) directions, making a total of \( 3pN \) degrees of freedom for the crystal. The number of allowed \( K \) values in a single branch is just \( N \) for one Brillouin zone.\(^1\) Thus the LA and the two TA branches have a total of \( 3N \) modes, thereby accounting for \( 3N \) of the total degrees of freedom. The remaining \((3p - 3)N\) degrees of freedom are accommodated by the optical branches.

We consider a cubic crystal where atoms of mass \( M_1 \) lie on one set of planes and atoms of mass \( M_2 \) lie on planes interleaved between those of the first set (Fig. 9). It is not essential that the masses be different, but either the force constants or the masses will be different if the two atoms of the basis are in nonequivalent sites. Let \( a \) denote the repeat distance of the lattice in the direction normal to the lattice planes considered. We treat waves that propagate in a symmetry direction for which a single plane contains only a single type of ion; such directions are [111] in the NaCl structure and [100] in the CsCl structure.

We write the equations of motion under the assumption that each plane interacts only with its nearest-neighbor planes and that the force constants are identical between all pairs of nearest-neighbor planes. We refer to Fig. 9 to obtain

\[
M_1 \frac{d^2 u_s}{dt^2} = C(v_s + v_{s-1} - 2u_s) ;
\]

\[
M_2 \frac{d^2 v_s}{dt^2} = C(u_{s+1} + u_s - 2v_s) .
\]

\(^1\)We show in Chapter 5 by application of periodic boundary conditions to the modes of the crystal of volume \( V \) that there is one \( K \) value in the volume \((2\pi)^3/V\) in Fourier space. The volume of a Brillouin zone is \((2\pi)^3/V_c\), where \( V_c \) is the volume of a crystal primitive cell. Thus the number of allowed \( K \) values in a Brillouin zone is \( V/V_c \), which is just \( N \), the number of primitive cells in the crystal.
Figure 7 Optical and acoustical branches of the dispersion relation for a diatomic linear lattice, showing the limiting frequencies at $K = 0$ and $K = K_{\text{max}} = \pi/a$. The lattice constant is $a$.

Figure 8a Phonon dispersion relations in the [111] direction in germanium at 80 K. The two TA phonon branches are horizontal at the zone boundary position, $K_{\text{max}} = (2\pi/a)(1/4)$). The LO and TO branches coincide at $K = 0$, this also is a consequence of the crystal symmetry of Ge. The results were obtained with neutron inelastic scattering by G. Nilsson and G. Nelis.

Figure 8b Dispersion curves in the [111] direction in KBr at 90 K, after A. D. B. Woods, B. N. Brockhouse, R. A. Cowley, and W. Cochran. The extrapolation to $K = 0$ of the TO, LO branches are called $\omega_T$, $\omega_L$; these are discussed in Chapter 10.

We look for a solution in the form of a traveling wave, now with different amplitudes $u$, $v$ on alternate planes:

$$u_s = u \exp(isKa) \exp(-i\omega t) ; \quad v_s = v \exp(isKa) \exp(-i\omega t).$$ (19)

Recall the definition of $a$ in Fig. 9 as the distance between nearest identical planes, not nearest-neighbor planes.
Figure 9  A diatomic crystal structure with masses $M_1, M_2$ connected by force constant $C$ between adjacent planes. The displacements of atoms $M_1$ are denoted by $u_{s-1}, u_s, u_{s+1}, \ldots$, and of atoms $M_2$ by $v_{s-1}, v_s, v_{s+1}$. The repeat distance is $a$ in the direction of the wavevector $K$. The atoms are shown in their undisplaced positions.

On substitution of (19) in (18) we have

$$-\omega^2 M_1 u = C[v[1 + \exp(-iKa)] - 2Cu] ;$$
$$-\omega^2 M_2 v = Cu[\exp(iKa) + 1] - 2Cv .$$

The homogeneous linear equations have a solution only if the determinant of the coefficients of the unknowns $u, v$ vanishes:

$$\begin{vmatrix}
2C - M_1 \omega^2 & -C[1 + \exp(-iKa)] \\
-C[1 + \exp(iKa)] & 2C - M_2 \omega^2
\end{vmatrix} = 0 ,$$

or

$$M_1 M_2 \omega^4 - 2C(M_1 + M_2)\omega^2 + 2C^2(1 - \cos Ka) = 0 .$$

We can solve this equation exactly for $\omega^2$, but it is simpler to examine the limiting cases $Ka \ll 1$ and $Ka = \pm \pi$ at the zone boundary. For small $Ka$ we have $\cos Ka \equiv 1 - \frac{1}{2}K^2a^2 + \ldots$, and the two roots are

$$\omega^2 \equiv 2C\left(\frac{1}{M_1} + \frac{1}{M_2}\right) \quad \text{(optical branch)} ;$$

$$\omega^2 \equiv \frac{\frac{1}{2}C}{M_1 + M_2} K^2a^2 \quad \text{(acoustical branch)} .$$

The extent of the first Brillouin zone is $-\pi/a \leq K \leq \pi/a$, where $a$ is the repeat distance of the lattice. At $K_{\max} = \pm \pi/a$ the roots are

$$\omega^2 = 2C/M_1 ; \quad \omega^2 = 2C/M_2 .$$

The dependence of $\omega$ on $K$ is shown in Fig. 7 for $M_1 > M_2$.

The particle displacements in the transverse acoustical (TA) and transverse optical (TO) branches are shown in Fig. 10. For the optical branch at $K = 0$ we find, on substitution of (23) in (20),

$$\frac{u}{v} = -\frac{M_2}{M_1} .$$
The atoms vibrate against each other, but their center of mass is fixed. If the two atoms carry opposite charges, as in Fig. 10, we may excite a motion of this type with the electric field of a light wave, so that the branch is called the optical branch. At a general $K$ the ratio $u/v$ will be complex, as follows from either of the equations (20). Another solution for the amplitude ratio at small $K$ is $u = v$, obtained as the $K = 0$ limit of (24). The atoms (and their center of mass) move together, as in long wavelength acoustical vibrations, whence the term acoustical branch.

Wavelike solutions do not exist for certain frequencies, here between $(2C/M_1)^{1/2}$ and $(2C/M_2)^{1/2}$. This is a characteristic feature of elastic waves in polyatomic lattices. There is a frequency gap at the boundary $K_{\text{max}} = \pm \pi/a$ of the first Brillouin zone. If we look for solutions in the gap with $\omega$ real, then the wavevector $K$ will be complex, so that the wave is damped in space.

**QUANTIZATION OF ELASTIC WAVES**

The energy of a lattice vibration is quantized. The quantum of energy is called a phonon in analogy with the photon of the electromagnetic wave. Elastic waves in crystals are made up of phonons. Thermal vibrations in crystals are thermally excited phonons, like the thermally excited photons of black-body electromagnetic radiation in a cavity.

The energy of an elastic mode of angular frequency $\omega$ is

$$\epsilon = (n + \frac{1}{2})h\omega$$

when the mode is excited to quantum number $n$; that is, when the mode is occupied by $n$ phonons. The term $\frac{1}{2}h\omega$ is the zero point energy of the mode. It occurs for both phonons and photons as a consequence of their equivalence to a quantum harmonic oscillator of frequency $\omega$, for which the energy eigenvalues are also $(n + \frac{1}{2})h\omega$. The quantum theory of phonons is developed in Appendix C.
We can easily quantize the mean square phonon amplitude. Consider the standing wave mode of amplitude

\[ u = u_0 \cos Kx \cos \omega t. \]

Here \( u \) is the displacement of a volume element from its equilibrium position at \( x \) in the crystal.

The energy in the mode, as in any harmonic oscillator, is half kinetic energy and half potential energy, when averaged over time. The kinetic energy density is \( \frac{1}{2} \rho \left( \frac{\partial u}{\partial t} \right)^2 \), where \( \rho \) is the mass density. In a crystal of volume \( V \), the volume integral of the kinetic energy is \( \frac{1}{2} \rho V \omega^2 u_0^2 \sin^2 \omega t \). The time average kinetic energy is

\[ \frac{1}{2} \rho V \omega^2 u_0^2 = \frac{1}{2} (n + \frac{1}{2}) \hbar \omega, \]  

and the square of the amplitude is

\[ u_0^2 = 4(n + \frac{1}{2}) \hbar / \rho V \omega. \]  

This relates the displacement in a given mode to the phonon occupancy \( n \) of the mode.

What is the sign of \( \omega \)? The equations of motion such as (2) are equations for \( \omega^2 \), and if this is positive then \( \omega \) can have either sign, + or −. But the energy of a phonon must be positive, so it is conventional and suitable to view \( \omega \) as positive. (For circularly polarized waves both signs are often used, to distinguish one sense of rotation from the other.) If the crystal structure is unstable, or becomes unstable through an unusual temperature dependence of the force constants (Chapter 13), then \( \omega^2 \) will be negative and \( \omega \) will be imaginary.

A mode with \( \omega \) imaginary will be unstable, at least if the real part of \( \omega \) is negative. The crystal will transform spontaneously to a more stable structure. An optical mode with \( \omega \) close to zero is called a soft mode, and these are often involved in phase transitions, as in ferroelectric crystals.

**PHONON MOMENTUM**

A phonon of wavevector \( K \) will interact with particles such as photons, neutrons, and electrons as if it had a momentum \( \hbar K \). However, a phonon does not carry physical momentum.

The reason that phonons on a lattice do not carry momentum is that a phonon coordinate (except for \( K = 0 \)) involves relative coordinates of the atoms. Thus in an H\(_2\) molecule the internuclear vibrational coordinate \( r_1 - r_2 \) is a relative coordinate and does not carry linear momentum; the center of mass coordinate \( \frac{1}{2}(r_1 + r_2) \) corresponds to the uniform mode \( K = 0 \) and can carry linear momentum.
The physical momentum of a crystal is

\[ p = M(d\ell/dt) \sum u_s \]  \hspace{1cm} (30)

When the crystal carries a phonon \( K \),

\[ p = M(d\ell/dt) \sum_s \exp(isKa) = M(d\ell/dt)[1 - \exp(iNKa)] / [1 - \exp(iKa)] \]  \hspace{1cm} (31)

where \( s \) runs over the \( N \) atoms. We have used the series

\[ \sum_{s=0}^{N-1} x^s = (1 - x^N)/(1 - x) \]  \hspace{1cm} (32)

In the next chapter we enumerate the discrete values of \( K \) compatible with the boundary conditions, to find that \( K = \pm 2\pi r/Na \), where \( r \) is an integer. Thus \( \exp(iNKa) = \exp(\pm i2\pi r) = 1 \), and from (31) it follows that the crystal momentum is zero:

\[ p = M(d\ell/dt) \sum_s \exp(isKa) = 0 \]  \hspace{1cm} (33)

The only exception to (33) is the uniform mode \( K = 0 \), for which all \( u_s \) equal \( u \), so that \( p = NM(d\ell/dt) \). This mode represents a uniform translation of the crystal as a whole, and such a translation does carry momentum.

All the same, for most practical purposes a phonon acts as if its momentum were \( \hbar K \), sometimes called the crystal momentum. In crystals there exist wavevector selection rules for allowed transitions between quantum states. We saw that the elastic scattering of an x-ray photon by a crystal is governed by the wavevector selection rule

\[ k' = k + G \]  \hspace{1cm} (34)

where \( G \) is a vector in the reciprocal lattice; \( k \) is the wavevector of the incident photon, and \( k' \) is the wavevector of the scattered photon. In the reflection process the crystal as a whole will recoil with momentum \(-\hbar G\), but this uniform mode momentum is rarely considered explicitly.

Equation (34) is an example of the rule that the total wavevector of interacting waves is conserved in a periodic lattice, with the possible addition of a reciprocal lattice vector \( G \). The true momentum of the whole system always is rigorously conserved.

If the scattering of the photon is inelastic, with the creation of a phonon of wavevector \( K \), then the wavevector selection rule becomes

\[ k' + K = k + G \]  \hspace{1cm} (35)

If a phonon \( K \) is absorbed in the process, we have instead the relation

\[ k' = k + K + G \]  \hspace{1cm} (36)

Relations (35) and (36) are the natural extensions of (34).
We exhibit the mathematics involved in the wavevector selection rule. Suppose two phonons $K_1$, $K_2$ interact through cubic terms in the elastic energy to create a third phonon $K_3$. The probability of the collision will involve the product of the three phonon wave amplitudes, summed over all lattice sites:

$$\sum \exp(-iK_1 \cdot r_n) \exp(-iK_2 \cdot r_n) \exp(iK_3 \cdot r_n) = \sum \exp[i(K_3 - K_1 - K_2) \cdot r_n].$$

This sum in the limit of a large number of lattice sites approaches zero unless $K_3 = K_1 + K_2$ or $K_3 = K_1 + K_2 + G$. If either of these conditions is satisfied, of which the first is merely a special case of the second, the sum is equal to the number of lattice sites $N$. A similar sum was considered in Problem 2.4.

**INELASTIC SCATTERING BY PHONONS**

Phonon dispersion relations $\omega(K)$ are most often determined by the inelastic scattering of neutrons with the emission or absorption of a phonon. Further, the angular width of the scattered neutron beam gives information bearing on the lifetime of phonons.

A neutron sees the crystal lattice chiefly by interaction with the nuclei of the atoms. The kinematics of the scattering of a neutron beam by a crystal lattice are described by the general wavevector selection rule:

$$K + G = k' \pm K,$$  \hspace{1cm} (37)

and by the requirement of conservation of energy. Here $K$ is the wavevector of the phonon created (+) or absorbed (−) in the process, and $G$ is any reciprocal lattice vector. For a phonon we choose $G$ such that $K$ lies in the first Brillouin zone.

The kinetic energy of the incident neutron is $p^2/2M_n$, where $M_n$ is the mass of the neutron. The momentum $p$ is given by $\hbar k$, where $k$ is the wavevector of the neutron. Thus $\hbar^2 k^2/2M_n$ is the kinetic energy of the incident neutron. If $k'$ is the wavevector of the scattered neutron, the energy of the scattered neutron is $\hbar^2 k'^2/2M_n$. The statement of conservation of energy is

$$\frac{\hbar^2 k^2}{2M_n} = \frac{\hbar^2 k'^2}{2M_n} \pm \hbar \omega,$$  \hspace{1cm} (38)

where $\hbar \omega$ is the energy of the phonon created (+) or absorbed (−) in the process.

To determine the dispersion relation using (37) and (38) it is necessary in the experiment to find the energy gain or loss of the scattered neutrons as a function of the scattering direction $k - k'$. Results for samarium and KBr are
given in Fig. 8; results for sodium are given in Fig. 11. A spectrometer used for phonon studies is shown in Fig. 12.

Recently the concept of "mirror symmetry" has been introduced into the study of alkali-halide ion dynamics. The idea is to consider the crystal that would be formed if the signs of the ions in $A^+B^-$ were reversed. Now $A^-B^+$ does not exist, but a nearest mass pair that does exist will have, it turns out experimentally, a phonon dispersion relation remarkably similar to that of $A^+B^-$. Thus KF is the approximate mirror image of NaCl, for $K^+$ is isoelectronic with $Cl^-$ and $F^-$ is isoelectronic with $Na^+$.

SUMMARY

- The quantum unit of a crystal vibration is a phonon. If the angular frequency is \( \omega \), the energy of the phonon is \( \hbar \omega \).
- When a phonon of wavevector \( \mathbf{K} \) is created by the inelastic scattering of a photon or neutron from wavevector \( \mathbf{k} \) to \( \mathbf{k}' \), the wavevector selection rule that governs the process is
  \[
  \mathbf{k} = \mathbf{k}' + \mathbf{K} + \mathbf{G},
  \]
  where \( \mathbf{G} \) is a reciprocal lattice vector.
- All elastic waves can be described by wavevectors that lie within the first Brillouin zone in reciprocal space.
- If there are \( p \) atoms in the primitive cell, the phonon dispersion relation will have 3 acoustical phonon branches and \( 3p - 3 \) optical phonon branches.

Problems

1. Monatomic linear lattice. Consider a longitudinal wave
   \[
   u_x = u \cos(\omega t - sKa)
   \]
   which propagates in a monatomic linear lattice of atoms of mass \( M \), spacing \( a \), and nearest-neighbor interaction \( C \).
   (a) Show that the total energy of the wave is
   \[
   E = \frac{1}{2} M \sum_s (du_s/dt)^2 + \frac{1}{2} C \sum_s (u_s - u_{s+1})^2,
   \]
   where \( s \) runs over all atoms.
   (b) By substitution of \( u_s \) in this expression, show that the time-average total energy per atom is
   \[
   \frac{1}{2} M \omega^2 u^2 + \frac{1}{2} C(1 - \cos Ka)u^2 = \frac{1}{2} M \omega^2 u^2,
   \]
   where in the last step we have used the dispersion relation (9) for this problem.

2. Continuum wave equation. Show that for long wavelengths the equation of motion (2) reduces to the continuum elastic wave equation
   \[
   \frac{\partial^2 u}{\partial t^2} = v^2 \frac{\partial^2 u}{\partial x^2},
   \]
   where \( v \) is the velocity of sound.

3. Basis of two unlike atoms. For the problem treated by (18) to (26), find the amplitude ratios \( u/v \) for the two branches at \( K_{max} = \pi/a \). Show that at this value of \( K \) the two lattices act as if decoupled: one lattice remains at rest while the other lattice moves.
4. **Kohn anomaly.** We suppose that the interplanar force constant $C_p$ between planes $s$ and $s + p$ is of the form

$$C_p = A \frac{\sin pk_0 a}{pa},$$

where $A$ and $k_0$ are constants and $p$ runs over all integers. Such a form is expected in metals. Use this and Eq. (16a) to find an expression for $\omega^2$ and also for $\partial \omega^2 / \partial K$. Prove that $\partial \omega^2 / \partial K$ is infinite when $K = k_0$. Thus a plot of $\omega^2$ versus $K$ or of $\omega$ versus $K$ has a vertical tangent at $k_0$; there is a kink at $k_0$ in the phonon dispersion relation $\omega(K)$.

5. **Diatomic chain.** Consider the normal modes of a linear chain in which the force constants between nearest-neighbor atoms are alternately $C$ and $10C$. Let the masses be equal, and let the nearest-neighbor separation be $a/2$. Find $\omega(K)$ at $K = 0$ and $K = \pi a$. Sketch in the dispersion relation by eye. This problem simulates a crystal of diatomic molecules such as $H_2$.

6. **Atomic vibrations in a metal.** Consider point ions of mass $M$ and charge $e$ immersed in a uniform sea of conduction electrons. The ions are imagined to be in stable equilibrium when at regular lattice points. If one ion is displaced a small distance $r$ from its equilibrium position, the restoring force is largely due to the electric charge within the sphere of radius $r$ centered at the equilibrium position. Take the number density of ions (or of conduction electrons) as $3/4\pi R^3$, which defines $R$. (a) Show that the frequency of a single ion set into oscillation is $\omega = (e^2/4\pi R^3)^{1/2}$. (b) Estimate the value of this expression for sodium, roughly. (c) From (a), (b), and some common sense, estimate the order of magnitude of the velocity of sound in the metal.

*7. **Soft phonon modes.** Consider a line of ions of equal mass but alternating in charge, with $e_p = e(-1)^p$ as the charge on the $p$th ion. The interatomic potential is the sum of two contributions: (1) a short-range interaction of force constant $C_{1R} = \gamma$ that acts between nearest neighbors only, and (2) a coulomb interaction between all ions. (a) Show that the contribution of the coulomb interaction to the atomic force constants is $C_{pC} = 2(-1)^p e^2/p^3 a^3$, where $a$ is the equilibrium nearest-neighbor distance. (b) From (16a) show that the dispersion relation may be written as

$$\omega^2/\omega_0^2 = \sin^2 \frac{1}{2} Ka + \sigma \sum_{p=1}^{\infty} (-1)^p (1 - \cos pKa)p^{-3},$$

where $\omega_0^2 = 4\gamma/M$ and $\sigma = e^2/\gamma a^3$. (c) Show that $\omega^2$ is negative (unstable mode) at the zone boundary $Ka = \pi$ if $\sigma > 0.475$ or $4/7\zeta(3)$, where $\zeta$ is a Riemann zeta function. Show further that the speed of sound at small $Ka$ is imaginary if $\sigma > (2 \ln 2)^{-1} = 0.721$. Thus $\omega^2$ goes to zero and the lattice is unstable for some value of $Ka$ in the interval $(0, \pi)$ if $0.475 < \sigma < 0.721$. Notice that the phonon spectrum is not that of a diatomic lattice because the interaction of any ion with its neighbors is the same as that of any other ion.

*This problem is rather difficult.
References


5

Phonons II. Thermal Properties

PHONON HEAT CAPACITY
   Planck distribution  117
   Normal mode enumeration  118
   Density of states in one dimension  119
   Density of states in three dimensions  120
   Debye model for density of states  122
   Debye $T^3$ law  124
   Einstein model of the density of states  124
   General result for $D(\omega)$  127

ANHARMONIC CRYSTAL INTERACTIONS  129
   Thermal expansion  130

THERMAL CONDUCTIVITY  131
   Thermal resistivity of phonon gas  133
   Umklapp processes  135
   Imperfections  137

PROBLEMS  139
   1. Singularity in density of states  139
   2. Rms thermal dilation of crystal cell  139
   3. Zero point lattice displacement and strain  139
   4. Heat capacity of layer lattice  139
   5. Grüneisen constant  139

REFERENCES  140
Figure 1  Plot of Planck distribution function. At high temperatures the occupancy of a state is approximately linear in the temperature. The function $\langle n \rangle + \frac{1}{2}$, which is not plotted, approaches the dashed line as asymptote at high temperatures. The dashed line is the classical limit.
We discuss the heat capacity of a phonon gas and then the effects of anharmonic lattice interactions on the phonons and on the crystal. Thermal properties of metals are treated in Chapter 6, superconductors in Chapter 12, magnetic materials in Chapters 14 and 15, and noncrystalline solids in Chapter 17.

PHONON HEAT CAPACITY

By heat capacity we shall usually mean the heat capacity at constant volume, which is more fundamental than the heat capacity at constant pressure, which is what the experiments determine. The heat capacity at constant volume is defined as $C_V = \langle \partial U/\partial T \rangle_V$ where $U$ is the energy and $T$ the temperature.

The contribution of the phonons to the heat capacity of a crystal is called the lattice heat capacity and is denoted by $C_{lat}$.

The total energy of the phonons at a temperature $\tau (= k_B T)$ in a crystal may be written as the sum of the energies over all phonon modes, here indexed by the wavevector $K$ and polarization index $p$:

$$U = \sum_K \sum_p U_{K,p} = \sum_K \sum_p \langle n_{K,p} \rangle \hbar \omega_{K,p}, \quad \text{(1)}$$

where $\langle n_{K,p} \rangle$ is the thermal equilibrium occupancy of phonons of wavevector $K$ and polarization $p$. The form of $\langle n_{K,p} \rangle$ is given by the Planck distribution function:

$$\langle n \rangle = \frac{1}{\exp(\hbar \omega/\tau) - 1} \quad \text{(2)}$$

where the $\langle \cdot \cdot \cdot \rangle$ denotes the average in thermal equilibrium. A graph of $\langle n \rangle$ is given in Fig. 1.

Planck Distribution

Consider a set of identical harmonic oscillators in thermal equilibrium. The ratio of the number of oscillators in their $(n + 1)$th quantum state of excitation to the number in the $n$th quantum state is

$$N_{n+1}/N_n = \exp(-\hbar \omega/\tau), \quad \tau = k_B T, \quad \text{(3)}$$

1A thermodynamic relation gives $C_p - C_V = 9\alpha^2 BVT$, where $\alpha$ is the temperature coefficient of linear expansion, $V$ the volume, and $B$ the bulk modulus. The fractional difference between $C_p$ and $C_V$ is usually small and often may be neglected. As $T \to 0$ we see that $C_p \to C_V$, provided $\alpha$ and $B$ are constant.
by use of the Boltzmann factor. Thus the fraction of the total number of oscillators in the $n$th quantum state is

$$\frac{N_n}{\sum_{s=0}^{\infty} N_s} = \frac{\exp(-n\hbar\omega/\tau)}{\sum_{s=0}^{\infty} \exp(-s\hbar\omega/\tau)}.$$  \hspace{1cm} (4)

We see that the average excitation quantum number of an oscillator is

$$\langle n \rangle = \frac{\sum_{s} s \exp(-s\hbar\omega/\tau)}{\sum_{s} \exp(-s\hbar\omega/\tau)}.$$  \hspace{1cm} (5)

The summations in (5) are

$$\sum_{s} x^s = \frac{1}{1 - x}; \quad \sum_{s} sx^s = x \frac{d}{dx} \sum_{s} x^s = \frac{x}{(1 - x)^2},$$  \hspace{1cm} (6)

with $x = \exp(-\hbar\omega/\tau)$. Thus we may rewrite (5) as the Planck distribution:

$$\langle n \rangle = \frac{1}{1 - x} = \frac{1}{\exp(\hbar\omega/\tau) - 1}.$$  \hspace{1cm} (7)

**Normal Mode Enumeration**

The energy of a collection of oscillators of frequencies $\omega_{K,p}$ in thermal equilibrium is found from (1) and (2):

$$U = \sum_{K} \sum_{p} \frac{\hbar\omega_{K,p}}{\exp(\hbar\omega_{K,p}/\tau) - 1}.$$  \hspace{1cm} (8)

It is usually convenient to replace the summation over $K$ by an integral. Suppose that the crystal has $D_p(\omega)d\omega$ modes of a given polarization $p$ in the frequency range $\omega$ to $\omega + d\omega$. Then the energy is

$$U = \sum_{p} \int d\omega \ D_p(\omega) \frac{\hbar\omega}{\exp(\hbar\omega/\tau) - 1}.$$  \hspace{1cm} (9)

The lattice heat capacity is found by differentiation with respect to temperature. Let $x = \hbar\omega/\tau = \hbar\omega/k_B T$: then $\partial U/\partial T$ gives

$$C_{\text{lat}} = k_B \sum_{p} \int d\omega \ D_p(\omega) \frac{x^2 \exp x}{(\exp x - 1)^2}.$$  \hspace{1cm} (10)

The central problem is to find $D(\omega)$, the number of modes per unit frequency range. This function is called the density of states or, more often,
Figure 2. Elastic line of $N + 1$ atoms, with $N = 10$, for boundary conditions that the end atoms $s = 0$ and $s = 10$ are fixed. The particle displacements in the normal modes for either longitudinal or transverse displacements are of the form, $u_s \propto \sin sKa$. This form is automatically zero at the atom at the end $s = 0$, and we choose $K$ to make the displacement zero at the end $s = 10$.

Figure 3. The boundary condition $\sin sKa = 0$ for $s = 10$ can be satisfied by choosing $K = \pi/10a, 2\pi/10a, \ldots, 9\pi/10a$, where $10a$ is the length $L$ of the line. The present figure is in $K$ space. The dots are not atoms but are the allowed values of $K$. Of the $N + 1$ particles on the line, only $N - 1$ are allowed to move, and their most general motion can be expressed in terms of the $N - 1$ allowed values of $K$. This quantization of $K$ has nothing to do with quantum mechanics but follows classically from the boundary conditions that the end atoms be fixed.

Density of states. The best practical way to obtain the density of states is to measure the dispersion relation $\omega$ versus $K$ in selected crystal directions by inelastic neutron scattering and then to make a theoretical analytic fit to give the dispersion relation in a general direction, from which $D(\omega)$ may be calculated.

**Density of States in One Dimension**

Consider the boundary value problem for vibrations of a one-dimensional line (Fig. 2) of length $L$ carrying $N + 1$ particles at separation $a$. We suppose that the particles $s = 0$ and $s = N$ at the ends of the line are held fixed. Each normal vibrational mode of polarization $p$ has the form of a standing wave, where $u_s$ is the displacement of the particle $s$:

$$u_s = u(0) \exp(-i\omega_{K,p}t) \sin sKa,$$

where $\omega_{K,p}$ is related to $K$ by the appropriate dispersion relation.

As in Fig. 3, the wavevector $K$ is restricted by the fixed-end boundary conditions to the values

$$K = \frac{\pi}{L}, \quad \frac{2\pi}{L}, \quad \frac{3\pi}{L}, \ldots, \quad \frac{(N - 1)\pi}{L}.$$  

(12)

The solution for $K = \pi/L$ has

$$u_s \propto \sin (s\pi a/L)$$

(13)

and vanishes for $s = 0$ and $s = N$ as required.

The solution for $K = N\pi/L = \pi/a = K_{max}$ has $u_s \propto \sin s\pi$; this permits no motion of any atom, because $\sin s\pi$ vanishes at each atom. Thus there are $N - 1$ allowed independent values of $K$ in (12). This number is equal to the
number of particles allowed to move. Each allowed value of $K$ is associated with a standing wave. For the one-dimensional line there is one mode for each interval $\Delta K = \pi/L$, so that the number of modes per unit range of $K$ is $L/\pi$ for $K \leq \pi/a$, and 0 for $K > \pi/a$.

There are three polarizations $p$ for each value of $K$: in one dimension two of these are transverse and one longitudinal. In three dimensions the polarizations are this simple only for wavevectors in certain special crystal directions.

Another device for enumerating modes is often used that is equally valid. We consider the medium as unbounded, but require that the solutions be periodic over a large distance $L$, so that $u(sa) = u(sa + L)$. The method of periodic boundary conditions (Figs. 4 and 5) does not change the physics of the problem in any essential respect for a large system. In the running wave solution $u = u(0) \exp[i(sKa - \omega t)]$ the allowed values of $K$ are

$$K = 0, \pm \frac{2\pi}{L}, \pm \frac{4\pi}{L}, \pm \frac{6\pi}{L}, \ldots, \frac{N\pi}{L}. \quad (14)$$

This method of enumeration gives the same number of modes (one per mobile atom) as given by (12), but we have now both plus and minus values of $K$, with the interval $\Delta K = 2\pi/L$ between successive values of $K$. For periodic boundary conditions the number of modes per unit range of $K$ is $L/2\pi$ for $-\pi/a \leq K \leq \pi/a$, and 0 otherwise. The situation in a two-dimensional lattice is portrayed in Fig. 6.

We need to know $D(\omega)$, the number of modes per unit frequency range. The number of modes $D(\omega) d\omega$ in $d\omega$ at $\omega$ is given in one dimension by

$$D(\omega) d\omega = \frac{L}{\pi} \frac{dK}{d\omega} d\omega = \frac{L}{\pi} \frac{d\omega}{dK}. \quad (15)$$

We can obtain the group velocity $d\omega/dK$ from the dispersion relation $\omega$ versus $K$. There is a singularity in $D(\omega)$ whenever the dispersion relation $\omega(K)$ is horizontal; that is, whenever the group velocity is zero.

**Density of States in Three Dimensions**

We apply periodic boundary conditions over $N^3$ primitive cells within a cube of side $L$, so that $K$ is determined by the condition

$$\exp[i(K_x x + K_y y + K_z z)] = \exp[i[K_x(x + L) + K_y(y + L) + K_z(z + L)]], \quad (16)$$

whence

$$K_x, K_y, K_z = 0; \pm \frac{2\pi}{L}; \pm \frac{4\pi}{L}; \ldots; \frac{N\pi}{L}. \quad (17)$$

Therefore there is one allowed value of $K$ per volume $(2\pi/L)^3$ in $K$ space, or

$$\left(\frac{L}{2\pi}\right)^3 = \frac{V}{8\pi^3}. \quad (18)$$
Figure 4  Consider $N$ particles constrained to slide on a circular ring. The particles can oscillate if connected by elastic springs. In a normal mode the displacement $u_s$ of atom $s$ will be of the form $sKa$ or $sKb$; these are independent modes. By the geometrical periodicity of the ring the boundary condition is that $u_{N+s} = u_s$ for all $s$, so that $NKa$ must be an integral multiple of $2\pi$. For $N = 8$ the allowed independent values of $K$ are 0, $2\pi/8a$, $4\pi/8a$, $6\pi/8a$, and $8\pi/8a$. The value $K = 0$ is meaningless for the sine form, because $\sin s0a = 0$. The value $8\pi/8a$ has a meaning only for the cosine form, because $\sin (8\pi/a) = \sin s\pi = 0$. The three other values of $K$ are allowed for both the sine and cosine modes, giving a total of eight allowed modes for the eight particles. Thus the periodic boundary condition leads to one allowed mode per particle, exactly as for the fixed-end boundary condition of Fig. 3. If we had taken the modes in the complex form $\exp(isKa)$, the periodic boundary condition would lead to the eight modes with $K = 0$, $\pm 2\pi/Na$, $\pm 4\pi/Na$, $\pm 6\pi/Na$, and $8\pi/Na$, as in Eq. (14).

\[
\begin{array}{cccccccc}
N\pi & 6\pi & 4\pi & 2\pi & 0 & 2\pi & 4\pi & 6\pi & N\pi \\
L & L & L & L & L & L & L & L & L
\end{array}
\]

Figure 5  Allowed values of wavevector $K$ for periodic boundary conditions applied to a linear lattice of periodicity $N = 8$ atoms on a line of length $L$. The $K = 0$ solution is the uniform mode. The special points $\pm N\pi/L$ represent only a single solution because $\exp(is\pi)$ is identical to $\exp(-is\pi)$; thus there are eight allowed modes, with displacements of the $s$th atom proportional to $1$, $\exp(\pm is\pi/4)$, $\exp(\pm is\pi/2)$, $\exp(\pm 3is\pi/4)$, $\exp(is\pi)$.

Figure 6  Allowed values in Fourier space of the phonon wavevector $K$ for a square lattice of lattice constant $a$, with periodic boundary conditions applied over a square of side $L = 10a$. The uniform mode is marked with a cross. There is one allowed value of $K$ per area $(2\pi/10a)^2 = (2\pi L)^2$, so that within the circle of area $\pi K^2$ the smoothed number of allowed points is $\pi K^2(L/2\pi)^2$. 
allowed values of $K$ per unit volume of $K$ space, for each polarization and for each branch. The volume of the specimen is $V = L^3$.

The total number of modes with wavevector less than $K$ is found from (18) to be $(L/2\pi)^3$ times the volume of a sphere of radius $K$. Thus

$$N = (L/2\pi)^3(4\pi K^3/3)$$

(19)

for each polarization type. The density of states for each polarization is

$$D(\omega) = dN/d\omega = (VK^2/2\pi^2)(dK/d\omega).$$

(20)

**Debye Model for Density of States**

In the Debye approximation the velocity of sound is taken as constant for each polarization type, as it would be for a classical elastic continuum. The dispersion relation is written as

$$\omega = vK,$$  

(21)

with $v$ the constant velocity of sound.

The density of states (20) becomes

$$D(\omega) = V\omega^2/2\pi^2 v^3.$$

(22)

If there are $N$ primitive cells in the specimen, the total number of acoustic phonon modes is $N$. A cutoff frequency $\omega_D$ is determined by (19) as

$$\omega_D^3 = 6\pi^2 v^3 N/V.$$

(23)

To this frequency there corresponds a cutoff wavevector in $K$ space:

$$K_D = \omega_D/v = (6\pi^2 N/V)^{1/3}.$$

(24)

On the Debye model we do not allow modes of wavevector larger than $K_D$. The number of modes with $K \leq K_D$ exhausts the number of degrees of freedom of a monatomic lattice.

The thermal energy (9) is given by

$$U = \int d\omega D(\omega)(n(\omega))\hbar\omega = \int_0^{\omega_D} d\omega \left(\frac{V\omega^2}{2\pi^2 v^3}\right) \left(\frac{\hbar\omega}{e^{\hbar\omega/\tau} - 1}\right),$$

(25)

for each polarization type. For brevity we assume that the phonon velocity is independent of the polarization, so that we multiply by the factor 3 to obtain

$$U = \frac{3Vh}{2\pi^2 v^3} \int_0^{\omega_D} d\omega \frac{\omega^3}{e^{\hbar\omega/\tau} - 1} = \frac{3Vk_BT^4}{2\pi^2 v^3 \hbar^3} \int_0^{x_D} dx \frac{x^3}{e^x - 1},$$

(26)

where $x = \hbar\omega/\tau = \hbar\omega/k_BT$ and

$$x_D = \hbar\omega_D/k_BT = \theta/T.$$

(27)

This defines the **Debye temperature** $\theta$ in terms of $\omega_D$ defined by (23). We may express $\theta$ as

$$\theta = \frac{hV}{k_B} \cdot \left(\frac{6\pi^2 N}{V}\right)^{1/3},$$

(28)
so that the total phonon energy is
\[ U = 9Nk_BT \left( \frac{T}{\theta} \right)^3 \int_0^{x_D} dx \frac{x^3}{e^x - 1} , \tag{29} \]
where \( N \) is the number of atoms in the specimen and \( x_D = \theta/T \).

The heat capacity is found most easily by differentiating the middle expression of (26) with respect to temperature. Then
\[ C_V = \frac{3V\hbar^2}{2\pi^2v^3k_BT^2} \int_0^{\omega_D} d\omega \frac{\omega^4 e^{\hbar\omega/\tau}}{(e^{\hbar\omega/\tau} - 1)^2} = 9Nk_B \left( \frac{T}{\theta} \right)^3 \int_0^{x_D} dx \frac{x^4 e^x}{(e^x - 1)^2} . \tag{30} \]
The Debye heat capacity is plotted in Fig. 7. At \( T \gg \theta \) the heat capacity approaches the classical value of \( 3Nk_B \). Measured values for silicon and germanium are plotted in Fig. 8.
Debye $T^3$ Law

At very low temperatures we may approximate (29) by letting the upper limit go to infinity. We have

$$\int_0^\infty dx \frac{x^3}{e^x - 1} = \int_0^\infty dx \ x^3 \sum_{s=1}^\infty \exp(-sx) = 6 \sum_{s=1}^\infty \frac{1}{s^4} = \frac{\pi^4}{15},$$

(31)

where the sum over $s^{-4}$ is found in standard tables. Thus $U = 3\pi^4 N k_b T^4/5\theta^3$ for $T \ll \theta$, and

$$C_v \approx \frac{12\pi^4}{5} N k_b \left(\frac{T}{\theta}\right)^3 \approx 234 \ N k_b \left(\frac{T}{\theta}\right)^3,$$

(32)

which is the Debye $T^3$ approximation. Experimental results for argon are plotted in Fig. 9.

At sufficiently low temperatures the $T^3$ approximation is quite good; that is, when only long wavelength acoustic modes are thermally excited. These are just the modes that may be treated as an elastic continuum with macroscopic elastic constants. The energy of the short wavelength modes (for which this approximation fails) is too high for them to be populated significantly at low temperatures.

We understand the $T^3$ result by a simple argument (Fig. 10). Only those lattice modes having $\hbar \omega < k_b T$ will be excited to any appreciable extent at a low temperature $T$. The excitation of these modes will be approximately classical, each with an energy close to $k_b T$, according to Fig. 1.

Of the allowed volume in $\mathbf{K}$ space, the fraction occupied by the excited modes is of the order of $(\omega_r / \omega_D)^3$ or $(K_T / K_D)^3$, where $K_T$ is a "thermal" wavevector defined such that $\hbar \nu K_T = k_b T$ and $K_D$ is the Debye cutoff wavevector. Thus the fraction occupied is $(T/\theta)^3$ of the total volume in $\mathbf{K}$ space. There are of the order of $3N(T/\theta)^3$ excited modes, each having energy $k_b T$. The energy is $\sim 3Nk_b T(T/\theta)^3$, and the heat capacity is $\sim 12Nk_b (T/\theta)^3$.

For actual crystals the temperatures at which the $T^3$ approximation holds are quite low. It may be necessary to be below $T = 0/50$ to get reasonably pure $T^3$ behavior.

Selected values of $\theta$ are given in Table 1. Note, for example, in the alkali metals that the heavier atoms have the lowest $\theta$'s, because the velocity of sound decreases as the density increases.

Einstein Model of the Density of States

Consider $N$ oscillators of the same frequency $\omega_0$ and in one dimension. The Einstein density of states is $D(\omega) = N\delta(\omega - \omega_0)$, where the delta function is centered at $\omega_0$. The thermal energy of the system is

$$U = N\langle n \rangle \hbar \omega = \frac{N \hbar \omega}{e^{\hbar \omega/\tau} - 1},$$

(33)
Figure 9  Low temperature heat capacity of solid argon, plotted against $T^3$. In this temperature region the experimental results are in excellent agreement with the Debye $T^3$ law with $\theta = 92.0$ K. (Courtesy of L. Finegold and N. E. Phillips.)

Figure 10  To obtain a qualitative explanation of the Debye $T^3$ law, we suppose that all phonon modes of wavevector less than $K_T$ have the classical thermal energy $k_B T$ and that modes between $K_T$ and the Debye cutoff $K_D$ are not excited at all. Of the $3N$ possible modes, the fraction excited is $(K_T/K_D)^3 = (T/\theta)^3$, because this is the ratio of the volume of the inner sphere to the outer sphere. The energy is $U = k_B T \cdot 3N(T/\theta)^3$, and the heat capacity is $C_V = \frac{\partial U}{\partial T} = 12Nk_B(T/\theta)^3$. 
Table 1  Debye temperature and thermal conductivity^a

<table>
<thead>
<tr>
<th></th>
<th>Li</th>
<th>Be</th>
<th>Na</th>
<th>Mg</th>
<th>K</th>
<th>Ca</th>
<th>Sc</th>
<th>Ti</th>
<th>V</th>
<th>Cr</th>
<th>Mn</th>
<th>Fe</th>
<th>Co</th>
<th>Ni</th>
<th>Cu</th>
<th>Zn</th>
<th>Ga</th>
<th>Ge</th>
<th>As</th>
<th>Se</th>
<th>Br</th>
<th>Kr</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>344</td>
<td>1440</td>
<td>158</td>
<td>400</td>
<td>91</td>
<td>230</td>
<td>360</td>
<td>420</td>
<td>380</td>
<td>630</td>
<td>410</td>
<td>470</td>
<td>445</td>
<td>450</td>
<td>343</td>
<td>327</td>
<td>320</td>
<td>374</td>
<td>282</td>
<td>90</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.85</td>
<td>2.00</td>
<td>1.41</td>
<td>1.56</td>
<td>1.02</td>
<td>0.16</td>
<td>0.22</td>
<td>0.31</td>
<td>0.94</td>
<td>0.08</td>
<td>0.80</td>
<td>1.00</td>
<td>0.91</td>
<td>4.01</td>
<td>1.16</td>
<td>0.41</td>
<td>0.60</td>
<td>0.50</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Low temperature limit of \( \theta \), in Kelvin

Thermal conductivity at 300 K, in W cm\(^{-1}\)K\(^{-1}\)

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>C</th>
<th>N</th>
<th>O</th>
<th>F</th>
<th>Ne</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2230</td>
<td>0.27</td>
<td>1.29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Al</th>
<th>Si</th>
<th>P</th>
<th>S</th>
<th>Cl</th>
<th>Ar</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>428</td>
<td>645</td>
<td>2.37</td>
<td>1.48</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>In</th>
<th>Sn</th>
<th>Sb</th>
<th>Te</th>
<th>I</th>
<th>Xe</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>211</td>
<td>153</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Th</th>
<th>Pa</th>
<th>U</th>
<th>Np</th>
<th>Pu</th>
<th>Am</th>
<th>Cm</th>
<th>Bk</th>
<th>Cf</th>
<th>Es</th>
<th>Fm</th>
<th>Md</th>
<th>No</th>
<th>Lr</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>163</td>
<td>207</td>
<td>0.28</td>
<td>0.06</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^a Most of the \( \theta \) values were supplied by N. Pearlman; references are given in the A.I.P. Handbook, 3rd ed; the thermal conductivity values are from R. W. Powell and Y. S. Touloukian, Science 181, 999 (1973).
with $\omega$ now written in place of $\omega_0$, for convenience.

The heat capacity of the oscillators is

$$C_V = \left( \frac{\partial U}{\partial T} \right)_V = Nk_B \left( \frac{\hbar \omega}{\tau} \right)^2 \frac{e^{\hbar \omega/\tau}}{(e^{\hbar \omega/\tau} - 1)^2},$$  \hspace{1cm} (34)

as plotted in Fig. 11. This expresses the Einstein (1907) result for the contribution of $N$ identical oscillators to the heat capacity of a solid.

In three dimensions $N$ is replaced by $3N$. The high temperature limit of $C_V$ becomes $3Nk_B$, which is known as the Dulong and Petit value.

At low temperatures (34) decreases as $\exp(-\hbar \omega/\tau)$, whereas the experimental form of the phonon contribution is known to be $T^3$ as accounted for by the Debye model treated above. The Einstein model is often used to approximate the optical phonon part of the phonon spectrum.

**General Result for $D(\omega)$**

We want to find a general expression for $D(\omega)$, the number of states per unit frequency range, given the phonon dispersion relation $\omega(K)$. The number of allowed values of $K$ for which the phonon frequency is between $\omega$ and $\omega + d\omega$ is

$$D(\omega) \ d\omega = \left( \frac{L}{2\pi} \right)^3 \int_{\text{shell}} d^3K,$$  \hspace{1cm} (35)

where the integral is extended over the volume of the shell in $K$ space bounded by the two surfaces on which the phonon frequency is constant, one surface on which the frequency is $\omega$ and the other on which the frequency is $\omega + d\omega$.

The real problem is to evaluate the volume of this shell. We let $dS_{\omega}$ denote an element of area (Fig. 12) on the surface in $K$ space of the selected constant
frequency $\omega$. The element of volume between the constant frequency surfaces $\omega$ and $\omega + d\omega$ is a right cylinder of base $dS_\omega$ and altitude $dK_\perp$, so that

$$\int_{\text{shell}} d^3K = \int dS_\omega dK_\perp .$$

(36)

Here $dK_\perp$ is the perpendicular distance (Fig. 13) between the surface $\omega$ constant and the surface $\omega + d\omega$ constant. The value of $dK_\perp$ will vary from one point to another on the surface.

The gradient of $\omega$, which is $\nabla_\mathbf{K} \omega$, is also normal to the surface $\omega$ constant, and the quantity

$$| \nabla_\mathbf{K} \omega | dK_\perp = d\omega ,$$

is the difference in frequency between the two surfaces connected by $dK_\perp$. Thus the element of the volume is

$$dS_\omega dK_\perp = dS_\omega \frac{d\omega}{| \nabla_\mathbf{K} \omega |} = dS_\omega \frac{d\omega}{v_g} ,$$

where $v_g = | \nabla_\mathbf{K} \omega |$ is the magnitude of the group velocity of a phonon. For (35) we have

$$D(\omega) d\omega = \left( \frac{L}{2\pi} \right)^3 \int \frac{dS_\omega}{v_g} d\omega .$$

We divide both sides by $d\omega$ and write $V = L^3$ for the volume of the crystal: the result for the density of states is

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

(37)
The quantity $dK_\perp$ is the perpendicular distance between two constant frequency surfaces in $K$ space, one at frequency $\omega$ and the other at frequency $\omega + d\omega$.

![Diagram showing $dK_\perp$ between two constant frequency surfaces](image)

The integral is taken over the area of the surface $\omega$ constant, in $K$ space. The result refers to a single branch of the dispersion relation. We can use this result also in electron band theory.

There is a special interest in the contribution to $D(\omega)$ from points at which the group velocity is zero. Such critical points produce singularities (known as Van Hove singularities) in the distribution function (Fig. 14).

**ANHARMONIC CRYSTAL INTERACTIONS**

The theory of lattice vibrations discussed thus far has been limited in the potential energy to terms quadratic in the interatomic displacements. This is the harmonic theory; among its consequences are:

- Two lattice waves do not interact; a single wave does not decay or change form with time.
- There is no thermal expansion.
- Adiabatic and isothermal elastic constants are equal.
- The elastic constants are independent of pressure and temperature.
- The heat capacity becomes constant at high temperatures $T > \theta$. 
In real crystals none of these consequences is satisfied accurately. The deviations may be attributed to the neglect of anharmonic (higher than quadratic) terms in the interatomic displacements. We discuss some of the simpler aspects of anharmonic effects.

Beautiful demonstrations of anharmonic effects are the experiments on the interaction of two phonons to produce a third phonon at a frequency $\omega_3 = \omega_1 + \omega_2$. Shiren described an experiment in which a beam of longitudinal phonons of frequency 9.20 GHz interacts in an MgO crystal with a parallel beam of longitudinal phonons at 9.18 GHz. The interaction of the two beams produced a third beam of longitudinal phonons at 9.20 + 9.18 = 18.38 GHz.

Three-phonon processes are caused by third-order terms in the lattice potential energy. A typical term might be $U_3 = A e_{xx} e_{yy} e_{zz}$, where the $e$'s are strain components and $A$ is a constant. The $A$'s have the same dimensions as elastic stiffness constants but may have values perhaps an order of magnitude larger. The physics of the phonon interaction can be stated simply: the presence of one phonon causes a periodic elastic strain which (through the anharmonic interaction) modulates in space and time the elastic constant of the crystal. A second phonon perceives the modulation of the elastic constant and thereupon is scattered to produce a third phonon, just as from a moving three-dimensional grating.

**Thermal Expansion**

We may understand thermal expansion by considering for a classical oscillator the effect of anharmonic terms in the potential energy on the mean separation of a pair of atoms at a temperature $T$. We take the potential energy of the atoms at a displacement $x$ from their equilibrium separation at absolute zero as

$$U(x) = cx^2 - gx^3 - fx^4,$$

(38)

with $c$, $g$, and $f$ all positive. The term in $x^3$ represents the asymmetry of the mutual repulsion of the atoms and the term in $x^4$ represents the softening of the vibration at large amplitudes. The minimum at $x = 0$ is not an absolute minimum, but for small oscillations the form is an adequate representation of an interatomic potential.

We calculate the average displacement by using the Boltzmann distribution function, which weights the possible values of $x$ according to their thermodynamic probability:

$$\langle x \rangle = \frac{\int_{-\infty}^{\infty} dx \ x \ \exp[-\beta U(x)]}{\int_{-\infty}^{\infty} dx \ \exp[-\beta U(x)]} ,$$
with $\beta \equiv 1/k_B T$. For displacements such that the anharmonic terms in the energy are small in comparison with $k_B T$, we may expand the integrands as

$$\int dx \exp(-\beta U) \equiv \int dx [\exp(-\beta cx^2)](x + \beta gx^4 + \beta fx^5) = (3\pi^{1/2}/4)(g/c^{5/2})\beta^{-3/2};$$

$$\int dx \exp(-\beta U) \equiv \int dx \exp(-\beta cx^2) = (\pi/\beta c)^{1/2},$$

whence the thermal expansion is

$$\langle x \rangle = \frac{3g}{4\pi^2} k_B T$$

in the classical region. Note that in (39) we have left $cx^2$ in the exponential, but we have expanded $\exp(\beta gx^3 + \beta fx^4) \approx 1 + \beta gx^3 + \beta fx^4 + \cdots$.

Measurements of the lattice constant of solid argon are shown in Fig. 15. The slope of the curve is proportional to the thermal expansion coefficient. The expansion coefficient vanishes as $T \to 0$, as we expect from Problem 5. In lowest order the thermal expansion does not involve the symmetric term $fx^4$ in $U(x)$, but only the antisymmetric term $gx^3$.

**THERMAL CONDUCTIVITY**

The thermal conductivity coefficient $K$ of a solid is defined with respect to the steady-state flow of heat down a long rod with a temperature gradient $dT/dx$:

$$j_U = -K \frac{dT}{dx},$$

where $j_U$ is the flux of thermal energy, or the energy transmitted across unit area per unit time.
This form implies that the process of thermal energy transfer is a random process. The energy does not simply enter one end of the specimen and proceed directly in a straight path to the other end, but diffuses through the specimen, suffering frequent collisions. If the energy were propagated directly through the specimen without deflection, then the expression for the thermal flux would not depend on the temperature gradient, but only on the difference in temperature $\Delta T$ between the ends of the specimen, regardless of the length of the specimen. The random nature of the conductivity process brings the temperature gradient and, as we shall see, a mean free path into the expression for the thermal flux.

From the kinetic theory of gases we find below in a certain approximation the following expression for the thermal conductivity:

$$K = \frac{1}{3} C v \ell,$$

where $C$ is the heat capacity per unit volume, $v$ is the average particle velocity, and $\ell$ is the mean free path of a particle between collisions. This result was applied first by Debye to describe thermal conductivity in dielectric solids, with $C$ as the heat capacity of the phonons, $v$ the phonon velocity, and $\ell$ the phonon mean free path. Several representative values of the mean free path are given in Table 2.

We give the elementary kinetic theory which leads to (42). The flux of particles in the $x$ direction is $\frac{1}{2} n \langle v_x \rangle$, where $n$ is the concentration of molecules; in equilibrium there is a flux of equal magnitude in the opposite direction. The $\langle \cdots \rangle$ denote average value.

If $c$ is the heat capacity of a particle, then in moving from a region at local temperature $T + \Delta T$ to a region at local temperature $T$ a particle will give up energy $c \Delta T$. Now $\Delta T$ between the ends of a free path of the particle is given by

$$\Delta T = \frac{dT}{dx} \ell_x = \frac{dT}{dx} v_x \tau,$$

where $\tau$ is the average time between collisions.

The net flux of energy (from both senses of the particle flux) is therefore

$$j_U = -n \langle v_x^2 \rangle c \tau \frac{dT}{dx} = -\frac{1}{3} n \langle v_x^2 \rangle c \tau \frac{dT}{dx}.$$

If, as for phonons, $v$ is constant, we may write (43) as

$$j_U = -\frac{1}{3} C v \ell \frac{dT}{dx},$$

with $\ell = v \tau$ and $C = nc$. Thus $K = \frac{1}{3} C v \ell$. 


Table 2 Phonon mean free paths

[Calculated from (44), taking \( v = 5 \times 10^5 \text{ cm/sec} \) as a representative sound velocity. The \( \ell \)'s obtained in this way refer to umklapp processes.]

<table>
<thead>
<tr>
<th>Crystal</th>
<th>( T, ^\circ\text{C} )</th>
<th>( C, \text{ in } \text{J cm}^{-2}\text{deg}^{-1} )</th>
<th>( K, \text{ in } \text{W cm}^{-1}\text{deg}^{-1} )</th>
<th>( \ell, \text{ in } \text{Å} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quartz*</td>
<td>0</td>
<td>2.00</td>
<td>0.13</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>-190</td>
<td>0.55</td>
<td>0.50</td>
<td>540</td>
</tr>
<tr>
<td>NaCl</td>
<td>0</td>
<td>1.88</td>
<td>0.07</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>-190</td>
<td>1.00</td>
<td>0.27</td>
<td>100</td>
</tr>
</tbody>
</table>

*Parallel to optic axis.

Thermal Resistivity of Phonon Gas

The phonon mean free path \( \ell \) is determined principally by two processes, geometrical scattering and scattering by other phonons. If the forces between atoms were purely harmonic, there would be no mechanism for collisions between different phonons, and the mean free path would be limited solely by collisions of a phonon with the crystal boundary, and by lattice imperfections. There are situations where these effects are dominant.

With anharmonic lattice interactions, there is a coupling between different phonons which limits the value of the mean free path. The exact states of the anharmonic system are no longer like pure phonons.

The theory of the effect of anharmonic coupling on thermal resistivity predicts that \( \ell \) is proportional to \( 1/T \) at high temperatures,\(^2\) in agreement with many experiments. We can understand this dependence in terms of the number of phonons with which a given phonon can interact; at high temperature the total number of excited phonons is proportional to \( T \). The collision frequency of a given phonon should be proportional to the number of phonons with which it can collide, whence \( \ell \propto 1/T \).

To define a thermal conductivity there must exist mechanisms in the crystal whereby the distribution of phonons may be brought locally into thermal equilibrium. Without such mechanisms we may not speak of the phonons at one end of the crystal as being in thermal equilibrium at a temperature \( T_2 \) and those at the other end in equilibrium at \( T_1 \).

It is not sufficient to have only a way of limiting the mean free path, but there must also be a way of establishing a local thermal equilibrium distribution of phonons. Phonon collisions with a static imperfection or a crystal boundary

Figure 16a Flow of gas molecules in a state of drifting equilibrium down a long open tube with frictionless walls. Elastic collision processes among the gas molecules do not change the momentum or energy flux of the gas because in each collision the velocity of the center of mass of the colliding particles and their energy remain unchanged. Thus energy is transported from left to right without being driven by a temperature gradient. Therefore the thermal resistivity is zero and the thermal conductivity is infinite.

Figure 16b The usual definition of thermal conductivity in a gas refers to a situation where no mass flow is permitted. Here the tube is closed at both ends, preventing the escape or entrance of molecules. With a temperature gradient the colliding pairs with above-average center of mass velocities will tend to be directed to the right, those with below-average velocities will tend to be directed to the left. A slight concentration gradient, high on the right, will be set up to enable the net mass transport to be zero while allowing a net energy transport from the hot to the cold end.

will not by themselves establish thermal equilibrium, because such collisions do not change the energy of individual phonons: the frequency $\omega_2$ of the scattered phonon is equal to the frequency $\omega_1$ of the incident phonon.

It is rather remarkable also that a three-phonon collision process

$$K_1 + K_2 = K_3$$  (45)

will not establish equilibrium, but for a subtle reason: the total momentum of the phonon gas is not changed by such a collision. An equilibrium distribution of phonons at a temperature $T$ can move down the crystal with a drift velocity which is not disturbed by three-phonon collisions of the form (45). For such collisions the phonon momentum

$$J = \sum_K n_K \hbar K$$  (46)

is conserved, because on collision the change in $J$ is $K_3 - K_2 - K_1 = 0$. Here $n_K$ is the number of phonons having wavevector $K$. 
Figure 16c In a crystal we may arrange to create phonons chiefly at one end, as by illuminating the left end with a lamp. From that end there will be a net flux of phonons toward the right end of the crystal. If only \( N \) processes \( (K_1 + K_2 = K_3) \) occur, the phonon flux is unchanged in momentum on collision and some phonon flux will persist down the length of the crystal. On arrival of phonons at the right end we can arrange in principle to convert most of their energy to radiation, thereby creating a sink for the phonons. Just as in (a) the thermal resistivity is zero.

Figure 16d In \( U \) processes there is a large net change in phonon momentum in each collision event. An initial net phonon flux will rapidly decay as we move to the right. The ends may act as sources and sinks. Net energy transport under a temperature gradient occurs as in (b).

For a distribution with \( J \neq 0 \), collisions such as (45) are incapable of establishing complete thermal equilibrium because they leave \( J \) unchanged. If we start a distribution of hot phonons down a rod with \( J \neq 0 \), the distribution will propagate down the rod with \( J \) unchanged. Therefore there is no thermal resistance. The problem as illustrated in Fig. 16 is like that of the collisions between molecules of a gas in a straight tube with frictionless walls.

**Umklapp Processes**

The important three-phonon processes that cause thermal resistivity are not of the form \( K_1 + K_2 = K_3 \) in which \( K \) is conserved, but are of the form

\[
K_1 + K_2 = K_3 + G,
\]

where \( G \) is a reciprocal lattice vector (Fig. 17). These processes, discovered by Peierls, are called umklapp processes. We recall that \( G \) may occur in all momentum conservation laws in crystals. In all allowed processes of the form of (46) and (47), energy is conserved.

We have seen examples of wave interaction processes in crystals for which the total wavevector change need not be zero, but may be a reciprocal lattice
vector. Such processes are always possible in periodic lattices. The argument is particularly strong for phonons: the only meaningful phonon \( \mathbf{K} \)'s lie in the first Brillouin zone, so that any longer \( \mathbf{K} \) produced in a collision must be brought back into the first zone by addition of \( \mathbf{G} \). A collision of two phonons both with a negative value of \( \mathbf{K} \), can by an umklapp process (\( \mathbf{G} \neq 0 \)) create a phonon with positive \( \mathbf{K} \). Umklapp processes are also called \( \mathbf{U} \) processes.

Collisions in which \( \mathbf{G} = 0 \) are called normal processes or \( \mathbf{N} \) processes. At high temperatures \( T > \theta \) all phonon modes are excited because \( k_B T > \hbar \omega_{\text{max}} \). A substantial proportion of all phonon collisions will then be \( \mathbf{U} \) processes, with the attendant high momentum change in the collision. In this regime we can estimate the thermal resistivity without particular distinction between \( \mathbf{N} \) and \( \mathbf{U} \) processes; by the earlier argument about nonlinear effects we expect to find a lattice thermal resistivity \( \propto T \) at high temperatures.

The energy of phonons \( \mathbf{K}_1, \mathbf{K}_2 \) suitable for umklapp to occur is of the order of \( \frac{1}{2} k_B \theta \), because each of the phonons 1 and 2 must have wavevectors of the order of \( \frac{1}{2} \mathbf{G} \) in order for the collision (47) to be possible. If both phonons have low \( \mathbf{K} \), and therefore low energy, there is no way to get from their collision a phonon of wavevector outside the first zone. The umklapp process must conserve energy, just as for the normal process. At low temperatures the number of suitable phonons of the high energy \( \frac{1}{2} k_B \theta \) required may be expected to vary roughly as \( \exp(-\theta/2T) \), according to the Boltzmann factor. The exponential form is in good agreement with experiment. In summary, the phonon mean
free path which enters (42) is the mean free path for umklapp collisions between phonons and not for all collisions between phonons.

Imperfections

Geometrical effects may also be important in limiting the mean free path. We must consider scattering by crystal boundaries, the distribution of isotopic masses in natural chemical elements, chemical impurities, lattice imperfections, and amorphous structures.

When at low temperatures the mean free path $\ell$ becomes comparable with the width of the test specimen, the value of $\ell$ is limited by the width, and the thermal conductivity becomes a function of the dimensions of the specimen. This effect was discovered by de Haas and Biermasz. The abrupt decrease in thermal conductivity of pure crystals at low temperatures is caused by the size effect.

At low temperatures the umklapp process becomes ineffective in limiting the thermal conductivity, and the size effect becomes dominant, as shown in Fig. 18. One would expect then that the phonon mean free path would be constant and of the order of the diameter $D$ of the specimen, so that

$$K \approx CvD.$$  (48)

The only temperature-dependent term on the right is $C$, the heat capacity, which varies as $T^3$ at low temperatures. We expect the thermal conductivity to vary as $T^3$ at low temperatures. The size effect enters whenever the phonon mean free path becomes comparable with the diameter of the specimen.
Dielectric crystals may have thermal conductivities as high as metals. Synthetic sapphire (Al₂O₃) has one of the highest values of the conductivity: nearly 200 W cm⁻¹ K⁻¹ at 30 K. The maximum of the thermal conductivity in sapphire is greater than the maximum of 100 W cm⁻¹ K⁻¹ in copper. Metallic gallium, however, has a conductivity of 845 W cm⁻¹ K⁻¹ at 1.8 K. The electronic contribution to the thermal conductivity of metals is treated in Chapter 6.

In an otherwise perfect crystal, the distribution of isotopes of the chemical elements often provides an important mechanism for phonon scattering. The random distribution of isotopic mass disturbs the periodicity of the density as seen by an elastic wave. In some substances scattering of phonons by isotopes is comparable in importance to scattering by other phonons. Results for germanium are shown in Fig. 19. Enhanced thermal conductivity has been observed also in isotopically pure silicon and diamond; the latter has device importance as a heat sink for laser sources.
Problems

1. Singularity in density of states. (a) From the dispersion relation derived in Chapter 4 for a monatomic linear lattice of N atoms with nearest neighbor interactions, show that the density of modes is

\[ D(\omega) = \frac{2N}{\pi} \cdot \frac{1}{(\omega_m^2 - \omega^2)^{1/2}}, \]

where \( \omega_m \) is the maximum frequency. (b) Suppose that an optical phonon branch has the form \( \omega(K) = \omega_0 - AK^2 \), near \( K = 0 \) in three dimensions. Show that \( D(\omega) = (L/2\pi)^3(2\pi/3\omega_0^2)(\omega_0^2 - \omega^2)^{1/2} \) for \( \omega < \omega_0 \) and \( D(\omega) = 0 \) for \( \omega > \omega_0 \). Here the density of modes is discontinuous.

2. Rms thermal dilation of crystal cell. (a) Estimate for 300 K the root mean square thermal dilation \( \Delta V/V \) for a primitive cell of sodium. Take the bulk modulus as \( 7 \times 10^{10} \text{ erg cm}^{-3} \). Note that the Debye temperature 158 K is less than 300 K, so that the thermal energy is of the order of \( k_B T \). (b) Use this result to estimate the root mean square thermal fluctuation \( \Delta a/a \) of the lattice parameter.

3. Zero point lattice displacement and strain. (a) In the Debye approximation, show that the mean square displacement of an atom at absolute zero is \( \langle R^2 \rangle = 3\hbar\omega_D^2/8\pi^2\rho c^2 \), where \( v \) is the velocity of sound. Start from the result (4.29) summed over the independent lattice modes: \( \langle R^2 \rangle = (\hbar/2\rho v)\Sigma_{\omega}^{-1} \). We have included a factor of \( 1/3 \) to go from mean square amplitude to mean square displacement. (b) Show that \( \Sigma_{\omega}^{-1} \) and \( \langle R^2 \rangle \) diverge for a one-dimensional lattice, but that the mean square strain is finite. Consider \( \langle (\partial R/\partial x)^2 \rangle = \frac{1}{2}\Sigma_{\omega} \) as the mean square strain, and show that it is equal to \( \hbar\omega_D^2 L/4MNc^2 \) for a line of \( N \) atoms each of mass \( M \), counting longitudinal modes only. The divergence of \( R^2 \) is not significant for any physical measurement.

4. Heat capacity of layer lattice. (a) Consider a dielectric crystal made up of layers of atoms, with rigid coupling between layers so that the motion of the atoms is restricted to the plane of the layer. Show that the phonon heat capacity in the Debye approximation in the low temperature limit is proportional to \( T^2 \). (b) Suppose instead, as in many layer structures, that adjacent layers are very weakly bound to each other. What form would you expect the phonon heat capacity to approach at extremely low temperatures?

5. Gruneisen constant. (a) Show that the free energy of a phonon mode of frequency \( \omega \) is \( k_B T \ln [2 \sinh (\hbar \omega / 2k_B T)] \). It is necessary to retain the zero-point energy \( \frac{1}{2} \hbar \omega \) to obtain this result. (b) If \( \Delta \) is the fractional volume change, then the free energy of the crystal may be written as

\[ F(\Delta, T) = \frac{1}{2}B\Delta^2 + k_B T \sum \ln [2 \sinh (\hbar \omega \kappa / 2k_B T)] \]

*This problem is somewhat difficult.
where $B$ is the bulk modulus. Assume that the volume dependence of $\omega_K$ is $\delta\omega/\omega = -\gamma\Delta$, where $\gamma$ is known as the Grünneisen constant. If $\gamma$ is taken as independent of the mode $K$, show that $F$ is a minimum with respect to $\Delta$ when $B\Delta = \gamma\Sigma \hbar \omega \coth (\hbar \omega/2k_B T)$, and show that this may be written in terms of the thermal energy density as $\Delta = \gamma U(T)/B$. (c) Show that on the Debye model $\gamma = -\partial \ln \theta/\partial \ln V$. Note: Many approximations are involved in this theory: the result (a) is valid only if $\omega$ is independent of temperature; $\gamma$ may be quite different for different modes.

References


THERMAL CONDUCTIVITY

Free Electron Fermi Gas

ENERGY LEVELS IN ONE DIMENSION 144
EFFECT OF TEMPERATURE ON THE FERMI-DIRAC DISTRIBUTION 146
FREE ELECTRON GAS IN THREE DIMENSIONS 146
HEAT CAPACITY OF THE ELECTRON GAS 151
   Experimental heat capacity of metals 155
   Heavy fermions 156
ELECTRICAL CONDUCTIVITY AND OHM'S LAW 156
   Experimental electrical resistivity of metals 159
   Umklapp scattering 162
MOTION IN MAGNETIC FIELDS 163
   Hall effect 164
THERMAL CONDUCTIVITY OF METALS 166
   Ratio of thermal to electrical conductivity 166
NANOSTRUCTURES 168
PROBLEMS 169
1. Kinetic energy of electron gas 169
2. Pressure and bulk modulus of an electron gas 169
3. Chemical potential in two dimensions 169
4. Fermi gases in astrophysics 169
5. Liquid He$^3$ 170
6. Frequency dependence of the electrical conductivity 170
7. Dynamic magnetoconductivity tensor for free electrons 170
8. Cohesive energy of free electron Fermi gas 170
9. Static magnetoconductivity tensor 171
10. Maximum surface resistance 171
11. Small metal spheres 171
12. Density of states—nanometric wire 172
13. Quantization of conductance 172
REFERENCES 172
Figure 1 Schematic model of a crystal of sodium metal. The atomic cores are Na$^+$ ions; they are immersed in a sea of conduction electrons. The conduction electrons are derived from the 3s valence electrons of the free atoms. The atomic cores contain 10 electrons in the configuration 1s$^2$2s$^2$2p$^6$. In an alkali metal the atomic cores occupy a relatively small part (~15 percent) of the total volume of the crystal, but in a noble metal (Cu, Ag, Au) the atomic cores are relatively larger and may be in contact with each other. The common crystal structure at room temperature is bcc for the alkali metals and fcc for the noble metals.
In a theory which has given results like these, there must certainly be a great deal of truth.

H. A. Lorentz

We can understand many physical properties of metals, and not only of the simple metals, in terms of the free electron model. According to this model, the valence electrons of the constituent atoms become conduction electrons and move about freely through the volume of the metal.

Even in metals for which the free electron model works best, the charge distribution of the conduction electrons reflects the strong electrostatic potential of the ion cores. The utility of the free electron model is greatest for properties that depend essentially on the kinetic properties of the conduction electrons. The interaction of the conduction electrons with the ions of the lattice is treated in Chapter 7.

The simplest metals are the alkali metals—lithium, sodium, potassium, cesium, and rubidium. In a free atom of sodium the valence electron is in a 3s state; in the metal this electron becomes a conduction electron. We speak of the 3s conduction band.

A monovalent crystal which contains $N$ atoms will have $N$ conduction electrons and $N$ positive ion cores. The Na$^+$ ion core contains 10 electrons that occupy the 1s, 2s, and 2p shells of the free ion, with a spatial distribution that is essentially the same when in the metal as in the free ion.

The ion cores fill only about 15 percent of the volume of a sodium crystal, as in Fig. 1. The radius of the free Na$^+$ ion is 0.98 Å, whereas one-half of the nearest-neighbor distance of the metal is 1.83 Å.

The interpretation of metallic properties in terms of the motion of free electrons was developed long before the invention of quantum mechanics. The classical theory had several conspicuous successes, notably the derivation of the form of Ohm’s law and the relation between the electrical and thermal conductivity. The classical theory fails to explain the heat capacity and the magnetic susceptibility of the conduction electrons. (These are not failures of the free electron model, but failures of the Maxwell distribution function.)

There is a further difficulty. From many types of experiments it is clear that a conduction electron in a metal can move freely in a straight path over many atomic distances, undeflected by collisions with other conduction electrons or by collisions with the atom cores. In a very pure specimen at low temperatures the mean free path may be as long as $10^8$ interatomic spacings (more than 1 cm).

Why is condensed matter so transparent to conduction electrons? The answer to the question contains two parts: (a) A conduction electron is not
deflected by ion cores arranged on a periodic lattice because matter waves propagate freely in a periodic structure. (b) A conduction electron is scattered only infrequently by other conduction electrons. This property is a consequence of the Pauli exclusion principle. By a free electron Fermi gas, we mean a gas of free electrons subject to the Pauli principle.

ENERGY LEVELS IN ONE DIMENSION

Consider a free electron gas in one dimension, taking account of quantum theory and of the Pauli principle. An electron of mass \( m \) is confined to a length \( L \) by infinite barriers (Fig. 2). The wavefunction \( \psi_n(x) \) of the electron is a solution of the Schrödinger equation \( \mathcal{H}\psi = \epsilon \psi \); with the neglect of potential energy we have \( \mathcal{H} = p^2/2m \), where \( p \) is the momentum. In quantum theory \( p \) may be represented by \(-i\hbar \, d/dx\), so that

\[
\mathcal{H}\psi_n = -\frac{\hbar^2}{2m} \frac{d^2\psi_n}{dx^2} = \epsilon_n \psi_n ,
\]

where \( \epsilon_n \) is the energy of the electron in the orbital.

We use the term orbital to denote a solution of the wave equation for a system of only one electron. The term allows us to distinguish between an exact quantum state of the wave equation of a system of \( N \) electrons and an approximate quantum state which we construct by assigning the \( N \) electrons to \( N \) different orbitals, where each orbital is a solution of a wave equation for one electron. The orbital model is exact only if there are no interactions between electrons.

The boundary conditions are \( \psi_n(0) = 0; \psi_n(L) = 0 \), as imposed by the infinite potential energy barriers. They are satisfied if the wavefunction is sinelike with an integral number \( n \) of half-wavelengths between 0 and \( L \):

\[
\psi_n = A \sin \left( \frac{2\pi}{\lambda_n} x \right) ; \quad \frac{1}{2} n\lambda_n = L ,
\]

where \( A \) is a constant. We see that (2) is a solution of (1), because

\[
\frac{d\psi_n}{dx} = A \left( \frac{n\pi}{L} \right) \cos \left( \frac{n\pi}{L} x \right) ; \quad \frac{d^2\psi_n}{dx^2} = -A \left( \frac{n\pi}{L} \right)^2 \sin \left( \frac{n\pi}{L} x \right) ,
\]

whence the energy \( \epsilon_n \) is given by

\[
\epsilon_n = \frac{\hbar^2}{2m} \left( \frac{n\pi}{L} \right)^2 .
\]

We want to accommodate \( N \) electrons on the line. According to the Pauli exclusion principle no two electrons can have all their quantum numbers iden-
tical. That is, each orbital can be occupied by at most one electron. This applies to electrons in atoms, molecules, or solids.

In a linear solid the quantum numbers of a conduction electron orbital are \( n \) and \( m_s \), where \( n \) is any positive integer and the magnetic quantum number \( m_s = \pm \frac{1}{2} \), according to spin orientation. A pair of orbitals labeled by the quantum number \( n \) can accommodate two electrons, one with spin up and one with spin down.

If there are six electrons, then in the ground state of the system the filled orbitals are those given in the table:

<table>
<thead>
<tr>
<th>( n )</th>
<th>( m_s )</th>
<th>Electron occupancy</th>
<th>( n )</th>
<th>( m_s )</th>
<th>Electron occupancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>↑</td>
<td>1</td>
<td>3</td>
<td>↑</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>↓</td>
<td>1</td>
<td>3</td>
<td>↓</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>↑</td>
<td>1</td>
<td>4</td>
<td>↑</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>↓</td>
<td>1</td>
<td>4</td>
<td>↓</td>
<td>0</td>
</tr>
</tbody>
</table>

More than one orbital may have the same energy. The number of orbitals with the same energy is called the degeneracy.

Let \( n_F \) denote the topmost filled energy level, where we start filling the levels from the bottom (\( n = 1 \)) and continue filling higher levels with electrons until all \( N \) electrons are accommodated. It is convenient to suppose that \( N \) is an even number. The condition \( 2n_F = N \) determines \( n_F \), the value of \( n \) for the uppermost filled level.

The Fermi energy \( \epsilon_F \) is defined as the energy of the topmost filled level in the ground state of the \( N \) electron system. By (3) with \( n = n_F \) we have in one dimension:

\[
\epsilon_F = \frac{\hbar^2}{2m} \left( \frac{n_F \pi}{L} \right)^2 = \frac{\hbar^2}{2m} \left( \frac{N \pi}{2L} \right)^2
\]
EFFECT OF TEMPERATURE ON THE FERMI-DIRAC DISTRIBUTION

The ground state is the state of the $N$ electron system at absolute zero. What happens as the temperature is increased? This is a standard problem in elementary statistical mechanics, and the solution is given by the Fermi-Dirac distribution function (Appendix D and TP, Chapter 7).

The kinetic energy of the electron gas increases as the temperature is increased: some energy levels are occupied which were vacant at absolute zero, and some levels are vacant which were occupied at absolute zero (Fig. 3). The Fermi-Dirac distribution gives the probability that an orbital at energy $\epsilon$ will be occupied in an ideal electron gas in thermal equilibrium:

$$ f(\epsilon) = \frac{1}{\exp[(\epsilon - \mu)/k_B T] + 1} \quad (5) $$

The quantity $\mu$ is a function of the temperature; $\mu$ is to be chosen for the particular problem in such a way that the total number of particles in the system comes out correctly—that is, equal to $N$. At absolute zero $\mu = \epsilon_F$, because in the limit $T \to 0$ the function $f(\epsilon)$ changes discontinuously from the value 1 (filled) to the value 0 (empty) at $\epsilon = \epsilon_F = \mu$. At all temperatures $f(\epsilon)$ is equal to $\frac{1}{2}$ when $\epsilon = \mu$, for then the denominator of (5) has the value 2.

The quantity $\mu$ is the chemical potential (TP, Chapter 5), and we see that at absolute zero the chemical potential is equal to the Fermi energy, defined as the energy of the topmost filled orbital at absolute zero.

The high energy tail of the distribution is that part for which $\epsilon - \mu \gg k_B T$; here the exponential term is dominant in the denominator of (5), so that $f(\epsilon) \approx \exp[(\mu - \epsilon)/k_B T]$. This limit is called the Boltzmann or Maxwell distribution.

FREE ELECTRON GAS IN THREE DIMENSIONS

The free-particle Schrödinger equation in three dimensions is

$$ -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi_k(r) = \epsilon_k \psi_k(r) \quad (6) $$

If the electrons are confined to a cube of edge $L$, the wavefunction is the standing wave

$$ \psi_n(r) = A \sin \left( \pi n_x x / L \right) \sin \left( \pi n_y y / L \right) \sin \left( \pi n_z z / L \right) \quad (7) $$

where $n_x$, $n_y$, $n_z$ are positive integers. The origin is at one corner of the cube.

It is convenient to introduce wavefunctions that satisfy periodic boundary conditions, as we did for phonons in Chapter 5. We require the wavefunc-
Figure 3 Fermi-Dirac distribution function at various temperatures, for $T_F = \varepsilon_F/k_B = 50,000$ K. The results apply to a gas in three dimensions. The total number of particles is constant, independent of temperature. The chemical potential at each temperature may be read off the graph as the energy at which $f = 0.5$.

functions to be periodic in $x, y, z$ with period $L$. Thus

$$
\psi(x + L, y, z) = \psi(x, y, z) ,
$$

and similarly for the $y$ and $z$ coordinates. Wavefunctions satisfying the free-particle Schrödinger equation and the periodicity condition are of the form of a traveling plane wave:

$$
\psi_k(r) = \exp (i k \cdot r) ,
$$

provided that the components of the wavevector $k$ satisfy

$$
k_x = 0 ; \quad \pm \frac{2\pi}{L} ; \quad \pm \frac{4\pi}{L} ; \quad \ldots ,
$$

and similarly for $k_y$ and $k_z$.

Any component of $k$ is of the form $2n\pi/L$, where $n$ is a positive or negative integer. The components of $k$ are the quantum numbers of the problem, along with the quantum number $m_s$ for the spin direction. We confirm that these values of $k_x$ satisfy (8), for

$$
\exp[ik_x(x + L)] = \exp[i2n\pi(x + L)/L]
$$

$$
= \exp(i2n\pi x/L) \exp(i2n\pi) = \exp(i2n\pi x/L) = \exp(ik_x x) .
$$
On substituting (9) in (6) we have the energy $\epsilon_k$ of the orbital with wave-vector $k$:
\[ \epsilon_k = \frac{\hbar^2}{2m} k^2 = \frac{\hbar^2}{2m} \left( k_x^2 + k_y^2 + k_z^2 \right). \]
\hspace{1cm} (12)

The magnitude of the wavevector is related to the wavelength $\lambda$ by $k = 2\pi/\lambda$.

The linear momentum $p$ may be represented in quantum mechanics by the operator $p = -i\hbar \nabla$, whence for the orbital (9)
\[ p\psi_k(r) = -i\hbar \nabla \psi_k(r) = i\hbar \psi_k(r), \]
so that the plane wave $\psi_k$ is an eigenfunction of the linear momentum with the eigenvalue $i\hbar k$. The particle velocity in the orbital $k$ is given by $v = \hbar k/m$.

In the ground state of a system of $N$ free electrons the occupied orbitals may be represented as points inside a sphere in $k$ space. The energy at the surface of the sphere is the Fermi energy; the wavevectors at the Fermi surface have a magnitude $k_F$ such that (Fig. 4):
\[ \epsilon_F = \frac{\hbar^2}{2m} k_F^2. \]
\hspace{1cm} (14)

From (10) we see that there is one allowed wavevector—that is, one distinct triplet of quantum numbers $k_x$, $k_y$, $k_z$—for the volume element $(2\pi/L)^3$ of $k$ space. Thus in the sphere of volume $4\pi k_F^3/3$ the total number of orbitals is
\[ 2 \cdot \frac{4\pi k_F^3}{3} = \frac{V}{3\pi^2} k_F^3 = N, \]
where the factor 2 on the left comes from the two allowed values of $m_s$, the spin quantum number, for each allowed value of $k$. Then
\[ k_F = \left( \frac{3\pi^2 N}{V} \right)^{1/3}, \]
\hspace{1cm} (16)

which depends only on the particle concentration.

Using (14),
\[ \epsilon_F = \frac{\hbar^2}{2m} \left( \frac{3\pi^2 N}{V} \right)^{2/3}. \]
\hspace{1cm} (17)

This relates the Fermi energy to the electron concentration $N/V$. The electron velocity $v_F$ at the Fermi surface is
\[ v_F = \left( \frac{\hbar k_F}{m} \right) = \left( \frac{\hbar}{m} \right) \left( \frac{3\pi^2 N}{V} \right)^{1/3}. \]
\hspace{1cm} (18)

Calculated values of $k_F$, $v_F$, and $\epsilon_F$ are given in Table I for selected metals; also given are values of the quantity $T_F$ which is defined as $\epsilon_F/k_B$. (The quantity $T_F$ has nothing to do with the temperature of the electron gas!)
Figure 4  In the ground state of a system of \( N \) free electrons the occupied orbitals of the system fill a sphere of radius \( k_F \), where \( \epsilon_F = \hbar^2 k_F^2 / 2m \) is the energy of an electron having a wavevector \( k_F \).

Figure 5  Density of single-particle states as a function of energy, for a free electron gas in three dimensions. The dashed curve represents the density \( f(\epsilon, T)D(\epsilon) \) of filled orbitals at a finite temperature, but such that \( k_B T \) is small in comparison with \( \epsilon_F \). The shaded area represents the filled orbitals at absolute zero. The average energy is increased when the temperature is increased from 0 to \( T \), for electrons are thermally excited from region 1 to region 2.

We now find an expression for the number of orbitals per unit energy range, \( D(\epsilon) \), called the density of states.\(^1\) We use (17) for the total number of orbitals of energy \( \leq \epsilon \):

\[
N = \frac{V}{3\pi^2} \left( \frac{2m\epsilon}{\hbar^2} \right)^{3/2},
\]

so that the density of states (Fig. 5) is

\[
D(\epsilon) \equiv \frac{dN}{d\epsilon} = \frac{V}{2\pi^2} \cdot \left( \frac{2m}{\hbar^2} \right)^{3/2} \cdot \epsilon^{1/2}.
\]

\(^1\)Strictly, \( D(\epsilon) \) is the density of one-particle states, or density of orbitals. Often one sets the volume \( V = 1 \) in expressions for \( D(\epsilon) \).
Table 1  Calculated free electron Fermi surface parameters for metals at room temperature
(Except for Na, K, Rb, Cs at 5 K and Li at 78 K)

<table>
<thead>
<tr>
<th>Valency</th>
<th>Metal</th>
<th>Electron concentration, in cm(^{-3})</th>
<th>Radius(^a) parameter (r_s)</th>
<th>Fermi wavevector, in cm(^{-1})</th>
<th>Fermi velocity, in cm s(^{-1})</th>
<th>Fermi energy, in eV</th>
<th>Fermi temperature (T_F = \varepsilon_F/k_B), in deg K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Li</td>
<td>4.70 \times 10^{23}</td>
<td>3.25</td>
<td>1.11 \times 10^8</td>
<td>1.29 \times 10^8</td>
<td>4.72</td>
<td>5.48 \times 10^4</td>
</tr>
<tr>
<td></td>
<td>Na</td>
<td>2.65</td>
<td>3.93</td>
<td>0.92</td>
<td>1.07</td>
<td>3.23</td>
<td>3.75</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>1.40</td>
<td>4.86</td>
<td>0.75</td>
<td>0.86</td>
<td>2.12</td>
<td>2.46</td>
</tr>
<tr>
<td></td>
<td>Rb</td>
<td>1.15</td>
<td>5.20</td>
<td>0.70</td>
<td>0.81</td>
<td>1.85</td>
<td>2.15</td>
</tr>
<tr>
<td></td>
<td>Cs</td>
<td>0.91</td>
<td>5.63</td>
<td>0.64</td>
<td>0.75</td>
<td>1.58</td>
<td>1.83</td>
</tr>
<tr>
<td></td>
<td>Cu</td>
<td>8.45</td>
<td>2.67</td>
<td>1.36</td>
<td>1.57</td>
<td>7.00</td>
<td>8.12</td>
</tr>
<tr>
<td></td>
<td>Ag</td>
<td>5.85</td>
<td>3.02</td>
<td>1.20</td>
<td>1.39</td>
<td>5.48</td>
<td>6.36</td>
</tr>
<tr>
<td></td>
<td>Au</td>
<td>5.90</td>
<td>3.01</td>
<td>1.20</td>
<td>1.39</td>
<td>5.51</td>
<td>6.39</td>
</tr>
<tr>
<td>2</td>
<td>Be</td>
<td>24.2</td>
<td>1.88</td>
<td>1.93</td>
<td>2.23</td>
<td>14.14</td>
<td>16.41</td>
</tr>
<tr>
<td></td>
<td>Mg</td>
<td>8.60</td>
<td>2.65</td>
<td>1.37</td>
<td>1.58</td>
<td>7.13</td>
<td>8.27</td>
</tr>
<tr>
<td></td>
<td>Ca</td>
<td>4.60</td>
<td>3.27</td>
<td>1.11</td>
<td>1.28</td>
<td>4.68</td>
<td>5.43</td>
</tr>
<tr>
<td></td>
<td>Sr</td>
<td>3.56</td>
<td>3.56</td>
<td>1.02</td>
<td>1.18</td>
<td>3.95</td>
<td>4.58</td>
</tr>
<tr>
<td></td>
<td>Ba</td>
<td>3.20</td>
<td>3.69</td>
<td>0.98</td>
<td>1.13</td>
<td>3.65</td>
<td>4.24</td>
</tr>
<tr>
<td></td>
<td>Zn</td>
<td>13.10</td>
<td>2.31</td>
<td>1.57</td>
<td>1.82</td>
<td>9.39</td>
<td>10.90</td>
</tr>
<tr>
<td></td>
<td>Cd</td>
<td>9.28</td>
<td>2.59</td>
<td>1.40</td>
<td>1.62</td>
<td>7.46</td>
<td>8.66</td>
</tr>
<tr>
<td>3</td>
<td>Al</td>
<td>18.06</td>
<td>2.07</td>
<td>1.75</td>
<td>2.02</td>
<td>11.63</td>
<td>13.49</td>
</tr>
<tr>
<td></td>
<td>Ca</td>
<td>15.30</td>
<td>2.19</td>
<td>1.65</td>
<td>1.91</td>
<td>10.35</td>
<td>12.01</td>
</tr>
<tr>
<td></td>
<td>In</td>
<td>11.49</td>
<td>2.41</td>
<td>1.50</td>
<td>1.74</td>
<td>8.60</td>
<td>9.98</td>
</tr>
<tr>
<td>4</td>
<td>Pb</td>
<td>13.20</td>
<td>2.30</td>
<td>1.57</td>
<td>1.82</td>
<td>9.37</td>
<td>10.87</td>
</tr>
<tr>
<td></td>
<td>Sn((\nu))</td>
<td>14.48</td>
<td>2.23</td>
<td>1.62</td>
<td>1.88</td>
<td>10.03</td>
<td>11.64</td>
</tr>
</tbody>
</table>

\(^a\)The dimensionless radius parameter is defined as \(r_s = r_0/a_B\), where \(a_B\) is the first Bohr radius and \(r_0\) is the radius of a sphere that contains one electron.
This result may be obtained and expressed more simply by writing (19) as
\[
\ln N = \frac{3}{2} \ln \varepsilon + \text{constant}; \quad \frac{dN}{N} = \frac{3}{2} \frac{d\varepsilon}{\varepsilon},
\]
whence
\[
D(\varepsilon) = \frac{dN}{d\varepsilon} = \frac{3N}{2\varepsilon}.
\]

Within a factor of the order of unity, the number of orbitals per unit energy range at the Fermi energy is the total number of conduction electrons divided by the Fermi energy, just as we would expect.

**HEAT CAPACITY OF THE ELECTRON GAS**

The question that caused the greatest difficulty in the early development of the electron theory of metals concerns the heat capacity of the conduction electrons. Classical statistical mechanics predicts that a free particle should have a heat capacity of \(\frac{3}{2}k_B\), where \(k_B\) is the Boltzmann constant. If \(N\) atoms each give one valence electron to the electron gas, and the electrons are freely mobile, then the electronic contribution to the heat capacity should be \(\frac{3}{2}Nk_B\), just as for the atoms of a monatomic gas. But the observed electronic contribution at room temperature is usually less than 0.01 of this value.

This discrepancy distracted the early workers, such as Lorentz: how can the electrons participate in electrical conduction processes as if they were mobile, while not contributing to the heat capacity? The question was answered only upon the discovery of the Pauli exclusion principle and the Fermi distribution function. Fermi found the correct equation, and he wrote, "One recognizes that the specific heat vanishes at absolute zero and that at low temperatures it is proportional to the absolute temperature."

When we heat the specimen from absolute zero not every electron gains an energy \(-k_BT\) as expected classically, but only those electrons in orbitals within an energy range \(k_BT\) of the Fermi level are excited thermally; these electrons gain an energy which is itself of the order of \(k_BT\), as in Fig. 5. This gives an immediate qualitative solution to the problem of the heat capacity of the conduction electron gas. If \(N\) is the total number of electrons, only a fraction of the order of \(T/T_F\) can be excited thermally at temperature \(T\), because only these lie within an energy range of the order of \(k_BT\) of the top of the energy distribution.

Each of these \(NT/T_F\) electrons has a thermal energy of the order of \(k_BT\). The total electronic thermal kinetic energy \(U\) is of the order of
\[
U = (NT/T_F)k_BT.
\]
The electronic heat capacity is given by

$$C_{el} = \frac{\partial U}{\partial T} = N k_B (T/T_F)$$  \hspace{1cm} (23)$$

and is directly proportional to $T$, in agreement with the experimental results discussed in the following section. At room temperature $C_{el}$ is smaller than the classical value $\frac{3}{2} N k_B$ by a factor of the order of 0.01 or less, for $T_F \sim 5 \times 10^4$ K.

We now derive a quantitative expression for the electronic heat capacity valid at low temperatures $k_B T \ll \varepsilon_F$. The increase $\Delta U = U(T) - U(0)$ in the total energy (Fig. 5) of a system of $N$ electrons when heated from 0 to $T$ is

$$\Delta U = \int_0^\infty d\varepsilon \varepsilon D(\varepsilon) f(\varepsilon) - \int_0^{\varepsilon_F} d\varepsilon \varepsilon D(\varepsilon).$$  \hspace{1cm} (24)$$

Here $f(\varepsilon)$ is the Fermi-Dirac function and $D(\varepsilon)$ is the number of orbitals per unit energy range. We multiply the identity

$$N = \int_0^\infty d\varepsilon \varepsilon D(\varepsilon) f(\varepsilon) = \int_0^{\varepsilon_F} d\varepsilon \varepsilon D(\varepsilon)$$  \hspace{1cm} (25)$$

by $\varepsilon_F$ to obtain

$$\left(\int_0^{\varepsilon_F} + \int_{\varepsilon_F}^\infty\right) d\varepsilon \varepsilon_F f(\varepsilon) D(\varepsilon) = \int_0^{\varepsilon_F} d\varepsilon \varepsilon_F D(\varepsilon).$$  \hspace{1cm} (26)$$

We use (26) to rewrite (24) as

$$\Delta U = \int_{\varepsilon_F}^\infty d\varepsilon (\varepsilon - \varepsilon_F) f(\varepsilon) D(\varepsilon) + \int_0^{\varepsilon_F} d\varepsilon (\varepsilon_F - \varepsilon) [1 - f(\varepsilon)] D(\varepsilon).$$  \hspace{1cm} (27)$$

The first integral on the right-hand side of (27) gives the energy needed to take electrons from $\varepsilon_F$ to the orbitals of energy $\varepsilon > \varepsilon_F$, and the second integral gives the energy needed to bring the electrons to $\varepsilon_F$ from orbitals below $\varepsilon_F$. Both contributions to the energy are positive.

The product $f(\varepsilon) D(\varepsilon) d\varepsilon$ in the first integral is the number of electrons elevated to orbitals in the energy range $d\varepsilon$ at an energy $\varepsilon$. The factor $[1 - f(\varepsilon)]$ in the second integral is the probability that an electron has been removed from an orbital $\varepsilon$. The function $\Delta U$ is plotted in Fig. 6. In Fig. 3 we plotted the Fermi-Dirac distribution function versus $\varepsilon$ for six values of the temperature. The electron concentration of the Fermi gas was taken such that $\varepsilon_F/k_B = 50,000$ K, characteristic of the conduction electrons in a metal.

The heat capacity of the electron gas is found on differentiating $\Delta U$ with respect to $T$. The only temperature-dependent term in (27) is $f(\varepsilon)$, whence we can group terms to obtain

$$C_{el} = \frac{dU}{dT} = \int_0^\infty d\varepsilon (\varepsilon - \varepsilon_F) \frac{df}{dT} D(\varepsilon).$$  \hspace{1cm} (28)$$

At the temperatures of interest in metals $T/\varepsilon_F < 0.01$, and we see from Fig. 3 that $(\varepsilon - \varepsilon_F) df/dT$ has large positive peaks at energies near $\varepsilon_F$. It is a good
Figure 6  Temperature dependence of the energy of a noninteracting fermion gas in three dimensions. The energy is plotted in normalized form as $\Delta U/N\varepsilon_F$, where $N$ is the number of electrons. The temperature is plotted as $k_B T/\varepsilon_F$.

Figure 7  Plot of the chemical potential $\mu$ versus temperature $k_B T$ for a gas of noninteracting fermions in three dimensions. For convenience in plotting, the units of $\mu$ and $k_B T$ are $0.763\varepsilon_F$.

approximation to evaluate the density of states $D(\varepsilon)$ at $\varepsilon_F$ and take it outside of the integral:

$$C_{el} \equiv D(\varepsilon_F) \int_{0}^{\infty} d\varepsilon(\varepsilon - \varepsilon_F) \frac{df}{dT}.$$  \hspace{1cm} (29)

Examination of the graphs in Figs. 7 and 8 of the variation of $\mu$ with $T$ suggests that when $k_B T \ll \varepsilon_F$ we ignore the temperature dependence of the chemical potential $\mu$ in the Fermi-Dirac distribution function and replace $\mu$ by
the constant $\epsilon_F$. We have then, with $\tau = k_B T$,

$$
\frac{df}{d\tau} = \frac{\epsilon - \epsilon_F}{\tau^2} \cdot \frac{\exp[(\epsilon - \epsilon_F)/\tau]}{\exp[(\epsilon - \epsilon_F)/\tau] + 1}^2.
$$

We set

$$
x = (\epsilon - \epsilon_F)/\tau,
$$

and it follows from (29) and (30) that

$$
C_{el} = k_B T D(\epsilon_F) \int_{-\epsilon_F/\tau}^{\infty} dx \frac{x^2 e^x}{(e^x + 1)^2}.
$$

We may safely replace the lower limit by $-\infty$ because the factor $e^x$ in the integrand is already negligible at $x = -\epsilon_F/\tau$ if we are concerned with low temperatures such that $\epsilon_F/\tau \sim 100$ or more. The integral$^2$ becomes

$$
\int_{-\infty}^{\infty} dx \frac{x^2 e^x}{(e^x + 1)^2} = \frac{\pi^2}{3},
$$

$^2$The integral is not elementary, but may be evaluated from the more familiar result

$$
\int_0^{\infty} dx \frac{x}{e^{ax} + 1} = \frac{\pi^2}{12a^2},
$$
on differentiation of both sides with respect to the parameter $a$.  

Figure 8. Variation with temperature of the chemical potential $\mu$, for free electron Fermi gases in one and three dimensions. In common metals $\pi/\epsilon_F \approx 0.01$ at room temperature, so that $\mu$ is closely equal to $\epsilon_F$. These curves were calculated from series expansions of the integral for the number of particles in the system.
whence the heat capacity of an electron gas is

$$C_{el} = \frac{1}{3} \pi^2 D(\varepsilon_F) k_B T .$$

(34)

From (21) we have

$$D(\varepsilon_F) = \frac{3N}{2\varepsilon_F} = \frac{3N}{2k_B T_F}$$

(35)

for a free electron gas with $k_B T_F = \varepsilon_F$. Thus (34) becomes

$$C_{el} = \frac{1}{3} \pi^2 N k_B T / T_F .$$

(36)

Recall that although $T_F$ is called the Fermi temperature, it is not an actual temperature, but only a convenient reference notation.

**Experimental Heat Capacity of Metals**

At temperatures much below both the Debye temperature and the Fermi temperature, the heat capacity of metals may be written as the sum of electron and phonon contributions: $C = \gamma T + A T^3$, where $\gamma$ and $A$ are constants characteristic of the material. The electronic term is linear in $T$ and is dominant at sufficiently low temperatures. It is convenient to exhibit the experimental values of $C$ as a plot of $C/T$ versus $T^2$:

$$C/T = \gamma + A T^2 ,$$

(37)

for then the points should lie on a straight line with slope $A$ and intercept $\gamma$. Such a plot for potassium is shown in Fig. 9. Observed values of $\gamma$, called the Sommerfeld parameter, are given in Table 2.

The observed values of the coefficient $\gamma$ are of the expected magnitude, but often do not agree very closely with the value calculated for free electrons of mass $m$ by use of (34). It is common practice to express the ratio of the observed to the free electron values of the electronic heat capacity as a ratio of
a thermal effective mass \( m_{\text{th}} \) to the electron mass \( m \), where \( m_{\text{th}} \) is defined by the relation
\[
\frac{m_{\text{th}}}{m} = \frac{\gamma(\text{observed})}{\gamma(\text{free})} .
\] (38)

This form arises in a natural way because \( \epsilon_F \) is inversely proportional to the mass of the electron, whence \( \gamma \propto m \). Values of the ratio are given in Table 2. The departure from unity involves three separate effects:

- The interaction of the conduction electrons with the periodic potential of the rigid crystal lattice. The effective mass of an electron in this potential is called the band effective mass and is treated later.
- The interaction of the conduction electrons with phonons. An electron tends to polarize or distort the lattice in its neighborhood, so that the moving electron tries to drag nearby ions along, thereby increasing the effective mass of the electron.
- The interaction of the conduction electrons with themselves. A moving electron causes an inertial reaction in the surrounding electron gas, thereby increasing the effective mass of the electron.

**Heavy Fermions.** Several metallic compounds have been discovered that have enormous values, two or three orders of magnitude higher than usual, of the electronic heat capacity. The heavy fermion compounds include UBe\(_{13}\), CeAl\(_3\), and CeCu\(_2\)Si\(_2\). It has been suggested that \( f \) electrons in these compounds may have inertial masses as high as 1000 \( m \), because of the weak overlap of wavefunctions of \( f \) electrons on neighboring ions (see Chapter 9, "tight binding"). References are given by Z. Fisk, J. L. Smith, and H. R. Ott, Physics Today, 38, S-20 (January, 1985). The heavy fermion compounds form a class of superconductors known as "exotic superconductors."

**ELECTRICAL CONDUCTIVITY AND OHM'S LAW**

The momentum of a free electron is related to the wavevector by \( mv = \hbar k \). In an electric field \( \mathbf{E} \) and magnetic field \( \mathbf{B} \) the force \( \mathbf{F} \) on an electron of charge \( -e \) is \(-e[\mathbf{E} + (1/c)\mathbf{v} \times \mathbf{B}]\), so that Newton's second law of motion becomes

\[
(CGS) \quad \mathbf{F} = m \frac{d\mathbf{v}}{dt} = \hbar \frac{d\mathbf{k}}{dt} = -e \left( \mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{B} \right) .
\] (39)

In the absence of collisions the Fermi sphere (Fig. 10) in \( \mathbf{k} \) space is displaced at a uniform rate by a constant applied electric field. We integrate with \( \mathbf{B} = 0 \) to obtain
\[
k(t) - k(0) = -eEt/\hbar .
\] (40)
Table 2  Experimental and free electron values of electronic heat capacity constant $\gamma$ of metals

(From compilations kindly furnished by N. Phillips and N. Pearlman. The thermal effective mass is defined by Eq. (38).)

|     | Li  | Be  | Na  | Mg  | K   | Ca  | Sc  | Ti  | V   | Cr  | Mn($\gamma$) | Fe  | Co  | Ni  | Cu  | Zn  | Ga  | Ge  | As  | B   | C   | N   | Al  | Si  | P   |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1.63| 0.17|     | 1.38| 1.3 | 1.668| 1.511|     |     |     |     |             |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 0.749| 0.500|     | 1.094| 0.992| 1.25 | 1.9 |     |     |     |     |             |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 2.18 | 0.34 |     | 1.26 | 1.3 |     |     |     |     |     |     |             |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

Observed $\gamma$ in mJ mol$^{-1}$ K$^{-2}$.

Calculated free electron $\gamma$ in mJ mol$^{-1}$ K$^{-2}$

$m_{\text{eff}}/m = \text{(observed } \gamma/\text{(free electron } \gamma)$.}
Figure 10 (a) The Fermi sphere encloses the occupied electron orbitals in k space in the ground state of the electron gas. The net momentum is zero, because for every orbital k there is an occupied orbital at -k. (b) Under the influence of a constant force F acting for a time interval t every orbital has its k vector increased by $\delta k = Ft/\hbar$. This is equivalent to a displacement of the whole Fermi sphere by $\delta k$. The total momentum is $N\hbar\delta k$, if there are N electrons present. The application of the force increases the energy of the system by $N(h\delta k)^2/2m$.

If the field is applied at time $t = 0$ to an electron gas that fills the Fermi sphere centered at the origin of k space, then at a later time $t$ the sphere will be displaced to a new center at

$$\delta k = -eEt/\hbar.$$  \hspace{1cm} (41)

Notice that the Fermi sphere is displaced as a whole.

Because of collisions of electrons with impurities, lattice imperfections, and phonons, the displaced sphere may be maintained in a steady state in an electric field. If collision time is $\tau$, the displacement of the Fermi sphere in the steady state is given by (41) with $t = \tau$. The incremental velocity is $v = -eE\tau/m$. If in a constant electric field $E$ there are $n$ electrons of charge $q = -e$ per unit volume, the electric current density is

$$j = nqv = ne^2\tau E/m.$$  \hspace{1cm} (42)

This is Ohm's law.

The electrical conductivity $\sigma$ is defined by $j = \sigma E$, so that

$$\sigma = \frac{ne^2\tau}{m}.$$  \hspace{1cm} (43)

The electrical resistivity $\rho$ is defined as the reciprocal of the conductivity, so that

$$\rho = m/ne^2\tau.$$  \hspace{1cm} (44)
Values of the electrical conductivity and resistivity of the elements are given in Table 3. It is useful to remember that $\sigma$ in Gaussian units has the dimensions of frequency.

It is easy to understand the result (43) for the conductivity. We expect the charge transported to be proportional to the charge density $ne$; the factor $elm$ enters because the acceleration in a given electric field is proportional to $e$ and inversely proportional to the mass $m$. The time $\tau$ describes the free time during which the field acts on the carrier. Closely the same result for the electrical conductivity is obtained for a classical (Maxwellian) gas of electrons, as realized at low carrier concentration in many semiconductor problems. The mathematics of this similarity is developed in the section on transport theory in TP, Chapter 14.

It is possible to obtain crystals of copper so pure that their conductivity at liquid helium temperatures (4 K) is nearly $10^5$ times that at room temperature; for these conditions $\tau \approx 2 \times 10^{-9}$ s at 4 K. The mean free path $\ell$ of a conduction electron is defined as

$$\ell = v_F \tau,$$

where $v_F$ is the velocity at the Fermi surface, because all collisions involve only electrons near the Fermi surface. From Table 1 we have $v_F = 1.57 \times 10^8$ cm s$^{-1}$ for Cu; thus the mean free path is

$$\ell(4 \text{ K}) \approx 0.3 \text{ cm} ; \quad \ell(300 \text{ K}) \approx 3 \times 10^{-6} \text{ cm}.$$

Mean free paths as long as 10 cm have been observed in very pure metals in the liquid helium temperature range. For electron–electron collisions, see Eq. (10.63).

**Experimental Electrical Resistivity of Metals**

The electrical resistivity of most metals is dominated at room temperature (300 K) by collisions of the conduction electrons with lattice phonons and at liquid helium temperature (4 K) by collisions with impurity atoms and mechanical imperfections in the lattice (Fig. 11). The rates of these collisions are often independent to a good approximation, so that if the electric field were switched off the momentum distribution would relax back to its ground state with the net relaxation time

$$\frac{1}{\tau} = \frac{1}{\tau_L} + \frac{1}{\tau_i},$$

where $\tau_L$ and $\tau_i$ are the collision times for scattering by phonons and by imperfections, respectively.

The net resistivity is given by

$$\rho = \rho_L + \rho_i,$$

(47)
<table>
<thead>
<tr>
<th>Li</th>
<th>Be</th>
<th>Al</th>
<th>Si</th>
<th>P</th>
<th>S</th>
<th>Cl</th>
<th>Ar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.07</td>
<td>3.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.32</td>
<td>3.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Na</th>
<th>Mg</th>
<th>K</th>
<th>Ca</th>
<th>Sc</th>
<th>Ti</th>
<th>V</th>
<th>Cr</th>
<th>Mn</th>
<th>Fe</th>
<th>Co</th>
<th>Ni</th>
<th>Cu</th>
<th>Zn</th>
<th>Ga</th>
<th>Ge</th>
<th>As</th>
<th>Se</th>
<th>Br</th>
<th>Kr</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.11</td>
<td>2.33</td>
<td>1.39</td>
<td>2.78</td>
<td>0.21</td>
<td>0.23</td>
<td>0.50</td>
<td>0.78</td>
<td>0.072</td>
<td>1.02</td>
<td>1.72</td>
<td>1.43</td>
<td>5.88</td>
<td>1.69</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.75</td>
<td>4.30</td>
<td>7.19</td>
<td>3.6</td>
<td>46.8</td>
<td>43.1</td>
<td>19.9</td>
<td>12.9</td>
<td>139.</td>
<td>9.8</td>
<td>5.8</td>
<td>7.0</td>
<td>1.70</td>
<td>5.92</td>
<td>1.48</td>
<td>1.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rb</th>
<th>Sr</th>
<th>Y</th>
<th>Zr</th>
<th>Nb</th>
<th>Mo</th>
<th>Tc</th>
<th>Ru</th>
<th>Rh</th>
<th>Pd</th>
<th>Ag</th>
<th>Cd</th>
<th>In</th>
<th>Sn</th>
<th>Sb</th>
<th>Te</th>
<th>I</th>
<th>Xe</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80</td>
<td>0.47</td>
<td>0.17</td>
<td>0.24</td>
<td>0.69</td>
<td>1.89</td>
<td>0.07</td>
<td>0.78</td>
<td>0.072</td>
<td>1.02</td>
<td>1.72</td>
<td>1.43</td>
<td>5.88</td>
<td>1.69</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td>21.5</td>
<td>58.5</td>
<td>42.4</td>
<td>14.5</td>
<td>5.3</td>
<td>1.4</td>
<td>7.4</td>
<td>4.8</td>
<td>10.5</td>
<td>1.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cs</th>
<th>Ba</th>
<th>La</th>
<th>Hf</th>
<th>Ta</th>
<th>W</th>
<th>Re</th>
<th>Os</th>
<th>Ir</th>
<th>Pt</th>
<th>Au</th>
<th>Hg</th>
<th>In</th>
<th>Tl</th>
<th>Pb</th>
<th>Bi</th>
<th>Po</th>
<th>At</th>
<th>Rn</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>0.26</td>
<td>0.13</td>
<td>0.33</td>
<td>0.76</td>
<td>1.89</td>
<td>0.54</td>
<td>1.16</td>
<td>1.96</td>
<td>0.96</td>
<td>4.55</td>
<td>0.10</td>
<td>0.61</td>
<td>0.48</td>
<td>0.086</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.0</td>
<td>39.</td>
<td>79.</td>
<td>30.6</td>
<td>13.1</td>
<td>5.3</td>
<td>18.6</td>
<td>9.1</td>
<td>5.1</td>
<td>10.4</td>
<td>2.20</td>
<td>95.9</td>
<td>16.4</td>
<td>21.0</td>
<td>116.</td>
<td>46.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Fr  | Ra  | Ac   | Ce   | Pr   | Nd   | Pm   | Sm   | Eu   | Gd   | Tb   | Dy   | Ho   | Er   | Tm   | Yb   | Lu   | Th   | Pa   | U    | Np   | Pu   | Am   | Cm   | Bk   | Cf   | Es   | Fm   | Md   | No   | Lr   |      |
|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 0.12| 0.17|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 81. | 67. | 59.  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |

| Th  | Pa  | U    | Np   | Pu   | Am   | Cm   | Bk   | Cf   | Es   | Fm   | Md   | No   | Lr   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 0.66| 0.39| 0.085| 0.070|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 15.2| 25.7| 118. | 143. |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
Figure 11 Electrical resistivity in most metals arises from collisions of electrons with irregularities in the lattice, as in (a) by phonons and in (b) by impurities and vacant lattice sites.

Figure 12 Resistance of potassium below 20 K, as measured on two specimens by D. MacDonald and K. Mendelssohn. The different intercepts at 0 K are attributed to different concentrations of impurities and static imperfections in the two specimens.

where $\rho_L$ is the resistivity caused by the thermal phonons, and $\rho_i$ is the resistivity caused by scattering of the electron waves by static defects that disturb the periodicity of the lattice. Often $\rho_L$ is independent of the number of defects when their concentration is small, and often $\rho_i$ is independent of temperature. This empirical observation expresses Matthiessen’s rule, which is convenient in analyzing experimental data (Fig. 12).

The residual resistivity, $\rho_r(0)$, is the extrapolated resistivity at 0 K because $\rho_L$ vanishes as $T \to 0$. The lattice resistivity, $\rho_L(T) = \rho - \rho_r(0)$, is the same for different specimens of a metal, even though $\rho_r(0)$ may itself vary widely. The
resistivity ratio of a specimen is usually defined as the ratio of its resistivity at room temperature to its residual resistivity. It is a convenient approximate indicator of sample purity: for many materials an impurity in solid solution creates a residual resistivity of about 1 μohm-cm \((1 \times 10^{-6} \text{ ohm-cm})\) per atomic percent of impurity. A copper specimen with a resistivity ratio of 1000 will have a residual resistivity of \(1.7 \times 10^{-3} \text{ μohm-cm}\), corresponding to an impurity concentration of about 20 ppm. In exceptionally pure specimens the resistivity ratio may be as high as \(10^6\), whereas in some alloys (e.g., manganin) it is as low as 1.1.

The temperature-dependent part of the electrical resistivity is proportional to the rate at which an electron collides with thermal phonons and thermal electrons (Chapter 10). The collision rate with phonons is proportional to the concentration of thermal phonons. One simple limit is at temperatures over the Debye temperature \(\theta\): here the phonon concentration is proportional to the temperature \(T\), so that \(\rho \propto T\) for \(T > \theta\). A sketch of the theory is given in Appendix J.

**Umklapp Scattering**

Umklapp scattering of electrons by phonons (Chapter 5) accounts for most of the electrical resistivity of metals at low temperatures. These are electron-phonon scattering processes in which a reciprocal lattice vector \(\mathbf{G}\) is involved, so that electron momentum change in the process may be much larger than in a normal electron-phonon scattering process at low temperatures. (In an umklapp process the wavevector of one particle may be “flipped over.”)

Consider a section perpendicular to [100] through two adjacent Brillouin zones in bcc potassium, with the equivalent Fermi spheres inscribed within each (Fig. 13). The lower half of the figure shows the normal electron-phonon collision \(k' = k + q\), while the upper half shows a possible scattering process \(k' = k + q + G\) involving the same phonon and terminating outside the first Brillouin zone, at the point \(A\). This point is exactly equivalent to the point \(A'\) inside the original zone, where \(AA'\) is a reciprocal lattice vector \(G\).

This scattering is an umklapp process, in analogy to phonons. Such collisions are strong scatterers because the scattering angle can be close to \(\pi\), and a single collision can practically restore the electron to its ground orbital.

When the Fermi surface does not intersect the zone boundary, there is some minimum phonon wavevector \(q_0\) for umklapp scattering. At low enough temperatures the number of phonons available for umklapp scattering falls as \(\exp(-\theta_U/T)\), where \(\theta_U\) is a characteristic temperature calculable from the geometry of the Fermi surface inside the Brillouin zone. For a spherical Fermi surface with one electron orbital per atom inside the bcc Brillouin zone, one can show by geometry that \(q_0 = 0.267 \ k_F\).
Figure 13 Two Fermi spheres in adjacent zones: a construction to show the role of phonon umklapp processes in electrical resistivity.

The experimental data for potassium have the expected exponential form with $\theta_U = 23$ K compared with the Debye $\theta = 91$ K. At the very lowest temperatures (below about 2 K in potassium) the number of umklapp processes is negligible and the lattice resistivity is then caused only by small angle scattering, which is the normal scattering.

Bloch obtained an analytic result for the normal scattering, with $\rho_L \propto T^5/\theta^6$ at very low temperatures. This is a classic limiting result. These normal processes contribute to the resistivity in all metals, but they have not yet been clearly isolated for any metal because of the large competing effects of imperfection scattering, electron-electron scattering, and umklapp scattering.

MOTION IN MAGNETIC FIELDS

By the argument of (39) and (41) we are led to the equation of motion for the displacement $\delta k$ of a Fermi sphere of particles acted on by a force $F$ and by friction as represented by collisions:

$$\hbar \left( \frac{d}{dt} + \frac{1}{\tau} \right) \delta k = F.$$  \hspace{1cm} (48)

The free particle acceleration term is $(\hbar d/dt) \delta k$ and the effect of collisions (the friction) is represented by $\hbar \delta k/\tau$, where $\tau$ is the collision time.

Consider now the motion of the system in a uniform magnetic field $B$. The Lorentz force on an electron is

(CGS) $$F = -e \left( E + \frac{1}{c} v \times B \right);$$  \hspace{1cm} (49)

(SI) $$F = -e (E + v \times B)$$

If $m v = \hbar \delta k$, then the equation of motion is

(CGS) $$m \left( \frac{d}{dt} + \frac{1}{\tau} \right) v = -e \left( E + \frac{1}{c} v \times B \right).$$  \hspace{1cm} (50)
An important situation is the following: let a static magnetic field $\mathbf{B}$ lie along the $z$ axis. Then the equations of motion are

$$(\text{CGS}) \quad m \left( \frac{d}{dt} + \frac{1}{\tau} \right) v_x = -e \left( E_x + \frac{B}{c} v_y \right) ;$$

$$(\text{CGS}) \quad m \left( \frac{d}{dt} + \frac{1}{\tau} \right) v_y = -e \left( E_y - \frac{B}{c} v_x \right) ;$$

$$(\text{CGS}) \quad m \left( \frac{d}{dt} + \frac{1}{\tau} \right) v_z = -eE_z .$$

The results in SI are obtained by replacing $c$ by 1.

In the steady state in a static electric field the time derivatives are zero, so that the drift velocity is

$$v_x = -\frac{e\tau}{m} E_x - \omega_c \tau v_y ; \quad v_y = -\frac{e\tau}{m} E_y + \omega_c \tau v_x ; \quad v_z = -\frac{e\tau}{m} E_z ,$$

(52)

where $\omega_c = eB/mc$ is the cyclotron frequency, as discussed in Chapter 8 for cyclotron resonance in semiconductors.

**Hall Effect**

The Hall field is the electric field developed across two faces of a conductor, in the direction $\mathbf{j} \times \mathbf{B}$, when a current $\mathbf{j}$ flows across a magnetic field $\mathbf{B}$. Consider a rod-shaped specimen in a longitudinal electric field $E_x$ and a transverse magnetic field, as in Fig. 14. If current cannot flow out of the rod in the $y$ direction we must have $\delta v_y = 0$. From (52) this is possible only if there is a transverse electric field

$$(\text{CGS}) \quad E_y = -\omega_c \tau E_x = -\frac{eB\tau}{mc} E_x ;$$

(53)

$$(\text{SI}) \quad E_y = -\omega_c \tau E_x = -\frac{eB\tau}{m} E_x .$$

The quantity defined by

$$R_H = \frac{E_y}{j_x B}$$

(54)

is called the Hall coefficient. To evaluate it on our simple model we use $j_x = ne^2 \tau E_x / m$ and obtain

$$(\text{CGS}) \quad R_H = -\frac{eB\tau E_x/mc}{ne^2 \tau E_x B/m} = -\frac{1}{ne} ;$$

(55)

$$(\text{SI}) \quad R_H = -\frac{1}{ne} .$$
This is negative for free electrons, for $e$ is positive by definition.

The lower the carrier concentration, the greater the magnitude of the Hall coefficient. Measuring $R_H$ is an important way of measuring the carrier concentration.

The symbol $R_H$ denotes the Hall coefficient (54), but it is sometimes used with a different meaning, that of Hall resistance in two-dimensional problems. When we treat such problems in Chapter 19, we shall instead let

$$\rho_H = BR_H = E_y/j_x$$

(55a)

denote the Hall resistance, where $j_x$ is the surface current density.

The simple result (55) follows from the assumption that all relaxation times are equal, independent of the velocity of the electron. A numerical factor of order unity enters if the relaxation time is a function of the velocity. The expression becomes somewhat more complicated if both electrons and holes contribute to the conductivity. The theory of the Hall effect again becomes simple in high magnetic fields such that $\omega_c \tau \gg 1$, where $\omega_c$ is the cyclotron frequency and $\tau$ the relaxation time. (See QTS, pp. 241–244.)
In Table 4 observed values of the Hall coefficient are compared with values calculated from the carrier concentration. The most accurate measurements are made by the method of helicon resonance which is treated as a problem in Chapter 10. In the table "conv." stands for "conventional."

The accurate values for sodium and potassium are in excellent agreement with values calculated for one conduction electron per atom, using (55). Notice, however, the experimental values for the trivalent elements aluminum and indium: these agree with values calculated for one positive charge carrier per atom and thus disagree in magnitude and sign with values calculated for the expected three negative charge carriers.

The problem of an apparent positive sign for the charge carriers arises also for Be and As, as seen in the table. The anomaly of the sign was explained by Peierls (1928). The motion of carriers of apparent positive sign, which Heisenberg later called "holes," cannot be explained by a free electron gas, but finds a natural explanation in terms of the energy band theory developed in Chapters 7–9. Band theory also accounts for the occurrence of very large values of the Hall coefficient, as for As, Sb, and Bi.

**THERMAL CONDUCTIVITY OF METALS**

In Chapter 5 we found an expression \( K = \frac{1}{3} C v \ell \) for the thermal conductivity of particles of velocity \( v \), heat capacity \( C \) per unit volume, and mean free path \( \ell \). The thermal conductivity of a Fermi gas follows from (36) for the heat capacity, and with \( \varepsilon_F = \frac{1}{2} m v_F^2 \):

\[
K_{el} = \frac{\pi^2}{3} \frac{n k_B^2 T}{m v_F^2} \cdot v_F \cdot \ell = \frac{\pi^2 n k_B^2 T \tau}{3m} \quad (56)
\]

Here \( \ell = v_F \tau \); the electron concentration is \( n \), and \( \tau \) is the collision time.

Do the electrons or the phonons carry the greater part of the heat current in a metal? In pure metals the electronic contribution is dominant at all temperatures. In impure metals or in disordered alloys, the electron mean free path is reduced by collision with impurities, and the phonon contribution may be comparable with the electronic contribution.

**Ratio of Thermal to Electrical Conductivity**

The **Wiedemann-Franz law** states that for metals at not too low temperatures the ratio of the thermal conductivity to the electrical conductivity is directly proportional to the temperature, with the value of the constant of proportionality independent of the particular metal. This result was important in the
Table 4 Comparison of observed Hall coefficients with free electron theory

[The experimental values of $R_H$ as obtained by conventional methods are summarized from data at room temperature presented in the Landolt-Bornstein tables. The values obtained by the helicon wave method at $4 \, K$ are by J. M. Goodman. The values of the carrier concentration $n$ are from Table 1.4 except for Na, K, Al, In, where Goodman’s values are used. To convert the value of $R_H$ in CGS units to the value in volt-cm/amp-gauss, multiply by $9 \times 10^{11}$; to convert $R_H$ in CGS to m$^3$/coulomb, multiply by $9 \times 10^{13}$.]

<table>
<thead>
<tr>
<th>Metal</th>
<th>Method</th>
<th>Experimental $R_H$, in $10^{-24}$ CGS units</th>
<th>Assumed carriers per atom</th>
<th>Calculated $-1/nec$, in $10^{-24}$ CGS units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li</td>
<td>conv.</td>
<td>$-1.89$</td>
<td>1 electron</td>
<td>$-1.48$</td>
</tr>
<tr>
<td>Na</td>
<td>helicon</td>
<td>$-2.619$</td>
<td>1 electron</td>
<td>$-2.603$</td>
</tr>
<tr>
<td></td>
<td>conv.</td>
<td>$-2.3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>helicon</td>
<td>$-4.946$</td>
<td>1 electron</td>
<td>$-4.944$</td>
</tr>
<tr>
<td></td>
<td>conv.</td>
<td>$-4.7$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rb</td>
<td>conv.</td>
<td>$-5.6$</td>
<td>1 electron</td>
<td>$-6.04$</td>
</tr>
<tr>
<td>Cu</td>
<td>conv.</td>
<td>$-0.6$</td>
<td>1 electron</td>
<td>$-0.82$</td>
</tr>
<tr>
<td>Ag</td>
<td>conv.</td>
<td>$-1.0$</td>
<td>1 electron</td>
<td>$-1.19$</td>
</tr>
<tr>
<td>Au</td>
<td>conv.</td>
<td>$-0.8$</td>
<td>1 electron</td>
<td>$-1.18$</td>
</tr>
<tr>
<td>Be</td>
<td>conv.</td>
<td>$+2.7$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg</td>
<td>conv.</td>
<td>$-0.92$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al</td>
<td>helicon</td>
<td>$+1.136$</td>
<td>1 hole</td>
<td>$+1.135$</td>
</tr>
<tr>
<td>In</td>
<td>helicon</td>
<td>$+1.774$</td>
<td>1 hole</td>
<td>$+1.780$</td>
</tr>
<tr>
<td>As</td>
<td>conv.</td>
<td>$+50.$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sb</td>
<td>conv.</td>
<td>$-22.$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bi</td>
<td>conv.</td>
<td>$-6000.$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5  Experimental Lorenz numbers

<table>
<thead>
<tr>
<th></th>
<th>$L \times 10^8$ watt-ohm/deg$^2$</th>
<th></th>
<th>$L \times 10^8$ watt-ohm/deg$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0°C</td>
<td>100°C</td>
<td>Metal</td>
</tr>
<tr>
<td>Ag</td>
<td>2.31</td>
<td>2.37</td>
<td>Pb</td>
</tr>
<tr>
<td>Au</td>
<td>2.35</td>
<td>2.40</td>
<td>Pt</td>
</tr>
<tr>
<td>Cd</td>
<td>2.42</td>
<td>2.43</td>
<td>Sn</td>
</tr>
<tr>
<td>Cu</td>
<td>2.23</td>
<td>2.33</td>
<td>W</td>
</tr>
<tr>
<td>Mo</td>
<td>2.61</td>
<td>2.79</td>
<td>Zn</td>
</tr>
</tbody>
</table>

history of the theory of metals, for it supported the picture of an electron gas. It can be explained by using (43) for $\sigma$ and (56) for $K$:

$$
\frac{K}{\sigma} = \frac{\pi^2 k_B^2 T n \tau / 3 m}{e^2 \tau / m} = \frac{\pi^2}{3} \left(\frac{k_B}{e}\right)^2 T.
$$

(57)

The Lorenz number $L$ is defined as

$$
L = K/\sigma T,
$$

(58)

and according to (57) should have the value

$$
L = \frac{\pi^2}{3} \left(\frac{k_B}{e}\right)^2 = 2.72 \times 10^{-13} \text{ (erg/esu – deg)}^2
$$

$$
= 2.45 \times 10^{-8} \text{ watt-ohm/deg}^2.
$$

(59)

This remarkable result involves neither $n$ nor $m$. It does not involve $\tau$ if the relaxation times are identical for electrical and thermal processes. Experimental values of $L$ at 0°C and at 100°C as given in Table 5 are in good agreement with (59). At low temperatures ($T \ll \theta$) the Lorenz number tends to decrease; see the book by Ziman.

NANOSTRUCTURES

The term nanostructure denotes a condensed matter structure having a minimum dimension approximately between 1 nm (10 Å) and 10 nm (100 Å). These structures may be fine particles, fine wires, or thin films. Fine particles typically contain between 10 and 1,000 atoms. Semiconductor technology (see section “superlattices” in Chapter 8) has made it possible to fabricate small pools of electrons called in various ways: single-electron transistors, quantum dots, artificial atoms, Coulomb islands, or quantum corrals (Chapter 19). The unusual physical properties of nanostructures compared with bulk solids are attributed to several factors involving properties treated above and in later chapters:
• The ratio of the number of atoms on the surface to the number of atoms in the interior may be of the order of unity.
• The ratio of surface energy to total energy may be of the order of unity.
• The conduction or valence electrons are confined to a small length or volume, so that the quantum wavelength of the lowest electronic state is constricted and consequently the minimum wavelength is shorter than in the bulk solid.
• A wavelength or boundary condition shift will change the optical absorption spectrum (Chapter 11).
• Assemblies of nanoclusters of metals may have great hardness and yield strength, because it is difficult to create and to move dislocations (Chapter 20) in spatially confined regions.
• Magnetic monolayers, as of alternating films of ferromagnetic iron and of paramagnetic chromium, may have the magnetization (Chapter 15) of the iron films coupled by tunneling of the magnetization through the chromium barriers.

Problems

1. Kinetic energy of electron gas. Show that the kinetic energy of a three-dimensional gas of \( N \) free electrons at 0 K is

\[
U_0 = \frac{3}{2} N e_F .
\]  

(60)

2. Pressure and bulk modulus of an electron gas. (a) Derive a relation connecting the pressure and volume of an electron gas at 0 K. Hint: Use the result of Problem 1 and the relation between \( e_F \) and electron concentration. The result may be written as

\[
p = \frac{3}{2} (U_0 V).
\]

(b) Show that the bulk modulus \( B = -V(\partial p/\partial V) \) of an electron gas at 0 K is \( B = 5p/3 = 10U_0/9V \). (c) Estimate for potassium, using Table 1, the value of the electron gas contribution to \( B \).

3. Chemical potential in two dimensions. Show that the chemical potential of a Fermi gas in two dimensions is given by:

\[
\mu(T) = k_B T \ln \left[ \exp(\pi n \hbar^2 / mk_B T) - 1 \right] ,
\]  

(61)

for \( n \) electrons per unit area. Note: The density of orbitals of a free electron gas in two dimensions is independent of energy: \( D(\epsilon) = nL \pi \hbar^2 \), per unit area of specimen.

4. Fermi gases in astrophysics. (a) Given \( M_\odot = 2 \times 10^{33} \text{ g} \) for the mass of the Sun, estimate the number of electrons in the Sun. In a white dwarf star this number of electrons may be ionized and contained in a sphere of radius \( 2 \times 10^9 \text{ cm} \); find the Fermi energy of the electrons in electron volts. (b) The energy of an electron in the relativistic limit \( \epsilon \gg mc^2 \) is related to the wavevector as \( \epsilon = pc = \hbar c k \). Show that the Fermi energy in this limit is \( \epsilon_F = \hbar c(N/V)^{1/3} \), roughly. (c) If the above number of electrons were contained within a pulsar of radius 10 km, show that the Fermi energy would be \( \approx 10^6 \text{ eV} \). This value explains why pulsars are believed to be com-
posed largely of neutrons rather than of protons and electrons, for the energy release in the reaction \( n \rightarrow p + e^- \) is only \( 0.8 \times 10^6 \) eV, which is not large enough to enable many electrons to form a Fermi sea. The neutron decay proceeds only until the electron concentration builds up enough to create a Fermi level of \( 0.8 \times 10^6 \) eV, at which point the neutron, proton, and electron concentrations are in equilibrium.

5. **Liquid He\(^3\).** The atom He\(^3\) has spin \( \frac{1}{2} \) and is a fermion. The density of liquid He\(^3\) is 0.081 g cm\(^{-3}\) near absolute zero. Calculate the Fermi energy \( \varepsilon_F \) and the Fermi temperature \( T_F \).

6. **Frequency dependence of the electrical conductivity.** Use the equation \( m(\partial v/\partial t + v/\tau) = -eE \) for the electron drift velocity \( v \) to show that the conductivity at frequency \( \omega \) is

\[
\sigma(\omega) = \sigma(0) \left( \frac{1 + i\omega\tau}{1 + (\omega\tau)^2} \right),
\]

where \( \sigma(0) = ne^2\tau/m \).

7. **Dynamic magnetoconductivity tensor for free electrons.** A metal with a concentration \( n \) of free electrons of charge \(-e\) is in a static magnetic field \( B\hat{z} \). The electric current density in the \( xy \) plane is related to the electric field by

\[
\begin{align*}
J_x &= \sigma_{xx}E_x + \sigma_{xy}E_y; \\
J_y &= \sigma_{yx}E_x + \sigma_{yy}E_y.
\end{align*}
\]

Assume that the frequency \( \omega \gg \omega_c \) and \( \omega \gg 1/\tau \), where \( \omega_c = eB/mc \) and \( \tau \) is the collision time. (a) Solve the drift velocity equation (51) to find the components of the magnetoconductivity tensor:

\[
\begin{align*}
\sigma_{xx} &= \sigma_{yy} = i\omega_p^2/4\pi\omega; \\
\sigma_{xy} &= -\sigma_{yx} = \omega_c\omega_p^2/4\pi\omega^2,
\end{align*}
\]

where \( \omega_p^2 = 4\pi ne^2/m \). (b) Note from a Maxwell equation that the dielectric function tensor of the medium is related to the conductivity tensor as \( \varepsilon = 1 + i(4\pi\omega)\sigma \). Consider an electromagnetic wave with wavevector \( \mathbf{k} = k\hat{z} \). Show that the dispersion relation for this wave in the medium is

\[
\varepsilon^2k^2 = \omega^2 - \omega_p^2 \pm \omega_c\omega_p^2/\omega.
\]

At a given frequency there are two modes of propagation with different wavevectors and different velocities. The two modes correspond to circularly polarized waves. Because a linearly polarized wave can be decomposed into two circularly polarized waves, it follows that the plane of polarization of a linearly polarized wave will be rotated by the magnetic field.

8. **Cohesive energy of free electron Fermi gas.** We define the dimensionless length \( r_s \) as \( r_0/a_H \), where \( r_0 \) is the radius of a sphere that contains one electron, and \( a_H \) is the Bohr radius \( h^2/e^2m \). (a) Show that the average kinetic energy per electron in a free electron Fermi gas at 0 K is \( 2.21/r_s^2 \), where the energy is expressed in rydbergs.

*This problem is somewhat difficult.*
with $1 \text{ Ry} = me^4/2\hbar^2$. (b) Show that the coulomb energy of a point positive charge $e$ interacting with the uniform electron distribution of one electron in the volume of radius $r_0$ is $-3e^2/2r_0$, or $-3/r_0$ in rydbergs. (c) Show that the coulomb self-energy of the electron distribution in the sphere is $3e^2/5r_0$, or $6/5r_0$ in rydbergs. (d) The sum of (b) and (c) gives $-1.80/r_0$ for the total coulomb energy per electron. Show that the equilibrium value of $r_0$ is 2.45. Will such a metal be stable with respect to separated H atoms?

9. **Static magnetoconductivity tensor.** For the drift velocity theory of (51), show that the static current density can be written in matrix form as

$$\begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} = \frac{\sigma_0}{1 + (\omega_c\tau)^2} \begin{pmatrix} 1 & -\omega_c\tau & 0 \\ \omega_c\tau & 1 & 0 \\ 0 & 0 & 1 + (\omega_c\tau)^2 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}.$$  \hspace{1cm} (64)

In the high magnetic field limit of $\omega_c\tau \gg 1$, show that

$$\sigma_{yx} = \frac{nec}{B} = -\sigma_{xy}.$$  \hspace{1cm} (65)

In this limit $\sigma_{xx} = 0$, to order $1/\omega_c\tau$. The quantity $\sigma_{yx}$ is called the Hall conductivity.

10. **Maximum surface resistance.** Consider a square sheet of side $L$, thickness $d$, and electrical resistivity $\rho$. The resistance measured between opposite edges of the sheet is called the surface resistance: $R_{sq} = \rho L/Ld = \rho/d$, which is independent of the area $L^2$ of the sheet. ($R_{sq}$ is called the resistance per square and is expressed in ohms per square, because $\rho/d$ has the dimensions of ohms.) If we express $\rho$ by (44), then $R_{sq} = m/nde^2\tau$. Suppose now that the minimum value of the collision time is determined by scattering from surfaces of the sheet, so that $\tau = d/v_F$, where $v_F$ is the Fermi velocity. Thus the maximum surface resistivity is $R_{sq} = mve^2/nde^2$. Show for a monatomic metal sheet one atom in thickness that $R_{sq} = h/e^2 = 4.1 \text{ k}\Omega$, where 1 k\Omega is $10^3$ ohms.

*11. **Small metal spheres.** Consider free electrons in a spherical square well potential of radius $a$, with an infinitely high boundary. (a) Show that the wave function of an orbital of angular momentum $\ell$ and projection $m$ has the form

$$\psi = R_{k\ell}(r)Y_{\ell m}(\theta,\phi),$$  \hspace{1cm} (66)

where the radial wave function has the form

$$R_{k\ell}(r) = (\pi/2kr)^{1/2} J_{\ell + 1/2}(kr),$$

and $Y$ is a spherical harmonic. Here $J$ is a Bessel function of half-integral order and satisfies the boundary condition $J_{\ell + 1/2}(ka) = 0$. The roots give the energy eigenvalues $\epsilon$ of the levels above the bottom of the well, where $\epsilon = h^2k^2/2m$. (b) Show that the order of the levels above the ground orbital is

$$1s, 1p, 1d, 2s, 1f, 2p, 1g, 2d, 1h, 3s, 2f, \ldots,$$

where $s, p, d, f, g, h$ denote $\ell = 0, 1, 2, 3, 4, 5$. 

12. **Density of states—nanometric wire.** (a) Consider a nanometric wire in the form of a rectangular parallelepiped, with two sides \( L_x = L_y = 1 \text{ nm} \) and the long axis \( L_z = 1 \text{ cm} \). The single particle eigenstates of the system may be written as

\[
\psi = \sin \left( \frac{n_x \pi x}{L_x} \right) \sin \left( \frac{n_y \pi y}{L_y} \right) \exp[i2\pi Nz] .
\]

The energy of the eigenstate is, with \( n = n_x, n_y \):

\[
E(n, N) - \epsilon(n) = \frac{(2\pi \hbar N)^2}{2m} = \Delta N^2 = \frac{\hbar^2 v^2}{2m} ,
\]

where \( v \) is the electron velocity along the \( z \)-axis. Here \( \Delta = \frac{(2\pi \hbar)^2}{2m} \) and \( N = \sqrt{\frac{(E - \epsilon)}{\epsilon/\hbar}} \). Then \( \Delta E = 2\Delta N \Delta N \), show that the density of states \( D_n \) at fixed \( n \), with account of the two spin orientations and the two \( \pm \) values of \( N \), is

\[
D_n(E) = 4\Delta N/\Delta E = 2/AN = 2/\sqrt{[(E - \epsilon)/\hbar^2]}.\]

(b) Sum over the values of \( n \) for which \( E \geq \epsilon_n \) to show that

\[
D(E) = \Sigma D_n(E) = \Sigma (2/\pi \hbar) (1/\sqrt{[(E - \epsilon_n)/\hbar^2]}).\]

where \( \Theta(x) \) is the Heaviside unit function, zero for \( x < 0 \) and unity for \( x > 0 \).

13. **Quantization of conductance.** The current in the nanometric wire of Problem 12 is \( I = (N_+ - N_-)eV_F \), where \( N_+ - N_- = D(E_F)eV \), where \( V \) is the bias voltage. Show that the current may be written as

\[
I = D(E_F)eV eV_F = \left(2e^2/\pi \hbar^2\right)n_{\text{occ}} V ,
\]

where \( n_{\text{occ}} \) is the number of occupied states \( n_x, n_y \), hence the quantized conductance is \( \left(2e^2/\pi \hbar^2\right)n_{\text{occ}} \).

---

**References**


### Energy Bands

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEARLY FREE ELECTRON MODEL</td>
<td>176</td>
</tr>
<tr>
<td>Origin of the energy gap</td>
<td>177</td>
</tr>
<tr>
<td>Magnitude of the energy gap</td>
<td>179</td>
</tr>
<tr>
<td>BLOCH FUNCTIONS</td>
<td>179</td>
</tr>
<tr>
<td>KRONIG-PENNEY MODEL</td>
<td>180</td>
</tr>
<tr>
<td>WAVE EQUATION OF ELECTRON IN A PERIODIC POTENTIAL</td>
<td>183</td>
</tr>
<tr>
<td>Restatement of the Bloch theorem</td>
<td>185</td>
</tr>
<tr>
<td>Crystal momentum of an electron</td>
<td>186</td>
</tr>
<tr>
<td>Solution of the central equation</td>
<td>186</td>
</tr>
<tr>
<td>Kronig-Penney model in reciprocal space</td>
<td>187</td>
</tr>
<tr>
<td>Empty lattice approximation</td>
<td>188</td>
</tr>
<tr>
<td>Approximate solution near a zone boundary</td>
<td>189</td>
</tr>
<tr>
<td>NUMBER OF ORBITALS IN A BAND</td>
<td>192</td>
</tr>
<tr>
<td>Metals and insulators</td>
<td>193</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>195</td>
</tr>
<tr>
<td>PROBLEMS</td>
<td>195</td>
</tr>
<tr>
<td>1. Square lattice, free electron energies</td>
<td>195</td>
</tr>
<tr>
<td>2. Free electron energies in reduced zone</td>
<td>195</td>
</tr>
<tr>
<td>3. Kronig-Penney model</td>
<td>195</td>
</tr>
<tr>
<td>4. Potential energy in the diamond structure</td>
<td>195</td>
</tr>
<tr>
<td>5. Complex wavevectors in the energy gap</td>
<td>196</td>
</tr>
<tr>
<td>6. Square lattice</td>
<td>196</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>196</td>
</tr>
</tbody>
</table>
Figure 1  Schematic electron occupancy of allowed energy bands for an insulator, metal, semimetal, and semiconductor. The vertical extent of the boxes indicates the allowed energy regions; the shaded areas indicate the regions filled with electrons. In a semimetal (such as bismuth) one band is almost filled and another band is nearly empty at absolute zero, but a pure semiconductor (such as silicon) becomes an insulator at absolute zero. The left of the two semiconductors shown is at a finite temperature, with carriers excited thermally. The other semiconductor is electron-deficient because of impurities.
CHAPTER 7: ENERGY BANDS

When I started to think about it, I felt that the main problem was to explain how the electrons could sneak by all the ions in a metal. . . . By straight Fourier analysis I found to my delight that the wave differed from the plane wave of free electrons only by a periodic modulation.

F. Bloch

The free electron model of metals gives us good insight into the heat capacity, thermal conductivity, electrical conductivity, magnetic susceptibility, and electrodynamics of metals. But the model fails to help us with other large questions: the distinction between metals, semimetals, semiconductors, and insulators; the occurrence of positive values of the Hall coefficient; the relation of conduction electrons in the metal to the valence electrons of free atoms; and many transport properties, particularly magnetotransport. We need a less naïve theory, and fortunately it turns out that almost any simple attempt to improve upon the free electron model is enormously profitable.

The difference between a good conductor and a good insulator is striking. The electrical resistivity of a pure metal may be as low as $10^{-10}$ ohm-cm at a temperature of 1 K, apart from the possibility of superconductivity. The resistivity of a good insulator may be as high as $10^{22}$ ohm-cm. This range of $10^{32}$ may be the widest of any common physical property of solids.

Every solid contains electrons. The important question for electrical conductivity is how the electrons respond to an applied electric field. We shall see that electrons in crystals are arranged in energy bands (Fig. 1) separated by regions in energy for which no wavelike electron orbitals exist. Such forbidden regions are called energy gaps or band gaps, and result from the interaction of the conduction electron waves with the ion cores of the crystal.

The crystal behaves as an insulator if the allowed energy bands are either filled or empty, for then no electrons can move in an electric field. The crystal behaves as a metal if one or more bands are partly filled, say between 10 and 90 percent filled. The crystal is a semiconductor or a semimetal if one or two bands are slightly filled or slightly empty.

To understand the difference between insulators and conductors, we must extend the free electron model to take account of the periodic lattice of the solid. The possibility of a band gap is the most important new property that emerges.

We shall encounter other quite remarkable properties of electrons in crystals. For example, they respond to applied electric or magnetic fields as if the electrons were endowed with an effective mass $m^*$, which may be larger or smaller than the free electron mass, or may even be negative. Electrons in crystals respond to applied fields as if endowed with negative or positive
charges, $-e$ or $+e$, and herein lies the explanation of the negative and positive values of the Hall coefficient.

**NEARLY FREE ELECTRON MODEL**

On the free electron model the allowed energy values are distributed essentially continuously from zero to infinity. We saw in Chapter 6 that

$$
\epsilon_k = \frac{k^2}{2m} (k_x^2 + k_y^2 + k_z^2) ,
$$

where, for periodic boundary conditions over a cube of side $L$,

$$
k_x, k_y, k_z = 0 ; \pm \frac{2\pi}{L} ; \pm \frac{4\pi}{L} ; \ldots .
$$

The free electron wavefunctions are of the form

$$
\psi_n(r) = \exp(ik \cdot r) ;
$$

they represent running waves and carry momentum $p = \hbar k$.

The band structure of a crystal can often be explained by the nearly free electron model for which the band electrons are treated as perturbed only weakly by the periodic potential of the ion cores. This model answers almost all the qualitative questions about the behavior of electrons in metals.

We know that Bragg reflection is a characteristic feature of wave propagation in crystals. Bragg reflection of electron waves in crystals is the cause of energy gaps. (At Bragg reflection wavelike solutions of the Schrödinger equation do not exist, as in Fig. 2.) These energy gaps are of decisive significance in determining whether a solid is an insulator or a conductor.

We explain physically the origin of energy gaps in the simple problem of a linear solid of lattice constant $a$. The low energy portions of the band structure are shown qualitatively in Fig. 2, in (a) for entirely free electrons and in (b) for electrons that are nearly free, but with an energy gap at $k = \pm \pi/a$. The Bragg condition $(k + G)^2 = k^2$ for diffraction of a wave of wavevector $k$ becomes in one dimension

$$
k = \pm \frac{1}{2}G = \pm n\pi/a ,
$$

where $G = 2\pi n/a$ is a reciprocal lattice vector and $n$ is an integer. The first reflections and the first energy gap occur at $k = \pm \pi/a$. The region in $k$ space between $-\pi/a$ and $\pi/a$ is the first Brillouin zone of this lattice. Other energy gaps occur for other values of the integer $n$.

The wavefunctions at $k = \pm \pi/a$ are not the traveling waves $\exp(i\pi x/a)$ or $\exp(-i\pi x/a)$ of free electrons. At these special values of $k$ the wavefunctions are
made up of equal parts of waves traveling to the right and to the left. When the Bragg reflection condition \( k = \pm \pi/a \) is satisfied by the wavevector, a wave traveling to the right is Bragg-reflected to travel to the left, and vice versa. Each subsequent Bragg reflection will reverse the direction of travel of the wave. A wave that travels neither to the right nor to the left is a standing wave: it doesn’t go anywhere.

The time-independent state is represented by standing waves. We can form two different standing waves from the two traveling waves \( \exp(\pm i\pi x/a) \), namely

\[
\psi(+) = \exp(i\pi x/a) + \exp(-i\pi x/a) = 2 \cos(\pi x/a) ;
\]

\[
\psi(-) = \exp(i\pi x/a) - \exp(-i\pi x/a) = 2i \sin(\pi x/a) .
\]

The standing waves are labeled \((+)\) or \((-)\) according to whether or not they change sign when \(-x\) is substituted for \(x\). Both standing waves are composed of equal parts of right- and left-directed traveling waves.

**Origin of the Energy Gap**

The two standing waves \( \psi(+) \) and \( \psi(-) \) pile up electrons at different regions, and therefore the two waves have different values of the potential energy. This is the origin of the energy gap. The probability density \( \rho \) of a particle is \( \psi^*\psi = |\psi|^2 \). For a pure traveling wave \( \exp(ikx) \), we have \( \rho = \exp(-ikx) \exp(ikx) = 1 \), so that the charge density is constant. The charge density is not constant for linear combinations of plane waves. Consider the standing wave \( \psi(+) \) in (5); for this we have

\[
\rho(+) = |\psi(+)|^2 \propto \cos^2 \frac{\pi x}{a} .
\]

This function piles up electrons (negative charge) on the positive ions centered at \( x = 0, a, 2a, \ldots \) in Fig. 3, where the potential energy is lowest.
Figure 3a pictures the variation of the electrostatic potential energy of a conduction electron in the field of the positive ion cores. The ion cores bear a net positive charge because the atoms are ionized in the metal, with the valence electrons taken off to form the conduction band. The potential energy of an electron in the field of a positive ion is negative, so that the force between them is attractive.

For the other standing wave $\psi(-)$ the probability density is

$$\rho(-) = |\psi(-)|^2 \propto \sin^2 \frac{\pi x}{a},$$

which concentrates electrons away from the ion cores. In Fig. 3b we show the electron concentration for the standing waves $\psi(+), \psi(-),$ and for a traveling wave.

When we calculate the average or expectation values of the potential energy over these three charge distributions, we find that the potential energy of

---

Figure 3  (a) Variation of potential energy of a conduction electron in the field of the ion cores of a linear lattice. (b) Distribution of probability density $\rho$ in the lattice for $|\psi(-)|^2 \propto \sin^2 \frac{\pi x}{a}; |\psi(+)|^2 \propto \cos^2 \frac{\pi x}{a},$ and for a traveling wave. The wavefunction $\psi(+)\) piles up electronic charge on the cores of the positive ions, thereby lowering the potential energy in comparison with the average potential energy seen by a traveling wave. The wavefunction $\psi(-)$ piles up charge in the region between the ions, thereby raising the potential energy in comparison with that seen by a traveling wave. This figure is the key to understanding the origin of the energy gap.
\( \rho(+) \) is lower than that of the traveling wave, whereas the potential energy of \( \rho(-) \) is higher than the traveling wave. We have an energy gap of width \( E_g \) if the energies of \( \rho(-) \) and \( \rho(+) \) differ by \( E_g \). Just below the energy gap at points A in Fig. 2 the wavefunction is \( \psi(+) \), and just above the gap at points B the wavefunction is \( \psi(-) \).

**Magnitude of the Energy Gap**

The wavefunctions at the Brillouin zone boundary \( k = \pi/a \) are \( \sqrt{2} \cos \pi x/a \) and \( \sqrt{2} \sin \pi x/a \), normalized over unit length of line. We write the potential energy of an electron in the crystal at point \( x \) as

\[
U(x) = U \cos 2\pi x/a .
\]

The first-order energy difference between the two standing wave states is

\[
E_g = \int_0^1 dx \ U(x) \left| |\psi(+)|^2 - |\psi(-)|^2 \right| = 2 \int_0^1 dx \ U \cos(2\pi x/a) \left( \cos^2 \pi x/a - \sin^2 \pi x/a \right) = U .
\]

We see that the gap is equal to the Fourier component of the crystal potential.

**BLOCH FUNCTIONS**

F. Bloch proved the important theorem that the solutions of the Schrödinger equation for a periodic potential must be of a special form:

\[
\psi_k(\mathbf{r}) = u_k(\mathbf{r}) \exp(ik \cdot \mathbf{r}) ,
\]

where \( u_k(\mathbf{r}) \) has the period of the crystal lattice with \( u_k(\mathbf{r}) = u_k(\mathbf{r} + \mathbf{T}) \). The result (7) expresses the Bloch theorem:

The eigenfunctions of the wave equation for a periodic potential are the product of a plane wave \( \exp(ik \cdot \mathbf{r}) \) times a function \( u_k(\mathbf{r}) \) with the periodicity of the crystal lattice.

A one-electron wavefunction of the form (7) is called a Bloch function and can be decomposed into a sum of traveling waves, as we see later. Bloch functions can be assembled into localized wave packets to represent electrons that propagate freely through the potential field of the ion cores.

We give now a restricted proof of the Bloch theorem, valid when \( \psi_k \) is nondegenerate. That is, when there is no other wavefunction with the same energy and wave vector as \( \psi_k \). The general case will be treated later. We con-
sider $N$ identical lattice points on a ring of length $Na$. The potential energy is periodic in $a$, with $U(x) = U(x + sa)$, where $s$ is an integer.

We are guided by the symmetry of the ring to look for solutions of the wave equation such that

$$
\psi(x + a) = C\psi(x) ,
$$
\[ (8) \]

where $C$ is a constant. Then, on going once around the ring,

$$
\psi(x + Na) = \psi(x) = C^N \psi(x) ,
$$

because $\psi(x)$ must be single-valued. It follows that $C$ is one of the $N$ roots of unity, or

$$
C = \exp(i2\pi s/N) ; \quad s = 0, 1, 2, \ldots, N - 1 .
$$
\[ (9) \]

We see that

$$
\psi(x) = u_k(x) \exp(i2\pi sx/Na)
$$
\[ (10) \]
satisfies (8), provided that $u_k(x)$ has the periodicity $a$, so that $u_k(x) = u_k(x + a)$. With $k = 2\pi s/Na$, we have the Bloch result (7). For another derivation, see (29).

**KRONIG-PENNEY MODEL**

A periodic potential for which the wave equation can be solved in terms of elementary functions is the square-well array of Fig. 4. The wave equation is

$$
-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + U(x)\psi = \epsilon\psi ,
$$
\[ (11) \]

where $U(x)$ is the potential energy and $\epsilon$ is the energy eigenvalue.

In the region $0 < x < a$ in which $U = 0$, the eigenfunction is a linear combination,

$$
\psi = Ae^{ikx} + Be^{-ikx} ,
$$
\[ (12) \]
of plane waves traveling to the right and to the left, with energy

$$
\epsilon = \hbar^2 k^2/2m .
$$
\[ (13) \]

In the region $-b < x < 0$ within the barrier the solution is of the form

$$
\psi = Ce^{Qx} + De^{-Qx} ,
$$
\[ (14) \]

with

$$
U_0 - \epsilon = \hbar^2 Q^2/2m .
$$
\[ (15) \]
We want the complete solution to have the Bloch form (7). Thus the solution in the region \( a < x < a + b \) must be related to the solution (14) in the region \(-b < x < 0\) by the Bloch theorem:

\[
\psi(a < x < a + b) = \psi(-b < x < 0) e^{ik(a+b)} ,
\]

which serves to define the wavevector \( k \) used as an index to label the solution.

The constants \( A, B, C, D \) are chosen so that \( \psi \) and \( d\psi/dx \) are continuous at \( x = 0 \) and \( x = a \). These are the usual quantum mechanical boundary conditions in problems that involve square potential wells. At \( x = 0 \),

\[
A + B = C + D ;
\]

\[
iK(A - B) = Q(C - D) .
\]

At \( x = a \), with the use of (16) for \( \psi(a) \) under the barrier in terms of \( \psi(-b) \),

\[
Ae^{ikA} + Be^{-ikA} = (Ce^{-Qb} + De^{Qb}) e^{ik(a+b)} ;
\]

\[
iK(Ae^{ikA} - Be^{-ikA}) = Q(Ce^{-Qb} - De^{Qb}) e^{ik(a+b)} .
\]

The four equations (17) to (20) have a solution only if the determinant of the coefficients of \( A, B, C, D \) vanishes, or if

\[
[(Q^2 - K^2)/2QK] \sinh Qb \sin Ka + \cosh Qb \cos Ka = \cos k(a + b) .\]

It is rather tedious to obtain this equation.

The result is simplified if we represent the potential by the periodic delta function obtained when we pass to the limit \( b = 0 \) and \( U_0 = \infty \) in such a way that \( Q^2ba/2 = P \), a finite quantity. In this limit \( Q \gg K \) and \( Qb \ll 1 \). Then (21a) reduces to

\[
(P/Ka)\sin Ka + \cos Ka = \cos ka .
\]

The ranges of \( K \) for which this equation has solutions are plotted in Fig. 5, for the case \( P = 3\pi/2 \). The corresponding values of the energy are plotted in Fig. 6. Note the energy gaps at the zone boundaries. The wavevector \( k \) of the Bloch function is the important index, not the \( K \) in (12), which is related to the energy by (13). A treatment of this problem in wavevector space is given later in this chapter.
Figure 5  Plot of the function \((P/Ka) \sin Ka + \cos Ka\), for \(P = 3\pi/2\). The allowed values of the energy \(\epsilon\) are given by those ranges of \(Ka = (2\pi \hbar^2) 1/2 a\) for which the function lies between \(\pm 1\). For other values of the energy there are no traveling wave or Bloch-like solutions to the wave equation, so that forbidden gaps in the energy spectrum are formed.

Figure 6  Plot of energy vs. wavenumber for the Kronig-Penney potential, with \(P = 3\pi/2\). Notice the energy gaps at \(ka = \pi, 2\pi, 3\pi\ldots\).
WAVE EQUATION OF ELECTRON IN A PERIODIC POTENTIAL

We considered in Fig. 3 the approximate form we expect for the solution of the Schrödinger equation if the wavevector is at a zone boundary, as at $k = \pi/a$. We treat in detail the wave equation for a general potential, at general values of $k$. Let $U(x)$ denote the potential energy of an electron in a linear lattice of lattice constant $a$. We know that the potential energy is invariant under a crystal lattice translation: $U(x) = U(x + a)$. A function invariant under a crystal lattice translation may be expanded as a Fourier series in the reciprocal lattice vectors $G$. We write the Fourier series for the potential energy as

$$U(x) = \sum_{G} U_G e^{iGx} . \quad (22)$$

The values of the coefficients $U_G$ for actual crystal potentials tend to decrease rapidly with increasing magnitude of $G$. For a bare coulomb potential $U_G$ decreases as $1/G^2$.

We want the potential energy $U(x)$ to be a real function:

$$U(x) = \sum_{G>0} U_G(e^{iGx} + e^{-iGx}) = 2 \sum_{G>0} U_G \cos Gx . \quad (23)$$

For convenience we have assumed that the crystal is symmetric about $x = 0$ and that $U_0 = 0$.

The wave equation of an electron in the crystal is $\mathcal{H}\psi = \epsilon\psi$, where $\mathcal{H}$ is the hamiltonian and $\epsilon$ is the energy eigenvalue. The solutions $\psi$ are called eigenfunctions or orbitals or Bloch functions. Explicitly, the wave equation is

$$\left( \frac{1}{2m}p^2 + U(x) \right) \psi(x) = \left( \frac{1}{2m}p^2 + \sum_G U_G e^{iGx} \right) \psi(x) = \epsilon\psi(x) . \quad (24)$$

Equation (24) is written in the one-electron approximation in which the orbital $\psi(x)$ describes the motion of one electron in the potential of the ion cores and in the average potential of the other conduction electrons.

The wavefunction $\psi(x)$ may be expressed as a Fourier series summed over all values of the wavevector permitted by the boundary conditions, so that

$$\psi = \sum_k C(k) e^{ikx} , \quad (25)$$

where $k$ is real. (We could equally well write the index $k$ as a subscript on $C$, as in $C_k$.)

The set of values of $k$ has the form $2\pi n/L$, because these values satisfy periodic boundary conditions over length $L$. Here $n$ is any integer, positive or negative. We do not assume, nor is it generally true, that $\psi(x)$ itself is periodic
in the fundamental lattice translation \( a \). The translational properties of \( \psi(x) \) are determined by the Bloch theorem (7).

Not all wavevectors of the set \( 2\pi n/L \) enter the Fourier expansion of any one Bloch function. If one particular wavevector \( k \) is contained in a \( \psi \), then all other wavevectors in the Fourier expansion of this \( \psi \) will have the form \( k + G \), where \( G \) is any reciprocal lattice vector. We prove this result in (29) below.

We can label a wavefunction \( \psi \) that contains a component \( k \) as \( \psi_k \) or, equally well, as \( \psi_{k+G} \), because if \( k \) enters the Fourier expansion then \( k + G \) may enter. The wavevectors \( k + G \) running over \( G \) are a restricted subset of the set \( 2\pi n/L \), as shown in Fig. 7.

We shall usually choose as a label for the Bloch function that \( k \) which lies within the first Brillouin zone. When other conventions are used, we shall say so. This situation differs from the phonon problem, where there are no components of the ion motion outside the first zone. The electron problem is like the x-ray diffraction problem because the electromagnetic field exists everywhere within the crystal and not only at the ions.

To solve the wave equation, substitute (25) in (24) to obtain a set of linear algebraic equations for the Fourier coefficients. The kinetic energy term is

\[
\frac{1}{2m} p^2 \psi(x) = \frac{1}{2m} \left( -i \hbar \frac{d}{dx} \right)^2 \psi(x) = -\frac{\hbar^2}{2m} \frac{d^2 \psi}{dx^2} = -\frac{\hbar^2}{2m} \sum_k k^2 C(k) \ e^{ikx} ;
\]

and the potential energy term is

\[
\left( \sum_G U_G \ e^{igx} \right) \psi(x) = \sum_G \sum_k U_G \ e^{igx} C(k) \ e^{ikx} .
\]

The wave equation is obtained as the sum:

\[
\sum_k \frac{\hbar^2}{2m} k^2 C(k) \ e^{ikx} + \sum_G \sum_k U_G C(k) \ e^{i(k+G)x} = \epsilon \sum_k C(k) \ e^{ikx} . \tag{26}
\]

Each Fourier component must have the same coefficient on both sides of the equation. Thus

\[
(\lambda_k - \epsilon) C(k) + \sum_G U_G C(k-G) = 0 . \tag{27}
\]

with the notation

\[
\lambda_k = \hbar^2 k^2/2m . \tag{28}
\]

Equation (27) is a useful form of the wave equation in a periodic lattice, although unfamiliar because a set of algebraic equations has taken the place of
the usual differential equation. The set appears unpleasant and formidable because there are, in principle, an infinite number of $C(k - G)$ to be determined. In practice a small number will often suffice, perhaps two or four. It takes some experience to appreciate the practical advantages of the algebraic approach.

Restatement of the Bloch Theorem

Once we determine the $C$’s from (27), the wavefunction (25) is given as

$$
\psi_k(x) = \sum_C C(k - G) e^{i(k-G)x} ,
$$

which may be rearranged as

$$
\psi_k(x) = \left( \sum_C C(k - G) e^{-iGx} \right) e^{ikx} = e^{ikx} u_k(x) ,
$$

with the definition

$$
u_k(x) = \sum_C C(k - G) e^{-iGx} .
$$

Because $u_k(x)$ is a Fourier series over the reciprocal lattice vectors, it is invariant under a crystal lattice translation $T$, so that $u_k(x) = u_k(x + T)$. We verify this directly by evaluating $u_k(x + T)$:

$$
u_k(x + T) = \sum C(k - G) e^{-iG(x+T)} = e^{-iT} \left[ \sum C(k - G) e^{-iGx} \right] = e^{-iT} u_k(x) .
$$

Because $\exp(-iT) = 1$ by (2.17), it follows that $u_k(x + T) = u_k(x)$, thereby establishing the periodicity of $u_k$. This is an alternate and exact proof of the Bloch theorem and is valid even when the $\psi_k$ are degenerate.
Crystal Momentum of an Electron

What is the significance of the wavevector $k$ used to label the Bloch function? It has several properties:

- Under a crystal lattice translation which carries $r$ to $r + T$ we have

$$\psi_k(r + T) = e^{ik \cdot T} e^{ik \cdot r} u_k(r + T) = e^{ik \cdot T} \psi_k(r),$$

(30)

because $u_k(r + T) = u_k(r)$. Thus $\exp(ik \cdot T)$ is the phase factor\(^1\) by which a Bloch function is multiplied when we make a crystal lattice translation $T$.

- If the lattice potential vanishes, the central equation (27) reduces to $(\lambda_k - \epsilon)C(k) = 0$, so that all $C(k - G)$ are zero except $C(k)$, and thus $u_k(r)$ is constant. We have $\psi_k(r) = e^{ik \cdot r}$, just as for a free electron. (This assumes we have had the foresight to pick the “right” $k$ as the label. For many purposes other choices of $k$, differing by a reciprocal lattice vector, will be more convenient.)

- The quantity $k$ enters in the conservation laws that govern collision processes in crystals. (The conservation laws are really selection rules for transitions.) Thus $\hbar k$ is called the crystal momentum of an electron. If an electron $k$ absorbs in a collision a phonon of wavevector $q$, the selection rule is $k + q = k' + G$. In this process the electron is scattered from a state $k$ to a state $k'$, with $G$ a reciprocal lattice vector. Any arbitrariness in labeling the Bloch functions can be absorbed in the $G$ without changing the physics of the process.

Solution of the Central Equation

Equation (27) may be called the central equation:

$$(\lambda_k - \epsilon)C(k) + \sum_G U_{G'} C(k - G) = 0$$

(31)

represents a set of simultaneous linear equations that connect the coefficients $C(k - G)$ for all reciprocal lattice vectors $G$. It is a set because there are as many equations as there are coefficients $C$. These equations are consistent if the determinant of the coefficients vanishes.

Let us write out the equations for an explicit problem. We let $g$ denote the shortest $G$. We suppose that the potential energy $U(x)$ contains only a single

---

\(^1\)We may also say that $\exp(ik \cdot T)$ is the eigenvalue of the crystal translation operation $T$, and $\psi_k$ is the eigenvector. That is, $T\psi_k(x) = \psi_k(x + T) = \exp(ik \cdot T)\psi_k(x)$, so that $k$ is a suitable label for the eigenvalue. Here we have used the Bloch theorem.
Fourier component $U_g = U_{-g}$, denoted by $U$. Then a block of the determinant of the coefficients is given by:

$$
\begin{array}{cccc}
\lambda_{k+2g} - \epsilon & U & 0 & 0 \\
U & \lambda_{k-g} - \epsilon & U & 0 \\
0 & U & \lambda_k - \epsilon & U \\
0 & 0 & U & \lambda_{k+g} - \epsilon \\
0 & 0 & 0 & U & \lambda_{k+2g} - \epsilon
\end{array}
$$

(32)

To see this, write out five successive equations of the set (31). The determinant in principle is infinite in extent, but it will often be sufficient to set equal to zero the portion we have shown.

At a given $k$, each root $\epsilon$ or $\epsilon_k$ lies on a different energy band, except in case of coincidence. The solution of the determinant (32) gives a set of energy eigenvalues $\epsilon_{nk}$, where $n$ is an index for ordering the energies and $k$ is the wavevector that labels $C_k$.

Most often $k$ will be taken in the first zone, to reduce possible confusion in the labeling. If we chose a $k$ different from the original by some reciprocal lattice vector, we would have obtained the same set of equations in a different order—but having the same energy spectrum.

**Kronig-Penney Model in Reciprocal Space**

As an example of the use of the central equation (31) for a problem that is exactly solvable, we use the Kronig-Penney model of a periodic delta-function potential:

$$
U(x) = 2 \sum_{G>0} U_G \cos Gx = A a \sum_s \delta(x - sa) ,
$$

(33)

where $A$ is a constant and $a$ the lattice spacing. The sum is over all integers $s$ between 0 and $1/a$. The boundary conditions are periodic over a ring of unit length, which means over $1/a$ atoms. Thus the Fourier coefficients of the potential are

$$
U_G = \int_0^1 dx \, U(x) \cos Gx = A a \sum_s \int_0^1 dx \, \delta(x - sa) \cos Gx
$$

(34)

$$
= A a \sum_s \cos Gsa = A .
$$

---

*2This treatment was suggested by Surjit Singh, Am. J. Phys. 51, 179 (1983).*
We write the central equation with \( k \) as the Bloch index. Thus (31) becomes
\[
(\lambda_k - \epsilon) C(k) + A \sum_n C(k - 2\pi n/a) = 0,
\]
(35)
where \( \lambda_k = \hbar^2 k^2 / 2m \) and the sum is over all integers \( n \). We want to solve (35) for \( \epsilon(k) \).

We define
\[
f(k) = \sum_n C(k - 2\pi n/a),
\]
(36)
so that (35) becomes
\[
C(k) = -\frac{(2mA/\hbar^2)f(k)}{k^2 - (2me/\hbar^2)}.
\]
(37)
Because the sum (36) is over all coefficients \( C \), we have, for any \( n \),
\[
f(k) = f(k - 2\pi n/a).
\]
(38)
This relation lets us write
\[
C(k - 2\pi n/a) = -(2mA/\hbar^2)f(k)[(k - 2\pi n/a)^2 - (2me/\hbar^2)]^{-1}.
\]
(39)
We sum both sides over all \( n \) to obtain, using (36) and cancelling \( f(k) \) from both sides,
\[
(\hbar^2/2mA) = -\sum_n [(k - 2\pi n/a)^2 - (2me/\hbar^2)]^{-1}.
\]
(40)
The sum can be carried out with the help of the standard relation
\[
\text{ctn } x = \sum_n \frac{1}{n\pi + x}.
\]
(41)
After trigonometric manipulations in which we use relations for the difference of two cotangents and the product of two sines, the sum in (40) becomes
\[
\frac{a^2 \sin Ka}{4Ka(\cos ka - \cos Ka)},
\]
(42)
where we write \( K^2 = 2me/\hbar^2 \) as in (13).

The final result for (40) is
\[
(mAa^2/2\hbar^2)(Ka)^{-1} \sin Ka + \cos Ka = \cos ka,
\]
(43)
which agrees with the Kronig-Penney result (21b) with \( P \) written for \( mAa^2/2\hbar^2 \).

**Empty Lattice Approximation**

Actual band structures are usually exhibited as plots of energy versus wavevector in the first Brillouin zone. When wavevectors happen to be given outside the first zone, they are carried back into the first zone by subtracting a
suitable reciprocal lattice vector. Such a translation can always be found. The operation is helpful in visualization and economical of graph paper.

When band energies are approximated fairly well by free electron energies
\[ \varepsilon_k = \hbar^2 k^2 / 2m, \]
it is advisable to start a calculation by carrying the free electron energies back into the first zone. The procedure is simple enough once you get the hang of it. We look for a \( G \) such that a \( k' \) in the first zone satisfies

\[ k' + G = k, \]

where \( k \) is unrestricted and is the true free electron wavevector in the empty lattice. (Once the plane wave is modulated by the lattice, there is no single "true" wavevector for the state \( \psi \).)

If we drop the prime on \( k' \) as unnecessary baggage, the free electron energy can always be written as

\[ \varepsilon(k_x, k_y, k_z) = (\hbar^2 / 2m)(k + G)^2 \]

\[ = (\hbar^2 / 2m) [(k_x + G_x)^2 + (k_y + G_y)^2 + (k_z + G_z)^2], \]

with \( k \) in the first zone and \( G \) allowed to run over the appropriate reciprocal lattice points.

We consider as an example the low-lying free electron bands of a simple cubic lattice. Suppose we want to exhibit the energy as a function of \( k \) in the [100] direction. For convenience, choose units such that \( \hbar^2 / 2m = 1 \). We show several low-lying bands in this empty lattice approximation with their energies \( \varepsilon(000) \) at \( k = 0 \) and \( \varepsilon(k,00) \) along the \( k_x \) axis in the first zone:

<table>
<thead>
<tr>
<th>Band</th>
<th>( G \alpha / 2\pi )</th>
<th>( \varepsilon(000) )</th>
<th>( \varepsilon(k,00) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>000</td>
<td>0</td>
<td>( k_x^2 )</td>
</tr>
<tr>
<td>2,3</td>
<td>100,100</td>
<td>( (2\pi/\alpha)^2 )</td>
<td>( (k_x \pm 2\pi/\alpha)^2 )</td>
</tr>
<tr>
<td>4,5,6,7</td>
<td>010,010,001,001,011</td>
<td>( (2\pi/\alpha)^2 )</td>
<td>( k_x^2 + (2\pi/\alpha)^2 )</td>
</tr>
<tr>
<td>8,9,10,11</td>
<td>110,101,110,101</td>
<td>( 2(2\pi/\alpha)^2 )</td>
<td>( (k_x + 2\pi/\alpha)^2 + (2\pi/\alpha)^2 )</td>
</tr>
<tr>
<td>12,13,14,15</td>
<td>110,101,110,101</td>
<td>( 2(2\pi/\alpha)^2 )</td>
<td>( (k_x - 2\pi/\alpha)^2 + (2\pi/\alpha)^2 )</td>
</tr>
<tr>
<td>16,17,18,19</td>
<td>011,011,011,011</td>
<td>( 2(2\pi/\alpha)^2 )</td>
<td>( k_x^2 + 2(2\pi/\alpha)^2 )</td>
</tr>
</tbody>
</table>

These free electron bands are plotted in Fig. 8. It is a good exercise to plot the same bands for \( k \) parallel to the [111] direction of wavevector space.

**Approximate Solution Near a Zone Boundary**

We suppose that the Fourier components \( U_G \) of the potential energy are small in comparison with the kinetic energy of a free electron at the zone boundary. We first consider a wavevector exactly at the zone boundary at \( \frac{1}{4} G \), that is, at \( \pi/\alpha \). Here

\[ k^2 = (\frac{1}{4} G)^2; \quad (k - G)^2 = (\frac{1}{4} G - G)^2 = (\frac{1}{4} G)^2, \]
so that at the zone boundary the kinetic energy of the two component waves $k = \pm \frac{1}{2}G$ are equal.

If $C(\frac{1}{2}G)$ is an important coefficient in the orbital (29) at the zone boundary, then $C(-\frac{1}{2}G)$ is also an important coefficient. This result also follows from the discussion of (5). We retain only those equations in the central equation that contain both coefficients $C(\frac{1}{2}G)$ and $C(-\frac{1}{2}G)$, and neglect all other coefficients.

One equation of (31) becomes, with $k = \frac{1}{2}G$ and $\lambda = \hbar^2(\frac{1}{2}G)^2/2m$,

$$ (\lambda - \epsilon)C(\frac{1}{2}G) + UC(-\frac{1}{2}G) = 0 \ . $$

Another equation of (31) becomes

$$ (\lambda - \epsilon)C(-\frac{1}{2}G) + UC(\frac{1}{2}G) = 0 \ . $$

These two equations have nontrivial solutions for the two coefficients if the energy $\epsilon$ satisfies

$$ \begin{vmatrix} \lambda - \epsilon & U \\ U & \lambda - \epsilon \end{vmatrix} = 0 \ . $$

Figure 8 Low-lying free electron energy bands of the empty sc lattice, as transformed to the first Brillouin zone and plotted vs. $(k,0,0)$. The free electron energy is $\hbar^2(k + G)^2/2m$, where the $G$'s are given in the second column of the table. The bold curves are in the first Brillouin zone, with $-\pi/a \leq k_x \leq \pi/a$. Energy bands drawn in this way are said to be in the reduced zone scheme.
whence,

$$(\lambda - \epsilon)^2 = U^2 ; \quad \epsilon = \lambda \pm U = \frac{\hbar^2}{2m}(\frac{1}{2}G)^2 \pm U \ . \quad (47)$$

The energy has two roots, one lower than the free electron kinetic energy by $U$, and one higher by $U$. Thus the potential energy $2U \cos Gx$ has created an energy gap $2U$ at the zone boundary.

The ratio of the C's may be found from either (44) or (45):

$$\frac{C(-\frac{1}{2}G)}{C(\frac{1}{2}G)} = \frac{\epsilon - \lambda}{U} = \pm 1 \ , \quad (48)$$

where the last step uses (47). Thus the Fourier expansion of $\psi(x)$ at the zone boundary has the two solutions

$$\psi(x) = \exp(iGx/2) \pm \exp(-iGx/2) \ .$$

These orbitals are identical to (5).

One solution gives the wavefunction at the bottom of the energy gap; the other gives the wavefunction at the top of the gap. Which solution has the lower energy depends on the sign of $U$.

We now solve for orbitals with wavevector $k$ near the zone boundary $\frac{1}{2}G$. We use the same two-component approximation, now with a wavefunction of the form

$$\psi(x) = C(k) e^{i{k}x} + C(k - G) e^{i(k-G)x} \ . \quad (49)$$

As directed by the central equation (31), we solve the pair of equations

$$(\lambda_k - \epsilon)C(k) + UC(k - G) = 0 \ ;$$

$$(\lambda_k - G - \epsilon)C(k - G) + UC(k) = 0 \ ,$$

with $\lambda_k$ defined as $\hbar^2k^2/2m$. These equations have a solution if the energy $\epsilon$ satisfies

$$\begin{vmatrix}
\lambda_k - \epsilon & U \\
U & \lambda_k - G - \epsilon
\end{vmatrix} = 0 \ ,$$

whence $\epsilon^2 - \epsilon(\lambda_k - \epsilon) + \lambda_k - U^2 = 0$.

The energy has two roots:

$$\epsilon = \frac{1}{2} (\lambda_k - \epsilon) \pm \left[\frac{1}{4}(\lambda_k - \epsilon) + U^2\right]^{1/2} \ , \quad (50)$$

and each root describes an energy band, plotted in Fig. 9. It is convenient to expand the energy in terms of a quantity $\tilde{K}$ (the mark over the $K$ is called a tilde), which measures the difference $\tilde{K} = k - \frac{1}{2}G$ in wavevector between $k$ and the zone boundary:

$$\epsilon_k = (\hbar^2/2m)(\frac{1}{4}G^2 + \tilde{K}^2) \pm \left[4\lambda(\hbar^2\tilde{K}^2/2m) + U^2\right]^{1/2} \ ,$$

$$= (\hbar^2/2m)(\frac{1}{4}G^2 + \tilde{K}^2) \pm U[1 + 2(\lambda/U^2)(\hbar^2\tilde{K}^2/2m)] \ , \quad (51)$$

in the region $\hbar^2C\tilde{K}/2m \ll |U|$. Here $\lambda = (\hbar^2/2m)(\frac{1}{2}G)^2$ as before.
Figure 9  Solutions of (50) in the periodic zone scheme, in the region near a boundary of the first Brillouin zone. The units are such that $U = -0.45, \ G = 2,$ and $\hbar^2/2m = 1.$ The free electron curve is drawn for comparison. The energy gap at the zone boundary is 0.90. The value of $U$ has deliberately been chosen large for this illustration, too large for the two-term approximation to be accurate.

Writing the two zone boundary roots of (47) as $\epsilon(\pm)$, we may write (51) as

$$\epsilon_k(\pm) = \epsilon(\pm) + \frac{k^2 \hbar^2}{2m} \left(1 + \frac{2\lambda}{U}\right).$$  \hspace{1cm} (52)

These are the roots for the energy when the wavevector is very close to the zone boundary at $\frac{1}{2}G$.

Note the quadratic dependence of the energy on the wavevector $\tilde{k}$. For $U$ negative, the solution $\epsilon(-)$ corresponds to the upper of the two bands, and $\epsilon(+)$ to the lower of the two bands. The two $C$'s are plotted in Fig. 10.

### NUMBER OF ORBITALS IN A BAND

Consider a linear crystal constructed of an even number $N$ of primitive cells of lattice constant $a$. In order to count states we apply periodic boundary conditions to the wavefunctions over the length of the crystal. The allowed values of the electron wavevector $k$ in the first Brillouin zone are given by (2):

$$k = 0 \ ; \ \pm \frac{2\pi}{L} \ ; \ \pm \frac{4\pi}{L} \ ; \ \ldots \ ; \ \frac{N\pi}{L}.$$  \hspace{1cm} (53)

We cut the series off at $N\pi/L = \pi/a$, for this is the zone boundary. The point $-N\pi/L = -\pi/a$ is not to be counted as an independent point because it is
connected by a reciprocal lattice vector with $\pi/a$. The total number of points is exactly $N$, the number of primitive cells.

Each primitive cell contributes exactly one independent value of $k$ to each energy band. This result carries over into three dimensions. With account taken of the two independent orientations of the electron spin, there are $2N$ independent orbitals in each energy band. If there is a single atom of valence one in each primitive cell, the band can be half filled with electrons. If each atom contributes two valence electrons to the band, the band can be exactly filled. If there are two atoms of valence one in each primitive cell, the band can also be exactly filled.

**Metals and Insulators**

If the valence electrons exactly fill one or more bands, leaving others empty, the crystal will be an insulator. An external electric field will not cause current flow in an insulator. (We suppose that the electric field is not strong enough to disrupt the electronic structure.) Provided that a filled band is separated by an energy gap from the next higher band, there is no continuous way to change the total momentum of the electrons if every accessible state is filled. Nothing changes when the field is applied. This is quite unlike the situation for free electrons for which $k$ increases uniformly in a field (Chapter 6).
A crystal can be an insulator only if the number of valence electrons in a primitive cell of the crystal is an even integer. (An exception must be made for electrons in tightly bound inner shells which cannot be treated by band theory.) If a crystal has an even number of valence electrons per primitive cell, it is necessary to consider whether or not the bands overlap in energy. If the bands overlap in energy, then instead of one filled band giving an insulator, we can have two partly filled bands giving a metal (Fig. 11).

The alkali metals and the noble metals have one valence electron per primitive cell, so that they have to be metals. The alkaline earth metals have two valence electrons per primitive cell; they could be insulators, but the bands overlap in energy to give metals, but not very good metals. Diamond, silicon, and germanium each have two atoms of valence four, so that there are eight valence electrons per primitive cell; the bands do not overlap, and the pure crystals are insulators at absolute zero.
SUMMARY

- The solutions of the wave equation in a periodic lattice are of the Bloch form \( \psi_k(r) = e^{ik \cdot r} u_k(r) \), where \( u_k(r) \) is invariant under a crystal lattice translation.

- There are regions of energy for which no Bloch function solutions of the wave equation exist. These energies form forbidden regions in which the wavefunctions are damped in space and the values of the \( k \)'s are complex, as pictured in Fig. 12. The existence of forbidden regions of energy is prerequisite to the existence of insulators.

- Energy bands may often be approximated by one or two plane waves: for example, \( \psi_k(x) \equiv C(k)e^{ikx} + C(k - G)e^{i(k - G)x} \) near the zone boundary at \( \frac{1}{2} G \).

- The number of orbitals in a band is \( 2N \), where \( N \) is the number of primitive cells in the specimen.

Problems

1. **Square lattice, free electron energies.** (a) Show for a simple square lattice (two dimensions) that the kinetic energy of a free electron at a corner of the first zone is higher than that of an electron at midpoint of a side face of the zone by a factor of 2. (b) What is the corresponding factor for a simple cubic lattice (three dimensions)? (c) What bearing might the result of (b) have on the conductivity of divalent metals?

2. **Free electron energies in reduced zone.** Consider the free electron energy bands of an fcc crystal lattice in the approximation of an empty lattice, but in the reduced zone scheme in which all \( k \)'s are transformed to lie in the first Brillouin zone. Plot roughly in the [111] direction the energies of all bands up to six times the lowest band energy at the zone boundary at \( k = (2\pi/a)(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}) \). Let this be the unit of energy. This problem shows why band edges need not necessarily be at the zone center. Several of the degeneracies (band crossings) will be removed when account is taken of the crystal potential.

3. **Kronig-Penney model.** (a) For the delta-function potential and with \( P \ll 1 \), find at \( k = 0 \) the energy of the lowest energy band. (b) For the same problem find the band gap at \( k = \pi/a \).

4. **Potential energy in the diamond structure.** (a) Show that for the diamond structure the Fourier component \( U_G \) of the crystal potential seen by an electron is equal to zero for \( G = 2A \), where \( A \) is a basis vector in the reciprocal lattice referred to the conventional cubic cell. (b) Show that in the usual first-order approximation to the solutions of the wave equation in a periodic lattice the energy gap vanishes at the zone boundary plane normal to the end of the vector \( A \).
Figure 12 In the energy gap there exist solutions of the wave equation for complex values of the wavevector. At the boundary of the first zone the real part of the wavevector is $\frac{1}{2}G$. The imaginary part of $k$ in the gap is plotted in the approximation of two plane waves, for $U = 0.01 \frac{\hbar^2 G^2}{2m}$. In an infinite unbounded crystal the wavevector must be real, or else the amplitude will increase without limit. But on a surface or at a junction there can exist solutions with complex wavevector.

**5. Complex wavevectors in the energy gap.** Find an expression for the imaginary part of the wavevector in the energy gap at the boundary of the first Brillouin zone, in the approximation that led to Eq. (46). Give the result for the $\text{Im}(k)$ at the center of the energy gap. The result for small $\text{Im}(k)$ is

$$\left(\frac{\hbar^2}{2m}\right) |\text{Im}(k)|^2 = 2mU^2/\hbar^2 G^2.$$  

The form as plotted in Fig. 12 is of importance in the theory of Zener tunneling from one band to another in the presence of a strong electric field.

6. **Square lattice.** Consider a square lattice in two dimensions with the crystal potential

$$U(x,y) = -4U \cos(2\pi x/a) \cos(2\pi y/a).$$

Apply the central equation to find approximately the energy gap at the corner point $(\pi/a, \pi/a)$ of the Brillouin zone. It will suffice to solve a $2 \times 2$ determinantal equation.

**References**


(Further references on band theory are given at the end of Chapter 9.)

*This problem is somewhat difficult.*
8

Semiconductor Crystals

BAND GAP 199

EQUATIONS OF MOTION 203
    Physical derivation of $\hbar k = F$ 205
    Holes 206
    Effective mass 209
    Physical interpretation of the effective mass 210
    Effective masses in semiconductors 212
    Silicon and germanium 214

INTRINSIC CARRIER CONCENTRATION 216
    Intrinsic mobility 220

IMPURITY CONDUCTIVITY 221
    Donor states 222
    Acceptor states 224
    Thermal ionization of donors and acceptors 226

THERMOELECTRIC EFFECTS 227

SEMIMETALS 228

SUPERLATTICES 228

SUMMARY 231

PROBLEMS 231
    1. Impurity orbits 231
    2. Ionization of donors 231
    3. Hall effect with two carrier types 231
    4. Cyclotron resonance 232
    5. Magnetoresistance with two carrier types 232

REFERENCES 232

NOTE: The discussion of carrier orbits in applied fields is continued in Chapter 9. Amorphous semiconductors are treated in Chapter 17. Junctions and barriers are treated in Chapter 19.
Figure 1 Carrier concentrations for metals, semimetals, and semiconductors. The semiconductor range may be extended upward by increasing the impurity concentration, and the range can be extended downward to merge eventually with the insulator range.
Carrier concentrations representative of metals, semimetals, and semiconductors are shown in Fig. 1. Semiconductors are generally classified by their electrical resistivity at room temperature, with values in the range of $10^{-2}$ to $10^{9}$ ohm-cm, and strongly dependent on temperature. At absolute zero a pure, perfect crystal of most semiconductors will be an insulator, if we arbitrarily define an insulator as having a resistivity above $10^{14}$ ohm-cm.

Devices based on semiconductors include transistors, switches, diodes, photovoltaic cells, detectors, and thermistors. These may be used as single circuit elements or as components of integrated circuits. We discuss in this chapter the central physical features of the classical semiconductor crystals, particularly silicon, germanium, and gallium arsenide.

Some useful nomenclature: the semiconductor compounds of chemical formula $AB$, where $A$ is a trivalent element and $B$ is a pentavalent element, are called III-V (three-five) compounds. Examples are indium antimonide and gallium arsenide. Where $A$ is divalent and $B$ is hexavalent, the compound is called a II-VI compound; examples are zinc sulfide and cadmium sulfide. Silicon and germanium are sometimes called diamond-type semiconductors, because they have the crystal structure of diamond. Diamond itself is more an insulator rather than a semiconductor. Silicon carbide SiC is a IV-IV compound.

A highly purified semiconductor exhibits intrinsic conductivity, as distinguished from the impurity conductivity of less pure specimens. In the intrinsic temperature range the electrical properties of a semiconductor are not essentially modified by impurities in the crystal. An electronic band scheme leading to intrinsic conductivity is indicated in Fig. 2. The conduction band is vacant at absolute zero and is separated by an energy gap $E_g$ from the filled valence band.

The band gap is the difference in energy between the lowest point of the conduction band and the highest point of the valence band. The lowest point in the conduction band is called the conduction band edge; the highest point in the valence band is called the valence band edge.

As the temperature is increased, electrons are thermally excited from the valence band to the conduction band (Fig. 3). Both the electrons in the conduction band and the vacant orbitals or holes left behind in the valence band contribute to the electrical conductivity.

**BAND GAP**

The intrinsic conductivity and intrinsic carrier concentrations are largely controlled by $E_g/k_BT$, the ratio of the band gap to the temperature. When this ratio is large, the concentration of intrinsic carriers will be low and the conductivity will be low. Band gaps of representative semiconductors are given in Table 1. The best values of the band gap are obtained by optical absorption.
Figure 2 Band scheme for intrinsic conductivity in a semiconductor. At 0 K the conductivity is zero because all states in the valence band are filled and all states in the conduction band are vacant. As the temperature is increased, electrons are thermally excited from the valence band to the conduction band, where they become mobile.

Figure 3 Intrinsic electron concentration as a function of temperature for (a) germanium and (b) silicon. Under intrinsic conditions the hole concentration is equal to the electron concentration. The intrinsic concentration at a given temperature is higher in Ge than in Si because the energy gap is narrower in Ge (0.66 eV) than in Si (1.11 eV). (After W. C. Dunlap.)
<table>
<thead>
<tr>
<th>Crystal</th>
<th>Gap</th>
<th>$E_v$, eV</th>
<th></th>
<th>Crystal</th>
<th>Gap</th>
<th>$E_v$, eV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0 K</td>
<td>300 K</td>
<td></td>
<td></td>
<td>0 K</td>
<td>300 K</td>
</tr>
<tr>
<td>Diamond</td>
<td>$i$</td>
<td>5.4</td>
<td></td>
<td>HgTe$^a$</td>
<td>$d$</td>
<td>-0.30</td>
<td></td>
</tr>
<tr>
<td>Si</td>
<td>$i$</td>
<td>1.17</td>
<td>1.11</td>
<td>PbS</td>
<td>$d$</td>
<td>0.286</td>
<td>0.34–0.37</td>
</tr>
<tr>
<td>Ge</td>
<td>$i$</td>
<td>0.744</td>
<td>0.66</td>
<td>PbSe</td>
<td>$i$</td>
<td>0.165</td>
<td>0.27</td>
</tr>
<tr>
<td>$\alpha$Sn</td>
<td>$d$</td>
<td>0.00</td>
<td>0.00</td>
<td>PbTe</td>
<td>$i$</td>
<td>0.190</td>
<td>0.29</td>
</tr>
<tr>
<td>InSb</td>
<td>$d$</td>
<td>0.23</td>
<td>0.17</td>
<td>CdS</td>
<td>$d$</td>
<td>2.582</td>
<td>2.42</td>
</tr>
<tr>
<td>InAs</td>
<td>$d$</td>
<td>0.43</td>
<td>0.36</td>
<td>CdSe</td>
<td>$d$</td>
<td>1.840</td>
<td>1.74</td>
</tr>
<tr>
<td>InP</td>
<td>$d$</td>
<td>1.42</td>
<td>1.27</td>
<td>CdTe</td>
<td>$d$</td>
<td>1.607</td>
<td>1.44</td>
</tr>
<tr>
<td>GaP</td>
<td>$i$</td>
<td>2.32</td>
<td>2.25</td>
<td>ZnO</td>
<td></td>
<td>3.436</td>
<td>3.2</td>
</tr>
<tr>
<td>GaAs</td>
<td></td>
<td>1.52</td>
<td>1.43</td>
<td>ZnS</td>
<td></td>
<td>3.91</td>
<td>3.6</td>
</tr>
<tr>
<td>GaSb</td>
<td>$d$</td>
<td>0.81</td>
<td>0.68</td>
<td>SnTe</td>
<td>$d$</td>
<td>0.3</td>
<td>0.18</td>
</tr>
<tr>
<td>AlSb</td>
<td>$i$</td>
<td>1.65</td>
<td>1.6</td>
<td>AgCl</td>
<td>—</td>
<td>—</td>
<td>3.2</td>
</tr>
<tr>
<td>SiC(hex)</td>
<td>$i$</td>
<td>3.0</td>
<td>—</td>
<td>AgI</td>
<td>—</td>
<td>—</td>
<td>2.8</td>
</tr>
<tr>
<td>Te</td>
<td>$d$</td>
<td>0.33</td>
<td>—</td>
<td>Cu$_2$O</td>
<td>$d$</td>
<td>2.172</td>
<td>—</td>
</tr>
<tr>
<td>ZnSb</td>
<td></td>
<td>0.56</td>
<td>0.56</td>
<td>TiO$_2$</td>
<td></td>
<td>3.03</td>
<td>—</td>
</tr>
</tbody>
</table>

$^a$HgTe is a semimetal; the bands overlap.

The threshold of continuous optical absorption at frequency $\omega_k$ determines the band gap $E_g = \hbar \omega_k$ in Figs. 4a and 5a. In the direct absorption process a photon is absorbed by the crystal with the creation of an electron and a hole.

In the indirect absorption process in Figs. 4b and 5b the minimum energy gap of the band structure involves electrons and holes separated by a substantial wavevector $k_c$. Here a direct photon transition at the energy of the minimum gap cannot satisfy the requirement of conservation of wavevector, because photon wavevectors are negligible at the energy range of interest. But if a phonon of wavevector $K$ and frequency $\Omega$ is created in the process, then we can have

$$k(\text{photon}) = k_c + K \equiv 0; \quad \hbar \omega = E_g + \hbar \Omega,$$

as required by the conservation laws. The phonon energy $\hbar \Omega$ will generally be much less than $E_g$: a phonon even of high wavevector is an easily accessible source of crystal momentum because the phonon energies are characteristically small (~0.01 to 0.03 eV) in comparison with the energy gap. If the temperature is high enough that the necessary phonon is already thermally excited in the crystal, it is possible also to have a photon absorption process in which the phonon is absorbed.
Figure 4 Optical absorption in pure insulators at absolute zero. In (a) the threshold determines the energy gap as $E_g = \hbar \omega_c$. In (b) the optical absorption is weaker near the threshold; at $\hbar \omega = E_g + \hbar \Omega$ a photon is absorbed with the creation of three particles: a free electron, a free hole, and a phonon of energy $\hbar \Omega$. In (b) the energy $E_{\text{seet}}$ marks the threshold for the creation of a free electron and a free hole, with no phonon involved. Such a transition is called vertical; it is similar to the direct transition in (a). These plots do not show absorption lines that sometimes are seen lying just to the low energy side of the threshold. Such lines are due to the creation of a bound electron-hole pair, called an exciton.

Figure 5 In (a) the lowest point of the conduction band occurs at the same value of $k$ as the highest point of the valence band. A direct optical transition is drawn vertically with no significant change of $k$, because the absorbed photon has a very small wavevector. The threshold frequency $\omega_c$ for absorption by the direct transition determines the energy gap $E_g = \hbar \omega_c$. The indirect transition in (b) involves both a photon and a phonon because the band edges of the conduction and valence bands are widely separated in $k$ space. The threshold energy for the indirect process in (b) is greater than the true band gap. The absorption threshold for the indirect transition between the band edges is at $\hbar \omega = E_g + \hbar \Omega$, where $\Omega$ is the frequency of an emitted phonon of wavevector $K = -k_c$. At higher temperatures phonons are already present; if a phonon is absorbed along with a photon, the threshold energy is $\hbar \omega = E_g - \hbar \Omega$. Note: The figure shows only the threshold transitions. Transitions occur generally between almost all points of the two bands for which the wavevectors and energy can be conserved.
The band gap may also be deduced from the temperature dependence of the conductivity or of the carrier concentration in the intrinsic range. The carrier concentration is obtained from measurements of the Hall voltage (Chapter 6), sometimes supplemented by conductivity measurements. Optical measurements determine whether the gap is direct or indirect. The band edges in Ge and in Si are connected by indirect transitions; the band edges in InSb are connected by a direct transition (Fig. 6). The gap in αSn is direct and is exactly zero; HgTe and HgSe are semimetals and have negative gaps—the bands overlap.

**EQUATIONS OF MOTION**

We derive the equation of motion of an electron in an energy band. We look at the motion of a wave packet in an applied electric field. Suppose that the wave packet is made up of wavefunctions near a particular wavevector $k$. The
group velocity is \( v_g = d\omega / dk \). The frequency associated with a wavefunction of energy \( \epsilon \) is \( \omega = \epsilon / \hbar \), and so
\[
v_g = \hbar^{-1} d\epsilon / dk \quad \text{or} \quad v = \hbar^{-1} \nabla_k \epsilon(k) .
\] (1)
The effects of the crystal on the electron motion are contained in the dispersion relation \( \epsilon(k) \).

The work \( \delta \epsilon \) done on the electron by the electric field \( E \) in the interval \( \delta t \) is
\[
\delta \epsilon = -eEv_g \delta t .
\] (2)
We observe that
\[
\delta \epsilon = (d\epsilon / dk) \delta k = \hbar v_g \delta k ,
\] (3)
using (1). On comparing (2) with (3) we have
\[
\delta k = -(eE / \hbar) \delta t ,
\] (4)
whence \( \hbar \delta k / dt = -eE \), the same relation as for free electrons.

We may write (4) in terms of the external force \( F \) as
\[
\hbar \frac{dk}{dt} = F .
\] (5)
This is an important relation: in a crystal \( \hbar \delta k / dt \) is equal to the external force on the electron. In free space \( d(mv) / dt \) is equal to the force. We have not overthrown Newton's second law of motion: the electron in the crystal is subject to forces from the crystal lattice as well as from external sources.

The force term in (5) also includes the Lorentz force on an electron in a magnetic field, under ordinary conditions where the magnetic field is not so strong that it breaks down the band structure. Thus the equation of motion of an electron of group velocity \( v \) in a constant magnetic field \( B \) is
\[
\text{(CGS)} \quad \hbar \frac{dk}{dt} = -\frac{e}{c} v \times B ;
\] (6)
\[
\text{(SI)} \quad \hbar \frac{dk}{dt} = -e v \times B ,
\] (6)
where the right-hand side is the Lorentz force on the electron. With the group velocity \( \hbar v = \text{grad}_k \epsilon \), the rate of change of the wavevector is
\[
\text{(CGS)} \quad \frac{dk}{dt} = -\frac{e}{\hbar^2 c} \nabla_k \epsilon \times B ;
\] (7)
\[
\text{(SI)} \quad \frac{dk}{dt} = -\frac{e}{\hbar^2} \nabla_k \epsilon \times B ,
\] (7)
where now both sides of the equation refer to the coordinates in \( k \) space.

We see from the vector cross-product in (7) that in a magnetic field an electron moves in \( k \) space in a direction normal to the direction of the gradient of the energy \( \epsilon \), so that the electron moves on a surface of constant energy.
The value of the section \( k_B \) of \( k \) on \( B \) is constant during the motion. The motion in \( k \) space is on a plane normal to the direction of \( B \), and the orbit is defined by the intersection of this plane with a surface of constant energy.

**Physical Derivation of \( \hbar \dot{k} = F \)**

We consider the Bloch eigenfunction \( \psi_k \) belonging to the energy eigenvalue \( \varepsilon_k \) and wavevector \( k \):

\[
\psi_k = \sum_G C(k + G) \exp[i(k + G) \cdot r] .
\]

The expectation value of the momentum of an electron in the state \( k \) is

\[
p_{el} = (k| -i\hbar \nabla |k) = \sum_G \hbar(k + G)|C(k + G)|^2 = \hbar(k + \sum_G G|C(k + G)|^2) ,
\]

using \( \Sigma|C(k + G)|^2 = 1 \).

We examine the transfer of momentum between the electron and the lattice when the state \( k \) of the electron is changed to \( k + \Delta k \) by the application of an external force. We imagine an insulating crystal electrostatically neutral except for a single electron in the state \( k \) of an otherwise empty band.

We suppose that a weak external force is applied for a time interval such that the total impulse given to the entire crystal system is \( J = \int F \, dt \). If the conduction electron were free \( (m^* = m) \), the total momentum imparted to the crystal system by the impulse would appear in the change of momentum of the conduction electron:

\[
J = \Delta p_{tot} = \Delta p_{el} = \hbar \Delta k .
\]

The neutral crystal suffers no net interaction with the electric field, either directly or indirectly through the free electron.

If the conduction electron interacts with the periodic potential of the crystal lattice, we must have

\[
J = \Delta p_{tot} = \Delta p_{lat} + \Delta p_{el} .
\]

From the result (9) for \( p_{el} \) we have

\[
\Delta p_{el} = \hbar \Delta k + \sum_G \hbar G[\nabla_k|C(k + G)|^2] \cdot \Delta k .
\]

The change \( \Delta p_{lat} \) in the lattice momentum resulting from the change of state of the electron may be derived by an elementary physical consideration. An electron reflected by the lattice transfers momentum to the lattice. If an incident electron with plane wave component of momentum \( \hbar k \) is reflected
with momentum $\hbar(k + G)$, the lattice acquires the momentum $-\hbar G$, as required by momentum conservation. The momentum transfer to the lattice when the state $\psi_k$ goes over to $\psi_{k+\Delta k}$ is

$$\Delta p_{\text{lat}} = -\hbar \sum G[(\nabla_k|C(k + G)|^2 \cdot \Delta k] ,$$  

(13)

as the portion

$$\nabla_k|C(k + G)|^2 \cdot \Delta k$$  

(14)

of each individual component of the initial state is reflected during the state change $\Delta k$.

The total momentum change is therefore

$$\Delta p_{\text{el}} + \Delta p_{\text{lat}} = J = \hbar \Delta k ,$$  

(15)

exactly as for free electrons, Eq. (10). Thus from the definition of $J$, we have

$$\hbar dk/dt = F ,$$  

(16)

derived in (5) by a different method. A rigorous derivation of (16) by an entirely different method is given in Appendix E.

**Holes**

The properties of vacant orbitals in an otherwise filled band are important in semiconductor physics and in solid state electronics. Vacant orbitals in a band are commonly called holes. A hole acts in applied electric and magnetic fields as if it has a positive charge $+e$. The reason is given in five steps in the boxes that follow.

1. $k_h = -k_e .$$\quad$  

(17)

The total wavevector of the electrons in a filled band is zero: $\Sigma k = 0$. This result follows from the geometrical symmetry of the Brillouin zone: every fundamental lattice type has symmetry under the inversion operation $r \rightarrow -r$ about any lattice point; it follows that the Brillouin zone of the lattice also has inversion symmetry. If the band is filled all pairs of orbitals $k$ and $-k$ are filled, and the total wavevector is zero.

If an electron is missing from an orbital of wavevector $k_e$, the total wavevector of the system is $-k_e$ and is attributed to the hole. This result is surprising: the electron is missing from $k_e$ and the position of the hole is usually indicated graphically as situated at $k_e$, as in Fig. 7. But the true wavevector $k_h$ of the hole is $-k_e$, which is the wavevector of the point $G$ if the hole is at $E$. The wavevector $-k_e$ enters into selection rules for photon absorption.
Figure 7  Absorption of a photon of energy $h\omega$ and negligible wavevector takes an electron from $E$ in the filled valence band to $Q$ in the conduction band. If $k_e$ was the wavevector of the electron at $E$, it becomes the wavevector of the electron at $Q$. The total wavevector of the valence band after the absorption is $-k_e$, and this is the wavevector we must ascribe to the hole if we describe the valence band as occupied by one hole. Thus $k_h = -k_e$; the wavevector of the hole is the same as the wavevector of the electron which remains at $G$. For the entire system the total wavevector after the absorption of the photon is $k_e + k_h = 0$, so that the total wavevector is unchanged by the absorption of the photon and the creation of a free electron and free hole.

The hole is an alternate description of a band with one missing electron, and we either say that the hole has wavevector $-k_e$ or that the band with one missing electron has total wavevector $-k_e$.

2.  

\[ \varepsilon_h(k_h) = -\varepsilon_e(k_e) \]  

(18)

Let the zero of energy of the valence band be at the top of the band. The lower in the band the missing electron lies, the higher the energy of the system. The energy of the hole is opposite in sign to the energy of the missing electron, because it takes more work to remove an electron from a low orbital than from a high orbital. Thus if the band is symmetric, \(^1\)  

\[ \varepsilon_e(k_e) = \varepsilon_e(-k_e) = -\varepsilon_h(-k_e) = -\varepsilon_h(k_h) \]. 

We construct in Fig. 8 a band scheme to represent the properties of a hole. This hole band is a helpful representation because it appears right side up.

\(^1\) Bands are always symmetric under the inversion $k \rightarrow -k$ if the spin-orbit interaction is neglected. Even with spin-orbit interaction, bands are always symmetric if the crystal structure permits the inversion operation. Without a center of symmetry, but with spin-orbit interaction, the bands are symmetric if we compare subbands for which the spin direction is reversed: $\varepsilon(k, \uparrow) = \varepsilon(-k, \downarrow)$, \( \varepsilon \) QTS, Chapter 9.
Figure 8. The upper half of the figure shows the hole band that simulates the dynamics of a hole, constructed by inversion of the valence band in the origin. The wavevector and energy of the hole are equal, but opposite in sign, to the wavevector and energy of the empty electron orbital in the valence band. We do not show the disposition of the electron removed from the valence band at \( k_e \).

3. \[ v_h = v_e \] (19)

The velocity of the hole is equal to the velocity of the missing electron.

From Fig. 8 we see that \( \nabla \epsilon_h(k_h) = -\nabla \epsilon_e(k_e) \), so that \( v_h(k_h) = \frac{\epsilon_e}{c} \).

4. \[ m_h = -m_e \] (20)

We show below that the effective mass is inversely proportional to the curvature \( d^2 \epsilon/dk^2 \), and for the hole band this has the opposite sign to that for an electron in the valence band. Near the top of the valence band \( m_e \) is negative, so that \( m_h \) is positive.

5. \[ \hbar \frac{dk_h}{dt} = e(E + \frac{1}{c} v_h \times B) \] (21)

This comes from the equation of motion

(CG5) \[ \hbar \frac{dk_e}{dt} = -e(E + \frac{1}{c} v_e \times B) \] (22)

that applies to the missing electron when we substitute \( -k_h \) for \( k_e \) and \( v_h \) for \( v_e \). The equation of motion for a hole is that of a particle of positive charge \( e \). The positive charge is consistent with the electric current carried by the valence band of Fig. 9: the current is carried by the unpaired electron in the orbital \( G \):

\[ j = (-e)\nabla(G) = (-e)[-\nabla(E)] = ev(E), \] (23)

which is just the current of a positive charge moving with the velocity ascribed to the missing electron at \( E \). The current is shown in Fig. 10.
Figure 9  (a) At $t = 0$ all states are filled except $F$ at the top of the band; the velocity $v_x$ is zero at $F$ because $de/dk_x = 0$. (b) An electric field $E_x$ is applied in the $+x$ direction. The force on the electrons is in the $-k_x$ direction and all electrons make transitions together in the $-k_x$-direction, moving the hole to the state $E$. (c) After a further interval the electrons move farther along in $k$ space and the hole is now at $D$.

Figure 10  Motion of electrons in the conduction band and holes in the valence band in the electric field $E$. The hole and electron drift velocities are in opposite directions, but their electric currents are in the same direction, the direction of the electric field.

**Effective Mass**

When we look at the energy-wavevector relation $\epsilon = (\hbar^2/2m)k^2$ for free electrons, we see that the coefficient of $k^2$ determines the curvature of $\epsilon$ versus $k$. Turned about, we can say that $1/m$, the reciprocal mass, determines the curvature. For electrons in a band there can be regions of unusually high curvature near the band gap at the zone boundary, as we see from the solutions of the wave equation near the zone boundary. If the energy gap is small in comparison with the free electron energy $\lambda$ at the boundary, the curvature is enhanced by the factor $\lambda/E_g$ and the reciprocal mass is enhanced by the same factor.

In semiconductors the band width, which is like the free electron energy, is of the order of 20 eV, while the band gap is of the order of 0.2 to 2 eV. Thus the reciprocal mass is enhanced by a factor 10 to 100, and the effective mass is reduced to 0.1–0.01 of the free electron mass. These values apply near the band gap; as we go away from the gap the curvatures are likely to approach those of free electrons.
To summarize the solutions of Chapter 7 for $U$ positive, an electron near the lower edge of the second band has an energy that may be written as

$$\epsilon(K) = \epsilon_e + (\hbar^2 / 2m_e)K^2; \quad m_e / m = 1/[(2\lambda / U) - 1] . \tag{24}$$

Here $K$ is the wavevector measured from the zone boundary, and $m_e$ denotes the effective mass of the electron near the edge of the second band. An electron near the top of the first band has the energy

$$\epsilon(K) = \epsilon_e - (\hbar^2 / 2m_h)K^2; \quad m_h / m = 1/[(2\lambda / U) + 1] . \tag{25}$$

The curvature and hence the mass will be negative near the top of the first band, but we have introduced a minus sign into (25) in order that the symbol $m_h$ for the hole mass will have a positive value—see (20) above.

The crystal does not weigh any less if the effective mass of a carrier is less than the free electron mass, nor is Newton's second law violated for the crystal taken as a whole, ions plus carriers. The important point is that an electron in a periodic potential is accelerated relative to the lattice in an applied electric or magnetic field as if the mass of the electron were equal to an effective mass which we now define.

We differentiate the result (1) for the group velocity to obtain

$$\frac{dv_g}{dt} = \hbar^{-1} \frac{d^2\epsilon}{dk \, dt} = \hbar^{-1} \left( \frac{d^2\epsilon}{dk^2} \frac{dk}{dt} \right) \tag{26}$$

We know from (5) that $dk / dt = F / \hbar$, whence

$$\frac{dv_g}{dt} = \left( \frac{1}{\hbar^2} \frac{d^2\epsilon}{dk^2} \right) F; \quad \text{or} \quad F = \frac{\hbar^2}{d^2\epsilon / dk^2} \frac{dv_g}{dt} . \tag{27}$$

If we identify $\hbar^2 / (d^2\epsilon / dk^2)$ as a mass, then (27) assumes the form of Newton's second law. We define the effective mass $m^*$ by

$$\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{d^2\epsilon}{dk^2} . \tag{28}$$

It is easy to generalize this to take account of an anisotropic energy surface, as for electrons in Si or Ge. We introduce the components of the reciprocal effective mass tensor

$$\left( \frac{1}{m^*} \right)_{\mu \nu} = \frac{1}{\hbar^2} \frac{d^2\epsilon_k}{dk_\mu \, dk_\nu}; \quad \frac{dv_\mu}{dt} = \left( \frac{1}{m^*} \right)_{\mu \nu} F_\nu , \tag{29}$$

where $\mu, \nu$ are Cartesian coordinates.

**Physical Interpretation of the Effective Mass**

How can an electron of mass $m$ when put into a crystal respond to applied fields as if the mass were $m^*$? It is helpful to think of the process of Bragg reflection of electron waves in a lattice. Consider the weak interaction approxi-
mation treated in Chapter 7. Near the bottom of the lower band the orbital is represented quite adequately by a plane wave \( \exp(ikx) \) with momentum \( \hbar k \); the wave component \( \exp[i(k - G)x] \) with momentum \( \hbar(k - G) \) is small and increases only slowly as \( k \) is increased, and in this region \( m^* = m \). An increase in the reflected component \( \exp[i(k - G)x] \) as \( k \) is increased represents momentum transfer to the electron from the lattice.

Near the boundary the reflected component is quite large; at the boundary it becomes equal in amplitude to the forward component, at which point the eigenfunctions are standing waves, rather than running waves. Here the momentum component \( \hbar(-\frac{1}{2}G) \) cancels the momentum component \( \hbar(\frac{1}{2}G) \).

It is not surprising to find negative values for \( m^* \) just below a zone boundary. A single electron in an energy band may have positive or negative effective mass: the states of positive effective mass occur near the bottom of a band because positive effective mass means that the band has upward curvature (\( d^2E/dk^2 \) is positive). States of negative effective mass occur near the top of the band. A negative effective mass means that on going from state \( k \) to state \( k + \Delta k \), the momentum transfer to the lattice from the electron is larger than the momentum transfer from the applied force to the electron. Although \( k \) is increased by \( \Delta k \) by the applied electric field, the approach to Bragg reflection can give an overall decrease in the forward momentum of the electron; when this happens the effective mass is negative (Fig. 11).

As we proceed in the second band away from the boundary, the amplitude of \( \exp[i(k - G)x] \) decreases rapidly and \( m^* \) assumes a small positive value. Here the increase in electron velocity resulting from a given external impulse is larger than that which a free electron would experience. The lattice makes up the difference through the reduced recoil it experiences when the amplitude of \( \exp[i(k - G)x] \) diminishes.
If the energy in a band depends only slightly on $k$, then the effective mass will be very large. That is, $m^*/m \gg 1$ when $d^2\epsilon/dk^2$ is very small. The tight-binding approximation discussed in Chapter 9 gives quick insight into the formation of narrow bands. If the wavefunctions centered on neighboring atoms overlap very little, then the overlap integral will be small; the width of the band narrow; and the effective mass large.

The overlap of wavefunctions centered on neighboring atoms is small for the inner or core electrons. The $4f$ electrons of the rare earth metals, for example, overlap very little. The overlap integral determines the rate of quantum tunneling of an electron from one ion to another. When the effective mass is large, the electron tunnels slowly from one ion to an adjacent ion in the lattice.

**Effective Masses in Semiconductors**

In many semiconductors it has been possible to determine by cyclotron resonance the form of the energy surfaces of the conduction and valence bands near the band edges. The determination of the energy surface is equivalent to a determination of the effective mass tensor (29). Cyclotron resonance in a semiconductor is carried out with centimeter wave or millimeter wave radiation at low carrier concentration.

The current carriers are accelerated in helical orbits about the axis of a static magnetic field. The angular rotation frequency $\omega_c$ is

\[
(CGS) \quad \omega_c = \frac{eB}{m^*c},
\]

\[
(SI) \quad \omega_c = \frac{eB}{m^*},
\]

where $m^*$ is the appropriate cyclotron effective mass. Resonant absorption of energy from an rf electric field perpendicular to the static magnetic field (Fig. 12) occurs when the rf frequency is equal to the cyclotron frequency. Holes and electrons rotate in opposite senses in a magnetic field.

We consider the experiment for $m^*/m = 0.1$. At $f_c = 24$ GHz, or $\omega_c = 1.5 \times 10^{11}$ s$^{-1}$, we have $B = 860$ G at resonance. The line width is determined by the collision relaxation time $\tau$, and to obtain a distinctive resonance it is necessary that $\omega_c\tau \approx 1$. The mean free path must be long enough to permit the average carrier to get one radian around a circle between collisions. The requirements are met with the use of higher frequency radiation and higher magnetic fields, with high purity crystals in liquid helium.

In direct-gap semiconductors with band edges at the center of the Brillouin zone, the bands have the structure shown in Fig. 13. The conduction band edge is spherical with the effective mass $m_e$:

\[
\epsilon_c = E_c + \frac{k^2}{2m_e},
\]

referred to the valence band edge. The valence bands are characteristically
Figure 12  Arrangement of fields in a cyclotron resonance experiment in a semiconductor. The sense of the circulation is opposite for electrons and holes.

Figure 13  Simplified view of the band edge structure of a direct gap semiconductor.
Table 2 Effective masses of electrons and holes in direct gap semiconductors

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Electron $m_e/m$</th>
<th>Heavy hole $m_{hh}/m$</th>
<th>Light hole $m_{lh}/m$</th>
<th>Split-off hole $m_{so}/m$</th>
<th>Spin-orbit $\Delta$, eV</th>
</tr>
</thead>
<tbody>
<tr>
<td>InSb</td>
<td>0.015</td>
<td>0.39</td>
<td>0.021</td>
<td>(0.11)</td>
<td>0.82</td>
</tr>
<tr>
<td>InAs</td>
<td>0.026</td>
<td>0.41</td>
<td>0.025</td>
<td>0.08</td>
<td>0.43</td>
</tr>
<tr>
<td>InP</td>
<td>0.073</td>
<td>0.4</td>
<td>(0.078)</td>
<td>(0.15)</td>
<td>0.11</td>
</tr>
<tr>
<td>GaSb</td>
<td>0.047</td>
<td>0.3</td>
<td>0.06</td>
<td>(0.14)</td>
<td>0.80</td>
</tr>
<tr>
<td>GaAs</td>
<td>0.066</td>
<td>0.5</td>
<td>0.082</td>
<td>0.17</td>
<td>0.34</td>
</tr>
<tr>
<td>Cu$_2$O</td>
<td>0.99</td>
<td></td>
<td>0.58</td>
<td>0.69</td>
<td>0.13</td>
</tr>
</tbody>
</table>

threefold near the edge, with the heavy hole $hh$ and light hole $lh$ bands degenerate at the center, and a band $soh$ split off by the spin-orbit splitting $\Delta$:

$$
\epsilon_e(hh) = -\hbar^2 k^2/2m_{hh}; \quad \epsilon_e(lh) = -\hbar^2 k^2/2m_{lh}; \quad \epsilon_e(soh) = -\Delta - \hbar^2 k^2/2m_{soh}
$$

Values of the mass parameters are given in Table 2. The forms (32) are only approximate, because even close to $k = 0$ the heavy and light hole bands are not spherical—see the discussion below for Ge and Si.

The perturbation theory of band edges (Problem 9.8) suggests that the electron effective mass should be proportional to the band gap, approximately, for a direct gap crystal. We use Tables 1 and 2 to find the fairly constant values $m_e/(m_E) = 0.063$, 0.060, and 0.051 in (eV)$^{-1}$ for the series InSb, InAs, and InP, in agreement with this suggestion.

**Silicon and Germanium**

The conduction and valence bands of germanium are shown in Fig. 14, based on a combination of theoretical and experimental results. The valence band edge in both crystals is at $k = 0$ and is derived from $p_{3/2}$ and $p_{1/2}$ states of the free atoms, as is clear from the tight-binding approximation (Chapter 9) to the wavefunctions.

The $p_{3/2}$ level is fourfold degenerate as in the atom; the four states correspond to $m_f$ values $\pm \frac{3}{2}$ and $\pm \frac{1}{2}$. The $p_{1/2}$ level is doubly degenerate, with $m_f = \pm \frac{1}{2}$. The $p_{3/2}$ states are higher in energy than the $p_{1/2}$ states; the energy difference $\Delta$ is a measure of the spin-orbit interaction.

The valence band edges are not simple. Holes near the band edge are characterized by two effective masses, light and heavy. These arise from the two bands formed from the $p_{3/2}$ level of the atom. There is also a band formed from the $p_{1/2}$ level, split off from the $p_{3/2}$ level by the spin-orbit interaction. The energy surfaces are not spherical, but warped (QTS, p 271):

$$
\epsilon(k) = AK^2 \pm [B^2 k^4 + C^2(k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2)]^{1/2}
$$
Figure 14 Calculated band structure of germanium, after C. Y. Fong. The general features are in good agreement with experiment. The four valence bands are shown in gray. The fine structure of the valence band edge is caused by spin-orbit splitting. The energy gap is indirect; the conduction band edge is at the point \( (2\pi/a)(1 1 1) \). The constant energy surfaces around this point are ellipsoidal.
The choice of sign distinguishes the two masses. The split-off band has \( \epsilon(k) = -\Delta + AK^2 \). The experiments give, in units \( \hbar^2/2m \),

\[
\begin{align*}
\text{Si:} & \quad A = -4.29 ; \quad |B| = 0.68 ; \quad |C| = 4.87 ; \quad \Delta = 0.044 \text{ eV} \\
\text{Ge:} & \quad A = -13.38 ; \quad |B| = 8.48 ; \quad |C| = 13.15 ; \quad \Delta = 0.29 \text{ eV}
\end{align*}
\]

Roughly, the light and heavy holes in germanium have masses 0.043 \( m \) and 0.34 \( m \); in silicon 0.16 \( m \) and 0.52 \( m \); in diamond 0.7 \( m \) and 2.12 \( m \).

The conduction band edges in Ge are at the equivalent points \( L \) of the Brillouin zone, Fig. 15a. Each band edge has a spheroidal energy surface oriented along a \( <111> \) crystal axis, with a longitudinal mass \( m_l = 1.59 \) \( m \) and a transverse mass \( m_t = 0.082 \) \( m \). For a static magnetic field at an angle \( \theta \) with the longitudinal axis of a spheroid, the effective cyclotron mass \( m_c \) is

\[
\frac{1}{m_c^2} = \frac{\cos^2\theta}{m_l^2} + \frac{\sin^2\theta}{m_t m_l}
\]

(34)

Results for Ge are shown in Fig. 16.

In silicon the conduction band edges are spheroids oriented along the equivalent \( <100> \) directions in the Brillouin zone, with mass parameters \( m_l = 0.92 \) \( m \) and \( m_t = 0.19 \) \( m \), as in Fig. 17a. The band edges lie along the lines labeled \( \Delta \) in the zone of Fig. 15a, a little way in from the boundary points \( X \).

In GaAs we have \( A = -6.98, B = -4.5, |C| = 6.2, \Delta = 0.341 \text{ eV} \). The band structure is shown in Fig. 17b.

**INTRINSIC CARRIER CONCENTRATION**

We want the concentration of intrinsic carriers in terms of the band gap. We do the calculation for simple parabolic band edges. We first calculate in terms of the chemical potential \( \mu \) the number of electrons excited to the conduction band at temperature \( T \). In semiconductor physics \( \mu \) is called the Fermi level. At the temperatures of interest we may suppose for the conduction band of a semiconductor that \( \epsilon - \mu \gg k_B T \), and the Fermi-Dirac distribution function reduces to

\[
f_e = \exp[(\mu - \epsilon)/k_B T].
\]

(35)

This is the probability that a conduction electron orbital is occupied, in an approximation valid when \( f_e \ll 1 \).

The energy of an electron in the conduction band is

\[
\epsilon_k = E_c + \hbar^2 k^2 / 2m_e,
\]

(36)

where \( E_c \) is the energy at the conduction band edge, as in Fig. 18. Here \( m_e \) is the effective mass of an electron. Thus from (6.20) the density of states at \( \epsilon \) is

\[
D_e(\epsilon) = \frac{1}{2\pi^2} \left( \frac{2m_e}{\hbar^2} \right)^{3/2} (\epsilon - E_c)^{1/2}.
\]

(37)
Figure 15 Standard labels of the symmetry points and axes of the Brillouin zones of the fcc and bcc lattices. The zone centers are $\Gamma$. In (a) the boundary point at $(2\pi/a)(100)$ is $X$; the boundary point at $(2\pi/a)(\frac{1}{2} \frac{1}{2} \frac{1}{2})$ is $L$; the line $\Delta$ runs between $\Gamma$ and $X$. In (b) the corresponding symbols are $H$, $P$ and $\Delta$.

Figure 16 Effective cyclotron mass of electrons in germanium at 4 K for magnetic field directions in a 110 plane. There are four independent mass spheroids in Ge, one along each 111 axis, but viewed in the 110 plane two spheroids always appear equivalent. (After Dresselhaus, Kip, and Kittel.)
The concentration of electrons in the conduction band is

\[ n = \int_{E_c}^{\infty} D_c(\epsilon)f_c(\epsilon)d\epsilon = \frac{1}{2\pi^2} \left( \frac{2m_e}{\hbar^2} \right)^{3/2} \exp(\mu/k_B T) \times \int_{E_c}^{\infty} (\epsilon - E_c)^{1/2} \exp(-\epsilon/k_B T) d\epsilon , \]  

which integrates to give

\[ n = 2 \left( \frac{m_e k_B T}{2\pi \hbar^2} \right)^{3/2} \exp[(\mu - E_c)/k_B T] . \]  

The problem is solved for \( n \) when \( \mu \) is known. It is useful to calculate the equilibrium concentration of holes \( p \). The distribution function \( f_h \) for holes is related to the electron distribution function \( f_e \) by \( f_h = 1 - f_e \), because a hole is the absence of an electron. We have

\[ f_h = 1 - \frac{1}{\exp[(\epsilon - \mu)/k_B T] + 1} = \frac{1}{\exp[(\mu - \epsilon)/k_B T] + 1} \]

\[ \approx \exp[(\epsilon - \mu)/k_B T] , \]  

provided \( (\mu - \epsilon) \gg k_B T \).

If the holes near the top of the valence band behave as particles with effective mass \( m_h \), the density of hole states is given by

\[ D_h(\epsilon) = \frac{1}{2\pi^2} \left( \frac{2m_h}{\hbar^2} \right)^{3/2} (E_v - \epsilon)^{1/2} , \]  

where \( E_v \) is the energy at the top of the valence band.
where $E_v$ is the energy at the valence band edge. Proceeding as in (38) we obtain

$$p = \int_{-\infty}^{E_v} D_h(\epsilon)f_h(\epsilon)d\epsilon = 2\left(\frac{m_h k_B T}{2\pi \hbar^2}\right)^{3/2} \exp[(E_v - \mu)/k_B T]$$  

for the concentration $p$ of holes in the valence band.

We multiply together the expressions for $n$ and $p$ to obtain the equilibrium relation, with the energy gap $E_g = E_c - E_v$,

$$np = 4\left(\frac{k_B T}{2\pi \hbar^2}\right)^3 (m_e m_h)^{3/2} \exp(-E_g/k_B T).$$  

This useful result does not involve the Fermi level $\mu$. It is an expression of the law of mass action.

We have nowhere assumed in the derivation that the material is intrinsic: the result holds in the presence of impurities as well. The only assumption made is that the distance of the Fermi level from the edge of both bands is large in comparison with $k_B T$. At 300 K the value of $np$ is $2.10 \times 10^{19} \text{ cm}^{-6}$, $2.89 \times 10^{26} \text{ cm}^{-6}$, and $6.55 \times 10^{12} \text{ cm}^{-6}$, for the actual band structures of Si, Ge, and GaAs, respectively.

A simple kinetic argument shows why the product $np$ is constant at a given temperature. Suppose that the equilibrium population of electrons and holes is maintained by black-body photon radiation at temperature $T$. The photons generate electron-hole pairs at a rate $A(T)$, while $B(T)np$ is the rate of the recombina-
nation reaction \( e + h = \text{photon} \). Then

\[
dn/dt = A(T) - B(T)np = dp/dt
\]  

(44)

In equilibrium \( dn/dt = 0 \); \( dp/dt = 0 \), whence \( np = A(T)B(T) \).

Because the product of the electron and hole concentrations is a constant independent of impurity concentration at a given temperature, the introduction of a small proportion of a suitable impurity to increase \( n \), say, must decrease \( p \). This result is important in practice—we can reduce the total carrier concentration \( n + p \) in an impure crystal, sometimes enormously, by the controlled introduction of suitable impurities. Such a reduction is called compensation.

In an intrinsic semiconductor the number of electrons is equal to the number of holes, because the thermal excitation of an electron leaves behind a hole in the valence band. Thus from (43) we have, letting the subscript \( i \) denote intrinsic and \( E_i = E_c - E_v \),

\[
n_i = p_i = 2\left(\frac{k_B T}{2\pi\hbar^2}\right)^{3/2} (m_e m_h)^{3/4} \exp(-E_i/2k_B T)
\]  

(45)

The intrinsic carrier depends exponentially on \( E_i/2k_B T \), where \( E_i \) is the energy gap. We set (39) equal to (42) to obtain, for the Fermi level as measured from the top of the valence band,

\[
\exp(2\mu/k_B T) = (m_h/m_e)^{3/2} \exp(E_i/k_B T);
\]  

(46)

\[
\mu = \frac{1}{2}E_i + \frac{3}{2}k_B T \ln (m_h/m_e)
\]  

(47)

If \( m_h = m_e \), then \( \mu = \frac{1}{2}E_i \) and the Fermi level is in the middle of the forbidden gap. A thorough treatment of the statistical physics of semiconductors is given in TP, Chapter 13.

**Intrinsic Mobility**

The mobility is the magnitude of the drift velocity per unit electric field:

\[
\mu = |v|/E.
\]  

(48)

The mobility is defined to be positive for both electrons and holes, although their drift velocities are opposite. By writing \( \mu_e \) or \( \mu_h \) for the electron or hole mobility we can avoid any confusion between \( \mu \) as the chemical potential and as the mobility.

The electrical conductivity is the sum of the electron and hole contributions:

\[
\sigma = (n e \mu_e + p e \mu_h),
\]  

(49)

where \( n \) and \( p \) are the concentrations of electrons and holes. In Chapter 6 the drift velocity of a charge \( q \) was found to be \( v = q\tau E/m \), whence

\[
\mu_e = e\tau_e/m_e; \quad \mu_h = e\tau_h/m_h.
\]  

(50)
Table 3 Carrier mobilities at room temperature, in cm²/V-s

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Electrons</th>
<th>Holes</th>
<th>Crystal</th>
<th>Electrons</th>
<th>Holes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond</td>
<td>1800</td>
<td>1200</td>
<td>GaAs</td>
<td>8000</td>
<td>300</td>
</tr>
<tr>
<td>Si</td>
<td>1350</td>
<td>480</td>
<td>GaSb</td>
<td>5000</td>
<td>1000</td>
</tr>
<tr>
<td>Ge</td>
<td>3600</td>
<td>1800</td>
<td>PbS</td>
<td>550</td>
<td>600</td>
</tr>
<tr>
<td>InSb</td>
<td>800</td>
<td>450</td>
<td>PbSe</td>
<td>1020</td>
<td>930</td>
</tr>
<tr>
<td>InAs</td>
<td>30000</td>
<td>450</td>
<td>PbTe</td>
<td>2500</td>
<td>1000</td>
</tr>
<tr>
<td>InP</td>
<td>4500</td>
<td>100</td>
<td>AgCl</td>
<td>50</td>
<td>—</td>
</tr>
<tr>
<td>AlAs</td>
<td>280</td>
<td>—</td>
<td>KBr (100 K)</td>
<td>100</td>
<td>—</td>
</tr>
<tr>
<td>AlSb</td>
<td>900</td>
<td>400</td>
<td>SiC</td>
<td>100</td>
<td>10–20</td>
</tr>
</tbody>
</table>

The mobilities depend on temperature as a modest power law. The temperature dependence of the conductivity in the intrinsic region will be dominated by the exponential dependence \( \exp(-E_g/2k_BT) \) of the carrier concentration, Eq. (45).

Table 3 gives experimental values of the mobility at room temperature. The mobility in SI units is expressed in m²/V-s and is \( 10^{-4} \) of the mobility in practical units. For most substances the values quoted are limited by the scattering of carriers by thermal phonons. The hole mobilities typically are smaller than the electron mobilities because of the occurrence of band degeneracy at the valence band edge at the zone center, thereby making possible interband scattering processes that reduce the mobility considerably.

In some crystals, particularly in ionic crystals, the holes are essentially immobile and get about only by thermally-activated hopping from ion to ion. The principal cause of this "self-trapping" is the lattice distortion associated with the Jahn-Teller effect of degenerate states (Chapter 14). The necessary orbital degeneracy is much more frequent for holes than for electrons.

There is a tendency for crystals with small energy gaps at direct band edges to have high values of the electron mobility. By (9.41) small gaps lead to light effective masses, which by (50) favor high mobilities. The highest mobility observed in a semiconductor is \( 5 \times 10^6 \) cm²/V-s in PbTe at 4 K, where the gap is 0.19 eV.

**IMPURITY CONDUCTIVITY**

Certain impurities and imperfections drastically affect the electrical properties of a semiconductor. The addition of boron to silicon in the proportion of 1 boron atom to \( 10^5 \) silicon atoms increases the conductivity of pure silicon by a factor of \( 10^3 \) at room temperature. In a compound semiconductor a stoichiometric deficiency of one constituent will act as an impurity; such semiconductors
are known as deficit semiconductors. The deliberate addition of impurities to a semiconductor is called doping.

We consider the effect of impurities in silicon and germanium. These elements crystallize in the diamond structure. Each atom forms four covalent bonds, one with each of its nearest neighbors, corresponding to the chemical valence four. If an impurity atom of valence five, such as phosphorus, arsenic, or antimony, is substituted in the lattice in place of a normal atom, there will be one valence electron from the impurity atom left over after the four covalent bonds are established with the nearest neighbors, that is, after the impurity atom has been accommodated in the structure with as little disturbance as possible.

**Donor States.** The structure in Fig. 19 has a positive charge on the impurity atom (which has lost one electron). Lattice constant studies have verified that the pentavalent impurities enter the lattice by substitution for normal atoms, and not in interstitial positions. Impurity atoms that can give up an electron are called donors. The crystal as a whole remains neutral because the electron remains in the crystal.

The electron moves in the coulomb potential $\frac{e^2}{r}$ of the impurity ion, where $e$ in a covalent crystal is the static dielectric constant of the medium. The factor $\frac{1}{\epsilon}$ takes account of the reduction in the coulomb force between charges caused by the electronic polarization of the medium. This treatment is valid for orbits large in comparison with the distance between atoms, and for slow motions of the electron such that the orbital frequency is low in comparison with the frequency $\omega_e$ corresponding to the energy gap. These conditions are satisfied quite well in Ge and Si by the donor electron of P, As, or Sb.

We estimate the ionization energy of the donor impurity. The Bohr theory of the hydrogen atom may be modified to take into account the dielectric constant of the medium and the effective mass of an electron in the periodic potential of the crystal. The ionization energy of atomic hydrogen is $-e^4m/2\hbar^2$ in CGS and $-e^4m/2(4\pi\varepsilon\varepsilon_0h)^2$ in SI.

In the semiconductor we replace $\epsilon$ by $\epsilon^2/e$ and $m$ by the effective mass $m_e$ to obtain

\[
\text{(CGS)} \quad E_d = \frac{e^4m_e}{2\epsilon^2\hbar^2} = \left(\frac{13.6}{\epsilon^2/m} \right) \text{ eV} ;
\]

\[
\text{(SI)} \quad E_d = \frac{e^4m_e}{2(4\pi\varepsilon\varepsilon_0h)^2} ;
\]

as the donor ionization energy of the semiconductor.

The Bohr radius of the ground state of hydrogen is $\hbar^2/m_e\epsilon^2$ in CGS or $4\pi\varepsilon\varepsilon_0\hbar^2/m_e\epsilon^2$ in SI. Thus the Bohr radius of the donor is

\[
\text{(CGS)} \quad a_d = \frac{e\hbar^2}{m_e\epsilon^2} = \left(\frac{0.53\epsilon}{m_e/m} \right) \text{ Å} ;
\]

\[
\text{(SI)} \quad a_d = \frac{4\pi\varepsilon\varepsilon_0\hbar^2}{m_e\epsilon^2} .
\]
Figure 19 Charges associated with an arsenic impurity atom in silicon. Arsenic has five valence electrons, but silicon has only four valence electrons. Thus four electrons on the arsenic form tetrahedral covalent bonds similar to silicon, and the fifth electron is available for conduction. The arsenic atom is called a donor because when ionized it donates an electron to the conduction band.

The application to germanium and silicon is complicated by the anisotropic effective mass of the conduction electrons. But the dielectric constant has the more important effect on the donor energy because it enters as the square, whereas the effective mass enters only as the first power.

To obtain a general impression of the impurity levels we use $m_e \approx 0.1 \, m$ for electrons in germanium and $m_e \approx 0.2 \, m$ in silicon. The static dielectric constant is given in Table 4. The ionization energy of the free hydrogen atom is 13.6 eV. For germanium the donor ionization energy $E_d$ on our model is 5 meV, reduced with respect to hydrogen by the factor $m_e/m_e^2 = 4 \times 10^{-4}$. The corresponding result for silicon is 20 meV. Calculations using the correct anisotropic mass tensor predict 9.05 meV for germanium and 29.8 meV for silicon. Observed values of donor ionization energies in Si and Ge are given in Table 5. Recall that 1 meV = $10^{-3}$ eV. In GaAs donors have $E_d \approx 6$ meV.

The radius of the first Bohr orbit is increased by $\epsilon m/m_e$ over the value

<table>
<thead>
<tr>
<th>Crystal</th>
<th>$\epsilon$</th>
<th>Crystal</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond</td>
<td>5.5</td>
<td>GaSb</td>
<td>15.69</td>
</tr>
<tr>
<td>Si</td>
<td>11.7</td>
<td>GaAs</td>
<td>13.13</td>
</tr>
<tr>
<td>Ge</td>
<td>15.8</td>
<td>AlAs</td>
<td>10.1</td>
</tr>
<tr>
<td>InSb</td>
<td>17.88</td>
<td>AlSb</td>
<td>10.3</td>
</tr>
<tr>
<td>InAs</td>
<td>14.55</td>
<td>SiC</td>
<td>10.2</td>
</tr>
<tr>
<td>InP</td>
<td>12.37</td>
<td>Cu$_2$O</td>
<td>7.1</td>
</tr>
</tbody>
</table>
Table 5  Donor ionization energies $E_d$ of pentavalent impurities in germanium and silicon, in meV

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>As</th>
<th>Sb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>45.</td>
<td>49.</td>
<td>39.</td>
</tr>
<tr>
<td>Ge</td>
<td>12.0</td>
<td>12.7</td>
<td>9.6</td>
</tr>
</tbody>
</table>

0.53 Å for the free hydrogen atom. The corresponding radius is $(160)(0.53) = 80$ Å in germanium and $(60)(0.53) = 30$ Å in silicon. These are large radii, so that donor orbits overlap at relatively low donor concentrations, compared to the number of host atoms. With appreciable overlap, an “impurity band” is formed from the donor states: see the discussion of the metal-insulator transition in Chapter 10.

The semiconductor can conduct in the impurity band by electrons hopping from donor to donor. The process of impurity band conduction sets in at lower donor concentration levels if there are also some acceptor atoms present, so that some of the donors are always ionized. It is easier for a donor electron to hop to an ionized (unoccupied) donor than to an occupied donor atom, so that two electrons will not have to occupy the same site during charge transport.

**Acceptor States.** A hole may be bound to a trivalent impurity in germanium or silicon (Fig. 20), just as an electron is bound to a pentavalent impurity. Trivalent impurities such as B, Al, Ga, and In are called acceptors because they accept electrons from the valence band in order to complete the covalent bonds with neighboring atoms, leaving holes in the band.

When an acceptor is ionized a hole is freed, which requires an input of energy. On the usual energy band diagram, an electron rises when it gains energy, whereas a hole sinks in gaining energy.

Experimental ionization energies of acceptors in germanium and silicon are given in Table 6. The Bohr model applies qualitatively for holes just as for electrons, but the degeneracy at the top of the valence band complicates the effective mass problem.

The tables show that donor and acceptor ionization energies in Si are comparable with $k_B T$ at room temperature (26 meV), so that the thermal ionization of donors and acceptors is important in the electrical conductivity of silicon at room temperature. If donor atoms are present in considerably greater numbers than acceptors, the thermal ionization of donors will release electrons into the conduction band. The conductivity of the specimen then will be controlled by electrons (negative charges), and the material is said to be $n$ type.

If acceptors are dominant, holes will be released into the valence band and the conductivity will be controlled by holes (positive charges): the material is $p$
Table 6  Acceptor ionization energies $E_a$ of trivalent impurities in germanium and silicon, in meV

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Al</th>
<th>Ga</th>
<th>In</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>45.</td>
<td>57.</td>
<td>65.</td>
<td>157.</td>
</tr>
<tr>
<td>Ge</td>
<td>10.4</td>
<td>10.2</td>
<td>10.8</td>
<td>11.2</td>
</tr>
</tbody>
</table>

Figure 20  Boron has only three valence electrons; it can complete its tetrahedral bonds only by taking an electron from a Si-Si bond, leaving behind a hole in the silicon valence band. The positive hole is then available for conduction. The boron atom is called an acceptor because when ionized it accepts an electron from the valence band. At 0 K the hole is bound; remember that holes float.

The sign of the Hall voltage is a rough test for $n$ or $p$ type. Another handy laboratory test is the sign of the thermoelectric potential, discussed below.

The numbers of holes and electrons are equal in the intrinsic regime. The intrinsic electron concentration $n_i$ at 300 K is $1.7 \times 10^{13}$ cm$^{-3}$ in germanium and $4.6 \times 10^9$ cm$^{-3}$ in silicon. The electrical resistivity of intrinsic material is 43 ohm-cm for germanium and $2.6 \times 10^5$ ohm-cm for silicon.

Germanium has $4.42 \times 10^{22}$ atoms per cm$^3$. The purification of Ge has been carried further than any other element.\(^2\) The concentration of the common electrically active impurities—the shallow donor and acceptor impurities—has been reduced below 1 impurity atom in $10^{11}$ Ge atoms (Fig. 21). For example, the concentration of P in Ge can be reduced below $4 \times 10^{10}$ cm$^{-3}$. The experimental sensitivity for detection of these impurities is $10^7$ cm$^{-3}$ by the method of photothermal ionization spectroscopy.\(^3\) There are impurities (H, O,


Si, C) whose concentrations in Ge cannot usually be reduced below $10^{12}$--$10^{14}$ cm$^{-3}$, but these do not affect electrical measurements and therefore may be hard to detect.

**Thermal Ionization of Donors and Acceptors**

The calculation of the equilibrium concentration of conduction electrons from ionized donors is identical with the standard calculation in statistical mechanics of the thermal ionization of hydrogen atoms (TP, p. 369). If there are no acceptors present, the result in the low temperature limit $k_B T \ll E_d$ is

$$n \equiv (n_0 N_d)^{1/2} \exp(-E_d/2k_B T), \quad (53)$$

with $n_0 = 2(m_e k_B T/2\pi\hbar^2)^{3/2}$, here $N_d$ is the concentration of donors. To obtain (53) we apply the laws of chemical equilibria to the concentration ratio $[e]/[N_d^+]$, and then set $[N_d^+] = [e] = n$. Identical results hold for acceptors, under the assumption of no donor atoms.

If the donor and acceptor concentrations are comparable, affairs are complicated and the equations are solved by numerical methods. However, the law of mass action (43) requires the $np$ product to be constant at a given temperature. An excess of donors will increase the electron concentration and decrease the hole concentration; the sum $n + p$ will increase. The conductivity will increase as $n + p$ if the mobilities are equal, as in Fig. 2.2.
Figure 22 Electrical conductivity and hole concentration \( p \) calculated as a function of electron concentration \( n \) for a semiconductor at a temperature such that \( np = 10^{20} \text{ cm}^{-6} \). The conductivity is symmetrical about \( n = 10^{10} \text{ cm}^{-3} \). For \( n > 10^{10} \) the specimen is \( n \) type; for \( n < 10^{10} \), it is \( p \) type. We have taken \( \mu_e = \mu_h \) for the mobilities.

THERMOELECTRIC EFFECTS

Consider a semiconductor maintained at a constant temperature while an electric field drives through it an electric current density \( j_q \). If the current is carried only by electrons, the charge flux is

\[
j_q = n(-e)(-\mu_e)E = n e \mu_e E \ ,
\]

where \( \mu_e \) is the electron mobility. The average energy transported by an electron is, referred to the Fermi level \( \mu \),

\[
(E_e - \mu) + \frac{3}{2}k_B T
\]

where \( E_e \) is the energy at the conduction band edge. We refer the energy to the Fermi level because different conductors in contact have the same Fermi level. The energy flux that accompanies the charge flux is

\[
j_U = n(E_e - \mu + \frac{3}{2}k_B T)(-\mu_e)E \ .
\]

The Peltier coefficient \( \Pi \) is defined by the relation \( j_U = \Pi j_q \); it is the energy carried per unit charge. For electrons,

\[
\Pi_e = -(E_e - \mu + \frac{3}{2}k_B T)/e
\]
and is negative because the energy flux is opposite to the charge flux. For holes

\[ j_q = p e \mu_h E ; \quad j_v = p (\mu - E_v + \frac{3}{2} k_B T) \mu_v E \]

(57)

where \( E_v \) is the energy at the valence band edge. Thus

\[ \Pi_h = (\mu - E_v + \frac{3}{2} k_B T)/e \]

(58)

and is positive. Equations (56) and (58) are the result of our simple drift velocity theory; a treatment by the Boltzmann transport equation gives minor numerical differences.\(^4\)

The absolute thermoelectric power \( Q \) is defined from the open circuit electric field created by a temperature gradient:

\[ E = Q \text{ grad } T \]

(59)

The Peltier coefficient \( \Pi \) is related to the thermoelectric power \( Q \) by

\[ \Pi = QT \]

(60)

This is the famous Kelvin relation of irreversible thermodynamics.\(^5\) A measurement of the sign of the voltage across a semiconductor specimen, one end of which is heated, is a rough and ready way to tell if the specimen is \( n \) type or \( p \) type (Fig. 23).

**SEMIMETALS**

In semimetals the conduction band edge is very slightly lower in energy than the valence band edge. A small overlap in energy of the conduction and valence bands leads to small concentration of holes in the valence band and of electrons in the conduction band (Table 7). Three of the semimetals, arsenic, antimony, and bismuth, are in group V of the periodic table.

Their atoms associate in pairs in the crystal lattice, with two ions and ten valence electrons per primitive cell. The even number of valence electrons would allow these elements to be insulators. Like semiconductors, the semimetals may be doped with suitable impurities to vary the relative numbers of holes and electrons. Their concentrations may also be varied with pressure, for the band edge overlap varies with pressure.

**SUPERLATTICES**

Consider a multilayer crystal of alternating thin layers of different compositions. Coherent layers on a nanometer thickness scale may be deposited by


molecular-beam epitaxy or metal-organic vapor deposition, thus building up a superperiodic structure on a large scale. Systems of alternate layers of GaAs and GaAlAs have been studied to 50 periods or more, with lattice spacing $A$ of perhaps 5 nm (50 Å). A superperiodic crystal potential arises from the superperiodic structure and acts on the conduction electrons and holes to create new (small) Brillouin zones and mini energy bands superposed on the band structures of the constituent layers. Here we treat the motion of an electron in a superlattice in an applied electric field; magnetic fields are treated in Chapter 19.

**Bloch Oscillator**

Consider a collisionless electron in a periodic lattice in one dimension, with motion normal to the planes of the superlattice. The equation of motion in
a constant electric field parallel to \( k \) is \( \hbar dk/dt = -eE \) or, for motion across a Brillouin zone with reciprocal lattice vector \( G = 2\pi/\Lambda \), we have \( \hbar G = \hbar 2\pi/\Lambda = eET \). The Bloch frequency of the motion is \( \omega_B = 2\pi/\Lambda = eEA/\hbar \). The electron accelerates from \( k = 0 \) toward the zone boundary; when it reaches \( k = \pi/\Lambda \) it reappears (as by an Umklapp process) at the zone boundary at the identical point \( -\pi/\Lambda \), using the argument of Chapter 2.

We consider the motion in a model system in real space. We suppose that the electron lies in a simple energy band of width \( \epsilon_0 \):

\[
\epsilon = \epsilon_0 (1 - \cos k\Lambda) .
\]  

(61)
The velocity in \( k \)-space (momentum space) is

\[
v = \hbar^{-1} d\epsilon/dk = (\Lambda \epsilon_0/\hbar) \sin k\Lambda ,
\]

(62)
and the position of the electron in real space, with the initial condition \( z = 0 \) at \( t = 0 \), is given by

\[
z = \int v \, dt = \int dk \ v(k)(dt/dk) = (\Lambda \epsilon_0/\hbar) \int dk \ (-\hbar/eE) \sin k\Lambda
\]

\[
= (-\epsilon_0/eE)(\cos k\Lambda - 1) = (-\epsilon_0/eE)(\cos(-eEA/\hbar) - 1) .
\]

(63)
This confirms that the oscillation frequency in real space is \( \omega_B = eEA/\hbar \). The motion in the periodic lattice is quite different from the motion in free space, for which the acceleration is constant.

**Wannier Ladder**

The electrostatic energy of the system is increased by \( eEA \) if the initial position of the electron is displaced by one lattice constant from \( z = 0 \) to \( z = \Lambda \). For a displacement of the initial position by \( n \) lattice constants, the energy eigenvalues are

\[
\epsilon_n = neEA + \epsilon_0 (1 - \cos k\Lambda) .
\]

(64)
This set of solutions defines the Wannier Ladder of energy levels, with adjacent level separation \( \Delta \epsilon = eEA = \hbar \omega_B \). The exact solution of the wave equation gives a continuum of states, with density peaked sharply at the eigenvalues (64). The ladders are observed in optical experiments. The Bloch frequency has not been observed directly (as of mid-1993).

**Estimates.** Let \( E = 3 \times 10^4 \) V cm\(^{-1} \), \( \epsilon_0 = 0.1 \) eV, and \( \Lambda = 10^{-6} \) cm; then \( \omega_B = 4.6 \times 10^{13} \) s\(^{-1} \) and the localization length is \( \epsilon_0/eE = 3 \) nm. For narrow bands in strong fields it may be possible to localize the electron wavefunction dominantly on a single atom; changing the index \( n \) in \( \epsilon_n \) will change the localization site of the electron.

**Zener Tunneling**

Thus far we have considered the effect of the electrostatic potential \( -eEz \) (or \( -eEnA \)) on one energy band; the potential tilts the whole band. Higher
bands will also be tilted similarly, creating the possibility of crossing between ladder levels of different bands. The interaction between different band levels at the same energy opens the possibility for an electron in one band at \( n \) to cross to another band at \( n' \). This field-induced interband tunneling is an example of Zener breakdown, met most often at a single junction as in the Zener diode.

**SUMMARY**

- The motion of a wave packet centered at wavevector \( \mathbf{k} \) is described by \( \mathbf{F} = \hbar \partial \mathbf{k} / \partial t \), where \( \mathbf{F} \) is the applied force. The motion in real space is obtained from the group velocity \( \mathbf{v}_g = \hbar^{-1} \nabla_k \epsilon(k) \).
- The smaller the energy gap, the smaller is \( |m^*| \) near the gap.
- A crystal with one hole has one empty electron state in an otherwise filled band. The properties of the hole are those of the \( N - 1 \) electrons:
  (a) If the electron is missing from the state of wavevector \( \mathbf{k}_e \), then the wavevector of the hole is \( \mathbf{k}_h = -\mathbf{k}_e \).
  (b) The rate of change of \( \mathbf{k}_h \) in an applied field requires the assignment of a positive charge to the hole: \( e_h = e = -e_e \).
  (c) If \( \mathbf{v}_e \) is the velocity an electron would have in the state \( \mathbf{k}_e \), then the velocity to be ascribed to the hole of wavevector \( \mathbf{k}_h = -\mathbf{k}_e \) is \( \mathbf{v}_h = \mathbf{v}_e \).
  (d) The energy of the hole referred to zero for a filled band is positive and is \( \epsilon_h(k_h) = -\epsilon(k_e) \).
  (e) The effective mass of a hole is opposite to the effective mass of an electron at the same point on the energy band: \( m_h = -m_e \).

**Problems**

1. **Impurity orbits.** Indium antimonide has \( E_g = 0.23 \) eV; dielectric constant \( \epsilon = 18 \); electron effective mass \( m_e = 0.015 \) m. Calculate (a) the donor ionization energy; (b) the radius of the ground state orbit. (c) At what minimum donor concentration will appreciable overlap effects between the orbits of adjacent impurity atoms occur? This overlap tends to produce an impurity band—a band of energy levels which permit conductivity presumably by a hopping mechanism in which electrons move from one impurity site to a neighboring ionized impurity site.

2. **Ionization of donors.** In a particular semiconductor there are \( 10^{13} \) donors/cm\(^3\) with an ionization energy \( E_d \) of 1 meV and an effective mass 0.01 m. (a) Estimate the concentration of conduction electrons at 4 K. (b) What is the value of the Hall coefficient? Assume no acceptor atoms are present and that \( E_g \gg k_B T \).

3. **Hall effect with two carrier types.** Assuming concentrations \( n, p \); relaxation times \( \tau_e, \tau_h \); and masses \( m_e, m_h \), show that the Hall coefficient in the drift velocity approximation is...
\[ R_H = \frac{1}{ec} \frac{p - nb^2}{(p + nb)^2}, \]

where \( b = \mu_e/\mu_h \) is the mobility ratio. In the derivation neglect terms of order \( B^2 \). In SI we drop the \( c \). Hint: In the presence of a longitudinal electric field, find the transverse electric field such that the transverse current vanishes. The algebra may seem tedious, but the result is worth the trouble. Use (6.64), but for two carrier types; neglect \( (\omega_c\tau)^2 \) in comparison with \( \omega_c \tau \).

4. **Cyclotron resonance for a spheroidal energy surface.** Consider the energy surface

\[ \epsilon(k) = \hbar \left( \frac{k_x^2 + k_y^2}{2m_e} + \frac{k_z^2}{2m_l} \right), \]

where \( m_l \) is the transverse mass parameter and \( m_l \) is the longitudinal mass parameter. A surface on which \( \epsilon(k) \) is constant will be a spheroid. Use the equation of motion (6), with \( v = \hbar^{-1} \nabla k \epsilon \), to show that \( \omega_c = eB/(m_e m_l)^{1/2} c \) when the static magnetic field \( B \) lies in the \( x, y \) plane. This result agrees with (34) when \( \theta = \pi/2 \). The result is in CGS; to obtain SI, omit the \( c \).

5. **Magnetoresistance with two carrier types.** Problem 6.9 shows that in the drift velocity approximation the motion of charge carriers in electric and magnetic fields does not lead to transverse magnetoresistance. The result is different with two carrier types. Consider a conductor with a concentration \( n \) of electrons of effective mass \( m_e \) and relaxation time \( \tau_e \); and a concentration \( p \) of holes of effective mass \( m_h \) and relaxation time \( \tau_h \). Treat the limit of very strong magnetic fields, \( \omega_c \tau \gg 1 \). (a) Show in this limit that \( \sigma_{yx} = (n - p)ec/B \). (b) Show that the Hall field is given by, with \( Q = \omega_c \tau \),

\[ E_y = -(n - p) \left( \frac{n}{Q_e} + \frac{p}{Q_h} \right)^{-1} E_x, \]

which vanishes if \( n = p \). (c) Show that the effective conductivity in the \( x \) direction is

\[ \sigma_{xx} = \frac{ec}{B} \left[ \left( \frac{n}{Q_e} + \frac{p}{Q_h} \right) + (n - p)^2 \left( \frac{n}{Q_e} + \frac{p}{Q_h} \right)^{-1} \right]. \]

If \( n = p \), \( \sigma \propto B^{-2} \). If \( n \neq p \), \( \sigma \) saturates in strong fields, that is, it approaches a limit independent of \( B \) as \( B \to \infty \).

**References**

9
Fermi Surfaces and Metals

Reduced zone scheme 235
Periodic zone scheme 237

CONSTRUCTION OF FERMI SURFACES 238
Nearly free electrons 240

ELECTRON ORBITS, HOLE ORBITS, AND OPEN ORBITS 242

CALCULATION OF ENERGY BANDS 244
  Tight binding method for energy bands 245
  Wigner-Seitz method 248
  Cohesive energy 249
  Pseudopotential methods 252

EXPERIMENTAL METHODS IN FERMI SURFACE STUDIES 255
  Quantization of orbits in a magnetic field 255
  De Haas-van Alphen effect 257
  Extremal orbits 260
  Fermi surface of copper 262
  Example: Fermi surface of gold 262
  Magnetic breakdown 264

SUMMARY 265

PROBLEMS 266
  1. Brillouin zones of rectangular lattice 266
  2. Brillouin zone, rectangular lattice 266
  3. Hexagonal close-packed structure 266
  4. Brillouin zones of two-dimensional divalent metal 266
  5. Open orbits 266
  6. Cohesive energy for a square well potential 266
  7. De Haas-van Alphen period of potassium 266
  8. Band edge structure on k · p perturbation theory 267
  9. Wannier functions 267
  10. Open orbits and magnetoresistance 267
  11. Landau levels 268

REFERENCES 268
Figure 1  Free electron Fermi surfaces for fcc metals with one (Cu) and three (Al) valence electrons per primitive cell. The Fermi surface shown for copper has been deformed from a sphere to agree with the experimental results. The second zone of aluminum is nearly half-filled with electrons. (A. R. Mackintosh.)
Few people would define a metal as "a solid with a Fermi surface." This may nevertheless be the most meaningful definition of a metal one can give today; it represents a profound advance in the understanding of why metals behave as they do. The concept of the Fermi surface, as developed by quantum physics, provides a precise explanation of the main physical properties of metals.

A. R. Mackintosh

The Fermi surface is the surface of constant energy $\epsilon_F$ in k space. The Fermi surface separates the unfilled orbitals from the filled orbitals, at absolute zero. The electrical properties of the metal are determined by the shape of the Fermi surface, because the current is due to changes in the occupancy of states near the Fermi surface.

The shape may be very intricate in a reduced zone scheme and yet have a simple interpretation when reconstructed to lie near the surface of a sphere. We exhibit in Fig. 1 the free electron Fermi surfaces constructed for two metals that have the face-centered cubic crystal structure: copper, with one valence electron, and aluminum, with three. The free electron Fermi surfaces were developed from spheres of radius $k_F$ determined by the valence electron concentration. How do we construct these surfaces from a sphere? The constructions require the reduced and periodic zone schemes.

**Reduced Zone Scheme**

It is always possible to select the wavevector index k of any Bloch function to lie within the first Brillouin zone. The procedure is known as mapping the band in the reduced zone scheme.

If we encounter a Bloch function written as $\psi_{k'}(r) = e^{ik' \cdot r} u_{k'}(r)$, with k' outside the first zone, as in Fig. 2, we may always find a suitable reciprocal lattice vector G such that $k = k' + G$ lies within the first Brillouin zone. Then

$$\psi_{k'}(r) = e^{ik' \cdot r} u_{k'}(r) = e^{ik \cdot r} e^{-iG \cdot r} u_k(r)$$

$$= e^{ik \cdot r} u_k(r) = \psi_k(r),$$

where $u_k(r) = e^{-iG \cdot r} u_{k'}(r)$. Both $e^{-iG \cdot r}$ and $u_k(r)$ are periodic in the crystal lattice, so $u_k(r)$ is also, whence $\psi_k(r)$ is of the Bloch form.

Even with free electrons it is useful to work in the reduced zone scheme, as in Fig. 3. Any energy $\epsilon_{k'}$ for k' outside the first zone is equal to an $\epsilon_k$ in the
Figure 2  First Brillouin zone of a square lattice of side $a$. The wavevector $k'$ can be carried into the first zone by forming $k' + G$. The wavevector at a point $A$ on the zone boundary is carried by $G$ to the point $A'$ on the opposite boundary of the same zone. Shall we count both $A$ and $A'$ as lying in the first zone? Because they can be connected by a reciprocal lattice vector, we count them as one identical point in the zone.

Figure 3  Energy-wavevector relation $\varepsilon_k = \hbar^2 k^2/2m$ for free electrons as drawn in the reduced zone scheme. This construction often gives a useful idea of the overall appearance of the band structure of a crystal. The branch $AC$ if displaced by $-2\pi/a$ gives the usual free electron curve for negative $k$, as suggested by the dashed curve. The branch $A'C$ if displaced by $2\pi/a$ gives the usual curve for positive $k$. A crystal potential $U(x)$ will introduce band gaps at the edges of the zone (as at $A$ and $A'$) and at the center of the zone (as at $C$). The point $C$ when viewed in the extended zone scheme falls at the edges of the second zone. The overall width and gross features of the band structure are often indicated properly by such free electron bands in the reduced zone scheme.
first zone, where \( \mathbf{k} = \mathbf{k}' + \mathbf{G} \). Thus we need solve for the energy only in the first Brillouin zone, for each band. An energy band is a single branch of the \( \epsilon_k \) versus \( k \) surface.

In the reduced zone scheme we should not be surprised to find different energies at the same value of the wavevector. Each different energy characterizes a different band.

Two wavefunctions at the same \( k \) but of different energies will be independent of each other: the wavefunctions will be made up of different combinations of the plane wave components \( \exp[i(k + G) \cdot r] \) in the expansion of (7.29). Because the values of the coefficients \( C(k + G) \) will differ for the different bands we should add a symbol, say \( n \), to the \( C \)'s to serve as a band index: \( C_n(k + G) \). Thus the Bloch function for a state of wavevector \( k \) in the band \( n \) can be written as

\[
\psi_{n,k} = \exp(i \mathbf{k} \cdot \mathbf{r}) u_{n,k}(\mathbf{r}) = \sum_{\mathbf{G}} C_n(k + G) \exp[i(k + G) \cdot \mathbf{r}].
\]

**Periodic Zone Scheme**

We can repeat a given Brillouin zone periodically through all of wavevector space. To repeat a zone, we translate the zone by a reciprocal lattice vector. If we can translate a band from other zones into the first zone, we can translate a band in the first zone into every other zone. In this scheme the energy \( \epsilon_k \) of a band is a periodic function in the reciprocal lattice:

\[
\epsilon_k = \epsilon_{k+G}.
\]  

(2)

Here \( \epsilon_{k+G} \) is understood to refer to the same energy band as \( \epsilon_k \).

The result of this construction is known as the **periodic zone scheme**. The periodic property of the energy also can be seen easily from the central equation (7.27).

Consider for example an energy band of a simple cubic lattice as calculated in the tight-binding approximation in (13) below:

\[
\epsilon_k = -\alpha - 2\gamma (\cos k_x a + \cos k_y a + \cos k_z a),
\]

(3)

where \( \alpha \) and \( \gamma \) are constants. A reciprocal lattice vector of the sc lattice is \( \mathbf{G} = (2\pi/a)\mathbf{t} \); if we add this vector to \( k \) the only change in (3) is

\[
\cos k_x a \rightarrow \cos (k_x + 2\pi/a) a = \cos (k_x a + 2\pi),
\]

but this is identically equal to \( \cos k_x a \). The energy is unchanged when the wavevector is increased by a reciprocal lattice vector, so that the energy is a periodic function of the wavevector.
Three different zone schemes are useful (Fig. 4):

- The extended zone scheme in which different bands are drawn in different zones in wavevector space.
- The reduced zone scheme in which all bands are drawn in the first Brillouin zone.
- The periodic zone scheme in which every band is drawn in every zone.

**CONSTRUCTION OF FERMI SURFACES**

We consider in Fig. 5 the analysis for a square lattice. The equation of the zone boundaries is \(2k \cdot G + G^2 = 0\) and is satisfied if \(k\) terminates on the plane normal to \(G\) at the midpoint of \(G\). The first Brillouin zone of the square lattice is
the area enclosed by the perpendicular bisectors of \( \mathbf{G}_1 \) and of the three reciprocal lattice vectors equivalent by symmetry to \( \mathbf{G}_1 \) in Fig. 5a. These four reciprocal lattice vectors are \( \pm (2\pi/a)\hat{k}_x \) and \( \pm (2\pi/a)\hat{k}_y \).

The second zone is constructed from \( \mathbf{G}_2 \) and the three vectors equivalent to it by symmetry, and similarly for the third zone. The pieces of the second and third zones are drawn in Fig. 5b.

To determine the boundaries of some zones we have to consider sets of several nonequivalent reciprocal lattice vectors. Thus the boundaries of section 3 of the third zone are formed from the perpendicular bisectors of three \( \mathbf{G}'s \), namely \((2\pi/a)\hat{k}_x; (4\pi/a)\hat{k}_y; \) and \((2\pi/a)(\hat{k}_x + \hat{k}_y)\).

The free electron Fermi surface for an arbitrary electron concentration is shown in Fig. 6. It is inconvenient to have sections of the Fermi surface that belong to the same zone appear detached from one another. The detachment can be repaired by a transformation to the reduced zone scheme.

We take the triangle labeled 2a and move it by a reciprocal lattice vector \( G = -(2\pi/a)\hat{k}_x \) such that the triangle reappears in the area of the first Brillouin zone (Fig. 7). Other reciprocal lattice vectors will shift the triangles 2b, 2c, 2d to other parts of the first zone, completing the mapping of the second zone into the reduced zone scheme. The parts of the Fermi surface falling in the second zone are now connected, as shown in Fig. 8.
Figure 6 Brillouin zones of a square lattice in two dimensions. The circle shown is a surface of constant energy for free electrons, it will be the Fermi surface for some particular value of the electron concentration. The total area of the filled region in k space depends only on the electron concentration and is independent of the interaction of the electrons with the lattice. The shape of the Fermi surface depends on the lattice interaction, and the shape will not be an exact circle in an actual lattice. The labels within the sections of the second and third zones refer to Fig. 7.

Figure 7 Mapping of the first, second, and third Brillouin zones in the reduced zone scheme. The sections of the second zone in Fig. 6 are put together into a square by translation through an appropriate reciprocal lattice vector. A different G is needed for each piece of a zone.

Figure 8 The free electron Fermi surface of Fig. 6, as viewed in the reduced zone scheme. The shaded areas represent occupied electron states. Parts of the Fermi surface fall in the second, third, and fourth zones. The fourth zone is not shown. The first zone is entirely occupied.

A third zone is assembled into a square in Fig. 8, but the parts of the Fermi surface still appear disconnected. When we look at it in the periodic zone scheme (Fig. 9), the Fermi surface forms a lattice of rosettes.

**Nearly Free Electrons**

How do we go from Fermi surfaces for free electrons to Fermi surfaces for nearly free electrons? We can make approximate constructions freehand by the use of four facts:
The interaction of the electron with the periodic potential of the crystal causes energy gaps at the zone boundaries.

Almost always the Fermi surface will intersect zone boundaries perpendicularly (see below).

The crystal potential will round out sharp corners in the Fermi surfaces.

The total volume enclosed by the Fermi surface depends only on the electron concentration and is independent of the details of the lattice interaction.

We cannot make quantitative statements without calculation, but qualitatively we expect the Fermi surfaces in the second and third zones of Fig. 8 to be changed as shown in Fig. 10.
Freehand impressions of the Fermi surfaces derived from free electron surfaces are useful. Fermi surfaces for free electrons are constructed by a procedure credited to Harrison, Fig. 11. The reciprocal lattice points are determined, and a free-electron sphere of radius appropriate to the electron concentration is drawn around each point. Any point in k space that lies within at least one sphere corresponds to an occupied state in the first zone. Points within at least two spheres correspond to occupied states in the second zone, and similarly for points in three or more spheres.

We said earlier that the alkali metals are the simplest metals, with weak interactions between the conduction electrons and the lattice. Because the alkalis have only one valence electron per atom, the first Brillouin zone boundaries are distant from the approximately spherical Fermi surface that fills one-half of the volume of the zone. It is known by calculation and experiment that the Fermi surface of Na is closely spherical, and that for Cs the Fermi surface is deformed by perhaps 10 percent from a sphere.

The divalent metals Be and Mg also have weak lattice interactions and nearly spherical Fermi surfaces. But because they have two valence electrons each, the Fermi surface encloses twice the volume in k space as for the alkalis. That is, the volume enclosed by the Fermi surface is exactly equal to that of a zone, but because the surface is spherical it extends out of the first zone and into the second zone.

**ELECTRON ORBITS, HOLE ORBITS, AND OPEN ORBITS**

We saw in Eq. (8.7) that electrons in a static magnetic field move on a curve of constant energy on a plane normal to \( \mathbf{B} \). An electron on the Fermi surface will move in a curve on the Fermi surface, because this is a surface of constant energy. Three types of orbits in a magnetic field are shown in Fig. 12.

The closed orbits of (a) and (b) are traversed in opposite senses. Because particles of opposite charge circulate in a magnetic field in opposite senses, we say that one orbit is electronlike and the other orbit is holelike. Electrons in holelike orbits move in a magnetic field as if endowed with a positive charge. This is consistent with the treatment of holes in Chapter 8.

In (c) the orbit is not closed: the particle on reaching the zone boundary at \( A \) is instantly umklapped back to \( B \), where \( B \) is equivalent to \( B' \) because they are connected by a reciprocal lattice vector. Such an orbit is called an open orbit. Open orbits have an important effect on the magnetoresistance.

Vacant orbitals near the top of an otherwise filled band give rise to holelike orbits, as in Figs. 13 and 14. A view of a possible energy surface in three dimensions is given in Fig. 15.

Orbits that enclose filled states are electron orbits. Orbits that enclose empty states are hole orbits. Orbits that move from zone to zone without closing are open orbits.
Figure 11 Harrison construction of free electron Fermi surfaces on the second, third, and fourth zones for a square lattice. The Fermi surface encloses the entire first zone, which therefore is filled with electrons.

Figure 12 Motion in a magnetic field of the wavevector of an electron on the Fermi surface, in (a) and (b) for Fermi surfaces topologically equivalent to those of Fig. 10. In (a) the wavevector moves around the orbit in a clockwise direction; in (b) the wavevector moves around the orbit in a counterclockwise direction. The direction in (b) is what we expect for a free electron of charge $-e$: the smaller $k$ values have the lower energy, so that the filled electron states lie inside the Fermi surface. We call the orbit in (b) electronlike. The sense of the motion in a magnetic field is opposite in (a) to that in (b), so that we refer to the orbit in (a) as holelike. A hole moves as a particle of positive charge $e$. In (c) for a rectangular zone we show the motion on an open orbit in the periodic zone scheme. This is topologically intermediate between a hole orbit and an electron orbit.

Figure 13 (a) Vacant states at the corners of an almost-filled band, drawn in the reduced zone scheme. (b) In the periodic zone scheme the various parts of the Fermi surface are connected. Each circle forms a holelike orbit. The different circles are entirely equivalent to each other, and the density of states is that of a single circle. (The orbits need not be true circles: for the lattice shown it is only required that the orbits have fourfold symmetry.)
CALCULATION OF ENERGY BANDS

Few masters of energy band calculation learned their methods entirely from books. Band calculation is a craft learned by experience, often developed in groups, and needing access to computers. Wigner and Seitz, who performed the first serious band calculations in 1933, refer to afternoons spent on the manual desk calculators of those days, using one afternoon for a trial wavefunction. Modern computers have eased the pain. However, the formulation of the problem requires great care, and the computer programs are not trivial.

Here we limit ourselves to three methods useful to beginners: the tight-binding method, useful for interpolation; the Wigner-Seitz method, useful for the visualization and understanding of the alkali metals; and the pseudopotential method, utilizing the general theory of Chapter 7, which shows the simplicity of many problems. Reviews of these and other methods are cited at the end of this chapter.
Figure 16  (a) Schematic drawing of wavefunctions of electrons on two hydrogen atoms at large separation. (b) Ground state wavefunction at closer separation. (c) Excited state wavefunction.

**Tight Binding Method for Energy Bands**

Let us start with neutral separated atoms and watch the changes in the atomic energy levels as the charge distributions of adjacent atoms overlap when the atoms are brought together to form a crystal. Consider two hydrogen atoms, each with an electron in the 1s ground state. The wavefunctions $\psi_A$, $\psi_B$ on the separated atoms are shown in Fig. 16a.

As the atoms are brought together, their wavefunctions overlap. We consider the two combinations $\psi_A \pm \psi_B$. Each combination shares an electron with the two protons, but an electron in the state $\psi_A + \psi_B$ will have a somewhat lower energy than in the state $\psi_A - \psi_B$.

In $\psi_A + \psi_B$ the electron spends part of the time in the region midway between the two protons, and in this region it is in the attractive potential of both protons at once, thereby increasing the binding energy. In $\psi_A - \psi_B$ the probability density vanishes midway between the nuclei; an extra binding does not appear.

As two atoms are brought together, two separated energy levels are formed for each level of the isolated atom. For $N$ atoms, $N$ orbitals are formed for each orbital of the isolated atom (Fig. 17).

As free atoms are brought together, the coulomb interaction between the atom cores and the electron splits the energy levels, spreading them into bands. Each state of given quantum number of the free atom is spread in the crystal into a band of energies. The width of the band is proportional to the strength of the overlap interaction between neighboring atoms.

There will be bands formed from $p$, $d$, $\ldots$ states ($l = 1, 2, \ldots$) of the free atoms. States degenerate in the free atom will form different bands. Each will not have the same energy as any other band over any substantial range of the wavevector. Bands may coincide in energy at certain values of $k$ in the Brillouin zone.
The approximation that starts out from the wavefunctions of the free atoms is known as the tight binding approximation or the LCAO (linear combination of atomic orbitals) approximation. The approximation is quite good for the inner electrons of atoms, but it is not often a good description of the conduction electrons themselves. It is used to describe approximately the $d$ bands of the transition metals and the valence bands of diamondlike and inert gas crystals.

Suppose that the ground state of an electron moving in the potential $U(r)$ of an isolated atom is $\varphi(r)$, an $s$ state. The treatment of bands arising from degenerate ($p, d, \ldots$) atomic levels is more complicated. If the influence of one atom on another is small, we obtain an approximate wavefunction for one electron in the whole crystal by taking

$$\psi(r) = \sum_j C_{kj} \varphi(r - r_j) ,$$

where the sum is over all lattice points. We assume the primitive basis contains one atom. This function is of the Bloch form (7.7) if $C_{kj} = N^{-1/2} e^{ik \cdot r_j}$, which gives, for a crystal of $N$ atoms,

$$\psi(r) = N^{-1/2} \sum_j \exp(ik \cdot r_j) \varphi(r - r_j) .$$

We prove (5) is of the Bloch form. Consider a translation $T$ connecting two lattice points:
\[ \psi_k(r + \mathbf{T}) = N^{-1/2} \sum_j \exp(ik \cdot r_j)\varphi(r + \mathbf{T} - r_j) = \exp(ik \cdot \mathbf{T}) N^{-1/2} \sum_j \exp[ik \cdot (r_j - \mathbf{T})] \varphi[r - (r_j - \mathbf{T})] = \exp(ik \cdot \mathbf{T})\psi_k(r), \]

exactly the Bloch condition.

We find the first-order energy by calculating the diagonal matrix elements of the Hamiltonian of the crystal:

\[ \langle k | H | k \rangle = N^{-1} \sum_j \sum_m \exp[ik \cdot (r_j - r_m)] \langle \varphi_m | H | \varphi_j \rangle, \]

where \( \varphi_m = \varphi(r - r_m) \). Writing \( \rho_m = r_m - r_j \),

\[ \langle k | H | k \rangle = \sum_m \exp(-ik \cdot \rho_m) \int dV \varphi^*(r - \rho_m)H\varphi(r). \]

We now neglect all integrals in (8) except those on the same atom and those between nearest neighbors connected by \( \rho \). We write

\[ \int dV \varphi^*(r)H\varphi(r) = -\alpha; \quad \int dV \varphi^*(r - \rho)H\varphi(r) = -\gamma; \]

and we have the first-order energy, provided \( \langle k | k \rangle = 1 \):

\[ \langle k | H | k \rangle = -\alpha - \gamma \sum_m \exp(-ik \cdot \rho_m) = \epsilon_k. \]

The dependence of the overlap energy \( \gamma \) on the interatomic separation \( \rho \) can be evaluated explicitly for two hydrogen atoms in 1s states. In rydberg energy units, \( \text{Ry} = \frac{m_e^4}{2\hbar^2} \), we have

\[ \gamma(\text{Ry}) = 2(1 + \rho/a_0) \exp(-\rho/a_0), \]

where \( a_0 = \frac{\hbar^2}{me^2} \). The overlap energy decreases exponentially with the separation.

For a simple cubic structure the nearest-neighbor atoms are at

\[ \rho_m = (\pm a, 0, 0); \quad (0, \pm a, 0); \quad (0, 0, \pm a), \]

so that (10) becomes

\[ \epsilon_k = -\alpha - 2\gamma(\cos k_x a + \cos k_y a + \cos k_z a). \]

Thus the energies are confined to a band of width \( 12\gamma \). The weaker the overlap, the narrower is the energy band. A constant energy surface is shown in Fig. 15.

For \( ka \ll 1 \), \( \epsilon_k = -\alpha - 6\gamma + \gamma k^2 a^2 \). The effective mass is \( m^* = \frac{\hbar^2}{2\gamma a^2} \). When
the overlap integral $\gamma$ is small, the band is narrow and the effective mass is high.

We considered one orbital of each free atom and obtained one band $\varepsilon_k$. The number of orbitals in the band that corresponds to a nondegenerate atomic level is $2N$, for $N$ atoms. We see this directly: values of $k$ within the first Brillouin zone define independent wavefunctions. The simple cubic zone has $-\pi/a < k_x < \pi/a$, etc. The zone volume is $8\pi^3a^3$. The number of orbitals (counting both spin orientations) per unit volume of $k$ space is $V/4\pi^3$, so that the number of orbitals is $2V/a^3$. Here $V$ is the volume of the crystal, and $1/a^3$ is the number of atoms per unit volume. Thus there are $2N$ orbitals.

For the bcc structure with eight nearest neighbors,

$$\varepsilon_k = -\alpha - 8\gamma \cos \frac{1}{2}k_xa \cos \frac{1}{2}k_ya \cos \frac{1}{2}k_za .$$  \hspace{1cm} (14)

For the fcc structure with 12 nearest neighbors,

$$\varepsilon_k = -\alpha - 4\gamma (\cos \frac{1}{2}k_ya \cos \frac{1}{2}k_za + \cos \frac{1}{2}k_xa \cos \frac{1}{2}k_ya + \cos \frac{1}{2}k_za \cos \frac{1}{2}k_xa) .$$ \hspace{1cm} (15)

A constant energy surface is shown in Fig. 18.

**Wigner-Seitz Method**

Wigner and Seitz showed that at least for the alkali metals there is no inconsistency between the electron wavefunctions of free atoms and the nearly free electron model of the band structure of a crystal. Over most of a band the energy may depend on the wavevector nearly as for a free electron. However, the Bloch wavefunction, unlike a plane wave, will pile up charge on the positive ion cores as in the atomic wavefunction.

A Bloch function satisfies the wave equation

$$\left(\frac{1}{2m}p^2 + U(r)\right) e^{ikr}u_k(r) = \varepsilon_k e^{ikr}u_k(r) .$$ \hspace{1cm} (16)

With $p = -i\hbar$ grad, we have

$$p e^{ikr}u_k(r) = \hbar k e^{ikr}u_k(r) + e^{ikr}pu_k(r) ;$$
$$p^2 e^{ikr}u_k(r) = (\hbar k)^2 e^{ikr}u_k(r) + e^{ikr} (2\hbar k \cdot p)u_k(r) + e^{ikr}p^2u_k(r) ;$$

thus (16) may be written as an equation for $u_k$:

$$\left(\frac{1}{2m}(p + \hbar k)^2 + U(r)\right) u_k(r) = \varepsilon_k u_k(r) .$$ \hspace{1cm} (17)

At $k = 0$ we have $\psi_0 = u_0(r)$, where $u_0(r)$ has the periodicity of the lattice, sees the ion cores, and near them will look like the wavefunction of the free atom.

It is much easier to find a solution at $k = 0$ than for a general $k$, because at $k = 0$ a nondegenerate solution will have the full symmetry of the crystal. We
can then use \( u_0(r) \) to construct the approximate solution

\[
\psi_k = \exp(ik \cdot r)u_0(r) .
\] (18)

This is of the Bloch form, but \( u_0 \) is not an exact solution of (17): it is a solution only if we drop the term in \( k \cdot p \). Often this term is treated as a perturbation, as in Problem 8. The \( k \cdot p \) perturbation theory developed there is especially useful in finding the effective mass \( m^* \) at a band edge.

Because it takes account of the ion core potential the function (18) is a much better approximation than a plane wave to the correct wavefunction. The energy of the approximate solution depends on \( k \) as \( (\hbar k)^2/2m \), exactly as for the plane wave, even though the modulation represented by \( u_0(r) \) may be very strong. Because \( u_0 \) is a solution of

\[
\left( \frac{1}{2m} p^2 + U(r) \right) u_0(r) = \epsilon_0 u_0(r) ,
\] (19)

the function (18) has the energy expectation value \( \epsilon_0 + (\hbar^2 k^2/2m) \). The function \( u_0(r) \) often will give us a good picture of the charge distribution within a cell.

Wigner and Seitz developed a simple and fairly accurate method of calculating \( u_0(r) \). Figure 19 shows the Wigner-Seitz wavefunction for \( k = 0 \) in the 3s conduction band of metallic sodium. The function is practically constant over 0.9 of the atomic volume. To the extent that the solutions for higher \( k \) may be approximated by \( \exp(ik \cdot r)u_0(r) \), the wavefunctions in the conduction band will be similar to plane waves over most of the atomic volume, but increase markedly and oscillate within the ion core.

**Cohesive Energy.** The stability of the simple metals with respect to free atoms is caused by the lowering of the energy of the Bloch orbital with \( k = 0 \) in the crystal compared to the ground valence orbital of the free atom. The effect is illustrated in Fig. 19 for sodium and in Fig. 20 for a linear periodic potential
Figure 19  Radial wavefunctions for the 3s orbital of free sodium atom and for 3s conduction band in sodium metal. The wavefunctions, which are not normalized here, are found by integrating the Schrödinger equation for an electron in the potential well of an Na$^+$ ion core. For the free atom the wavefunction is integrated subject to the usual Schrödinger boundary condition $\psi(r) \to 0$ as $r \to \infty$; the energy eigenvalue is $-5.15$ eV. The wavefunction for wavevector $k = 0$ in the metal is subject to the Wigner-Seitz boundary condition that $d\psi/dr = 0$ when $r$ is midway between neighboring atoms; the energy of this orbital is $-8.2$ eV, considerably lower than for the free atom. The orbitals at the zone boundary are not filled in sodium; their energy is $+2.7$ eV. (After E. Wigner and F. Seitz.)

Figure 20  Ground orbital ($k = 0$) energy for an electron in a periodic square well potential of depth $|U_0| = 2\hbar^2/m a^2$. The energy is lowered as the wells come closer together. Here $a$ is held constant and $b$ is varied. Large $b/a$ corresponds to separated atoms. (Courtesy of C. Y. Fong.)
of attractive square wells. The ground orbital energy is much lower at the actual spacing in the metal than for isolated atoms.

A decrease in ground orbital energy will increase the binding. The decrease in ground orbital energy is a consequence of the change in the boundary condition on the wavefunction: The Schrödinger boundary condition for the free atom is \( \psi(r) \to 0 \) as \( r \to \infty \). In the crystal the \( k = 0 \) wavefunction \( u_0(r) \) has the symmetry of the lattice and is symmetric about \( r = 0 \). To have this, the normal derivative of \( \psi \) must vanish across every plane midway between adjacent atoms.

In a spherical approximation to the shape of the smallest Wigner-Seitz cell we use the Wigner-Seitz boundary condition

\[
(d\psi/dr)_{r_0} = 0 ,
\]

(20)

where \( r_0 \) is the radius of a sphere equal in volume to a primitive cell of the lattice. In sodium, \( r_0 = 3.95 \) Bohr units, or 2.08 Å; the half-distance to a nearest neighbor is 1.86 Å. The spherical approximation is not bad for fcc and bcc structures. The boundary condition allows the ground orbital wavefunction to have much less curvature than the free atom boundary condition. Much less curvature means much less kinetic energy.

In sodium the other filled orbitals in the conduction band can be represented in a rough approximation by wavefunctions of the form (18), with

\[
\psi_k = e^{ikr}u_0(r) ; \quad \epsilon_k = \epsilon_0 + \frac{\hbar^2k^2}{2m} .
\]

The Fermi energy is 3.1 eV, from Table 6.1. The average kinetic energy per electron is 0.6 of the Fermi energy, or 1.9 eV. Because \( \epsilon_0 = -8.2 \) eV at \( k = 0 \), the average electron energy is \( \langle \epsilon_k \rangle = -8.2 + 1.9 = -6.3 \) eV, compared with -5.15 eV for the valence electron of the free atom, Fig. 21.

We therefore estimate that sodium metal is stable by about 1.1 eV with respect to the free atom. This result agrees well with the experimental value.
1.13 eV. We have neglected several corrections whose overall effect in sodium is small.

**Pseudopotential Methods**

Conduction electron wavefunctions are usually smoothly varying in the region between the ion cores, but have a complicated nodal structure in the region of the cores. This behavior is illustrated by the ground orbital of sodium, Fig. 19. It is helpful to view the nodes in the conduction electron wavefunction in the core region as created by the requirement that the function be orthogonal to the wavefunctions of the core electrons. This all comes out of the Schrödinger equation, but we can see that we need the flexibility of two nodes in the 3s conduction orbital of Na in order to be orthogonal both to the 1s core orbital with no nodes and the 2s core orbital with one node.

Outside the core the potential energy that acts on the conduction electron is relatively weak: the potential energy is only the coulomb potential of the singly-charged positive ion cores and is reduced markedly by the electrostatic screening of the other conduction electrons, Chapter 10. In this outer region the conduction electron wavefunctions are as smoothly varying as plane waves.

If the conduction orbitals in this outer region are approximately plane waves, the energy must depend on the wavevector approximately as $\epsilon_k = \hbar^2 k^2 / 2m$ as for free electrons. But how do we treat the conduction orbitals in the core region where the orbitals are not at all like plane waves?

What goes on in the core is largely irrelevant to the dependence of $\epsilon$ on $k$. Recall that we can calculate the energy by applying the hamiltonian operator to an orbital at any point in space. Applied in the outer region, this operation will give an energy nearly equal to the free electron energy.

This argument leads naturally to the idea that we might replace the actual potential energy (and filled shells) in the core region by an effective potential energy¹ that gives the same wavefunctions outside the core as are given by the actual ion cores. It is startling to find that the effective potential or pseudopotential that satisfies this requirement is nearly zero. This conclusion about pseudopotentials is supported by a large amount of empirical experience as well as by theoretical arguments. The result is referred to as the cancellation theorem.

The pseudopotential for a problem is not unique nor exact, but it may be very good. On the Empty Core Model (ECM) we can even take the unscreened pseudopotential to be zero inside some radius $R_e$:

$$U(r) = \begin{cases} 0 & \text{for } r < R_e; \\ -e^2/r & \text{for } r > R_e. \end{cases} \quad (21)$$

This potential should now be screened as described in Chapter 10. Each component $U(K)$ of $U(r)$ is to be divided by the dielectric constant $\varepsilon(K)$ of the electron gas. If, just as an example, we use the Thomas-Fermi dielectric function (10.33), we obtain the screened pseudopotential plotted in Fig. 22a.

The pseudopotential as drawn is much weaker than the true potential, but the pseudopotential was adjusted so that the wave function in the outer region is nearly identical to that for the true potential. In the language of scattering theory, we adjust the phase shifts of the pseudopotential to match those of the true potential.

Calculation of the band structure depends only on the Fourier components of the pseudopotential at the reciprocal lattice vectors. Usually only a few values of the coefficients $U(G)$ are needed to get a good band structure: see the $U(G)$ in Fig. 22b. These coefficients are sometimes calculated from model potentials, and sometimes they are obtained from fits of tentative band structures to the results of optical measurements. Good values of $U(0)$ can be estimated from first principles; it is shown in (10.43) that for a screened Coulomb potential $U(0) = -\frac{3}{2}e_F$.

In the remarkably successful Empirical Pseudopotential Method (EPM) the band structure is calculated using a few coefficients $U(G)$ deduced from theoretical fits to measurements of the optical reflectance and absorption of crystals, as discussed in Chapter 11. Tables of values of $U(G)$ are given in the review by M. L. Cohen and V. Heine.

Charge density maps can be plotted from the wavefunctions generated by the EPM—see Fig. 3.11. The results are in excellent agreement with x-ray diffraction determinations; such maps give an understanding of the bonding and have great predictive value for proposed new structures and compounds.

The EPM values of the coefficients $U(G)$ often are additive in the contributions of the several types of ions that are present. Thus it may be possible to construct the $U(G)$ for entirely new structures, starting from results on known structures. Further, the pressure dependence of a band structure may be determined when it is possible to estimate from the form of the $U(r)$ curve the dependence of $U(G)$ on small variations of $G$.

It is often possible to calculate band structures, cohesive energy, lattice constants, and bulk moduli from first principles. In such *ab initio* pseudopotential calculations the basic inputs are the crystal structure type and the atomic number, along with well-tested theoretical approximations to exchange energy
Figure 22a  Pseudopotential for metallic sodium, based on the empty core model and screened by the Thomas-Fermi dielectric function. The calculations were made for an empty core radius $R_e = 1.66a_0$, where $a_0$ is the Bohr radius, and for a screening parameter $k_e a_0 = 0.79$. The dashed curve shows the assumed unscreened potential, as from (21). The dotted curve is the actual potential of the ion core; other values of $U(r)$ are $-50.4$, $-11.6$, and $-4.6$, for $r = 0.15$, 0.4, and 0.7, respectively. Thus the actual potential of the ion (chosen to fit the energy levels of the free atom) is very much larger than the pseudopotential, over 200 times larger at $r = 0.15$.

Figure 22b  A typical reciprocal space pseudopotential. Values of $U(k)$ for wavevectors equal to the reciprocal lattice vectors, $G$, are indicated by the dots. For very small $k$ the potential approaches $(-2/3)$ times the Fermi energy, which is the screened-ion limit for metals. This limit is derived in Chapter 10 (After M. L. Cohen.)
terms. This is not the same as calculating from atomic number alone, but it is the most reasonable basis for a first-principles calculation. The results of M. T. Yin and M. L. Cohen, Phys. Rev. B 26, 5668 (1982), are compared with experiment in the table that follows.

<table>
<thead>
<tr>
<th></th>
<th>Lattice constant (Å)</th>
<th>Cohesive energy (eV)</th>
<th>Bulk modulus (Mbar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silicon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculated</td>
<td>5.45</td>
<td>4.84</td>
<td>0.96</td>
</tr>
<tr>
<td>Experimental</td>
<td>5.43</td>
<td>4.63</td>
<td>0.99</td>
</tr>
<tr>
<td>Germanium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculated</td>
<td>5.66</td>
<td>4.26</td>
<td>0.73</td>
</tr>
<tr>
<td>Experimental</td>
<td>5.65</td>
<td>3.85</td>
<td>0.77</td>
</tr>
<tr>
<td>Diamond</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculated</td>
<td>3.60</td>
<td>8.10</td>
<td>4.33</td>
</tr>
<tr>
<td>Experimental</td>
<td>3.57</td>
<td>7.35</td>
<td>4.43</td>
</tr>
</tbody>
</table>

**EXPERIMENTAL METHODS IN FERMI SURFACE STUDIES**

Powerful experimental methods have been developed for the determination of Fermi surfaces. The methods include magnetoresistance, anomalous skin effect, cyclotron resonance, magneto-acoustic geometric effects, the Shubnikow-de Haas effect, and the de Haas-van Alphen effect. Further information on the momentum distribution is given by positron annihilation, Compton scattering, and the Kohn effect.

We propose to study one method rather thoroughly. All the methods are useful, but all need detailed theoretical analysis and are not for beginners. We select the de Haas-van Alphen effect because it exhibits very well the characteristic periodicity in $1/B$ of the properties of a metal in a uniform magnetic field.

**Quantization of Orbits in a Magnetic Field**

The momentum $p$ of a particle in a magnetic field is the sum (Appendix G) of two parts, the kinetic momentum $p_{\text{kin}} = mv = \hbar k$ and the potential momentum or field momentum $p_{\text{field}} = qA/c$, where $q$ is the charge. The vector potential is related to the magnetic field by $B = \text{curl } A$. The total momentum is

\[
\text{CGS} \quad p = p_{\text{kin}} + p_{\text{field}} = \hbar k + qA/c .
\]

In SI the factor $c^{-1}$ is omitted.
Following the semiclassical approach of Onsager and Ishitz, we assume that the orbits in a magnetic field are quantized by the Bohr-Sommerfeld relation
\[ \oint p \cdot dr = (n + \gamma)2\pi\hbar , \] (23)
when \( n \) is an integer and \( \gamma \) is a phase correction that for free electrons has the value \( \frac{1}{2} \). Then
\[ \oint p \cdot dr = \oint \hbar k \cdot dr + \frac{q}{c} \oint A \cdot dr . \] (24)

The equation of motion of a particle of charge \( q \) in a magnetic field is
\[ \hbar \frac{dk}{dt} = \frac{q}{c} \frac{dr}{dt} \times B . \] (25a)

We integrate with respect to time to give
\[ \hbar k = \frac{q}{c} r \times B , \]

apart from an additive constant which does not contribute to the final result.

Thus one of the path integrals in (24) is
\[ \oint \hbar k \cdot dr = \frac{q}{c} \oint r \times B \cdot dr = -\frac{q}{c} B \cdot \oint r \times dr = -\frac{2q}{c} \Phi , \] (25b)
where \( \Phi \) is the magnetic flux contained within the orbit in real space. We have used the geometrical result that
\[ \oint r \times dr = 2 \times (\text{area enclosed by the orbit}) . \]

The other path integral in (24) is
\[ \frac{q}{c} \oint A \cdot dr = \frac{q}{c} \oint \text{curl} A \cdot d\sigma = \frac{q}{c} \oint B \cdot d\sigma = \frac{q}{c} \Phi , \] (25c)
by the Stokes theorem; here \( d\sigma \) is the area element in real space. The momentum path integral is the sum of (25b) and (25c):
\[ \oint p \cdot dr = -\frac{q}{c} \Phi = (n + \gamma)2\pi\hbar . \] (26)

It follows that the orbit of an electron is quantized in such a way that the flux through it is
\[ \Phi_n = (n + \gamma)(2\pi\hbar c)e . \] (27)

The flux unit \( 2\pi\hbar c/e = 4.14 \times 10^{-7} \) gauss cm\(^2\) or T m\(^2\).

In the de Haas-van Alphen effect discussed below we need the area of the orbit in wavevector space. We obtained in (27) the flux through the orbit in real space. By (25a) we know that a line element \( \Delta r \) in the plane normal to \( B \) is
related to $\Delta k$ by $\Delta \left( \frac{\hbar c}{eB} \right) \Delta k$, so that the area $S_n$ in $k$ space is related to the area $A_n$ of the orbit in $r$ space by

$$A_n = \left( \frac{\hbar c}{eB} \right)^2 S_n.$$  

(28)

It follows that

$$\Phi_n = \left( \frac{\hbar c}{e} \right)^2 \frac{1}{B} S_n = (n + \gamma) \frac{2\pi\hbar c}{e},$$

(29)

from (27), whence the area of an orbit in $k$ space will satisfy

$$S_n = (n + \gamma) \frac{2\pi e}{\hbar c} B.$$  

(30)

In Fermi surface experiments we may be interested in the increment $\Delta B$ for which two successive orbits, $n$ and $n + 1$, have the same area in $k$ space on the Fermi surface. The areas are equal when

$$S \left( \frac{1}{B_{n+1}} - \frac{1}{B_n} \right) = \frac{2\pi e}{\hbar c},$$

(31)

from (30). We have the important result that equal increments of $1/B$ reproduce similar orbits—this periodicity in $1/B$ is a striking feature of the magneto-oscillatory effects in metals at low temperatures: resistivity, susceptibility, heat capacity.

The population of orbits on or near the Fermi surface oscillates as $B$ is varied, causing a wide variety of effects. From the period of the oscillation we reconstruct the Fermi surface. The result (30) is independent of the gauge of the vector potential used in the expression (22) for momentum; that is, $p$ is not gauge invariant, but $S_n$ is. Gauge invariance is discussed further in Chapter 12 and in Appendix G.

**De Haas-van Alphen Effect**

The de Haas-van Alphen effect is the oscillation of the magnetic moment of a metal as a function of the static magnetic field intensity. The effect can be observed in pure specimens at low temperatures in strong magnetic fields; we do not want the quantization of the electron orbits to be blurred by collisions, and we do not want the population oscillations to be averaged out by thermal population of adjacent orbits.

The analysis of the dHvA effect is given for absolute zero in Fig. 23. The electron spin is neglected. The treatment is given for a two-dimensional (2D) system; in 3D we need only multiply the 2D wavefunction by plane wave factors $\exp(ik_z z)$, where the magnetic field is parallel to the $z$ axis. The area of an orbit in $k_x$, $k_y$ space is quantized as in (30). The area between successive orbits is

$$\Delta S = S_n - S_{n-1} = \frac{2\pi e B}{\hbar c}.$$  

(32)
Figure 23 Explanation of the de Haas-van Alphen effect for a free electron gas in two dimensions in a magnetic field. The filled orbitals of the Fermi sea in the absence of a magnetic field are shaded in a and d. The energy levels in a magnetic field are shown in b, c, and e. In b the field has a value \( B_1 \) such that the total energy of the electrons is the same as in the absence of a magnetic field: as many electrons have their energy raised as lowered by the orbital quantization in the magnetic field \( B_1 \). When we increase the field to \( B_2 \) the total electron energy is increased, because the uppermost electrons have their energy raised. In c for field \( B_3 \) the energy is again equal to that for the field \( B = 0 \). The total energy is a minimum at points such as \( B_1 \), \( B_3 \), \( B_5 \), \ldots, and a maximum near points such as \( B_2 \), \( B_4 \), \ldots.

The area in \( k \) space occupied by a single orbital is \((2\pi/L)^2\), neglecting spin, for a square specimen of side \( L \). Using (32) we find that the number of free electron orbitals that coalesce in a single magnetic level is

\[
D = (2\pi eB/\hbar c)(L/2\pi)^2 = \rho B ,
\]

where \( \rho = eL^2/2\pi\hbar c \), as in Fig. 24. Such a magnetic level is called a Landau level.

The dependence of the Fermi level on \( B \) is dramatic. For a system of \( N \) electrons at absolute zero the Landau levels are entirely filled up to a magnetic quantum number we identify by \( s \), where \( s \) is a positive integer. Orbitals at the next higher level \( s + 1 \) will be partly filled to the extent needed to accommodate the electrons. The Fermi level will lie in the Landau level \( s + 1 \) if there are electrons in this level; as the magnetic field is increased the electrons move to lower levels. When \( s + 1 \) is vacated, the Fermi level moves down abruptly to the next lower level \( s \).

The electron transfer to lower Landau levels can occur because their degeneracy \( D \) increases as \( B \) is increased, as shown in Fig. 25. As \( B \) is increased there occur values of \( B \) at which the quantum number of the uppermost filled
Figure 24 (a) Allowed electron orbitals in two dimensions in absence of a magnetic field. (b) In a magnetic field the points which represent the orbitals of free electrons may be viewed as restricted to circles in the former $k_x k_y$ plane. The successive circles correspond to successive values of the quantum number $n$ in the energy $(n - \frac{1}{2})k_0$. The area between successive circles is

(CGS) \[ \pi \Delta k^2 = 2 \pi k (\Delta k) = (2\pi n \hbar^2) \Delta \epsilon = 2\pi n \omega_c / \hbar = 2\pi B / \hbar c , \]

The angular position of the points has no significance. The number of orbitals on a circle is constant and is equal to the area between successive circles times the number of orbitals per unit area in (a), or $\left(2\pi B / \hbar c\right) (L/2\pi)^2 = L^2 eB / 2\pi \hbar c$, neglecting electron spin.

Figure 25 (a) The heavy line gives the number of particles in levels which are completely occupied in a magnetic field $B$, for a two-dimensional system with $N = 50$ and $\rho = 0.50$. The shaded area gives the number of particles in levels partially occupied. The value of $s$ denotes the quantum number of the highest level which is completely filled. Thus at $B = 40$ we have $s = 2$: the levels $n = 1$ and $n = 2$ are filled and there are 10 particles in the level $n = 3$. At $B = 50$ the level $n = 3$ is empty. (b) The periodicity in $1/B$ is evident when the same points are plotted against $1/B$.

level decreases abruptly by unity. At the critical magnetic fields labeled $B_s$ no level is partly occupied at absolute zero, so that

$$ spB_s = N . $$

The number of filled levels times the degeneracy at $B_s$ must equal the number of electrons $N$. 

(34)
To show the periodicity of the energy as $B$ is varied, we use the result that the energy of the Landau level of magnetic quantum number $n$ is $E_n = (n - \frac{1}{2})\hbar \omega_c$, where $\omega_c = eB/m^*c$ is the cyclotron frequency. The result for $E_n$ follows from the analogy between the cyclotron resonance orbits and the simple harmonic oscillator, but now we have found it convenient to start counting at $n = 1$ instead of at $n = 0$.

The total energy of the electrons in levels that are fully occupied is

$$\sum_{n=1}^{s} D\hbar \omega_c(n - \frac{1}{2}) = \frac{1}{2}D\hbar \omega_c s^2,$$  (35)

where $D$ is the number of electrons in each level. The total energy of the electrons in the partly occupied level $s + 1$ is

$$\hbar \omega_c(s + \frac{1}{2})(N - sD),$$  (36)

where $sD$ is the number of electrons in the lower filled levels. The total energy of the $N$ electrons is the sum of (35) and (36), as in Fig. 26.

The magnetic moment $\mu$ of a system at absolute zero is given by $\mu = -\partial U/\partial B$. The moment here is an oscillatory function of $1/B$, Fig. 27. This oscillatory magnetic moment of the Fermi gas at low temperatures is the de Haas-van Alphen effect. From (31) we see that the oscillations occur at equal intervals of $1/B$ such that

$$\Delta \left(\frac{1}{B}\right) = \frac{2\pi e}{\hbar c S},$$  (37)

where $S$ is the extremal area (see below) of the Fermi surface normal to the direction of $B$. From measurements of $\Delta(1/B)$, we deduce the corresponding extremal areas $S$; thereby much can be inferred about the shape and size of the Fermi surface.

**Extremal Orbits.** One point in the interpretation of the dHvA effect is subtle. For a Fermi surface of general shape the sections at different values of $k_B$ will have different periods. The response will be the sum of contributions from all sections or all orbits. But the dominant response of the system comes from orbits whose periods are stationary with respect to small changes in $k_B$. Such orbits are called extremal orbits. Thus in Fig. 28 the section $AA'$ dominates the observed cyclotron period.

The argument can be put in mathematical form, but we do not give the proof here (QTS, p. 223; Ziman, p. 322). Essentially it is a question of phase cancellation: the contributions of different nonextremal orbits cancel, but near the extrema the phase varies only slowly and there is a net signal from these orbits. Sharp resonances are obtained even from complicated Fermi surfaces because the experiment selects the extremal orbits.
Figure 26 The upper curve is the total electronic energy versus $1/B$. The oscillations in the energy $U$ may be detected by measurement of the magnetic moment, given by $-\partial U/\partial B$. The thermal and transport properties of the metal also oscillate as successive orbital levels cut through the Fermi level when the field is increased. The shaded region in the figure gives the contribution to the energy from levels that are only partly filled. The parameters for the figure are the same as for Fig. 25, and we have taken the units of $B$ such that $B = h\omega_c$.

Figure 27 At absolute zero the magnetic moment is given by $-\partial U/\partial B$. The energy plotted in Fig. 26 leads to the magnetic moment shown here, an oscillatory function of $1/B$. In impure specimens the oscillations are smudged out in part because the energy levels are no longer sharply defined.
**Fermi Surface of Copper.** The Fermi surface of copper is distinctly non-spherical: eight necks make contact with the hexagonal faces of the first Brillouin zone of the fcc lattice. The electron concentration in a monovalent metal with an fcc structure is \( n = 4/a^3 \); there are four electrons in a cube of volume \( a^3 \). The radius of a free electron Fermi sphere is

\[
k_F = (3\pi^2 n)^{1/3} = (12\pi^2/a^3)^{1/3} \approx (4.90/a),
\]

and the diameter is \( 9.80/a \).

The shortest distance across the Brillouin zone (the distance between hexagonal faces) is \( (2\pi/a)(3)^{1/2} = 10.88/a \), somewhat larger than the diameter of the free electron sphere. The sphere does not touch the zone boundary, but we know that the presence of a zone boundary tends to lower the band energy near the boundary. Thus it is plausible that the Fermi surface should neck out to meet the closest (hexagonal) faces of the zone (Figs. 18 and 29).

The square faces of the zone are more distant, with separation \( 12.57/a \), and the Fermi surface does not neck out to meet these faces.

---

**EXAMPLE: Fermi Surface of Gold.** In gold for quite a wide range of field directions Shoenberg finds the magnetic moment has a period of \( 2 \times 10^{-9} \) gauss\(^{-1} \). This period corresponds to an extremal orbit of area

\[
S = \frac{2\pi e/hc}{\Delta(1/B)} \approx \frac{9.55 \times 10^7}{2 \times 10^{-9}} \approx 4.8 \times 10^{16} \text{ cm}^{-2}.
\]

From Table 6.1, we have \( k_F = 1.2 \times 10^8 \) cm\(^{-1} \) for a free electron Fermi sphere for gold, or an extremal area of \( 4.5 \times 10^{16} \) cm\(^{-2} \), in general agreement with the experimental value. The actual periods reported by Shoenberg are \( 2.05 \times 10^{-9} \) gauss\(^{-1} \) for the orbit \( B_{111} \) of Fig. 23 and \( 1.95 \times 10^{-9} \) gauss\(^{-1} \) for \( B_{100} \). In the [111] direction in Au a large period of \( 6 \times 10^{-8} \) gauss\(^{-1} \) is also found; the corresponding orbital area is \( 1.6 \times 10^{15} \) cm\(^{-2} \). This is the “neck” orbit \( N \). Another extremal orbit, the “dog’s bone,” is shown in Fig. 30; its area in Au is about 0.4 of the belly area. Experimental results are shown in Fig. 31. To do the example in SI, drop \( c \) from the relation for \( S \) and use as the period \( 2 \times 10^{-5} \) tesla\(^{-1} \).
Figure 29  Fermi surface of copper, after Pippard. The Brillouin zone of the fcc structure is the truncated octahedron derived in Chapter 2. The Fermi surface makes contact with the boundary at the center of the hexagonal faces of the zone, in the [111] directions in k space. Two "belly" extremal orbits are shown, denoted by $B$; the extremal "neck" orbit is denoted by $N$.

Figure 30  Dog's bone orbit of an electron on the Fermi surface of copper or gold in a magnetic field. This orbit is classified as holelike because the energy increases toward the interior of the orbit.

Figure 31  De Haas-van Alphen effect in gold with $B \parallel [110]$. The oscillation is from the dog's bone orbit of Fig. 30. The signal is related to the second derivative of the magnetic moment with respect to field. The results were obtained by a field modulation technique in a high-homogeneity superconducting solenoid at about 1.2 K. (Courtesy of I. M. Templeton.)
The free electron Fermi sphere of aluminum fills the first zone entirely and has a large overlap into the second and third zones, Fig. 1. The third zone Fermi surface is quite complicated, even though it is just made up of certain pieces of the surface of the free electron sphere. The free electron model also gives small pockets of holes in the fourth zone, but when the lattice potential is taken into account these empty out, the electrons being added to the third zone. The general features of the predicted Fermi surface of aluminum are quite well verified by experiment. Figure 32 shows part of the free electron Fermi surface of magnesium.

**Magnetic Breakdown.** Electrons in sufficiently high magnetic fields will move in free particle orbits, the circular cyclotron orbits of Fig. 33a. Here the magnetic forces are dominant, and the lattice potential is a slight perturbation. In this limit the classification of the orbitals into bands may have little importance. However, we know that at low magnetic fields the motion is described
by (8.7) with the band structure $\epsilon_k$ that obtains in the absence of a magnetic field.

The eventual breakdown of this description as the magnetic field is increased is called magnetic breakdown.\(^2\) The passage to strong magnetic fields may drastically change the connectivity of the orbits, as in the figure. The onset of magnetic breakdown will be revealed by physical properties such as magneto-resistance that depend sensitively on the connectivity.

The condition for magnetic breakdown is that $\hbar \omega_c \epsilon_F > E_g^2$, approximately. Here $\epsilon_F$ is the free electron Fermi energy and $E_g$ is the energy gap. This condition is much milder, especially in metals with small gaps, than the naive condition that the magnetic splitting $\hbar \omega_c$ exceed the gap.

Small gaps may be found in hcp metals where the gap across the hexagonal face of the zone would be zero except for a small splitting introduced by the spin-orbit interaction. In Mg the splitting is of the order of $10^{-3}$ eV; for this gap and $\epsilon_F \sim 10$ eV the breakdown condition is $\hbar \omega_c > 10^{-5}$ eV, or $B > 1000$ G.

**SUMMARY**

- A Fermi surface is the surface in $k$ space of constant energy equal to $\epsilon_F$. The Fermi surface separates filled states from empty states at absolute zero. The form of the Fermi surface is usually exhibited best in the reduced zone scheme, but the connectivity of the surfaces is clearest in the periodic zone scheme.

- An energy band is a single branch of the $\epsilon_k$ versus $k$ surface.

- The cohesion of simple metals is accounted for by the lowering of energy of the $k = 0$ conduction band orbital when the boundary conditions on the wavefunction are changed from Schrödinger to Wigner-Seitz.

- The periodicity in the de Haas-van Alphen effect measures the extremal cross-section area $S$ in $k$ space of the Fermi surface, the cross section being taken perpendicular to $B$:

\[
\Delta \left( \frac{1}{B} \right) = \frac{2\pi e}{\hbar c S}.
\]

Problems

1. **Brillouin zones of rectangular lattice.** Make a plot of the first two Brillouin zones of a primitive rectangular two-dimensional lattice with axes $a$, $b = 3a$.

2. **Brillouin zone, rectangular lattice.** A two-dimensional metal has one atom of valency one in a simple rectangular primitive cell $a = 2 \text{ Å}; b = 4 \text{ Å}$. (a) Draw the first Brillouin zone. Give its dimensions, in cm$^{-1}$. (b) Calculate the radius of the free electron Fermi sphere, in cm$^{-1}$. (c) Draw this sphere to scale on a drawing of the first Brillouin zone. Make another sketch to show the first few periods of the free electron band in the periodic zone scheme, for both the first and second energy bands. Assume there is a small energy gap at the zone boundary.

3. **Hexagonal close-packed structure.** Consider the first Brillouin zone of a crystal with a simple hexagonal lattice in three dimensions with lattice constants $a$ and $c$. Let $G_e$ denote the shortest reciprocal lattice vector parallel to the $c$ axis of the crystal lattice. (a) Show that for a hexagonal close-packed crystal structure the Fourier component $U(G_e)$ of the crystal potential $U(\mathbf{r})$ is zero. (b) Is $U(2G_e)$ also zero? (c) Why is it possible in principle to obtain an insulator made up of divalent atoms at the lattice points of a simple hexagonal lattice? (d) Why is it not possible to obtain an insulator made up of monovalent atoms in a hexagonal-close-packed structure?

4. **Brillouin zones of two-dimensional divalent metal.** A two-dimensional metal in the form of a square lattice has two conduction electrons per atom. In the almost free electron approximation, sketch carefully the electron and hole energy surfaces. For the electrons choose a zone scheme such that the Fermi surface is shown as closed.

5. **Open orbits.** An open orbit in a monovalent tetragonal metal connects opposite faces of the boundary of a Brillouin zone. The faces are separated by $G = 2 \times 10^9 \text{ cm}^{-1}$. A magnetic field $B = 10^3$ gauss $= 10^{-1}$ tesla is normal to the plane of the open orbit. (a) What is the order of magnitude of the period of the motion in $k$ space? Take $v \approx 10^8$ cm/sec. (b) Describe in real space the motion of an electron on this orbit in the presence of the magnetic field.

6. **Cohesive energy for a square well potential.** (a) Find an expression for the binding energy of an electron in one dimension in a single square well of depth $U_0$ and width $a$. (This is the standard first problem in elementary quantum mechanics.) Assume that the solution is symmetric about the midpoint of the well. (b) Find a numerical result for the binding energy in terms of $U_0$ for the special case $|U_0| = 2\hbar^2ma^2$ and compare with the appropriate limit of Fig. 20. In this limit of widely separated wells the band width goes to zero, so the energy for $k = 0$ is the same as the energy for any other $k$ in the lowest energy band. Other bands are formed from the excited states of the well, in this limit.

7. **De Haas-van Alphen period of potassium.** (a) Calculate the period $\Delta(1/B)$ expected for potassium on the free electron model. (b) What is the area in real space of the extremal orbit, for $B = 10 \text{ kG} = 1 \text{ T}$? The same period applies to oscillations in the electrical resistivity, known as the Shubnikow-de Haas effect.
8. **Band edge structure on** $k \cdot p$ **perturbation theory.** Consider a nondegenerate orbital $\psi_{nk}$ at $k = 0$ in the band $n$ of a cubic crystal. Use second-order perturbation theory to find the result

$$
\epsilon_n(k) = \epsilon_n(0) + \frac{\hbar^2 k^2}{2m} + \frac{\hbar^2}{m^*} \sum' \frac{|(n0|k \cdot p|j0)|^2}{\epsilon_n(0) - \epsilon_j(0)},
$$

(39)

where the sum is over all other orbitals $\psi_{jk}$ at $k = 0$. The effective mass at this point is

$$
\frac{m}{m^*} = 1 + \frac{2}{m} \sum_j \frac{|(n0|p|j0)|^2}{\epsilon_n(0) - \epsilon_j(0)}.
$$

(40)

The mass at the conduction band edge in a narrow gap semiconductor is often dominated by the effect of the valence band edge, whence

$$
\frac{m}{m^*} \approx \frac{2}{mE_g} \sum_v |(v|p|v)|^2,
$$

(41)

where the sum is over the valence bands; $E_g$ is the energy gap. For given matrix elements, small gaps lead to small masses.

9. **Wannier functions.** The Wannier functions of a band are defined in terms of the Bloch functions of the same band by

$$
w(r - r_n) = N^{-1/2} \sum_k \exp(-ik \cdot r_n) \psi_k(r),
$$

(42)

where $r_n$ is a lattice point. 

(a) Prove that Wannier functions about different lattice points $n,m$ are orthogonal:

$$\int dV w^*(r - r_n)w(r - r_m) = 0, \quad n \neq m.
$$

(43)

This orthogonality property makes the functions often of greater use than atomic orbitals centered on different lattice sites, because the latter are not generally orthogonal. 

(b) The Wannier functions are peaked around the lattice sites. Show that for $\psi_k = N^{-1/2} e^{ikr} u_0(x)$ the Wannier function is

$$
w(x - x_n) = u_0(x) \frac{\sin \pi(x - x_n)/a}{\pi(x - x_n)/a},
$$

for $N$ atoms on a line of lattice constant $a$.

10. **Open orbits and magnetoresistance.** We considered the transverse magnetoresistance of free electrons in Problem 6.9 and of electrons and holes in Problem 8.5. In some crystals the magnetoresistance saturates except in special crystal orientations. An open orbit carries current only in a single direction in the plane normal to the magnetic field; such carriers are not deflected by the field. In the arrangement of Fig. 6.13, let the open orbits be parallel to $k_z$; in real space these orbits carry

*This problem is somewhat difficult.*
current parallel to the $y$ axis. Let $\sigma_{yy} = s\sigma_0$ be the conductivity of the open orbits; this defines the constant $s$. The magnetoconductivity tensor in the high field limit $\omega_c \tau \gg 1$ is

$$
\sigma_0 \begin{pmatrix}
Q^{-2} & -Q^{-1} & 0 \\
-Q^{-1} & s & 0 \\
0 & 0 & 1
\end{pmatrix},
$$

with $Q = \omega_c \tau$. (a) Show that the Hall field is $E_y = -E_x/sQ$. (b) Show that the effective resistivity in the $x$ direction is $\rho = (Q^2/\sigma_0)(s/s + 1)$, so that the resistivity does not saturate, but increases as $B^2$.

**11. Landau levels.** The vector potential of a uniform magnetic field $Bz$ is $A = -By\hat{x}$ in the Landau gauge. The Hamiltonian of a free electron without spin is

$$
H = -(\hbar^2/2m)(\partial^2/\partial y^2 + \partial^2/\partial z^2) + (1/2m)[-i\hbar \partial/\partial x - evB/c]^2.
$$

We will look for an eigenfunction of the wave equation $H\psi = \epsilon \psi$ in the form

$$
\psi = \chi(y) \exp[i(k_x x + k_z z)].
$$

(a) Show that $\chi(y)$ satisfies the equation

$$
(h^2/2m)\partial^2\chi/\partial y^2 + [\epsilon - (\hbar^2 k_x^2/2m) - \frac{1}{2}m\omega_c^2(y - y_0)^2]\chi = 0,
$$

where $\omega_c = eB/mc$ and $y_0 = chk_x/eB$. (b) Show that this is the wave equation of a harmonic oscillator with frequency $\omega_c$, where

$$
\epsilon_n = (n + \frac{1}{2})\hbar \omega_c + \hbar^2 k_x^2/2m.
$$

**References**


**FERMI SURFACES**


**PSEUDOPOTENTIALS**

# 10

**Plasmons, Polaritons, and Polaron**

<table>
<thead>
<tr>
<th>DIELECTRIC FUNCTION OF THE ELECTRON GAS</th>
<th>271</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma optics</td>
<td>272</td>
</tr>
<tr>
<td>Dispersion relation for electromagnetic waves</td>
<td>273</td>
</tr>
<tr>
<td>Transverse optical modes in a plasma</td>
<td>274</td>
</tr>
<tr>
<td>Transparency of alkali metals in the ultraviolet</td>
<td>274</td>
</tr>
<tr>
<td>Longitudinal plasma oscillations</td>
<td>276</td>
</tr>
</tbody>
</table>

**PLASMONS**

<table>
<thead>
<tr>
<th>ELECTROSTATIC SCREENING</th>
<th>280</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screened coulomb potential</td>
<td>282</td>
</tr>
<tr>
<td>Pseudopotential component U(0)</td>
<td>283</td>
</tr>
<tr>
<td>Mott metal-insulator transition</td>
<td>284</td>
</tr>
<tr>
<td>Screening and phonons in metals</td>
<td>286</td>
</tr>
</tbody>
</table>

**POLARITONS**

<table>
<thead>
<tr>
<th>ELECTRON-ELECTRON INTERACTION</th>
<th>294</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fermi liquid</td>
<td>294</td>
</tr>
</tbody>
</table>

**ELECTRON-PHONON INTERACTION: POLARONS**

<table>
<thead>
<tr>
<th>PEIERLS INSTABILITY OF LINEAR METALS</th>
<th>297</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMMAR Y</td>
<td>300</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PROBLEMS</th>
<th>301</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Surface plasmons</td>
<td>302</td>
</tr>
<tr>
<td>2. Interface plasmons</td>
<td>302</td>
</tr>
<tr>
<td>3. Alfvén waves</td>
<td>302</td>
</tr>
<tr>
<td>4. Helicon waves</td>
<td>303</td>
</tr>
<tr>
<td>5. Plasmon mode of a sphere</td>
<td>303</td>
</tr>
<tr>
<td>6. Magnetoplasma frequency</td>
<td>303</td>
</tr>
<tr>
<td>7. Photon branch at low wavevector</td>
<td>303</td>
</tr>
<tr>
<td>8. Plasma frequency and electrical conductivity</td>
<td>303</td>
</tr>
<tr>
<td>9. Bulk modulus of the Fermi gas</td>
<td>303</td>
</tr>
<tr>
<td>10. Response of electron gas</td>
<td>304</td>
</tr>
<tr>
<td>11. Cap plasmons and the van der Waals interaction</td>
<td>304</td>
</tr>
</tbody>
</table>

**REFERENCES**

NOTE: The text and problems of this chapter assume facility in the use of electromagnetic theory at the level of a good senior course.
Figure 1  Dielectric function $\varepsilon(\omega)$ of a free electron gas versus frequency in units of the plasma frequency $\omega_p$. Electromagnetic waves propagate without damping only when $\varepsilon$ is positive and real. Electromagnetic waves are totally reflected from the medium when $\varepsilon$ is negative.
CHAPTER 10: PLASMONS, POLARITONS, AND POLARONS

DIELECTRIC FUNCTION OF THE ELECTRON GAS

The dielectric function $\varepsilon(\omega, K)$ of the electron gas, with its strong dependence on frequency and wavevector, has significant consequences for the physical properties of solids. In one limit, $\varepsilon(\omega, 0)$ describes the collective excitations of the Fermi sea—the volume and surface plasmons. In another limit, $\varepsilon(0, K)$ describes the electrostatic screening of the electron-electron, electron-lattice, and electron-impurity interactions in crystals.

We use the dielectric function of an ionic crystal to derive the polariton spectrum. Later we discuss the properties of polarons. But first we are concerned with the electron gas in metals.

Definitions of the Dielectric Function. The dielectric constant $\varepsilon$ of electrostatics is defined in terms of the electric field $E$ and the polarization $P$, the dipole moment density:

\[(\text{CGS}) \quad D = E + 4\pi P = \varepsilon E ; \quad (\text{SI}) \quad D = \varepsilon_0 E + P = \varepsilon_0 \varepsilon E \quad (1)\]

Thus defined, $\varepsilon$ is also known as the relative permittivity.

The introduction of the displacement $D$ is motivated by the usefulness of this vector related to the external applied charge density $\rho_{\text{ext}}$ in the same way as $E$ is related to the total charge density $\rho = \rho_{\text{ext}} + \rho_{\text{ind}}$, where $\rho_{\text{ind}}$ is the charge density induced in the system by $\rho_{\text{ext}}$.

Thus the divergence relation of the electric field is

\[(\text{CGS}) \quad \text{div} \; D = \text{div} \; \varepsilon E = 4\pi \rho_{\text{ext}} \quad (\text{SI}) \quad \text{div} \; D = \text{div} \; \varepsilon \varepsilon_0 E = \rho_{\text{ext}} \quad (2)\]

\[(\text{CGS}) \quad \text{div} \; E = 4\pi \rho = 4\pi (\rho_{\text{ext}} + \rho_{\text{ind}}) \quad (\text{SI}) \quad \text{div} \; E = \rho/\varepsilon_0 = (\rho_{\text{ext}} + \rho_{\text{ind}})/\varepsilon_0 \quad (3)\]

Parts of this chapter will be written in CGS; to obtain results in SI, write $1/\varepsilon_0$ for $4\pi$. 
We shall need relations between the Fourier components of \( \mathbf{D}, \mathbf{E}, \rho, \) and the electrostatic potential \( \varphi. \) For brevity we do not exhibit here the frequency dependence. Define \( \varepsilon(K) \) such that
\[
\mathbf{D}(K) = \varepsilon(K)\mathbf{E}(K) ;
\]
then (3) becomes
\[
\text{div } \mathbf{E} = \text{div } \sum \varepsilon(K)\mathbf{E}(K) e^{ikr} = 4\pi \sum \rho(K) e^{ikr} ,
\]
and (2) becomes
\[
\text{div } \mathbf{D} = \text{div } \sum \varepsilon(K)\mathbf{E}(K) e^{ikr} = 4\pi \sum \rho_{\text{ext}}(K) e^{ikr} .
\]
Each of the equations must be satisfied term by term; we divide one by the other to obtain
\[
\varepsilon(K) = \frac{\rho_{\text{ext}}(K)}{\rho(K)} = 1 - \frac{\rho_{\text{ind}}(K)}{\rho(K)} .
\]
The electrostatic potential \( \varphi_{\text{ext}} \) defined by \( -\nabla \varphi_{\text{ext}} = \mathbf{D} \) satisfies the Poisson equation \( \nabla^2 \varphi_{\text{ext}} = -4\pi \rho_{\text{ext}} ; \) and the electrostatic potential \( \varphi \) defined by \( -\nabla \varphi = \mathbf{E} \) satisfies \( \nabla^2 \varphi = -4\pi \rho. \) The Fourier components of the potentials must therefore satisfy
\[
\frac{\varphi_{\text{ext}}(K)}{\varphi(K)} = \frac{\rho_{\text{ext}}(K)}{\rho(K)} = \varepsilon(K) ,
\]
by (3d). We use this relation in the treatment of the screened coulomb potential.

---

**Plasma Optics**

The long wavelength dielectric response \( \varepsilon(\omega,0) \) or \( \varepsilon(\omega) \) of an electron gas is obtained from the equation of motion of a free electron in an electric field:
\[
m \frac{d^2x}{dt^2} = -eE .
\]
If \( x \) and \( E \) have the time dependence \( e^{-i\omega t}, \) then
\[
-\omega^2 mx = -eE ; \quad x = eE/m\omega^2 .
\]
The dipole moment of one electron is \( -ex = -e^2E/m\omega^2, \) and the polarization, defined as the dipole moment per unit volume, is
\[
P = -nex = -\frac{ne^2}{m\omega^2}E ,
\]
where \( n \) is the electron concentration.
The dielectric function at frequency $\omega$ is

\begin{equation}
\varepsilon(\omega) = \frac{D(\omega)}{E(\omega)} = 1 + 4\pi \frac{P(\omega)}{E(\omega)} ;
\end{equation}

\text{(CGS)}

\begin{equation}
\varepsilon(\omega) = \frac{D(\omega)}{\varepsilon_0 E(\omega)} = 1 + \frac{P(\omega)}{\varepsilon_0 E(\omega)} .
\end{equation}

\text{(SI)}

The dielectric function of the free electron gas follows from (6) and (7):

\begin{equation}
\varepsilon(\omega) = 1 - \frac{4\pi ne^2}{m\omega^2} ;
\end{equation}

\text{(CGS)}

\begin{equation}
\varepsilon(\omega) = 1 - \frac{ne^2}{\varepsilon_0 m\omega^2} .
\end{equation}

\text{(SI)}

The plasma frequency $\omega_p$ is defined by the relation

\begin{equation}
\omega_p^2 = 4\pi ne^2/m ;
\end{equation}

\text{(CGS)}

\begin{equation}
\omega_p^2 = ne^2/\varepsilon_0 m .
\end{equation}

\text{(SI)}

A plasma is a medium with equal concentration of positive and negative charges, of which at least one charge type is mobile. In a solid the negative charges of the conduction electrons are balanced by an equal concentration of positive charge of the ion cores. We write the dielectric function (8) as

\begin{equation}
\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2} ,
\end{equation}

\text{(10)}

plotted in Fig. 1.

If the positive ion core background has a dielectric constant labeled $\varepsilon(\infty)$ essentially constant up to frequencies well above $\omega_p$, then (8) becomes

\begin{equation}
\varepsilon(\omega) = \varepsilon(\infty) - 4\pi ne^2/m\omega^2 = \varepsilon(\infty)[1 - \bar{\omega}_p^2/\omega^2] ,
\end{equation}

\text{(11)}

where $\bar{\omega}_p$ is defined as

\begin{equation}
\bar{\omega}_p^2 = 4\pi ne^2/\varepsilon(\infty)m .
\end{equation}

\text{(12)}

Notice that $\varepsilon = 0$ at $\omega = \bar{\omega}_p$.

Dispersion Relation for Electromagnetic Waves

In a nonmagnetic isotropic medium the electromagnetic wave equation is

\begin{equation}
\partial^2 D/\partial t^2 = c^2 \nabla^2 E ;
\end{equation}

\text{(CGS)}

\begin{equation}
\mu_0 \partial^2 D/\partial t^2 = \nabla^2 E .
\end{equation}

\text{(SI)}

We look for a solution with $E \propto \exp(-i\omega t) \exp(iK \cdot r)$ and $D = \varepsilon(\omega,K)E$; then we have the dispersion relation for electromagnetic waves:

\begin{equation}
\varepsilon(\omega,K)\omega^2 = c^2 K^2 ;
\end{equation}

\text{(CGS)}

\begin{equation}
\varepsilon(\omega,K)\omega^2 = K^2 ,
\end{equation}

\text{(SI)}

\begin{equation}
\omega^2 = \varepsilon_0 \mu_0 \omega^2 = K^2 .
\end{equation}
This relation tells us a great deal. Consider

- $\varepsilon$ real and $> 0$. For $\omega$ real, $K$ is real and a transverse electromagnetic wave propagates with the phase velocity $c/\varepsilon^{1/2}$.
- $\varepsilon$ real and $< 0$. For $\omega$ real, $K$ is imaginary and the wave is damped with a characteristic length $1/|K|$.
- $\varepsilon$ complex. For $\omega$ real, $K$ is complex and the waves are damped in space.
- $\varepsilon = \infty$. This means the system has a finite response in the absence of an applied force; thus the poles of $\varepsilon(\omega, K)$ define the frequencies of the free oscillations of the medium.
- $\varepsilon = 0$. We shall see that longitudinally polarized waves are possible only at the zeros of $\varepsilon$.

**Transverse Optical Modes in a Plasma**

The dispersion relation (14) becomes, with (11) for $\varepsilon(\omega)$,

\[
\varepsilon(\omega)\omega^2 = \varepsilon(\omega)(\omega^2 - \omega_p^2) = c^2K^2.
\]

(15) For $\omega < \omega_p$ we have $K^2 < 0$, so that $K$ is imaginary. The solutions of the wave equation are of the form $\exp(-|K|x)$ in the frequency region $0 < \omega \leq \omega_p$. Waves incident on the medium in this frequency region do not propagate, but will be totally reflected.

An electron gas is transparent when $\omega > \omega_p$, for here the dielectric function is positive real. The dispersion relation in this region may be written as

\[
\omega^2 = \omega_p^2 + c^2K^2/\varepsilon(\omega);
\]

(16) this describes transverse electromagnetic waves in a plasma (Fig. 2).

Values of the plasma frequency $\omega_p$ and of the free space wavelength $\lambda_p = 2\pi c/\omega_p$ for electron concentrations of interest are given below. A wave will propagate if its free space wavelength is less than $\lambda_p$; otherwise the wave is reflected.

<table>
<thead>
<tr>
<th>$n$, electrons/cm$^3$</th>
<th>$10^{22}$</th>
<th>$10^{18}$</th>
<th>$10^{14}$</th>
<th>$10^{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_p$, s$^{-1}$</td>
<td>$5.7 \times 10^{15}$</td>
<td>$5.7 \times 10^{13}$</td>
<td>$5.7 \times 10^{11}$</td>
<td>$5.7 \times 10^9$</td>
</tr>
<tr>
<td>$\lambda_p$, cm</td>
<td>$3.3 \times 10^{-5}$</td>
<td>$3.3 \times 10^{-3}$</td>
<td>0.33</td>
<td>33</td>
</tr>
</tbody>
</table>

**Transparency of Alkali Metals in the Ultraviolet.** From the preceding discussion of the dielectric function we conclude that simple metals should reflect light in the visible region and be transparent to ultraviolet light. The effect was discovered by Wood and explained by Zener. A comparison of calculated and observed cutoff wavelengths is given in Table I. The reflection of light from a metal is entirely similar to the reflection of radio waves from the ionosphere, for the free electrons in the ionosphere make the dielectric constant negative at low frequencies. Experimental results for InSb with $n = 4 \times 10^{18}$ cm$^{-3}$ are shown in Fig. 3, where the plasma frequency is near $0.09$ eV.
Figure 2 Dispersion relation for transverse electromagnetic waves in a plasma. The group velocity \( v_g = d\omega/dK \) is the slope of the dispersion curve. Although the dielectric function is between zero and one, the group velocity is less than the velocity of light in vacuum.

Figure 3 Reflectance of indium antimonide with \( n = 4 \times 10^{18} \) cm\(^{-3} \). (After J. N. Hodgson.)

Table 1 Ultraviolet transmission limits of alkali metals, in Å

<table>
<thead>
<tr>
<th></th>
<th>Li</th>
<th>Na</th>
<th>K</th>
<th>Rb</th>
<th>Cs</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_p ), calculated, mass ( m )</td>
<td>1550</td>
<td>2090</td>
<td>2870</td>
<td>3220</td>
<td>3620</td>
</tr>
<tr>
<td>( \lambda_p ), observed</td>
<td>1550</td>
<td>2100</td>
<td>3150</td>
<td>3400</td>
<td>—</td>
</tr>
</tbody>
</table>
Longitudinal Plasma Oscillations

The zeros of the dielectric function determine the frequencies of the longitudinal modes of oscillation. That is, the condition

$$\epsilon(\omega_L) = 0$$

(17)
determines the longitudinal frequency $\omega_L$ near $K = 0$.

By the geometry of a longitudinal polarization wave there is a depolarization field $E = -4\pi P$, discussed below. Thus $D = E + 4\pi P = 0$ for a longitudinal wave in a plasma or more generally in a crystal. In SI units, $D = \epsilon_0 E + P = 0$.

For an electron gas, at the zero of the dielectric function

$$\epsilon(\omega_L) = 1 - \omega_p^2/\omega_L^2 = 0,$$

whence $\omega_L = \omega_p$. Thus there is a free longitudinal oscillation mode (Fig. 4) of an electron gas at the plasma frequency described by (15) as the low-frequency cutoff of transverse electromagnetic waves.

A longitudinal plasma oscillation with $K = 0$ is shown in Fig. 5 as a uniform displacement of an electron gas in a thin metallic slab. The electron gas is moved as a whole with respect to the positive ion background. The displacement $u$ of the electron gas creates an electric field $E = 4\pi n e u$ that acts as a restoring force on the gas.

The equation of motion of a unit volume of the electron gas of concentration $n$ is

(CG) $$nm \frac{d^2 u}{dt^2} = -n e E = -4\pi n^2 e^2 u,$$

(19)
or

(CG) $$\frac{d^2 u}{dt^2} + \omega_p^2 u = 0; \quad \omega_p = \left(\frac{4\pi n e^2}{m}\right)^{1/2}.$$

(20)

This is the equation of motion of a simple harmonic oscillator of frequency $\omega_p$, the plasma frequency. The expression for $\omega_p$ is identical with (9), which arose in a different connection. In SI, the displacement $u$ creates the electric field $E = n e u/\epsilon_0$, whence $\omega_p = (n e^2/\epsilon_0 m)^{1/2}$.

A plasma oscillation of small wavevector has approximately the frequency $\omega_p$. The wavevector dependence of the dispersion relation for longitudinal oscillations in a Fermi gas is given by

$$\omega \equiv \omega_p \left(1 + 3k^2 v_F^2/10\omega_p^2 + \cdots\right),$$

(21)

where $v_F$ is the electron velocity at the Fermi energy.
Figure 4  A plasma oscillation. The arrows indicate the direction of displacement of the electrons.

Figure 5  In (a) is shown a thin slab or film of a metal. A cross section is shown in (b), with the positive ion cores indicated by + signs and the electron sea indicated by the gray background. The slab is electrically neutral. In (c) the negative charge has been displaced upward uniformly by a small distance $u$, shown exaggerated in the figure. As in (d), this displacement establishes a surface charge density $-neu$ on the upper surface of the slab and $+neu$ on the lower surface, where $n$ is the electron concentration. An electric field $E = 4\pi neu$ is produced inside the slab. This field tends to restore the electron sea to its equilibrium position (b). In SI units, $E = neu/\epsilon_0$. 
PLASMONS

A plasma oscillation in a metal is a collective longitudinal excitation of the conduction electron gas. A plasmon is a quantum of a plasma oscillation; we may excite a plasmon by passing an electron through a thin metallic film (Figs. 6 and 7) or by reflecting an electron or a photon from a film. The charge of the electron couples with the electrostatic field fluctuations of the plasma oscillations. The reflected or transmitted electron will show an energy loss equal to integral multiples of the plasmon energy.

Experimental excitation spectra for Al and Mg are shown in Fig. 8. A comparison of observed and calculated values of plasmon energies is given in Table 2; further data are given in the reviews by Raether and by Daniels. Recall that \( \tilde{\omega}_p \) as defined by (12) includes the ion core effects by use of \( \epsilon(\infty) \).

It is equally possible to excite collective plasma oscillations in dielectric films; results for several dielectrics are included. The calculated plasma energies of Si, Ge, and InSb are based on four valence electrons per atom. In a dielectric the plasma oscillation is physically the same as in a metal: the entire valence electron sea oscillates back and forth with respect to the ion cores.

Table 2  Volume plasmon energies, in eV

<table>
<thead>
<tr>
<th>Material</th>
<th>Observed</th>
<th>( \hbar\omega_p )</th>
<th>( \hbar\tilde{\omega}_p )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metals</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li</td>
<td>7.12</td>
<td>8.02</td>
<td>7.96</td>
</tr>
<tr>
<td>Na</td>
<td>5.71</td>
<td>5.95</td>
<td>5.58</td>
</tr>
<tr>
<td>K</td>
<td>3.72</td>
<td>4.29</td>
<td>3.86</td>
</tr>
<tr>
<td>Mg</td>
<td>10.6</td>
<td>10.9</td>
<td></td>
</tr>
<tr>
<td>Al</td>
<td>15.3</td>
<td>15.8</td>
<td></td>
</tr>
<tr>
<td><strong>Dielectrics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Si</td>
<td>16.4–16.9</td>
<td>16.0</td>
<td></td>
</tr>
<tr>
<td>Ge</td>
<td>16.0–16.4</td>
<td>16.0</td>
<td></td>
</tr>
<tr>
<td>InSb</td>
<td>12.0–13.0</td>
<td>12.0</td>
<td></td>
</tr>
</tbody>
</table>
Figure 6 Creation of a plasmon in a metal film by inelastic scattering of an electron. The incident electron typically has an energy 1 to 10 keV; the plasmon energy may be of the order of 10 eV. An event is also shown in which two plasmons are created.

Figure 7 A spectrometer with electrostatic analyzer for the study of plasmon excitation by electrons. (After J. Daniels et al.)

Figure 8 Energy loss spectra for electrons reflected from films of (a) aluminum and (b) magnesium, for primary electron energies of 2020 eV. The 12 loss peaks observed in Al are made up of combinations of 10.3 and 15.3 eV losses, where the 10.3 eV loss is due to surface plasmons and the 15.3 eV loss is due to volume plasmons. The ten loss peaks observed in Mg are made up of combinations of 7.1 eV surface plasmons and 10.6 eV volume plasmons. Surface plasmons are the subject of Problem 1. (After C. J. Powell and J. B. Swan.)
ELECTROSTATIC SCREENING

The electric field of a positive charge embedded in an electron gas falls off with increasing \( r \) faster than \( 1/r \), because the electron gas tends to gather around and thus to screen the positive charge. The static screening can be described by the wavevector dependence of the static dielectric function \( \varepsilon(0, K) \). We consider the response of the electrons to an applied external electrostatic field. We start with a uniform gas of electrons of charge concentration \( -n_0 e \) superimposed on a background of positive charge of concentration \( n_0 e \). Let the positive charge background be deformed mechanically to produce a sinusoidal variation of positive charge density in the \( x \) direction:

\[
\rho^+(x) = n_0 e + \rho_{\text{ext}}(K) \sin Kx.
\] (22)

The term \( \rho_{\text{ext}}(K) \sin Kx \) gives rise to an electrostatic field that we call the external field applied to the electron gas.

The electrostatic potential \( \varphi \) of a charge distribution is found from the Poisson equation \( \nabla^2 \varphi = -4\pi \rho \), by (3) with \( E = -\nabla \varphi \). For the positive charge we have

\[
\varphi = \varphi_{\text{ext}}(K) \sin Kx; \quad \rho = \rho_{\text{ext}}(K) \sin Kx.
\] (23)

The Poisson equation gives the relation

\[
K^2 \varphi_{\text{ext}}(K) = 4\pi \rho_{\text{ext}}(K).
\] (24)

The electron gas will be deformed by the combined influences of the electrostatic potential \( \varphi_{\text{ext}}(K) \) of the positive charge distribution and of the as yet unknown induced electrostatic potential \( \varphi_{\text{ind}}(K) \) of the deformation of the electron gas itself. The electron charge density is

\[
\rho^-(x) = -n_0 e + \rho_{\text{ind}}(K) \sin Kx,
\] (25)

where \( \rho_{\text{ind}}(K) \) is the amplitude of the charge density variation induced in the electron gas. We want to find \( \rho_{\text{ind}}(K) \) in terms of \( \rho_{\text{ext}}(K) \).

The amplitude of the total electrostatic potential \( \varphi(K) = \varphi_{\text{ext}}(K) + \varphi_{\text{ind}}(K) \) of the positive and negative charge distributions is related to the total charge density variation \( \rho(K) = \rho_{\text{ext}}(K) + \rho_{\text{ind}}(K) \) by the Poisson equation. Then, as in Eq. (24),

\[
K^2 \varphi(K) = 4\pi \rho(K).
\] (26)

To go further we need another equation that relates the electron concentration to the electrostatic potential. We develop this connection in what is called the Thomas-Fermi approximation. The approximation consists in assuming that a local internal chemical potential can be defined as a function of the electron concentration at that point. Now the total chemical potential of the electron gas is constant in equilibrium, independent of position. In a region where there is no electrostatic contribution to the chemical potential we have
Figure 9 In thermal and diffusive equilibrium the chemical potential is constant; to maintain it constant we increase the electron concentration in regions of space where the potential energy is low, and we decrease the concentration where the potential is high.

\[ \mu = \epsilon^0_F = \frac{\hbar^2}{2m} \left(3\pi^2 n_0\right)^{2/3} \]

at absolute zero, according to (6.17). In a region where the electrostatic potential is \( \varphi(x) \), the total chemical potential (Fig. 9) is constant and equal to

\[ \mu = \epsilon_F(x) - e\varphi(x) \equiv \frac{\hbar^2}{2m} \left[3\pi^2 n(x)\right]^{2/3} - e\varphi(x) \equiv \frac{\hbar^2}{2m} \left[3\pi^2 n_0\right]^{2/3} \]

where \( \epsilon_F(x) \) is the local value of the Fermi energy.

The expression (28) is valid for static electrostatic potentials that vary slowly compared with the wavelength of an electron at the Fermi level; specifically, the approximation is \( q \ll k_F \). By a Taylor series expansion of \( \epsilon_F \), Eq. (28) may be written as

\[ \frac{d\epsilon_F}{dn_0} [n(x) - n_0] \equiv e\varphi(x) \]

From (27) we have \( d\epsilon_F/dn_0 = 2\epsilon_F/3n_0 \), whence

\[ n(x) - n_0 \equiv \frac{3}{2} n_0 \frac{e\varphi(x)}{\epsilon_F} \]

The left-hand side is the induced part of the electron concentration; thus the Fourier components of this equation are

\[ \rho_{\text{ind}}(K) = -\left(3n_0e^2/2\epsilon_F\right)\varphi(K) \]

By (26) this becomes

\[ \rho_{\text{ind}}(K) = -\left(6\pi n_0e^2/\epsilon_F K^2\right)\rho(K) \]

By (3d) we have

\[ \epsilon(0,K) = 1 - \frac{\rho_{\text{ind}}(K)}{\rho(K)} = 1 + \frac{k_x^2}{K^2} \]

here, after some rearrangement,

\[ k_x^2 = 6\pi n_0e^2/\epsilon_F = 4(3/\pi)^{1/3} n_0^{1/3}/a_0 = 4\pi e^2 D(\epsilon_F) \]
where \( a_0 \) is the Bohr radius and \( D(\epsilon_F) \) is the density of states for a free electron gas. The approximation (33) for \( \epsilon(0,K) \) is called the Thomas-Fermi dielectric function, and \( 1/k_s \) is the Thomas-Fermi screening length, as in (40) below. For copper with \( n_0 = 8.5 \times 10^{22} \text{ cm}^{-3} \), the screening length is 0.55 Å.

We have derived two limiting expressions for the dielectric function of an electron gas:

\[
\epsilon(0,K) = 1 + \frac{k_s^2}{K^2}; \quad \epsilon(\omega,0) = 1 - \frac{\omega_p^2}{\omega^2}.
\]  

(35)

We notice that \( \epsilon(0,K) \) as \( K \to 0 \) does not approach the same limit as \( \epsilon(\omega,0) \) as \( \omega \to 0 \). This means that great care must be taken with the dielectric function near the origin of the \( \omega-K \) plane. The full theory for the general function \( \epsilon(\omega,K) \) is due to Lindhard.¹

**Screened Coulomb Potential.** We consider a point charge \( q \) placed in a sea of conduction electrons. The Poisson equation for the unscreened coulomb potential is

\[
\nabla^2 \varphi_0 = -4\pi q \delta(r),
\]  

(36)

and we know that \( \varphi_0 = q/r \). Let us write

\[
\varphi_0(r) = (2\pi)^{-3} \int dK \varphi_0(K) \exp(iK \cdot r).
\]  

(37)

We use in (36) the Fourier representation of the delta function:

\[
\delta(r) = (2\pi)^{-3} \int dK \exp(iK \cdot r),
\]  

(38)

whence \( K^2 \varphi_0(K) = 4\pi q \).

By (3e),

\[
\varphi_0(K)/\varphi(K) = \epsilon(K),
\]

where \( \varphi(K) \) is the total or screened potential. We use \( \epsilon(K) \) in the Thomas-Fermi form (33) to find

\[
\varphi(K) = \frac{4\pi q}{K^2 + k_s^2}.
\]  

(39)

The screened coulomb potential is the transform of \( \varphi(K) \):

\[
\varphi(r) = \frac{4\pi q}{(2\pi)^3} \int_0^\infty dK K^2 \frac{2\pi K^2}{K^2 + k_s^2} \int_{-1}^1 d(\cos \theta) \exp(iKr \cos \theta)
\]

\[
= \frac{2q}{\pi r} \int_0^\infty dK K \sin Kr \frac{K}{K^2 + k_s^2} = \frac{q}{r} \exp(-k_s r)
\]  

(40)

¹A good discussion of the Lindhard dielectric function is given by J. Ziman, Principles of the theory of solids, 2nd ed., Cambridge, 1972, Chapter 5. The algebraic steps in the evaluation of Ziman's equation (5.16) are given in detail by C. Kittel, Solid state physics 22, 1 (1968), Section 6.
as in Fig. 10a. The screening parameter $k_s$ is defined by (34). The exponential factor reduces the range of the coulomb potential. The bare potential $q/r$ is obtained on letting the charge concentration $n_0 \to 0$, for then $k_s \to 0$. In the vacuum limit $\varphi(K) = 4\pi q/K^2$.

One application of the screened interaction is to the resistivity of certain alloys. The atoms of the series Cu, Zn, Ga, Ge, As have valences 1, 2, 3, 4, 5. An atom of Zn, Ga, Ge, or As added substitutionally to metallic Cu has an excess charge, referred to Cu, of 1, 2, 3, or 4 if all the valence electrons join the conduction band of the host metal. The foreign atom scatters the conduction electrons, with an interaction given by the screened coulomb potential. This scattering contributes to the residual electrical resistivity, and calculations by Mott of the resistivity increase are in fair agreement with experiment.

**Pseudopotential Component $U(0)$**. In the legend to Fig. 9.22b we stated a result that is important in pseudopotential theory: "For very small $k$ the potential approaches $-\frac{3}{8} t$ times the Fermi energy." The result, which is known as the screened ion limit for metals, can be derived from Eq. (39). When converted to the potential energy of an electron of charge $e$ in a metal of valency $z$ with $n_0$ ions per unit volume, the potential energy component at $k = 0$ becomes

$$U(0) = -ezn_0\varphi(0) = -4\pi zn_0 e^2/k_s^2 .$$

(41)

The result (34) for $k_s^2$ in this situation reads

$$k_s^2 = 6\pi zn_0 e^2 / e_F ,$$

(42)

whence

$$U(0) = -\frac{3}{8} e_F .$$

(43)
Mott Metal-Insulator Transition

A crystal composed of one hydrogen atom per primitive cell should always be a metal, according to the independent-electron model, because there will always be a half-filled energy band within which charge transport can take place. A crystal with one hydrogen molecule per primitive cell is a different matter, because the two electrons can fill a band. Under extreme high pressure, as in the planet Jupiter, it is possible that hydrogen occurs in a metallic form.

But let us imagine a lattice of hydrogen atoms at absolute zero: will this be a metal or an insulator? The answer depends on the lattice constant, with small values of $a$ giving a metal and large values giving an insulator. Mott\(^2\) made an early estimate of the critical value $a_c$ of the lattice constant that separates the metallic state from the insulating state: $a_c = 4.5a_0$, where $a_0 = \hbar^2/m_e c^2$ is the radius of the first Bohr orbit of a hydrogen atom.

On one approach to the problem, we start in the metallic state where a conduction electron sees a screened coulomb interaction from each proton:

$$U(r) = -(e^2/r) \exp(-k_s r) ,$$

where $k_s^2 = 3.939n_0^{1/3}/a_0$, as in (34), where $n_0$ is the electron concentration. At high concentrations $k_s$ is large and the potential has no bound state, so that we must have a metal.

The potential is known\(^3\) to have a bound state when $k_s$ is smaller than $1.19/a_0$. With a bound state possible the electrons may condense about the protons to form an insulator. The inequality may be written in terms of $n_0$ as

$$3.939n_0^{1/3}/a_0 < 1.42/a_0^2 .$$

With $n_0 = 1/a^3$ for a simple cubic lattice, we may have an insulator when $a_c > 2.78a_0$, which is not far from the Mott result $4.5a_0$ found in a different way.

The term metal-insulator transition has come to denote situations where the electrical conductivity of a material changes from metal to insulator as a function of some external parameter, which may be composition, pressure, strain, or magnetic field. The metallic phase may usually be pictured in terms of an independent-electron model; the insulator phase may suggest important electron-electron interactions. Sites randomly occupied introduce new and interesting aspects to the problem, aspects that lie within percolation theory. The percolation transition is beyond the scope of our book.


When a semiconductor is doped with increasing concentrations of donor (or acceptor) atoms, a transition will occur to a conducting metallic phase. Experimental results for P atoms in silicon are shown in Figure 10b. Here the insulator-metal transition takes place when the concentration is so high that the ground state wavefunctions of electrons on neighboring impurity atoms overlap significantly.

The observed value of the critical concentration in the Si:P alloy system is $n_c = 3.74 \times 10^{18} \text{ cm}^{-3}$, as in the figure. If we take $32 \times 10^{-8} \text{ cm}$ as the radius of the ground state of a donor in Si in the spherical approximation, then by the Mott criterion $a_c = 1.44 \times 10^{-6} \text{ cm}$. The P atoms are believed to occupy lattice sites at random, but if instead their lattice were simple cubic, the critical Mott concentration would be

$$n_c = \frac{1}{a_c^3} = 0.33 \times 10^{18} \text{ cm}^{-3},$$

(46)

appreciably less than the observed value. It is usual in the semiconductor liter-
ature to refer to a heavily-doped semiconductor in the metallic range as a degenerate semiconductor.

**Screening and Phonons in Metals**

An interesting application of our two limiting forms of the dielectric function is to longitudinal acoustic phonons in metals. For longitudinal modes the total dielectric function, ions plus electrons, must be zero, by (17). Provided the sound velocity is less than the Fermi velocity of the electrons, we may use for the electrons the Thomas-Fermi dielectric function

$$\epsilon_{el}(\omega, K) = 1 + \frac{k_s^2}{K^2} . \quad (47)$$

Provided also that the ions are well-spaced and move independently, we may use for them the plasmon $\epsilon(\omega, 0)$ limit with the approximate mass $M$.

The total dielectric function, lattice plus electrons, but without the electronic polarizability of the ion cores, is

$$\epsilon(\omega, K) = 1 - \frac{4\pi ne^2}{M\omega^2} + \frac{k_s^2}{K^2} . \quad (48)$$

At low $K$ and $\omega$ we neglect the term 1. At a zero of $\epsilon(\omega, K)$ we have, with $\epsilon_F = \frac{1}{2}mv_F^2$,

$$\omega^2 = \frac{4\pi ne^2}{Mk_s^2}K^2 = \frac{4\pi ne^2}{M} \cdot \frac{\epsilon_F}{6\pi ne^2}K^2 = \frac{m}{3M}v_F^2K^2 , \quad (49)$$

or

$$\omega = vK ; \quad v = (m/3M)^{1/2}v_F . \quad (50)$$

This describes long wavelength acoustic phonons.

In the alkali metals the result is in quite good agreement with the observed longitudinal wave velocity. For potassium we calculate $v = 1.8 \times 10^5$ cm s$^{-1}$; the observed longitudinal sound velocity at 4 K in the [100] direction is $2.2 \times 10^5$ cm s$^{-1}$.

There is another zero of $\epsilon(\omega, K)$ for positive ions imbedded in an electron sea. For high frequencies we use the dielectric contribution $-\omega_p^2/\omega^2$ of the electron gas:

$$\epsilon(\omega, 0) = 1 - \frac{4\pi ne^2}{M\omega^2} - \frac{4\pi ne^2}{m\omega^2} . \quad (51)$$

and this function has a zero when

$$\omega^2 = \frac{4\pi ne^2}{\mu} ; \quad \frac{1}{\mu} = \frac{1}{M} + \frac{1}{m} . \quad (52)$$

This is the electron plasma frequency, but with the reduced mass correction for the motion of the positive ions.
Figure 11  A plot of the observed energies and wavevectors of the polaritons and of the LO phonons in GaP. The theoretical dispersion curves are shown by the solid lines. The dispersion curves for the uncoupled phonons and photons are shown by the short, dashed lines. (After C. H. Henry and J. J. Hopfield.)

POLARITONS

Longitudinal optical phonons and transverse optical phonons were discussed in Chapter 4, but we deferred treatment of the interaction of transverse optical phonons with transverse electromagnetic waves. At resonance the phonon-photon coupling entirely changes the character of the propagation, and a forbidden band is established for reasons that have nothing to do with the periodicity of the lattice.

By resonance we mean a condition in which the frequencies and wavevectors of both waves are approximately equal. The region of the crossover of the two dashed curves in Fig. 11 is the resonance region; the two dashed curves are the dispersion relations for photons and transverse optical phonons in the absence of any coupling between them. In reality, however, there always is coupling implicit in Maxwell’s equations and expressed by the dielectric function. The quantum of the coupled phonon-photon transverse wave field is called a polariton.

In this section we see how the coupling is responsible for the dispersion relations shown as solid curves in the figure. All takes place at very low values of the wavevector in comparison with a zone boundary, because at crossover \( \omega(\text{photon}) = c \kappa(\text{photon}) = \omega(\text{phonon}) \approx 10^{13} \text{ s}^{-1} \); thus \( \kappa \approx 300 \text{ cm}^{-1} \).

An early warning: although the symbol \( \omega_L \) will necessarily arise in the theory, the effects do not concern longitudinal optical phonons. Longitudinal phonons do not couple to transverse photons in the bulk of a crystal.
The coupling of the electric field $E$ of the photon with the dielectric polarization $P$ of the TO phonon is described by the electromagnetic wave equation:

\begin{equation}
(c^2 K^2 E = \omega^2 (E + 4\pi P))
\end{equation}

(53)

At low wavevectors the TO phonon frequency $\omega_T$ is independent of $K$. The polarization is proportional to the displacement of the positive ions relative to the negative ions, so that the equation of motion of the polarization is like that of an oscillator and may be written as, with $P = Nqu$,

\begin{equation}
-\omega^2 P + \omega_T^2 P = (Nq^2/M)E,
\end{equation}

(54)

where there are $N$ ion pairs of effective charge $q$ and reduced mass $M$, per unit volume. For simplicity we neglect the electronic contribution to the polarization.

The equations (53) and (54) have a solution when

\begin{equation}
\begin{vmatrix}
\omega^2 - c^2 K^2 & 4\pi \omega^2 \\
Nq^2/M & \omega^2 - \omega_T^2
\end{vmatrix} = 0.
\end{equation}

(55)

This gives the polariton dispersion relation, similar to that plotted in Figs. 11 and 12. At $K = 0$ there are two roots, $\omega = 0$ for the photon and

\begin{equation}
\omega^2 = \omega_T^2 + 4\pi Nq^2/M
\end{equation}

(56)

for the polariton. Here $\omega_T$ is the TO phonon frequency in the absence of coupling with photons.

The dielectric function obtained from (54) is:

\begin{equation}
\varepsilon(\omega) = 1 + 4\pi P/E = 1 + \frac{4\pi Nq^2/M}{\omega_T^2 - \omega^2}.
\end{equation}

(57)

If there is an optical electronic contribution to the polarization from the ion cores, this should be included. In the frequency range from zero up through the infrared, we write

\begin{equation}
\varepsilon(\omega) = \varepsilon(\infty) + \frac{4\pi Nq^2/M}{\omega_T^2 - \omega^2}
\end{equation}

(58)

in accord with the definition of $\varepsilon(\infty)$ as the optical dielectric constant, obtained as the square of the optical refractive index.

We set $\omega = 0$ to obtain the static dielectric function:

\begin{equation}
\varepsilon(0) = \varepsilon(\infty) + 4\pi Nq^2/M \omega_T^2,
\end{equation}

(59)

which is combined with (58) to obtain $\varepsilon(\omega)$ in terms of accessible parameters:
Figure 12  Coupled modes of photons and transverse optical phonons in an ionic crystal. The fine horizontal line represents oscillators of frequency $\omega_T$ in the absence of coupling to the electromagnetic field, and the fine line labeled $\omega = cK/\sqrt{\varepsilon(\infty)}$ corresponds to electromagnetic waves in the crystal, but uncoupled to the lattice oscillators $\omega_T$. The heavy lines are the dispersion relations in the presence of coupling between the lattice oscillators and the electromagnetic wave. One effect of the coupling is to create the frequency gap between $\omega_T$ and $\omega_T$: within this gap the wavevector is pure imaginary of magnitude given by the broken line in the figure. In the gap the wave attenuates as $\exp(-|K|x)$, and we see from the plot that the attenuation is much stronger near $\omega_T$ than near $\omega_L$. The character of the branches varies with $K$, there is a region of mixed electric-mechanical aspects near the nominal crossover. Note, finally, it is intuitively obvious that the group velocity of light in the medium is always $<c$, because the slope $\partial \omega/\partial K$ for the actual dispersion relations (heavy lines) is everywhere less than the slope $c$ for the uncoupled photon in free space.

$$\varepsilon(\omega) = \varepsilon(\infty) + \left[\varepsilon(0) - \varepsilon(\infty)\right] \frac{\omega_T^2}{\omega_T^2 - \omega^2} = \varepsilon(\infty) \left( \frac{\omega_L^2 - \omega^2}{\omega_T^2 - \omega^2} \right)$$

or

$$\varepsilon(\omega) = \frac{\omega_T^2 \varepsilon(0) - \omega^2 \varepsilon(\infty)}{\omega_T^2 - \omega^2} = \varepsilon(\infty) \left( \frac{\omega_L^2 - \omega^2}{\omega_T^2 - \omega^2} \right). \quad (60)$$

The zero of $\varepsilon(\omega)$ defines the frequency $\omega_L$, as the pole of $\varepsilon(\omega)$ defines $\omega_T$. The zero gives

$$\varepsilon(\infty) \omega_L^2 = \varepsilon(0) \omega_T^2. \quad (61)$$

Waves do not propagate in the frequency region for which $\varepsilon(\omega)$ is negative, between its pole at $\omega = \omega_T$ and its zero at $\omega = \omega_L$, as in Fig. 13. For negative $\varepsilon$, waves do not propagate because then $K$ is imaginary for real $\omega$, and
Figure 13a Plot of $\varepsilon(\omega)$ from (60) for $\varepsilon(\infty) = 2$ and $\varepsilon(0) = 3$. The dielectric constant is negative between $\omega = \omega_T$ and $\omega_L = (3/2)^{1/2} \omega_T$, that is, between the pole (infinity) of $\varepsilon(\omega)$ and the zero of $\varepsilon(\omega)$. Incident electromagnetic waves of frequencies $\omega_T < \omega < \omega_L$ will not propagate in the medium, but will be reflected at the boundary.

Figure 13b Dielectric function (real part) of SrF$_2$ measured over a wide frequency range, exhibiting the decrease of the ionic polarizability at high frequencies. (A. von Hippel.)
exp(iKx) → exp(−|K|x), damped in space. The zero of $\varepsilon(\omega)$, by our earlier argument, is the LO frequency at low $K$, Fig. 14. Just as with the plasma frequency $\omega_p$, the frequency $\omega_L$ has two meanings, one as the LO frequency at low $K$ and the other as the upper cutoff frequency of the forbidden band for propagation of an electromagnetic wave. The value of $\omega_L$ is identical at both frequencies.

**LST Relation**

We write (61) as

$$\frac{\omega_B^2}{\omega_R^2} = \frac{\varepsilon(0)}{\varepsilon(\infty)},$$

where $\varepsilon(0)$ is the static dielectric constant and $\varepsilon(\infty)$ is the high-frequency limit of the dielectric function, defined to include the core electron contribution. This result is the Lyddane-Sachs-Teller relation. The derivation assumed a cubic crystal with two atoms per primitive cell.\(^4\) For soft modes with $\omega_R \rightarrow 0$ we see that $\varepsilon(0) \rightarrow \infty$, a characteristic of ferroelectricity.

Undamped electromagnetic waves with frequencies within the gap cannot propagate in a thick crystal. The reflectivity of a crystal surface is expected to be high in this frequency region, as in Fig. 15.

For films of thickness less than a wavelength the situation is changed. Because for frequencies in the gap the wave attenuates as \( \exp(-|K|x) \), it is possible for the radiation to be transmitted through a film for the small values of \(|K|\) near \( \omega_L \), but for the large values of \(|K|\) near \( \omega_T \) the wave will be reflected. By reflection at nonnormal incidence the frequency \( \omega_L \) of longitudinal optical phonons can be observed, as in Fig. 16.

Experimental values of \( \varepsilon(0) \), \( \varepsilon(\infty) \), and \( \omega_T \) are given in Table 3, with values of \( \omega_L \) calculated using the LST relation, Eq. (62). We compare values of

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Static dielectric constant ( \varepsilon(0) )</th>
<th>Optical dielectric constant ( \varepsilon(\infty) )</th>
<th>( \omega_T ), in ( 10^{13} ) s(^{-1}) experimental</th>
<th>( \omega_L ), in ( 10^{12} ) s(^{-1}) LST relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiH</td>
<td>12.9</td>
<td>3.6</td>
<td>11.</td>
<td>21.</td>
</tr>
<tr>
<td>LiF</td>
<td>8.9</td>
<td>1.9</td>
<td>5.8</td>
<td>12.</td>
</tr>
<tr>
<td>LiCl</td>
<td>12.0</td>
<td>2.7</td>
<td>3.6</td>
<td>7.5</td>
</tr>
<tr>
<td>LiBr</td>
<td>13.2</td>
<td>3.2</td>
<td>3.0</td>
<td>6.1</td>
</tr>
<tr>
<td>NaF</td>
<td>5.1</td>
<td>1.7</td>
<td>4.5</td>
<td>7.8</td>
</tr>
<tr>
<td>NaCl</td>
<td>5.9</td>
<td>2.25</td>
<td>3.1</td>
<td>5.0</td>
</tr>
<tr>
<td>NaBr</td>
<td>6.4</td>
<td>2.6</td>
<td>2.5</td>
<td>3.9</td>
</tr>
<tr>
<td>KF</td>
<td>5.5</td>
<td>1.5</td>
<td>3.6</td>
<td>6.1</td>
</tr>
<tr>
<td>KCl</td>
<td>4.85</td>
<td>2.1</td>
<td>2.7</td>
<td>4.0</td>
</tr>
<tr>
<td>KI</td>
<td>5.1</td>
<td>2.7</td>
<td>1.9</td>
<td>2.6</td>
</tr>
<tr>
<td>RbF</td>
<td>6.5</td>
<td>1.9</td>
<td>2.9</td>
<td>5.4</td>
</tr>
<tr>
<td>RbI</td>
<td>5.5</td>
<td>2.6</td>
<td>1.4</td>
<td>1.9</td>
</tr>
<tr>
<td>CsCl</td>
<td>7.2</td>
<td>2.6</td>
<td>1.9</td>
<td>3.1</td>
</tr>
<tr>
<td>CsI</td>
<td>5.65</td>
<td>3.0</td>
<td>1.2</td>
<td>1.6</td>
</tr>
<tr>
<td>TlCl</td>
<td>31.9</td>
<td>5.1</td>
<td>1.2</td>
<td>3.0</td>
</tr>
<tr>
<td>TlBr</td>
<td>29.8</td>
<td>5.4</td>
<td>0.81</td>
<td>1.9</td>
</tr>
<tr>
<td>AgCl</td>
<td>12.3</td>
<td>4.0</td>
<td>1.9</td>
<td>3.4</td>
</tr>
<tr>
<td>AgBr</td>
<td>13.1</td>
<td>4.6</td>
<td>1.5</td>
<td>2.5</td>
</tr>
<tr>
<td>MgO</td>
<td>9.8</td>
<td>2.95</td>
<td>7.5</td>
<td>14.</td>
</tr>
<tr>
<td>CaP</td>
<td>10.7</td>
<td>8.5</td>
<td>6.9</td>
<td>7.6</td>
</tr>
<tr>
<td>GaAs</td>
<td>12.9</td>
<td>10.9</td>
<td>5.1</td>
<td>5.5</td>
</tr>
<tr>
<td>GaSb</td>
<td>16.1</td>
<td>14.4</td>
<td>4.3</td>
<td>4.6</td>
</tr>
<tr>
<td>InP</td>
<td>12.4</td>
<td>9.6</td>
<td>5.7</td>
<td>6.5</td>
</tr>
<tr>
<td>InAs</td>
<td>14.9</td>
<td>12.3</td>
<td>4.1</td>
<td>4.5</td>
</tr>
<tr>
<td>InSb</td>
<td>17.7</td>
<td>15.6</td>
<td>3.5</td>
<td>3.7</td>
</tr>
<tr>
<td>SiC</td>
<td>9.6</td>
<td>6.7</td>
<td>14.9</td>
<td>17.9</td>
</tr>
<tr>
<td>C</td>
<td>5.5</td>
<td>5.5</td>
<td>25.1</td>
<td>25.1</td>
</tr>
<tr>
<td>Si</td>
<td>11.7</td>
<td>11.7</td>
<td>9.9</td>
<td>9.9</td>
</tr>
<tr>
<td>Ge</td>
<td>15.8</td>
<td>15.8</td>
<td>5.7</td>
<td>5.7</td>
</tr>
</tbody>
</table>

Data largely due to E. Burstein.
Figure 15  Reflectance of a crystal of NaCl at several temperatures, versus wavelength. The nominal values of $\omega_k$ and $\omega_{tr}$ at room temperature correspond to wavelengths of 38 and $61 \times 10^{-4}$ cm, respectively. (After A. Mitsuishi et al.)

Figure 16  Reflectance versus wavelength of a LiF film backed by silver, for radiation incident near 30°. The longitudinal optical phonon absorbs strongly the radiation polarized ($p$) in the plane normal to the film, but absorbs hardly at all the radiation polarized ($s$) parallel to the film. (After D. W. Berreman.)
\( \omega_i/\omega_T \) obtained by inelastic neutron scattering with experimental values of 
\[ [\varepsilon(0)/\varepsilon(\infty)]^{1/2} \] 
obtained by dielectric measurements:

<table>
<thead>
<tr>
<th></th>
<th>NaI</th>
<th>KBr</th>
<th>GaAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega_i/\omega_T )</td>
<td>1.44 ± 0.05</td>
<td>1.39 ± 0.02</td>
<td>1.07 ± 0.02</td>
</tr>
<tr>
<td>[ [\varepsilon(0)/\varepsilon(\infty)]^{1/2} ]</td>
<td>1.45 ± 0.03</td>
<td>1.38 ± 0.03</td>
<td>1.08</td>
</tr>
</tbody>
</table>

The agreement with the LST relation is excellent.

**ELECTRON-ELECTRON INTERACTION**

*Fermi Liquid.* Because of the interaction of the conduction electrons with each other through their electrostatic interaction, the electrons suffer collisions. Further, a moving electron causes an inertial reaction in the surrounding electron gas, thereby increasing the effective mass of the electron. The effects of electron-electron interactions are usually described within the framework of the Landau theory of a Fermi liquid. The object of the theory is to give a unified account of the effect of interactions. A Fermi gas is a system of noninteracting fermions; the same system with interactions is a Fermi liquid.

Landau's theory gives a good account of the low-lying single particle excitations of the system of interacting electrons. These single particle excitations are called quasiparticles; they have a one-to-one correspondence with the single particle excitations of the free electron gas. A quasiparticle may be thought of as a single particle accompanied by a distortion cloud in the electron gas. One effect of the coulomb interactions between electrons is to change the effective mass of the electron; in the alkali metals the increase is roughly of the order of 25 percent.

*Electron-Electron Collisions.* It is an astonishing property of metals that conduction electrons, although crowded together only 2 Å apart, travel long distances between collisions with each other. The mean free paths for electron-electron collisions are longer than \( 10^4 \) Å at room temperature and longer than 10 cm at 1 K.

Two factors are responsible for these long mean free paths, without which the free electron model of metals would have little value. The most powerful factor is the exclusion principle (Fig. 17), and the second factor is the screening of the coulomb interaction between two electrons.

We show how the exclusion principle reduces the collision frequency of an electron that has a low excitation energy \( \epsilon_i \) outside a filled Fermi sphere (Fig. 18). We estimate the effect of the exclusion principle on the two-body collision \( 1 + 2 \rightarrow 3 + 4 \) between an electron in the excited orbital \( 1 \) and an

Figure 17  A collision between two electrons of wavevector $k_1$ and $k_2$. After the collision the particles have wavevector $k_3$ and $k_4$. The Pauli exclusion principle allows collisions only to final states $k_3$, $k_4$ which were vacant before the collision.

Figure 18  In (a) the electrons in initial orbitals 1 and 2 collide. If the orbitals 3 and 4 are initially vacant, the electrons 1 and 2 can occupy orbitals 3 and 4 after the collision. Energy and momentum are conserved. In (b) the electrons in initial orbitals 1 and 2 have no vacant final orbitals available that allow energy to be conserved in the collision. Orbitals such as 3 and 4 would conserve energy and momentum, but they are already filled with other electrons. In (c) we have denoted with $x$ the wavevector of the center of mass 1 and 2. All pairs of orbitals 3 and 4 conserve momentum and energy if they lie at opposite ends of a diameter of the small sphere. The small sphere was drawn from the center of mass to pass through 1 and 2. But not all pairs of points 3, 4 are allowed by the exclusion principle, for both 3, 4 must lie outside the Fermi sphere; the fraction allowed is $\varepsilon_i / \varepsilon_F$. 
electron in the filled orbital 2 in the Fermi sea. It is convenient to refer all energies to the Fermi level \( \mu \) taken as the zero of energy; thus \( \epsilon_1 \) will be positive and \( \epsilon_2 \) will be negative. Because of the exclusion principle the orbitals 3 and 4 of the electrons after collision must lie outside the Fermi sphere, all orbitals within the sphere being already occupied; thus both energies \( \epsilon_3, \epsilon_4 \) must be positive referred to zero on the Fermi sphere.

The conservation of energy requires that \( |\epsilon_2| < \epsilon_1 \), for otherwise \( \epsilon_3 + \epsilon_4 = \epsilon_1 + \epsilon_2 \) could not be positive. This means that collisions are possible only if the orbital 2 lies within a shell of thickness \( \epsilon_1 \) within the Fermi surface, as in Fig. 18a. Thus the fraction \( \approx \epsilon_1/\epsilon_F \) of the electrons in filled orbitals provides a suitable target for electron 1. But even if the target electron 2 is in the suitable energy shell, only a small fraction of the final orbitals compatible with conservation of energy and momentum are allowed by the exclusion principle. This gives a second factor of \( \epsilon_1/\epsilon_F \).

In Fig. 18c we show a small sphere on which all pairs of orbitals 3, 4 at opposite ends of a diameter satisfy the conservation laws, but collisions can occur only if both orbitals 3, 4 lie outside the Fermi sea. The product of the two fractions is \( (\epsilon_1/\epsilon_F)^2 \). If \( \epsilon_1 \) corresponds to 1 K and \( \epsilon_F \) to \( 5 \times 10^4 \) K, we have \( (\epsilon_1/\epsilon_F)^2 \approx 4 \times 10^{-10} \), the factor by which the exclusion principle reduces the collision rate.

The argument is not changed for a thermal distribution of electrons at a low temperature such that \( k_B T \ll \epsilon_F \). We replace \( \epsilon_1 \) by the thermal energy \( \approx k_B T \), and now the rate at which electron-electron collisions take place is reduced below the classical value by \( (k_B T/\epsilon_F)^2 \), so that the effective collision cross section \( \sigma \) is

\[
\sigma = (k_B T/\epsilon_F)^2 \sigma_0 ,
\]

where \( \sigma_0 \) is the cross section for the electron-electron interaction.

The interaction of one electron with another has a range of the order of the screening length \( 1/k_s \) as in (34). Numerical calculations give the effective cross section with screening for collisions between electrons as of the order of \( 10^{-15} \) cm\(^2\) or \( 10 \) \( \AA \(^2\) in typical metals. The effect of the electron gas background in electron-electron collisions is to reduce \( \sigma_0 \) below the value expected from the Rutherford scattering equation for the unscreened coulomb potential. However, much the greatest reduction in the cross section is caused by the Pauli factor \( (k_B T/\epsilon_F)^2 \).

At room temperature in a typical metal \( k_B T/\epsilon_F \) is \( \sim 10^{-2} \), so that \( \sigma \sim 10^{-4} \sigma_0 \sim 10^{-19} \) cm\(^2\). The mean free path for electron-electron collisions is \( \ell \approx 1/n\sigma \sim 10^{-4} \) cm at room temperature. This is longer than the mean free path due to electron-phonon collisions by at least a factor of 10, so that at room temperature collisions with phonons are likely to be dominant. At liquid helium temperatures a contribution proportional to \( T^2 \) has been found in the
resistivity of a number of metals, consistent with the form of the electron-electron scattering cross section (63). The mean free path of electrons in indium at 2 K is of the order of 30 cm, as expected from (63). Thus the Pauli principle explains one of the central questions of the theory of metals: how do the electrons travel long distances without colliding with each other?

ELECTRON-PHONON INTERACTION: POLARONS

The most common effect of the electron-phonon interaction is seen in the temperature dependence of the electrical resistivity, which for pure copper is 1.55 microhm-cm at 0°C and 2.28 microhm-cm at 100°C. The electrons are scattered by the phonons, and the higher the temperature, the more phonons there are and hence more scattering. Above the Debye temperature the number of thermal phonons is roughly proportional to the absolute temperature, and we find that the resistivity increases approximately as the absolute temperature in any reasonably pure metal in this temperature region.

A more subtle effect of the electron-phonon interaction is the apparent increase in electron mass that occurs because the electron drags the heavy ion cores along with it. In an insulator the combination of the electron and its strain field is known as a polaron, Fig. 19. The effect is large in ionic crystals because of the strong coulomb interaction between ions and electrons. In covalent crystals the effect is weak because neutral atoms have only a weak interaction with electrons.

The strength of the electron-lattice interaction is measured by the dimensionless coupling constant $\alpha$ given by

$$\frac{1}{2} \alpha = \frac{\text{deformation energy}}{\hbar \omega_L},$$

(64)

where $\omega_L$ is the longitudinal optical phonon frequency near zero wavevector. We view $\frac{1}{2} \alpha$ as "the number of phonons which surround a slowmoving electron in a crystal."

Values of $\alpha$ deduced from diverse experiments and theory are given in Table 4, after F. C. Brown. The values of $\alpha$ are high in ionic crystals and low in covalent crystals. The values of the effective mass $m_{\text{pol}}^*$ of the polaron are from cyclotron resonance experiments. The values given for the band effective mass $m^*$ were calculated from $m_{\text{pol}}^*$. The last row in the table gives the factor $m_{\text{pol}}^*/m^*$ by which the band mass is increased by the deformation of the lattice.

---

6See QTS, Chapter 7.
Table 4  Polaron coupling constants $\alpha$, masses $m_{\text{pol}}^*$, and band masses $m^*$ for electrons in the conduction band

<table>
<thead>
<tr>
<th>Crystal</th>
<th>KCl</th>
<th>KBr</th>
<th>AgCl</th>
<th>AgBr</th>
<th>ZnO</th>
<th>PbS</th>
<th>InSb</th>
<th>CaAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>3.97</td>
<td>3.52</td>
<td>2.00</td>
<td>1.69</td>
<td>0.85</td>
<td>0.16</td>
<td>0.014</td>
<td>0.06</td>
</tr>
<tr>
<td>$m_{\text{pol}}^*/m$</td>
<td>1.25</td>
<td>0.93</td>
<td>0.51</td>
<td>0.33</td>
<td>---</td>
<td>---</td>
<td>0.014</td>
<td>---</td>
</tr>
<tr>
<td>$m^*/m$</td>
<td>0.50</td>
<td>0.43</td>
<td>0.35</td>
<td>0.24</td>
<td>---</td>
<td>---</td>
<td>0.014</td>
<td>---</td>
</tr>
<tr>
<td>$m_{\text{pol}}^<em>/m^</em>$</td>
<td>2.5</td>
<td>2.2</td>
<td>1.5</td>
<td>1.4</td>
<td>---</td>
<td>---</td>
<td>1.0</td>
<td>---</td>
</tr>
</tbody>
</table>

Theory relates the effective mass of the polaron $m_{\text{pol}}^*$ to the effective band mass $m^*$ of the electron in the undeformed lattice by the relation

$$m_{\text{pol}}^* = m^* \left( \frac{1 - 0.0008\alpha^2}{1 - \frac{1}{4}\alpha + 0.0034\alpha^2} \right); \quad (65)$$

for $\alpha \ll 1$ this is approximately $m^*(1 + \frac{1}{4}\alpha)$. Because the coupling constant $\alpha$ is always positive, the polaron mass is greater than the bare mass, as we expect from the inertia of the ions.

It is common to speak of large and small polarons. The electron associated with a large polaron moves in a band, but the mass is slightly enhanced; these are the polarons we have discussed above. The electron associated with a small polaron spends most of its time trapped on a single ion. At high temperatures the electron moves from site to site by thermally activated hopping; at low temperatures the electron tunnels slowly through the crystal, as if in a band of large effective mass.

Holes or electrons can become self-trapped by inducing an asymmetric local deformation of the lattice. This is most likely to occur when the band edge is degenerate and the crystal is polar (such as an alkali halide or silver halide), with strong coupling of the particle to the lattice. The valence band edge is more often degenerate than the conduction band edge, so that holes are more likely to be self-trapped than are electrons. Holes appear to be self-trapped in all the alkali and silver halides.

Ionic solids at room temperature generally have very low conductivities for the motion of ions through the crystal, less than $10^{-6}$ (ohm-cm)$^{-1}$, but a family of compounds has been reported\(^7\) with conductivities of 0.2 (ohm-cm)$^{-1}$ at 20°C. The compounds have the composition $\text{MAg}_4\text{I}_5$, where M denotes K, Rb, or NH$_4$. The Ag$^+$ ions occupy only a fraction of the equivalent lattice sites available, and the ionic conductivity proceeds by the hopping of a silver ion from one site to a nearby vacant site. The crystal structures also have parallel open channels.

Figure 19  The formation of a polaron. (a) A conduction electron is shown in a rigid lattice of an ionic crystal, KCl. The forces on the ions adjacent to the electron are shown. (b) The electron is shown in an elastic or deformable lattice. The electron plus the associated strain field is called a polaron. The displacement of the ions increases the effective inertia and, hence, the effective mass of the electron; in KCl the mass is increased by a factor of 2.5 with respect to the band theory mass in a rigid lattice. In extreme situations, often with holes, the particle can become self-trapped (localized) in the lattice. In covalent crystals the forces on the atoms from the electron are weaker than in ionic crystals, so that polaron deformations are small in covalent crystals.
PEIERLS INSTABILITY OF LINEAR METALS

Consider a one-dimensional metal with an electron gas filling all conduction band orbitals out to the wavevector \( k_F \), at absolute zero of temperature. Peierls suggested that such a linear metal is unstable with respect to a static lattice deformation of wavevector \( G = 2k_F \). Such a deformation creates an energy gap at the Fermi surface, thereby lowering the energy of electrons below the energy gap, Fig. 20. The deformation proceeds until limited by the increase of elastic energy: the equilibrium deformation \( \Delta \) is given by the root of

\[
\frac{d}{d\Delta} (E_{\text{electronic}} + E_{\text{elastic}}) = 0 .
\]  

(66)

Consider the elastic strain \( \Delta \cos 2k_F x \). The spatial-average elastic energy per unit length is \( E_{\text{elastic}} = \frac{1}{4} C \Delta^2 (\cos^2 2k_F x) = \frac{1}{4} C \Delta^2 \), where \( C \) is the force constant of the linear metal. We next calculate \( E_{\text{electronic}} \). Suppose that the ion contribution to the lattice potential seen by a conduction electron is proportional to the deformation: \( U(x) = 2A\Delta \cos 2k_F x \). From (7.51) we have

\[
\epsilon_K = (\hbar^2/2m)(k_F^2 + K^2) \pm \left[ 4(\hbar^2 k_F^2/2m)(\hbar^2 K^2/2m) + A^2 \Delta^2 \right]^{1/2} .
\]

(67)

It is convenient to define

\[
x_K = \hbar^2 K^2/m ; \quad x_F = \hbar^2 k_F^2/m ; \quad x = \hbar^2 K k_F/m .
\]

We retain the + sign in (67) and form

\[
\frac{d\epsilon_K}{d\Delta} = \frac{-A^2 \Delta}{(x_F x_K + A^2 \Delta^2)^{1/2}} ,
\]

whence, with \( dK/\pi \) as the number of orbitals per unit length,

\[
\frac{dE_{\text{electronic}}}{d\Delta} = \frac{2}{\pi} \int_0^{k_F} dK \frac{d\epsilon_K}{d\Delta} = -(2A^2 \Delta/\pi) \int_0^{k_F} \frac{dK}{(x_F x_K + A^2 \Delta^2)^{1/2}}
\]

\[
= -(2A^2 \Delta/\pi)(k_F/x_F) \int_0^{x_F} \frac{dx}{(x^2 + A^2 \Delta^2)^{1/2}} = -(2A^2 \Delta/\pi)(k_F/x_F) \sinh^{-1}(x_F/A\Delta) .
\]

We put it all together. The equilibrium deformation is the root of

\[
\frac{1}{2} C \Delta - (2A^2 m \Delta/\pi \hbar^2 k_F) \sinh^{-1}(\hbar^2 k_F^2/m \Delta) = 0 .
\]

The root \( \Delta \) that corresponds to the minimum energy is given by

\[
\hbar^2 k_F^2/m \Delta = \sinh(-\hbar^2 k_F \pi C/4mA^2) ,
\]

(68)

whence

\[
|\Delta| = (2\hbar^2 k_F^2/m) \exp(-\hbar^2 k_F \pi C/4mA^2) ,
\]

(69)

if the argument of the \( \sinh \) in (68) is \( \gg 1 \). We assume \( k_F \leq \frac{1}{2} k_{\text{max}} \).
The result is of the form of the energy gap equation in the BCS theory of superconductivity, Chapter 12. The deformation $\Delta$ is a collective effect of all the electrons. If we set $W = \hbar^2 k_F^2 / 2m = \text{conduction band width}; N(0) = 2m/\pi \hbar^2 k_F = \text{density of orbitals at Fermi level}; V = 2A^2/C = \text{effective electron-electron interaction energy}$, then we can write (69) as

$$|\Delta| = 4W \exp[-1/N(0)V] ,$$

which is analogous to the BCS energy gap equation. An example of a Peierls insulator is $TaS_3$.

**SUMMARY**

- The dielectric function may be defined as

$$\epsilon(\omega, K) = \frac{\rho_{\text{ext}}(\omega, K)}{\rho_{\text{ext}}(\omega, K) + \rho_{\text{ind}}(\omega, K)} ,$$

in terms of the applied and induced charge density components at $\omega, K$.

- The plasma frequency $\tilde{\omega}_p = [4\pi ne^2/\epsilon(\omega)c]^{1/2}$ is the frequency of the uniform collective longitudinal oscillation of the electron gas against a background of fixed positive ions. It is also the low cutoff for propagation of transverse electromagnetic waves in the plasma.

- The poles of the dielectric function define $\omega_T$ and the zeroes define $\omega_L$.

---

8A review of charge density waves and superlattices in metallic layered structures is given by J. A. Wilson, F. J. Di Salvo and S. Mahajan, Advances in Physics 24, 117 (1975); this article contains some of the most beautiful diffraction photographs in solid state physics. P. B. Littlewood and V. Heine, J. Phys. C. 14, 2493 (1981) have shown how the Peierls effect is modified by the electron-electron interaction; see also J. E. Hirsch, Phys. Rev. Lett. 51, 296 (1983).
• In a plasma the coulomb interaction is screened; it becomes \((q/r) \exp(-k_z r)\), where the screening length \(1/k_s = (\epsilon_f / 6\pi m_0 e^2)^{1/2}\).

• A metal-insulator transition may occur when the nearest-neighbor separation \(a\) becomes of the order of \(4a_0\), where \(a_0\) is the radius of the first Bohr orbit in the insulator. The metallic phase exists at smaller values of \(a\).

• A polariton is a quantum of the coupled TO phonon-photon fields. The coupling is assured by the Maxwell equations. The spectral region \(\omega_T < \omega < \omega_L\) is forbidden to electromagnetic wave propagation.

• The Lyddane-Sachs-Teller relation is \(\omega_T^2/\omega_L^2 = \epsilon(0)/\epsilon(\infty)\).

**Problems**

1. **Surface plasmons.** Consider a semi-infinite plasma on the positive side of the plane \(z = 0\). A solution of Laplace's equation \(\nabla^2 \varphi = 0\) in the plasma is \(\varphi_i(x, z) = A \cos k x \, e^{-k_2 z}\), whence \(E_{zi} = k A \cos k x \, e^{-k_2 z}\), \(E_{z1} = k A \sin k x \, e^{-k_2 z}\). (a) Show that in the vacuum \(\varphi_0(x, z) = A \cos k x \, e^{k_2 z}\) for \(z < 0\) satisfies the boundary condition that the tangential component of \(E\) be continuous at the boundary; that is, find \(E_{z0}\). (b) Note that \(D_i = \epsilon(\omega) E_i\), \(D_e = E_0\). Show that the boundary condition that the normal component of \(D\) be continuous at the boundary requires that \(\epsilon(\omega) = -1\), whence from (10) we have the Stern-Ferrell result:

\[
\omega_s^2 = \frac{1}{2} \omega_p^2
\]

for the frequency \(\omega_s\) of a surface plasma oscillation.

2. **Interface plasmons.** We consider the plane interface \(z = 0\) between metal 1 at \(z > 0\) and metal 2 at \(z < 0\). Metal 1 has bulk plasmon frequency \(\omega_{p1}\); metal 2 has \(\omega_{p2}\). The dielectric constants in both metals are those of free electron gases. Show that surface plasmons associated with the interface have the frequency

\[
\omega = \left[ \frac{1}{2}(\omega_{p1}^2 + \omega_{p2}^2) \right]^{1/2}.
\]

3. **Alfvén waves.** Consider a solid with an equal concentration \(n\) of electrons of mass \(m_e\) and holes of mass \(m_h\). This situation may arise in a semimetal or in a compensated semiconductor. Place the solid in a uniform magnetic field \(B = B \hat{z}\). Introduce the coordinate \(\xi = x + i y\) appropriate for circularly polarized motion, with \(\xi\) having time dependence \(e^{-i\omega t}\). Let \(\omega_e = eB/m_e c\) and \(\omega_h = eB/m_h c\). (a) In CGS units, show that \(\xi_e = eE^+ / m_e \omega (\omega + \omega_e)\), \(\xi_h = -eE^+ / m_h \omega (\omega - \omega_h)\) are the displacements of the electrons and holes in the electric field \(E^+ \, e^{-i\omega t} = (E_x + iE_y) \, e^{-i\omega t}\). (b) Show that the dielectric polarization \(P^+ = ne(\xi_e - \xi_h)\) in the regime \(\omega \ll \omega_e, \omega_h\) may be written as \(P^+ = n c^2 (m_h + m_e) E^+ / B^2\), and the dielectric function \(\epsilon(\omega) = \epsilon_1 + 4\pi P^+ / E^+ = \epsilon_1 + 4\pi e^2 n / B^2\), where \(\epsilon_1\) is the dielectric constant of the host lattice and \(\rho = n(m_e + m_h)\) is the mass density of the carriers. If \(\epsilon_1\) may be neglected,
the dispersion relation $\omega^2 \varepsilon(\omega) = c^2 k^2$ becomes, for electromagnetic waves propagating in the $z$ direction, $\omega^2 = (B^2/4\pi\mu)k^2$. Such waves are known as Alfvén waves, they propagate with the constant velocity $B/(4\pi\mu)^{1/2}$. If $B = 10$ kG; $n = 10^{16}$ cm$^{-3}$; $m = 10^{-27}$ g, the velocity is $\sim 10^8$ cm s$^{-1}$. Alfvén waves have been observed in semimetals and in electron-hole drops in germanium (Chapter 11).

4. Helicon waves. (a) Employ the method of Problem 3 to treat a specimen with only one carrier type, say holes in concentration $p$, and in the limit $\omega \ll \omega_h = eB/m_e c$. Show that $\varepsilon(\omega) = 4\pi p e^2/m_e \omega_n$, where $D^+(\omega) = \varepsilon(\omega)E^+(\omega)$. The term $\varepsilon_i$ in $\varepsilon$ has been neglected. (b) Show further that the dispersion relation becomes $\omega = (Bc/4\pi pe)K^2$, the helicon dispersion relation, in CGS. For $K = 1$ cm$^{-1}$ and $B = 1000$ G, estimate the helicon frequency in sodium metal. (The frequency is negative; with circular-polarized modes the sign of the frequency refers to the sense of the rotation.)

5. Plasmon mode of a sphere. The frequency of the uniform plasmon mode of a sphere is determined by the depolarization field $E = -4\pi F/3$ of a sphere, where the polarization $P = -ner$, with $r$ as the average displacement of the electrons of concentration $n$. Show from $F = ma$ that the resonance frequency of the electron gas is $\omega_0^2 = 4\pi ne^2/3m$. Because all electrons participate in the oscillation, such an excitation is called a collective excitation or collective mode of the electron gas.

6. Magnetoplasma frequency. Use the method of Problem 5 to find the frequency of the uniform plasmon mode of a sphere placed in a constant uniform magnetic field $B$. Let $\mathbf{B}$ be along the $z$ axis. The solution should go to the cyclotron frequency $-\omega_c = -eB/m_c$ in one limit and to $\omega_0 = (4\pi ne^2/3m)^{1/2}$ in another limit. Take the motion in the $x - y$ plane.

7. Photon branch at low wavevector. (a) Find what (56) becomes when $\varepsilon(\infty)$ is taken into account. (b) Show that there is a solution of (55) which at low wavevector is $\omega = cK/\sqrt{\varepsilon(0)}$, which is what you expect for a photon in a crystal of refractive index $n^2 = \varepsilon$.

8. Plasma frequency and electrical conductivity. An organic conductor has recently been found by optical studies to have $\omega_p = 1.80 \times 10^{15}$ s$^{-1}$ for the plasma frequency, and $\tau = 2.83 \times 10^{-15}$ s for the electron relaxation time at room temperature. (a) Calculate the electrical conductivity from these data. The carrier mass is not known and is not needed here. Take $\varepsilon(\infty) = 1$. Convert the result to units (\Omega cm)$^{-1}$. (b) From the crystal and chemical structure, the conduction electron concentration is $4.7 \times 10^{21}$ cm$^{-3}$. Calculate the electron effective mass $m^*$. 

9. Bulk modulus of the Fermi gas. Show that the contribution of the kinetic energy to the bulk modulus of the electron gas at absolute zero is $B = 1/3 m v_F^2$. It is convenient to use (6.60). We can use our result for $B$ to find the velocity of sound, which in a compressible fluid is $v = (B/\rho)^{1/2}$, whence $v = (m/3M)^{1/2} v_F$, in agreement with (46). These estimates neglect attractive interactions.
10. **Response of electron gas.** It is sometimes stated erroneously in books on electromagnetism that the static conductivity $\sigma$, which in gaussian units has the dimensions of a frequency, measures the response frequency of a metal to an electric field suddenly applied. Criticize this statement as it might apply to copper at room temperature. The resistivity is $\sim 1$ $\mu$ohm-cm; the electron concentration is $8 \times 10^{22}$ cm$^{-3}$, the mean free path is $\sim 400$ Å; the Fermi velocity is $1.6 \times 10^5$ cm s$^{-1}$. You will not necessarily need all these data. Give the order of magnitude of the three frequencies $\sigma$, $\omega_p$, and $1/\tau$ that might be relevant in the problem. Set up and solve the problem of the response $x(t)$ of the system to an electric field $E(t < 0) = 0$; $E(t > 0) = 1$. The system is a sheet of copper; the field is applied normal to the sheet. Include the damping. Solve the differential equation by elementary methods.

*11. **Gap plasmons and the van der Waals interaction.** Consider two semi-infinite media with plane surfaces $z = 0, d$. The dielectric function of the identical media is $\varepsilon(\omega)$. Show that for surface plasmons symmetrical with respect to the gap the frequency must satisfy $\varepsilon(\omega) = -\tanh(Kd/2)$, where $K^2 = k_x^2 + k_y^2$. The electric potential will have the form

$$\varphi = f(z) \exp(i k_x x + i k_y y - i \omega t).$$

Look for nonretarded solutions—that is, solutions of the Laplace equation rather than of the wave equation. The sum of the zero-point energy of all gap modes is the nonretarded part of the van der Waals attraction between the two specimens—see N. G. van Kampen, B. R. A. Nijboer, and K. Schram, Physics Letters 26A, 307 (1966).

**References**


**SOLID STATE PLASMAS**


**POLARONS**


*This problem is somewhat difficult.*
# Optical Processes and Excitons

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Optical Reflectance</strong></td>
<td></td>
</tr>
<tr>
<td>Kramers-Kronig relations</td>
<td>307</td>
</tr>
<tr>
<td>Example: conductivity of collisionless electron gas</td>
<td>308</td>
</tr>
<tr>
<td>Electronic interband transitions</td>
<td>311</td>
</tr>
<tr>
<td><strong>Excitons</strong></td>
<td></td>
</tr>
<tr>
<td>Frenkel excitons</td>
<td>312</td>
</tr>
<tr>
<td>Alkali halides</td>
<td>316</td>
</tr>
<tr>
<td>Molecular crystals</td>
<td>318</td>
</tr>
<tr>
<td>Weakly bound (Mott-Wannier) excitons</td>
<td>318</td>
</tr>
<tr>
<td>Exciton condensation into electron-hole drops (EHD)</td>
<td>319</td>
</tr>
<tr>
<td><strong>Raman Effect in Crystals</strong></td>
<td></td>
</tr>
<tr>
<td>Electron spectroscopy with x-rays</td>
<td>322</td>
</tr>
<tr>
<td><strong>Energy Loss of Fast Particles in a Solid</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td>324</td>
</tr>
<tr>
<td><strong>Problems</strong></td>
<td></td>
</tr>
<tr>
<td>1. Causality and the response function</td>
<td>325</td>
</tr>
<tr>
<td>2. Dissipation sum rule</td>
<td>326</td>
</tr>
<tr>
<td>3. Reflection at normal incidence</td>
<td>329</td>
</tr>
<tr>
<td>4. Conductivity sum rule and superconductivity</td>
<td>330</td>
</tr>
<tr>
<td>5. Dielectric constant and the semiconductor energy gap</td>
<td>330</td>
</tr>
<tr>
<td>6. Hagen-Rubens relation for infrared reflectivity of metals</td>
<td>331</td>
</tr>
<tr>
<td>7. Davydov splitting of exciton lines</td>
<td>331</td>
</tr>
<tr>
<td><strong>References</strong></td>
<td>332</td>
</tr>
</tbody>
</table>
Raman scattering (generic term).
Brillouin scattering when acoustic phonon is involved, polaron scattering when optical phonon is involved.

\[ \omega = \omega' \pm \Omega \]
\[ k = k' \pm K \]

Signs:
\[ + \text{ for phonon emission (Stokes process)} \]
\[ - \text{ for phonon absorption (anti-Stokes)} \]

Two phonon infrared absorption:
the absorption is strongest for phonons near the zone boundary where the density of modes is a maximum.

Electron spectroscopy with x-rays (XPS):
incident x-ray photon ejects valence or core electron from solid.

Figure 1 There are many types of experiments in which light interacts with wavelike excitations in a crystal. Several processes are illustrated here.
CHAPTER 11: OPTICAL PROCESSES AND EXCITONS

The dielectric function $\epsilon(\omega, K)$ was introduced in the preceding chapter to describe the response of a crystal to an electromagnetic field. The dielectric function depends sensitively on the electronic band structure of a crystal, and studies of the dielectric function by optical spectroscopy are very useful in the determination of the overall band structure of a crystal. Indeed, in the last decade optical spectroscopy developed into the single most important experimental tool for band structure determination.

In the infrared, visible, and ultraviolet spectral regions the wavevector of the radiation is very small compared with the shortest reciprocal lattice vector, and therefore it may usually be taken as zero. We are concerned then with the real $\epsilon'$ and imaginary $\epsilon''$ parts of the dielectric function $\epsilon(\omega) = \epsilon'(\omega) + i\epsilon''(\omega)$, also written as $\epsilon_1(\omega) + i\epsilon_2(\omega)$.

However, the dielectric function is not directly accessible experimentally from optical measurements: the directly accessible functions are the reflectance $R(\omega)$, the refractive index $n(\omega)$, and the extinction coefficient $K(\omega)$. Our first objective is to relate the experimentally observable quantities to the real and imaginary parts of the dielectric function.

OPTICAL REFLECTANCE

The optical measurements that give the fullest information on the electronic system are measurements of the reflectivity of light at normal incidence on single crystals. The reflectivity coefficient $r(\omega)$ is a complex function defined at the crystal surface as the ratio of the reflected electric field $E(\text{refl})$ to the incident electric field $E(\text{inc})$:

$$E(\text{refl})/E(\text{inc}) = r(\omega) = \rho(\omega) \exp[i\theta(\omega)] ,$$  \hspace{1cm} (1)

where we have separated the amplitude $\rho(\omega)$ and phase $\theta(\omega)$ components of the reflectivity coefficient.

The refractive index $n(\omega)$ and the extinction coefficient $K(\omega)$ in the crystal are related to the reflectivity at normal incidence by

$$r(\omega) = \frac{n + iK - 1}{n + iK + 1} ,$$  \hspace{1cm} (2)

as derived in Problem 3 by elementary consideration of the continuity of the components of $E$ and $B$ parallel to the crystal surface. By definition $n(\omega)$ and $K(\omega)$ are related to the dielectric function $\epsilon(\omega)$ by

$$\sqrt{\epsilon(\omega)} = n(\omega) + iK(\omega) \equiv N(\omega) ,$$  \hspace{1cm} (3)
where \( N(\omega) \) is the complex refractive index. Do not confuse \( K(\omega) \) as used here with a wavevector.

If the incident traveling wave has the wavevector \( k \), then the \( y \) component of a wave traveling in the \( x \) direction is

\[
E_y(\text{inc}) = E_{y0} \exp[i(kx - \omega t)] .
\]  
(4)

The transmitted wave in the medium is attenuated because, by the dispersion relation for electromagnetic waves, the wavevector in the medium is related to the incident \( k \) in vacuum by \( (n + iK)k \):

\[
E_y(\text{trans}) \propto \exp[i((n + iK)kx - \omega t)] = \exp(-Kkx) \exp[i(nkx - \omega t)],
\]  
(5)

One quantity measured in experiments is the reflectance \( R \), defined as the ratio of the reflected intensity to the incident intensity:

\[
R = \frac{E^*(\text{refl})E(\text{refl})}{E^*(\text{inc})E(\text{inc})} = r*r = \rho^2 .
\]  
(6)

It is difficult to measure the phase \( \theta(\omega) \) of the reflected wave, but we show below that it can be calculated from the measured reflectance \( R(\omega) \) if this is known at all frequencies.

Once we know both \( R(\omega) \) and \( \theta(\omega) \), we can proceed by (2) to obtain \( n(\omega) \) and \( K(\omega) \). We use these in (3) to obtain \( \epsilon(\omega) = \epsilon'(\omega) + i\epsilon''(\omega) \), where \( \epsilon'(\omega) \) and \( \epsilon''(\omega) \) are the real and imaginary parts of the dielectric function. The inversion of (3) gives

\[
\epsilon'(\omega) = n^2 - K^2 ; \quad \epsilon''(\omega) = 2nK .
\]  
(7)

We now show how to find the phase \( \theta(\omega) \) as an integral over the reflectance \( R(\omega) \); by a similar method we relate the real and imaginary parts of the dielectric function. In this way we can find everything from the experimental \( R(\omega) \).

**Kramers-Kronig Relations**

The Kramers-Kronig relations enable us to find the real part of the response of a linear passive system if we know the imaginary part of the response at all frequencies, and vice versa. They are central to the analysis of optical experiments on solids.

The response of any linear passive system can be represented as the superposition of the responses of a collection of damped harmonic oscillators. Let the response function \( \alpha(\omega) = \alpha'(\omega) + i\alpha''(\omega) \) of the collection of oscillators be defined by

\[
x_\omega = \alpha(\omega) F_\omega ,
\]  
(8)

where the applied force field is the real part of \( F_\omega \exp(-i\omega t) \) and the total displacement \( x = \sum_j x_j \) is the real part of \( x_\omega \exp(-i\omega t) \). From the equation of motion,

\[
M_j(d^2x_j/dt^2 + \rho_jdx_j/dt + \omega_j^2 x_j) = F ,
\]
we have the complex response function of the oscillator system:

\[ \alpha(\omega) = \sum_j \frac{f_j}{\omega_j^2 - \omega^2 - i\omega \rho_j} = \sum f_j \frac{\omega_j^2 - \omega^2 + i\omega \rho_j}{(\omega_j^2 - \omega^2)^2 + \omega^2 \rho_j^2}, \tag{9} \]

where the constants \( f_j = 1/M_j \) and relaxation frequencies \( \rho_j \) are all positive for a passive system. The masses are \( M_j \).

If \( \alpha(\omega) \) is the dielectric polarizability of atoms in concentration \( n \), then \( f \) has the form of an oscillator strength times \( ne^2/m \); such a dielectric response function is said to be of the Kramers-Heisenberg form. The relations we develop also apply to the electrical conductivity \( \sigma(\omega) \) in Ohm's law, \( j_\omega = \sigma(\omega)E_\omega \).

We need not assume the specific form (9), but we make use of three properties of the response function viewed as a function of the complex variable \( \omega \). Any function with the following properties will satisfy the Kramers-Kronig relations (11):

(a) The poles of \( \alpha(\omega) \) are all below the real axis.

(b) The integral of \( \alpha(\omega)/\omega \) vanishes when taken around an infinite semicircle in the upper half of the complex \( \omega \)-plane. It suffices that \( \alpha(\omega) \to 0 \) uniformly as \( |\omega| \to \infty \).

(c) The function \( \alpha'(\omega) \) is even and \( \alpha''(\omega) \) is odd with respect to real \( \omega \).

Consider the Cauchy integral in the form

\[ \alpha(\omega) = \frac{1}{\pi i} \text{P} \int_{-\infty}^{\infty} \frac{\alpha(s)}{s - \omega} \, ds \tag{10}, \]

where \( \text{P} \) denotes the principal part of the integral, as discussed in the mathematical note that follows. The right-hand side is to be completed by an integral over the semicircle at infinity in the upper half-plane, but we have seen in (b) that this integral vanishes.

We equate the real parts of (10) to obtain

\[ \alpha'(\omega) = \frac{1}{\pi} \text{P} \int_{-\infty}^{\infty} \frac{\alpha''(s)}{s - \omega} \, ds = \frac{1}{\pi} \text{P} \left[ \int_{0}^{\infty} \frac{\alpha''(s)}{s - \omega} \, ds + \int_{-\infty}^{0} \frac{\alpha''(p)}{p - \omega} \, dp \right]. \]

In the last integral we substitute \( s \) for \( -p \) and use property (c) that \( \alpha''(-s) = -\alpha''(s) \); this integral then becomes

\[ \int_{0}^{\infty} \frac{\alpha''(s)}{s + \omega} \, ds, \]

and we have, with

\[ \frac{1}{s - \omega} + \frac{1}{s + \omega} = \frac{2s}{s^2 - \omega^2}, \]
the result

\[
\alpha'(\omega) = \frac{2}{\pi} \text{P} \int_0^\infty \frac{sa''(s)}{s^2 - \omega^2} \, ds .
\] (11a)

This is one of the Kramers-Kronig relations. The other relation follows on equating the imaginary parts of Eq. (10):

\[
\alpha''(\omega) = -\frac{1}{\pi} \text{P} \int_{-\infty}^\infty \frac{\alpha'(s)}{s - \omega} \, ds = -\frac{1}{\pi} \text{P} \left[ \int_0^\infty \frac{\alpha'(s)}{s - \omega} \, ds - \int_0^\infty \frac{\alpha'(s)}{s + \omega} \, ds \right],
\]

whence

\[
\alpha''(\omega) = -\frac{2\omega}{\pi} \text{P} \int_0^\infty \frac{\alpha'(s)}{s^2 - \omega^2} \, ds .
\] (11b)

These relations are applied below to the analysis of optical reflectance data; this is their most important application.

Let us apply the Kramers-Kronig relations to \(r(\omega)\) viewed as a response function between the incident and reflected waves in (1) and (6). We apply (11) to

\[
\ln r(\omega) = \ln R^{1/2}(\omega) + i\theta(\omega)
\] (12)

to obtain the phase in terms of the reflectance:

\[
\theta(\omega) = -\frac{\omega}{\pi} \text{P} \int_0^\infty \frac{\ln R(s)}{s^2 - \omega^2} \, ds .
\] (13)

We integrate by parts to obtain a form that gives insight into the contributions to the phase angle:

\[
\theta(\omega) = -\frac{1}{2\pi} \int_0^\infty \ln \left| \frac{s + \omega}{s - \omega} \right| \frac{d \ln R(s)}{ds} \, ds .
\] (14)

Spectral regions in which the reflectance is constant do not contribute to the integral; further, spectral regions \(s \gg \omega\) and \(s \ll \omega\) do not contribute much because the function \(\ln \lvert (s + \omega)/(s - \omega)\rvert\) is small in these regions.

**Mathematical Note.** To obtain the Cauchy integral (10) we take the integral \(\int \alpha(s)(s - \omega)^{-1} \, ds\) over the contour\(^1\) in Fig. 2. The function \(\alpha(s)\) is analytic in the upper half-plane, so that the value of the integral is zero. The contribution of segment 4 to the integral vanishes if the integrand \(\alpha(s)/s \to 0\) faster than \(|s|^{-1}\) as \(|s| \to \infty\). For the response function (9) the integrand \(\to 0\) as \(|s|^{-3}\); and for the

conductivity $\sigma(s)$ the integrand $\to 0$ as $|s|^{-2}$. The segment 2 contributes, in the limit as $u \to 0$,
\[ \int_{(2)} \frac{\alpha(s)}{s - \omega} ds \to \alpha(\omega) \int_{\pi}^{0} \frac{i u e^{i\theta} d\theta}{u e^{i\theta}} = -\pi i \alpha(\omega) \]
to the integral, where $s = \omega + u e^{i\theta}$. The segments 1 and 3 are by definition the principal part of the integral between $-\infty$ and $\infty$. Because the integral over $1 + 2 + 3 + 4$ must vanish,
\[ \int_{(1)} + \int_{(3)} = \text{P} \int_{-\infty}^{\infty} \frac{\alpha(s)}{s - \omega} ds = \pi i \alpha(\omega) \quad (15) \]
as in (10).

**EXAMPLE: Conductivity of Collisionless Electron Gas.** Consider a gas of free electrons in the limit as the collision frequency goes to zero. From (9) the response function is, with $f = 1/m$,
\[ \alpha(\omega) = -\frac{1}{m \omega} \lim_{\rho \to 0} \frac{1}{\omega + i\rho} = -\frac{1}{m \omega} \left[ \frac{1}{\omega} - i\pi \delta(\omega) \right] \quad (16) \]
by the Dirac identity. We confirm that the delta function in (16) satisfies the Kramers-Kronig relation (11a), by which
\[ \alpha'(\omega) = \frac{2}{m} \int_{0}^{\infty} \frac{\delta(s)}{s^2 - \omega^2} ds = -\frac{1}{m \omega^2} \quad (17) \]
in agreement with (16).

We obtain the electrical conductivity $\sigma(\omega)$ from the dielectric function
\[ \epsilon(\omega) - 1 = 4\pi P_\omega/E_\omega = -4\pi n e x_\omega/E_\omega = 4\pi n e^2 \alpha(\omega) \quad (18) \]
where $\alpha(\omega) = x_\omega/(-e)E_\omega$ is the response function. We use the equivalence
\[ \text{CGS} \quad \quad \sigma(\omega) = (-i\omega/4\pi) [\epsilon(\omega) - 1] \quad (19) \]
for the Maxwell equation can be written either as $\nabla \times \mathbf{H} = 4\pi \sigma(\omega) \mathbf{E} - i\omega \mathbf{E}$ or as $\nabla \times \mathbf{H} = -i\omega \varepsilon(\omega) \mathbf{E}$. We combine (16), (18), and (19) to find the conductivity of a collisionless electron gas:
\[ \sigma'(\omega) + i\sigma''(\omega) = \frac{ne^2}{m} \left[ \pi \delta(\omega) + \frac{i}{\omega} \right] \quad (20) \]
For collisionless electrons the real part of the conductivity has a delta function at $\omega = 0$. 

Electronic Interband Transitions

It came as a surprise that optical spectroscopy developed as an important experimental tool for the determination of band structure. First, the absorption and reflection bands of crystals are broad and apparently featureless functions of the photon energy when this is greater than the band gap. Second, direct interband absorption of a photon $\hbar \omega$ will occur at all points in the Brillouin zone for which energy is conserved:

$$\hbar \omega = \epsilon_c(k) - \epsilon_v(k) ,$$  \hspace{1cm} (21)

where $c$ is an empty band and $v$ is a filled band. The total absorption at given $\omega$ is an integral over all transitions in the zone that satisfy (21).

Three factors unraveled the spectra:

- The broad bands are not like a spectral line greatly broadened by damping, but the bands convey much intelligence which emerges when derivatives are taken of the reflectance (Fig. 3); derivatives with respect to wavelength, electric field, temperature, pressure, or uniaxial stress, for example. The spectroscopy of derivatives is called modulation spectroscopy.
- The relation (21) does not exclude spectral structure in a crystal, because transitions accumulate at frequencies for which the bands $c, v$ are parallel—that is, at frequencies where

$$\nabla_k [\epsilon_c(k) - \epsilon_v(k)] = 0 .$$  \hspace{1cm} (22)

At these critical points in $k$ space the joint density of states $D_c(\epsilon_c + \hbar \omega)D_v(\epsilon_v)$ is singular, according to the same argument we used in (5.37) to show that the density of phonon modes $D(\omega)$ is singular when $\nabla_k \omega$ is zero.
- The pseudopotential method for calculating energy bands helps identify the positions in the Brillouin zone of the critical points found in modulation spectra. Band-band energy differences can be calculated with an accuracy as good as 0.1 eV. The experimental results can then be fed back to give improvements in the pseudopotential calculations.

EXCITONS

Reflectance and absorption spectra often show structure for photon energies just below the energy gap, where we might expect the crystal to be transparent. This structure is caused by the absorption of a photon with the creation of a bound electron-hole pair. An electron and a hole may be bound together by their attractive coulomb interaction, just as an electron is bound to a proton to form a neutral hydrogen atom.

The bound electron-hole pair is called an exciton, Fig. 4. An exciton can move through the crystal and transport energy; it does not transport charge
because it is electrically neutral. It is similar to positronium, which is formed from an electron and a positron.

Excitons can be formed in every insulating crystal. When the band gap is indirect, excitons near the direct gap may be unstable with respect to decay into a free electron and free hole. All excitons are unstable with respect to the ultimate recombination process in which the electron drops into the hole. Excitons can also form complexes, such as a biexciton from two excitons.

We have seen that a free electron and free hole are created whenever a photon of energy greater than the energy gap is absorbed in a crystal. The threshold for this process is $\hbar \omega > E_g$ in a direct process. In the indirect phonon-assisted process of Chapter 8 the threshold is lower by the phonon energy $\hbar \Omega$. But in the formation of excitons the energy is lowered with respect to these thresholds by the binding energy of the exciton, which may be in the range 1 meV to 1 eV, as in Table 1.

Excitons can be formed by photon absorption at any critical point (22), for if $\nabla_k \epsilon_e = \nabla_k \epsilon_h$ the group velocities of electron and hole are equal and the particles may be bound by their coulomb attraction. Transitions leading to the formation of excitons below the energy gap are indicated in Figs. 5 and 6.
Figure 4a  An exciton is a bound electron-hole pair, usually free to move together through the crystal. In some respects it is similar to an atom of positronium, formed from a positron and an electron. The exciton shown is a Mott-Wannier exciton: it is weakly bound, with an average electron-hole distance large in comparison with a lattice constant.

Figure 4b  A tightly bound or Frenkel exciton shown localized on one atom in an alkali halide crystal. An ideal Frenkel exciton will travel as a wave throughout the crystal, but the electron is always close to the hole.

Table 1  Binding energy of excitons, in meV

<table>
<thead>
<tr>
<th></th>
<th>Binding Energy (meV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>14.7</td>
</tr>
<tr>
<td>Ge</td>
<td>4.15</td>
</tr>
<tr>
<td>GaAs</td>
<td>4.2</td>
</tr>
<tr>
<td>GaP</td>
<td>3.5</td>
</tr>
<tr>
<td>CdS</td>
<td>29.</td>
</tr>
<tr>
<td>CdSe</td>
<td>15.</td>
</tr>
<tr>
<td>BaO</td>
<td>56.</td>
</tr>
<tr>
<td>InP</td>
<td>4.0</td>
</tr>
<tr>
<td>InSb</td>
<td>(0.4)</td>
</tr>
<tr>
<td>KI</td>
<td>480.</td>
</tr>
<tr>
<td>KCl</td>
<td>400.</td>
</tr>
<tr>
<td>KBr</td>
<td>400.</td>
</tr>
<tr>
<td>RbCl</td>
<td>440.</td>
</tr>
<tr>
<td>LiF</td>
<td>(1000)</td>
</tr>
<tr>
<td>AgBr</td>
<td>20.</td>
</tr>
<tr>
<td>AgCl</td>
<td>30.</td>
</tr>
<tr>
<td>TlCl</td>
<td>11.</td>
</tr>
<tr>
<td>TlBr</td>
<td>6.</td>
</tr>
</tbody>
</table>

Data assembled by Frederick C. Brown and Arnold Schmidt.
Figure 5  Exciton levels in relation to the conduction band edge, for a simple band structure with both conduction and valence band edges at \( k = 0 \). An exciton can have translational kinetic energy. Excitons are unstable with respect to radiative recombination in which the electron drops into the hole in the valence band, accompanied by the emission of a photon or phonons.

Figure 6  Energy levels of an exciton created in a direct process. Optical transitions from the top of the valence band are shown by the arrows; the longest arrow corresponds to the energy gap. The binding energy of the exciton is \( E_{ex} \), referred to a free electron and free hole. The lowest frequency absorption line of the crystal at absolute zero is not \( E_{ex} \), but is \( E_g - E_{ex} \).

The binding energy of the exciton can be measured in three ways:

- In optical transitions from the valence band, by the difference between the energy required to create an exciton and the energy to create a free electron and free hole, Fig. 7.
- In recombination luminescence, by comparison of the energy of the free electron-hole recombination line with the energy of the exciton recombination line.
- By photo-ionization of excitons, to form free carriers. This experiment requires a high concentration of excitons.
Figure 7 Effect of an exciton level on the optical absorption of a semiconductor for photons of energy near the band gap $E_g$ in gallium arsenide at 21 K. The vertical scale is the intensity absorption coefficient, as in $I(x) = I_0 \exp(-\alpha x)$. The energy gap and exciton binding energy are deduced from the shape of the absorption curve: the gap $E_g$ is 1.521 eV and the exciton binding energy is 0.0034 eV. (After M. D. Sturge.)

We discuss excitons in two different limiting approximations, one by Frenkel in which the exciton is small and tightly bound, and the other by Mott and Wannier in which the exciton is weakly bound, with an electron-hole separation large in comparison with a lattice constant. Intermediate examples are known.

**Frenkel Excitons**

In a tightly bound exciton (Fig. 4b) the excitation is localized on or near a single atom: the hole is usually on the same atom as the electron although the pair may be anywhere in the crystal. A Frenkel exciton is essentially an excited state of a single atom, but the excitation can hop from one atom to another by virtue of the coupling between neighbors. The excitation wave travels through the crystal much as the reversed spin of a magnon travels through the crystal.

The crystalline inert gases have excitons which in their ground states correspond somewhat to the Frenkel model. Atomic krypton has its lowest strong atomic transition at 9.99 eV. The corresponding transition in the crystal is closely equal and is at 10.17 eV, Fig. 8. The energy gap in the crystal is 11.7 eV, so the exciton ground state energy is $11.7 - 10.17 = 1.5$ eV, referred to a free electron and free hole separated and at rest in the crystal.

The translational states of Frenkel excitons have the form of propagating waves, like all other excitations in a periodic structure. Consider a crystal of $N$
atoms on a line or ring. If $u_j$ is the ground state of atom $j$, the ground state of the crystal is

$$\psi_k = u_1 u_2 \cdots u_{N-1} u_N ,$$

if interactions between the atoms are neglected. If a single atom $j$ is in an excited state $v_j$, the system is described by

$$\varphi_j = u_1 u_2 \cdots u_{j-1} v_j u_{j+1} \cdots u_N .$$

This function has the same energy as the function $\varphi_l$ with any other atom $l$ excited. However, the functions $\varphi$ that describe a single excited atom and $N - 1$ atoms in their ground state are not the stationary quantum states of the problem. If there is any interaction between an excited atom and a nearby atom in its ground state, the excitation energy will be passed from atom to atom. The eigenstates will have a wavelike form, as we now show.

When the hamiltonian of the system operates on the function $\varphi_j$ with the $j$th atom excited, we obtain

$$\mathcal{H} \varphi_j = \epsilon \varphi_j + T(\varphi_{j-1} + \varphi_{j+1}) ,$$

where $\epsilon$ is the free atom excitation energy; the interaction $T$ measures the rate of transfer of the excitation from $j$ to its nearest neighbors, $j - 1$ and $j + 1$. The solutions of (25) are waves of the Bloch form:

$$\psi_k = \sum_j \exp(ik a) \varphi_j .$$

Figure 8 Absorption spectrum of solid krypton at 20 K. (After C. Baldini.)
To see this we let $\mathcal{H}$ operate on $\psi_k$:

$$\mathcal{H}\psi_k = \sum_j e^{ijk}a \mathcal{H}\varphi_j = \sum_j e^{ijk}[\epsilon\varphi_j + T(\varphi_{j-1} + \varphi_{j+1})] ,$$

(27)

from (25). We rearrange the right-hand side to obtain

$$\mathcal{H}\psi_k = \sum_j e^{ijk}[\epsilon + T(e^{ika} + e^{-ika})]\varphi_j = (\epsilon + 2T \cos ka)\psi_k ,$$

(28)

so that the energy eigenvalues of the problem are

$$E_k = \epsilon + 2T \cos ka ,$$

(29)

as in Fig. 9. The application of periodic boundary conditions determines the allowed values of the wavevector $k$:

$$k = 2\pi s/Na; \quad s = -\frac{1}{2}N, -\frac{1}{2}N + 1, \cdots, \frac{1}{2}N - 1 .$$

(30)

**Alkali Halides.** In alkali halide crystals the lowest-energy excitons are localized on the negative halogen ions, as in Fig. 4b. The negative ions have lower electronic excitation levels than do the positive ions. Pure alkali halide crystals are transparent in the visible spectral region, which means that the exciton energies do not lie in the visible, but the crystals show considerable excitonic absorption structure in the vacuum ultraviolet.

A doublet structure is particularly evident in sodium bromide, a structure similar to that of the lowest excited state of the krypton atom—which is isoelectronic with the Br$^-$ ion of KBr. The splitting is caused by the spin-orbit interaction. These excitons are Frenkel excitons.

**Molecular Crystals.** In molecular crystals the covalent binding within a molecule is strong in comparison with the van der Waals binding between molecules, so that the excitons are Frenkel excitons. Electronic excitation lines of an individual molecule appear in the crystalline solid as an exciton, often with little shift in frequency. At low temperatures the lines in the solid are quite sharp, although there may be more structure to the lines in the solid than in the molecule because of the Davydov splitting, as discussed in Problem 7.

**Weakly Bound (Mott-Wannier) Excitons**

Consider an electron in the conduction band and a hole in the valence band. The electron and hole attract each other by the coulomb potential

(CGS) \[ U(r) = -\frac{e^2}{\epsilon r} , \]

(31)

where $r$ is the distance between the particles and $\epsilon$ is the appropriate dielectric constant. (The lattice polarization contribution to the dielectric constant should not be included if the frequency of motion of the exciton is higher than the optical phonon frequencies.) There will be bound states of the exciton system having total energies lower than the bottom of the conduction band.
The problem is the hydrogen atom problem if the energy surfaces for the electron and hole are spherical and nondegenerate. The energy levels referred to the top of the valence band are given by a modified Rydberg equation

$E_n = E_g - \frac{\mu e^4}{2\hbar^2 \varepsilon n^2}$ \hspace{1cm} (32)

Here $n$ is the principal quantum number and $\mu$ is the reduced mass:

$\frac{1}{\mu} = \frac{1}{m_e} + \frac{1}{m_h}$ \hspace{1cm} (33)

formed from the effective masses $m_e$, $m_h$ of the electron and hole.

The exciton ground state energy is obtained on setting $n = 1$ in (32); this is the ionization energy of the exciton. Studies of the optical absorption lines in cuprous oxide, Cu$_2$O, at low temperatures give results for the exciton level spacing in good agreement with the Rydberg equation (32) except for transitions to the state $n = 1$. An empirical fit to the lines of Fig. 10 is obtained with the relation $\nu$(cm$^{-1}) = 17,508 - (800/n^2)$. Taking $\varepsilon = 10$, we find $\mu = 0.7$ m from the coefficient of $1/n^2$. The constant term 17,508 cm$^{-1}$ corresponds to an energy gap $E_g = 2.17$ eV.

**Exciton Condensation into Electron-Hole Drops (EHD)**

A condensed phase of an electron-hole plasma forms in Ge and Si when maintained at a low temperature and irradiated by light. The following sequence of events takes place when an electron-hole drop (EHD) is formed in Ge: The absorption of a photon of energy $\hbar \omega > E_g$ produces a free electron and
free hole, with high efficiency. These combine rapidly, perhaps in 1 ns, to form an exciton. The exciton may decay with annihilation of the e-h pair with a lifetime of 8 μs.

But if the exciton concentration is sufficiently high—over $10^{13}$ cm$^{-3}$ at 2 K—most of the excitons will condense into a drop. The drop lifetime is 40 μs, but in strained Ge may be as long as 600 μs. Within the drop the excitons dissolve into a degenerate Fermi gas of electrons and holes, with metallic properties: this state was predicted by L. V. Keldysh. The binding energy in Ge is 1.8 meV with respect to free excitons, and the concentration $n = p = 2.57 \times 10^{17}$ cm$^{-3}$.

Experimental studies of the condensed EHD phase have utilized the methods of recombination (luminescence) radiation, light scattering, plasma resonance, Alfvén wave resonance, and $p-n$ junction noise.

Figure 11 shows the recombination radiation in Ge from free excitons (714 meV) and from the EHD phase (709 meV). The width of the 714 meV line is accounted for by Doppler broadening, and the width of the 709 meV line is compatible with the kinetic energy distribution of electrons and holes in a Fermi gas of concentration $2 \times 10^{17}$ cm$^{-3}$. Figure 12 is a photograph of a large EHD.
Figure 11  Recombination radiation of free electrons and of electron-hole drops in Ge at 3.04 K. The Fermi energy in the drop is $\epsilon_F$ and the cohesive energy of the drop with respect to a free exciton is $\varphi_s$. (After T. K. Lo.)

Figure 12  Photograph of an electron-hole drop in a 4 mm disk of pure germanium. The drop is the intense spot adjacent to the set screw on the left of the disk. The photograph is the image of the drop obtained by focusing its electron-hole recombination luminescence onto the surface of an infrared-sensitive videcon image tube. (After J. P. Wolfe et al.)
<table>
<thead>
<tr>
<th>Crystal</th>
<th>Binding energy relative to free exciton, in meV</th>
<th>Concentration, n or p, in cm$^{-3}$</th>
<th>Critical temperature, K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ge</td>
<td>1.8</td>
<td>$2.57 \times 10^{17}$</td>
<td>6.7</td>
</tr>
<tr>
<td>Si</td>
<td>9.3</td>
<td>$3.5 \times 10^{18}$</td>
<td>23.</td>
</tr>
<tr>
<td>GaP</td>
<td>17.5</td>
<td>$8.6 \times 10^{18}$</td>
<td>45.</td>
</tr>
<tr>
<td>3C-SiC</td>
<td>17</td>
<td>$10. \times 10^{18}$</td>
<td>41.</td>
</tr>
</tbody>
</table>

Table 2  Electron-hole liquid parameters


The exciton phase diagram for silicon is plotted in the temperature-concentration plane in Fig. 13. The exciton gas is insulating at low pressures. At high pressures (at the right of the diagram) the exciton gas breaks up into a conducting plasma of unpaired electrons and holes. The transition from excitons to the plasma is an example of the Mott transition, Chapter 10. Further data are given in Table 2.

**RAMAN EFFECT IN CRYSTALS**

Raman scattering involves two photons—one in, one out—and is one step more complex than the one photon processes treated earlier in this chapter. In the Raman effect a photon is scattered inelastically by a crystal, with creation or annihilation of a phonon or magnon (Fig. 14). The process is identical to the inelastic scattering of x-rays and it is similar to the inelastic scattering of neutrons by a crystal.

The selection rules for the first-order Raman effect are

$$\omega = \omega' \pm \Omega \quad , \quad k = k' \pm K \quad , \quad (34)$$

where $\omega$, $k$ refer to the incident photon; $\omega'$, $k'$ refer to the scattered photon; and $\Omega$, $K$ refer to the phonon created or destroyed in the scattering event. In the second-order Raman effect, two phonons are involved in the inelastic scattering of the photon.

The Raman effect is made possible by the strain-dependence of the electronic polarizability. To show this, we suppose that the polarizability $\alpha$ associated with a phonon mode may be written as a power series in the phonon amplitude $u$:

$$\alpha = \alpha_0 = \alpha_1 u + \alpha_2 u^2 + \cdots . \quad (35)$$

If $u(t) = u_0 \cos \Omega t$ and the incident electric field is $E(t) = E_0 \cos \omega t$, then the induced electric dipole moment has a component

$$\alpha_1 E_0 u_0 \cos \omega t \cos \Omega t = \frac{1}{2} \alpha_1 E_0 u_0 [\cos(\omega + \Omega) t + \cos(\omega - \Omega) t] \quad . \quad (36)$$
Figure 13 Phase diagram for photoexcited electrons and holes in unstressed silicon. The diagram shows, for example, that with an average concentration near $10^{17} \text{ cm}^{-3}$ at 15 K, a free-exciton gas with saturated-gas concentration of $10^{16} \text{ cm}^{-3}$ coexists with a (variable) volume of liquid droplets, each with a density of $3 \times 10^{18} \text{ cm}^{-3}$. The liquid critical temperature is about 23 K. Theoretical and experimental values for the metal-insulator transition for excitons are also shown. (From J. P. Wolfe.)

Figure 14 Raman scattering of a photon with emission or absorption of a phonon. The process is called Brillouin scattering when an acoustic phonon is involved and polariton scattering when an optical phonon is involved. Similar processes occur with magnons (spin waves).

Thus photons at frequencies $\omega + \Omega$ and $\omega - \Omega$ can be emitted, accompanied by absorption or emission of a phonon of frequency $\Omega$.

The photon at $\omega - \Omega$ is called the Stokes line and that at $\omega + \Omega$ is the anti-Stokes line. The intensity of the Stokes line involves the matrix element for phonon creation, which is just the matrix element for the harmonic oscillator, as in Appendix C:

$$I(\omega - \Omega) \propto |\langle n_\mathbf{K} + 1 | u | n_\mathbf{K} \rangle|^2 \propto n_\mathbf{K} + 1 ,$$

where $n_\mathbf{K}$ is the initial population of phonon mode $\mathbf{K}$.
The anti-Stokes line involves phonon annihilation, with a photon intensity proportional to

$$I(\omega + \Omega) \propto \langle n_K - 1 | u | n_K \rangle^2 \propto n_K.$$  

(38)

If the phonon population is initially in thermal equilibrium at temperature $T$, the intensity ratio of the two lines is

$$\frac{I(\omega + \Omega)}{I(\omega - \Omega)} = \frac{\langle n_K \rangle}{\langle n_K \rangle + 1} = \exp(-\hbar \Omega / k_B T),$$  

(39)

with $\langle n_K \rangle$ given by the Planck distribution function $1/[\exp(\hbar \Omega / k_B T) - 1]$. We see that the relative intensity of the anti-Stokes lines vanishes as $T \to 0$, because here there are no thermal phonons available to be annihilated.

Observations on the $K = 0$ optical phonon in silicon are shown in Figs. 15 and 16. Silicon has two identical atoms in the primitive cell, and there is no electric dipole moment associated with the primitive cell in the absence of deformation by phonons. But $\alpha_{1u}$ does not vanish for silicon at $K = 0$, so that we can observe the mode by first-order Raman scattering of light.

The second-order Raman effect arises from the term $\alpha_{2u}^2$ in the polarizability. Inelastic scattering of light in this order is accompanied by the creation of two phonons, or the absorption of two phonons, or the creation of one and the absorption of another phonon. The phonons may have different frequencies. The intensity distribution in the scattered photon spectrum may be quite complicated if there are several atoms in the primitive cell because of the corresponding number of optical phonon modes.

Second-order Raman spectra have been observed and analyzed in numerous crystals. Measurements on gallium phosphide (GaP) are shown in Fig. 17.

**Electron Spectroscopy with X-Rays**

The next degree of complexity in optical processes involves a photon in and an electron out of the solid, as in Fig. 1. The important techniques of x-ray photoemission from solids (XPS) and ultraviolet photoemission (UPS) have recently been developed. In solid state physics they are used in band structure studies and surface physics, including catalysis and adsorption.

XPS and UPS spectra can be compared directly with valence band densities of states $D(e)$. The specimen is irradiated with highly monochromatized x-rays or ultraviolet photons. The photon is absorbed, with the emission of a photoelectron whose kinetic energy is equal to the energy of the incident photon minus the binding energy of the electron in the solid. The electrons come from a thin layer near the surface, typically 50 Å in depth.

The resolution of ESCA spectrometer systems is of the order of 0.6 eV, which permits refined studies of band structure, particularly of core electrons. ESCA stands for electron spectroscopy for chemical analysis.
Figure 15  First-order Raman spectra of the $K = 0$ optical mode of a silicon crystal observed at three temperatures. The incident photon has a wavelength of $5145$ Å. The optical phonon frequency is equal to the frequency shift, it depends slightly on the temperature. (After T. R. Hart, R. L. Aggarwal, and B. Lax.)

Figure 16  Intensity ratio anti-Stokes to Stokes lines as a function of temperature, for the observations of Fig. 15 on the optical mode of silicon. The observed temperature dependence is in good agreement with the prediction of Eq. (39); the solid curve is a plot of the function $\exp(-\hbar \Omega/k_B T)$.

Figure 17  Raman spectrum of GaP at $20$ K. The two highest peaks are the first-order Raman lines associated with the excitation of an LO phonon at $404$ cm$^{-1}$ and a TO phonon at $366$ cm$^{-1}$. All the other peaks involve two phonons. (After M. V. Hobden and J. P. Russell.)
The valence band structure of silver is shown by Fig. 18, with the zero of energy set at the Fermi level. Electrons in the first 3 eV below the Fermi level come from the 5s conduction band. The strong peak with structure below 3 eV is from the 4d valence electrons.

Excitations are also seen from deeper levels, often accompanied by excitation of plasmons. For example, in silicon the 2p electron with a binding energy close to 99.2 eV is observed in replica at 117 eV with single plasmon excitation and at 134.7 eV with two plasmon excitation. The plasmon energy is 18 eV.

ENERGY LOSS OF FAST PARTICLES IN A SOLID

So far we have used photons as probes of the electronic structure of solids. We can also use electron beams for the same purpose. The results also involve the dielectric function, now through the imaginary part of $1/\varepsilon(\omega)$. The dielectric function enters as $\textrm{Im}\{\varepsilon(\omega)\}$ into the energy loss by an electromagnetic wave in a solid, but as $-\textrm{Im}\{1/\varepsilon(\omega)\}$ into the energy loss by a charged particle that penetrates a solid.

Consider this difference. The general result from electromagnetic theory for the power dissipation density by dielectric losses is

$$
\mathcal{P} = \frac{1}{4\pi} \mathbf{E} \cdot (\partial \mathbf{D}/\partial t),
$$

per unit volume. With a transverse electromagnetic wave $\mathbf{E}e^{-i\omega t}$ in the crystal, we have $d\mathbf{D}/dt = -i\omega \varepsilon(\omega)\mathbf{E}e^{-i\omega t}$, whence the time-average power is

$$
\mathcal{P} = \frac{1}{4\pi} \langle \textrm{Re}\{\mathbf{E}e^{-i\omega t}\} \textrm{Re}\{-i\omega \varepsilon(\omega)\mathbf{E}e^{-i\omega t}\} \rangle
$$

$$
= \frac{1}{4\pi} \omega \varepsilon^2 \langle (\varepsilon'' \cos \omega t - \varepsilon' \sin \omega t) \cos \omega t \rangle = \frac{1}{8\pi} \omega \varepsilon''(\omega) \varepsilon^2,
$$

proportional to $\varepsilon''(\omega)$. The tangential component of $\mathbf{E}$ is continuous across the boundary of the solid.

If a particle of charge $e$ and velocity $v$ enters a crystal, the dielectric displacement is

$$
(DG) \quad \mathbf{D}(\mathbf{r},t) = -\nabla \frac{e}{|\mathbf{r} - \mathbf{v}t|},
$$

because by the Poisson equation it is $\mathbf{D}$, and not $\mathbf{E}$, that is related to the free charge. In an isotropic medium the Fourier component $E(\omega,\mathbf{k})$ is related to the Fourier component $D(\omega,\mathbf{k})$ of $\mathbf{D}(\mathbf{r},t)$ by $E(\omega,\mathbf{k}) = D(\omega,\mathbf{k})/\varepsilon(\omega,\mathbf{k})$. 
The time-average power dissipation associated with this Fourier component is

$$
\mathcal{P}(\omega,k) = \frac{1}{4\pi} \langle \Re\{\epsilon^{-1}(\omega,k)D(\omega,k)e^{-i\omega t}\}\Re\{-i\omega D(\omega,k)e^{-i\omega t}\} \rangle \\
= \frac{1}{4\pi} \omega D^2(\omega,k) \left\{ \left( \frac{1}{\epsilon} \right) ' \cos \omega t + \left( \frac{1}{\epsilon} \right) '' \sin \omega t \right\} [-\sin \omega t],
$$

whence

$$
\mathcal{P}(\omega,k) = -\frac{1}{8\pi} \omega \left( \frac{1}{\epsilon} \right) '' D^2(\omega,k) = \frac{1}{8\pi} \omega \frac{\epsilon''(\omega,k)}{|\epsilon|^2} D^2(\omega,k).
$$

(43)

The result is the motivation for the introduction of the energy loss function $-\Im\{1/\epsilon(\omega,k)\}$ and it is also a motivation for experiments on energy losses by fast electrons in thin films.

If the dielectric function is independent of $k$, the power loss is

$$
\mathcal{P}(\omega) = -\frac{2}{\pi} \frac{\epsilon^2}{\hbar v} \Im\{1/\epsilon(\omega)\} \ln(k_0v/\omega),
$$

(44)

where $\hbar k_0$ is the maximum possible momentum transfer from the primary particle to an electron of the crystal. Figure 19 shows the excellent experimental agreement between values of $\epsilon''(\omega)$ deduced from optical reflectivity measurements with values deduced from electron energy loss measurements.

Table 3 lists the acronyms for some of the principal experimental methods used in studies of energy band structure, particularly of the gross aspects of the structure. These methods differ from those used in Fermi surface studies, Chapter 9.
Figure 19 $\varepsilon''(\omega)$ for Cu and Au; the bold lines are from energy loss measurements by J. Daniels, and the other lines were calculated from optical measurements by D. Beaglehole, and L. R. Canfield et al.

Table 3 Acronyms of current experimental methods for band structure studies

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>Auger electron spectroscopy</td>
</tr>
<tr>
<td>ELS</td>
<td>Electron (energy) loss spectroscopy</td>
</tr>
<tr>
<td>ERM</td>
<td>Evaporation rate monitoring</td>
</tr>
<tr>
<td>ESCA</td>
<td>Electron spectroscopy for chemical analysis</td>
</tr>
<tr>
<td>EXAFS</td>
<td>Extended x-ray absorption fine structure</td>
</tr>
<tr>
<td>FDMS</td>
<td>Flash desorption mass spectrometry</td>
</tr>
<tr>
<td>FIM</td>
<td>Field ionization microscopy</td>
</tr>
<tr>
<td>IAP</td>
<td>Imaging atom probe</td>
</tr>
<tr>
<td>INS</td>
<td>Inelastic neutron scattering</td>
</tr>
<tr>
<td>IRM</td>
<td>Ion rate monitoring</td>
</tr>
<tr>
<td>ISS</td>
<td>Ion scattering spectrometry</td>
</tr>
<tr>
<td>ITD</td>
<td>Isothermal desorption spectrometry</td>
</tr>
<tr>
<td>LEED</td>
<td>Low energy electron diffraction</td>
</tr>
<tr>
<td>RHEED</td>
<td>Reflection high energy electron diffraction</td>
</tr>
<tr>
<td>SAM</td>
<td>Scanning Auger microscopy</td>
</tr>
<tr>
<td>SEM</td>
<td>Scanning electron microscopy</td>
</tr>
<tr>
<td>SES</td>
<td>Secondary electron spectroscopy</td>
</tr>
<tr>
<td>SIMS</td>
<td>Secondary ion mass spectrometry</td>
</tr>
<tr>
<td>SLEED</td>
<td>Spin-polarized low energy electron diffraction</td>
</tr>
<tr>
<td>TDMS</td>
<td>Thermal desorption mass spectrometry</td>
</tr>
<tr>
<td>TEM</td>
<td>Transmission electron microscopy</td>
</tr>
<tr>
<td>THEED</td>
<td>Transmission high energy electron diffraction</td>
</tr>
<tr>
<td>UPS</td>
<td>Ultraviolet photoelectron spectroscopy</td>
</tr>
<tr>
<td>XPS</td>
<td>X-ray photoelectron spectroscopy</td>
</tr>
</tbody>
</table>
SUMMARY

- The Kramers-Kronig relations connect the real and imaginary parts of a response function:

\[ \alpha'(\omega) = \frac{2}{\pi} \mathcal{P} \int_{0}^{\infty} \frac{s\alpha''(s)}{s^2 - \omega^2} \, ds \quad \alpha''(\omega) = -\frac{2\omega}{\pi} \mathcal{P} \int_{0}^{\infty} \frac{\alpha'(s)}{s^2 - \omega^2} \, ds. \]

- The complex refractive index \( N(\omega) = n(\omega) + iK(\omega) \), where \( n \) is the refractive index and \( K \) is the extinction coefficient; further, \( \epsilon(\omega) = N^2(\omega) \), whence \( \epsilon'(\omega) = n^2 - K^2 \) and \( \epsilon''(\omega) = 2nK \).

- The reflectance at normal incidence is

\[ R = \frac{(n - 1)^2 + K^2}{(n + 1)^2 + K^2}. \]

- The energy loss function \( -\text{Im}\{1/\epsilon(\omega)\} \) gives the energy loss by a charged particle moving in a solid.

Problems

1. **Causality and the response function.** The Kramers-Kronig relations are consistent with the principle that an effect not precede its cause. Consider a delta-function force applied at time \( t = 0 \):

\[ F(t) = \delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} \, d\omega, \]

whence \( F_\omega = 1/2\pi \). (a) Show by direct integration or by use of the KK relations that the oscillator response function

\[ \alpha(\omega) = (\omega_0^2 - \omega^2 - i\omega\rho)^{-1} \]

gives zero displacement, \( x(t) = 0 \), for \( t < 0 \) under the above force. For \( t < 0 \) the contour integral may be completed by a semicircle in the upper half-plane. (b) Evaluate \( x(t) \) for \( t > 0 \). Note that \( \alpha(\omega) \) has poles at \( \pm(\omega_0^2 - 4\rho^2)^{1/2} - i\rho \), both in the lower half-plane.

2. **Dissipation sum rule.** By comparison of \( \alpha'(\omega) \) from (9) and from (11a) in the limit \( \omega \to \infty \), show that the following sum rule for the oscillator strengths must hold:

\[ \sum f_j = \frac{2}{\pi} \int_{0}^{\infty} s\alpha''(s) \, ds \]
3. Reflection at normal incidence. Consider an electromagnetic wave in vacuum, with field components of the form

\[ E_y(\text{inc}) = B_z(\text{inc}) = A e^{i(kx - \omega t)} \, . \]

Let the wave be incident upon a medium of dielectric constant \( \varepsilon \) and permeability \( \mu = 1 \) that fills the half-space \( x > 0 \). Show that the reflectivity coefficient \( r(\omega) \) as defined by \( E(\text{refl}) = r(\omega)E(\text{inc}) \) is given by

\[ r(\omega) = \frac{n + iK - 1}{n + iK + 1} \, , \]

where \( n + iK = \varepsilon^{1/2} \), with \( n \) and \( K \) real. Show further that the reflectance is

\[ R(\omega) = \frac{(n - 1)^2 + K^2}{(n + 1)^2 + K^2} \, . \]

4. Conductivity sum rule and superconductivity. We write the electrical conductivity as \( \sigma(\omega) = \sigma'(\omega) + i\sigma''(\omega) \), where \( \sigma' \), \( \sigma'' \) are real. (a) Show by a Kramers-Kronig relation that

\[ \lim_{\omega \to \infty} \omega \sigma''(\omega) = \frac{2}{\pi} \int_0^\infty \sigma'(s) \, ds \, . \]

This result is used in the theory of superconductivity. If at very high frequencies (such as x-ray frequencies) \( \sigma''(\omega) \) is identical for the superconducting and normal states, then we must have

\[ \int_0^\infty \sigma'_n(\omega) \, d\omega = \int_0^\infty \sigma'_s(\omega) \, d\omega \, . \]

But at frequencies \( 0 < \omega < \omega_c \) within the superconducting energy gap the real part of the conductivity of a superconductor vanishes, so that in this region the integral on the left-hand side is lower by \( = \sigma'_n(\omega_c) \). There must be an additional contribution to \( \sigma'_s \) to balance this deficiency. (b) Show that if \( \sigma'_s(\omega < \omega_c) < \sigma'_n(\omega < \omega_c) \), as is observed experimentally, then \( \sigma'_s(\omega) \) can have a delta function contribution at \( \omega = 0 \), and from the delta function there is a contribution \( \sigma''_s(\omega) \approx \sigma'_n(\omega_c)/\omega \). The delta function corresponds to infinite conductivity at zero frequency. (c) By elementary consideration of the classical motion of conduction electrons at very high frequencies, show that

\[
(CGS) \int_0^\infty \sigma'(\omega) \, d\omega = \frac{\pi e^2}{2m} \, .
\]

5. Dielectric constant and the semiconductor energy gap. The effect on \( \varepsilon''(\omega) \) of an energy gap \( \omega_c \) in a semiconductor may be approximated very roughly by substituting \( \frac{1}{2} \delta(\omega - \omega_c) \) for \( \delta(\omega) \) in the response function (16); that is, we take \( \varepsilon''(\omega) = (2\pi n e^2/m\omega)\pi \delta(\omega - \omega_c) \). This is crude because it puts all the absorption at the gap

*This problem is somewhat difficult.
frequency. The factor 1/2 enters as soon as we move the delta function away from the origin, because the integral in the sum rule of Problem 2 starts at the origin. Show that the real part of the dielectric constant on this model is

\[ \varepsilon'(\omega) = 1 + \omega_p^2/(\omega_s^2 - \omega^2), \quad \omega_p = 4\pi ne^2/m. \]

It follows that the static dielectric constant is \( \varepsilon'(0) = 1 + \omega_p^2/\omega_s^2 \), widely used as a rule of thumb.

6. Hagen-Rubens relation for infrared reflectivity of metals. The complex refractive index \( n + iK \) of a metal for \( \omega \tau \ll 1 \) is given by

\[ \varepsilon(\omega) = (n + iK)^2 = 1 + 4\pi i\sigma_0/\omega, \]

where \( \sigma_0 \) is the conductivity for static fields. We assume here that intraband currents are dominant; interband transitions are neglected. Using the result of Problem 3 for the reflection coefficient at normal incidence, show that

\[ R = 1 - (2\omega/\pi\sigma_0)^{1/2}, \]

provided that \( \sigma_0 \gg \omega \). This is the Hagen-Rubens relation. For sodium at room temperature, \( \sigma_0 = 2.1 \times 10^{17} \text{ s}^{-1} \) in CGS and \( \tau = 3.1 \times 10^{-14} \text{ s} \), as deduced from \( \tau = \sigma_0 m/e^2 \). Radiation of 10 \( \mu \text{m} \) has \( \omega = 1.88 \times 10^{14} \text{ s}^{-1} \), so that the Hagen-Rubens result should apply: \( R = 0.976 \). The result calculated from experimental values of \( n \) and \( K \) is 0.987. Hint: If \( \sigma_0 \gg \omega \), then \( n^2 \approx K^2 \). This simplifies the algebra.

*7. Davydov splitting of exciton lines. The Frenkel exciton band of Fig. 9 is doubled when there are two atoms A, B in a primitive cell. Extend the theory of Eqs. (25) to (29) to a linear crystal AB.AB.AB.AB with transfer integrals \( T_1 \) between AB and \( T_2 \) between B.A. Find an equation for the two bands as functions of the wavevector. The splitting between the bands at \( k = 0 \) is called the Davydov splitting.

*This problem is somewhat difficult.
References

# Superconductivity

## EXPERIMENTAL SURVEY
- Occurrence of superconductivity: 335
- Destruction of superconductivity by magnetic fields: 337
- Meissner effect: 338
- Heat capacity: 338
- Energy gap: 342
- Microwave and infrared properties: 344
- Isotope effect: 345

## THEORETICAL SURVEY
- Thermodynamics of the superconducting transition: 346
- London equation: 347
- Coherence length: 349
- BCS theory of superconductivity: 352
- BCS ground state: 354
- Flux quantization in a superconducting ring: 355
- Duration of persistent currents: 356
- Type II superconductors: 359
- Vortex state: 360
- Estimation of $H_{c1}$ and $H_{c2}$: 361
- Single particle tunneling: 362
- Josephson superconductor tunneling: 364
- DC Josephson effect: 366
- AC Josephson effect: 368
- Macroscopic quantum interference: 369

## HIGH-TEMPERATURE SUPERCONDUCTORS
- Critical fields and critical currents: 371
- Hall number: 372
- Fullerenes: 373

## SUMMARY

**NOTATION:** In this chapter $B_a$ denotes the applied magnetic field. In the CGS system the critical value $B_{ac}$ of the applied field will be denoted by the symbol $H_c$ in accordance with the custom of workers in superconductivity. Values of $B_{ac}$ are given in gauss in CGS units and in teslas in SI units, with $1 \, \text{T} = 10^4 \, \text{G}$. In SI we have $B_{ac} = \mu_0 H_c$. 
PROBLEMS

1. Magnetic field penetration in a plate 375
2. Critical field of thin films 375
3. Two-fluid model of a superconductor 376
4. Structure of a vortex 376
5. London penetration depth 377
6. Diffraction effect of Josephson junction 377
7. Meissner effect in sphere 377

REFERENCES

APPENDICES RELEVANT TO SUPERCONDUCTIVITY

H Cooper Pairs 656
I Ginzburg-Landau Equation 658
J Electron-Phonon Collisions 662

Figure 1  Resistance in ohms of a specimen of mercury versus absolute temperature. This plot by Kamerlingh Onnes marked the discovery of superconductivity.
CHAPTER 12: SUPERCONDUCTIVITY

The electrical resistivity of many metals and alloys drops suddenly to zero when the specimen is cooled to a sufficiently low temperature, often a temperature in the liquid helium range. This phenomenon, called superconductivity, was observed first by Kamerlingh Onnes\(^1\) in Leiden in 1911, three years after he first liquefied helium. At a critical temperature \(T_c\) the specimen undergoes a phase transition from a state of normal electrical resistivity to a superconducting state, Fig. 1.

Superconductivity is now very well understood. It is a field with many practical and theoretical aspects. The length of this chapter and the relevant appendices reflect the richness and subtleties of the field.

EXPERIMENTAL SURVEY

In the superconducting state the dc electrical resistivity is zero, or so close to zero that persistent electrical currents have been observed to flow without attenuation in superconducting rings for more than a year, until at last the experimentalist wearied of the experiment.

The decay of supercurrents in a solenoid was studied by File and Mills\(^2\) using precision nuclear magnetic resonance methods to measure the magnetic field associated with the supercurrent. They concluded that the decay time of the supercurrent is not less than 100,000 years. We estimate the decay time below.

In some superconducting materials, particularly those used for superconducting magnets, finite decay times are observed because of an irreversible redistribution of magnetic flux in the material.

The magnetic properties exhibited by superconductors are as dramatic as their electrical properties. The magnetic properties cannot be accounted for by the assumption that a superconductor is a normal conductor with zero electrical resistivity.

\(^1\)H. Kamerlingh Onnes, Akad. van Wetenschappen (Amsterdam) 14, 113, 818 (1911). "The value of the mercury resistance used was 172.7 ohms in the liquid condition at 0°C; extrapolation from the melting point to 0°C by means of the temperature coefficient of solid mercury gives a resistance corresponding to this of 39.7 ohms in the solid state. At 4.3 K this had sunk to 0.084 ohms, that is, to 0.0021 times the resistance which the solid mercury would have at 0°C. At 3 K the resistance was found to have fallen below \(3 \times 10^{-6}\) ohms, that is to one ten-millionth of the value which it would have at 0°C. As the temperature sunk further to 1.5 K this value remained the upper limit of the resistance." Historical references are given by C. J. Gorter, Rev. Mod. Phys. 36, 1 (1964).

Table 1  Superconductivity parameters of the elements

An asterisk denotes an element superconducting only in thin films or under high pressure in a crystal modification not normally stable. Data courtesy of B. T. Matthias, revised by T. Geballe.

| Li | Be | 0.026 |
| Na | Mg |

| Transition temperature in K | 1.140 |
| Critical magnetic field at absolute zero in gauss ($10^4$ tesla) | 105 |

| K | Ca | Sc | Ti | V | Cr* | Mn | Fe | Co | Ni | Cu | Zn | Ga | Ge* | As* | Se* | Br | Kr |
|---|----|----|----|---|-----|----|----|----|----|----|----|----|-----|-----|----|----|----|----|
| 0.39 | 5.38 | 1420 |
| 100 | 1420 |

| Rb | Sr | Y* | Zr | Nb | Mo | Tc | Ru | Rh | Pd | Ag | Cd | In | Sn (w) | Sb* | Te* | I | Xe |
|---|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|----|----|----|----|
| 0.546 | 9.50 | 0.92 | 7.77 | 0.51 | 0.003 | 53 | 51 |
| 47 | 1980 | 95 | 1410 | 70 | 0.049 |

| Cs* | Ba* | La* | Hf | Ta | W | Re | Os | Ir | Pt | Au | Hg (n) | Tl | Pb | Bi* | Po | At | Rn |
|---|----|----|---|----|---|----|----|----|----|----|-----|----|----|-----|----|----|----|----|
| 6.00 | 1.2 | 4483 | 0.012 | 1.4 | 0.655 | 0.14 | 4153 | 2.39 | 7.193 |
| 1100 | 0.12 | 830 | 1.07 | 198 | 65 | 19 |

<table>
<thead>
<tr>
<th>Fr</th>
<th>Ra</th>
<th>Ac</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.368</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>1.62</td>
<td>1.4</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2. Meissner effect in a superconducting sphere cooled in a constant applied magnetic field, on passing below the transition temperature the lines of induction B are ejected from the sphere.

It is an experimental fact that a bulk superconductor in a weak magnetic field will act as a perfect diamagnet, with zero magnetic induction in the interior. When a specimen is placed in a magnetic field and is then cooled through the transition temperature for superconductivity, the magnetic flux originally present is ejected from the specimen. This is called the Meissner effect. The sequence of events is shown in Fig. 2. The unique magnetic properties of superconductors are central to the characterization of the superconducting state.

The superconducting state is an ordered state of the conduction electrons of the metal. The order is in the formation of loosely associated pairs of electrons. The electrons are ordered at temperatures below the transition temperature, and they are disordered above the transition temperature.

The nature and origin of the ordering was explained by Bardeen, Cooper, and Schrieffer. In the present chapter we develop as far as we can in an elementary way the physics of the superconducting state. We shall also discuss the basic physics of the materials used for superconducting magnets, but not their technology. Appendices H and I give deeper treatments of the superconducting state.

Occurrence of Superconductivity

Superconductivity occurs in many metallic elements of the periodic system and also in alloys, intermetallic compounds, and doped semiconductors. The range of transition temperatures best confirmed at present extends from 90.0 K for the compound YBa$_2$Cu$_3$O$_{6.9}$ to below 0.001 K for the element Rh. Several f-band superconductors, also known as “exotic superconductors,” are listed in Chapter 6. Several materials become superconducting only under high pressure; for example, Si has a superconducting form at 165 kbar, with $T_c = 8.3$ K. The elements known to be superconducting are displayed in Table 1, for zero pressure.

---

Will every nonmagnetic metallic element become a superconductor at sufficiently low temperatures? We do not know. In experimental searches for superconductors with ultralow transition temperatures it is important to eliminate from the specimen even trace quantities of foreign paramagnetic elements, because they can lower the transition temperature severely. One part of Fe in $10^4$ will destroy the superconductivity of Mo, which when pure has $T_c = 0.92$ K; and 1 at. percent of gadolinium lowers the transition temperature of lanthanum from 5.6 K to 0.6 K. Nonmagnetic impurities have no very marked effect on the transition temperature. The transition temperatures of a number of interesting superconducting compounds are listed in Table 2. Several organic compounds show superconductivity at fairly low temperatures.

**Destruction of Superconductivity by Magnetic Fields**

A sufficiently strong magnetic field will destroy superconductivity. The threshold or critical value of the applied magnetic field for the destruction of superconductivity is denoted by $H_c(T)$ and is a function of the temperature. At the critical temperature the critical field is zero: $H_c(T_c) = 0$. The variation of the critical field with temperature for several superconducting elements is shown in Fig. 3.

The threshold curves separate the superconducting state in the lower left of the figure from the normal state in the upper right. Note: We should denote the critical value of the applied magnetic field as $B_{ac}$, but this is not common practice among workers in superconductivity. In the CGS system we shall always understand that $H_c = B_{ac}$, and in the SI we have $H_c = B_{ac} / \mu_0$. The symbol $B_a$ denotes the applied magnetic field.

**Meissner Effect**

Meissner and Ochsenfeld (1933) found that if a superconductor is cooled in a magnetic field to below the transition temperature, then at the transition the lines of induction $B$ are pushed out (Fig. 2). The Meissner effect shows that a bulk superconductor behaves as if inside the specimen $B = 0$.

<table>
<thead>
<tr>
<th>Compound</th>
<th>$T_c$, in K</th>
<th>Compound</th>
<th>$T_c$, in K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nb$_3$Sn</td>
<td>18.05</td>
<td>V$_3$Ga</td>
<td>16.5</td>
</tr>
<tr>
<td>Nb$_3$Ge</td>
<td>23.2</td>
<td>V$_3$Si</td>
<td>17.1</td>
</tr>
<tr>
<td>Nb$_3$Al</td>
<td>17.5</td>
<td>YBa$_2$Cu$_3$O$_6.9$</td>
<td>90.0</td>
</tr>
<tr>
<td>NbN</td>
<td>16.0</td>
<td>Rb$_2$CsC$_6$</td>
<td>31.3</td>
</tr>
<tr>
<td>K$_3$C$_6$O</td>
<td>19.2</td>
<td>La$_3$In</td>
<td>10.4</td>
</tr>
</tbody>
</table>

Table 2 Superconductivity of selected compounds
We obtain a particularly useful form of this result if we limit ourselves to long thin specimens with long axis parallel to \( B_a \); now the demagnetizing field contribution (see Chapter 13) to \( B \) will be negligible, whence: \(^4\)

\[
\begin{align*}
\text{(CGS)} & \quad B = B_a + 4\pi M = 0 ; & \quad \frac{M}{B_a} = \frac{1}{4\pi} ; \\
\text{(SI)} & \quad B = B_a + \mu_0 M = 0 ; & \quad \frac{M}{B_a} = \frac{1}{\mu_0} = -\epsilon_0 c^2 .
\end{align*}
\]

The result \( B = 0 \) cannot be derived from the characterization of a superconductor as a medium of zero resistivity. From Ohm's law, \( E = \rho j \), we see that if the resistivity \( \rho \) goes to zero while \( j \) is held finite, then \( E \) must be zero. By a Maxwell equation \( dB/dt \) is proportional to curl \( E \), so that zero resistivity implies \( dB/dt = 0 \). This argument is not entirely transparent, but the result predicts that the flux through the metal cannot change on cooling through the transition. The Meissner effect contradicts this result and suggests that perfect diamagnetism is an essential property of the superconducting state.

We expect another difference between a superconductor and a perfect conductor, defined as a conductor in which the electrons have an infinite mean free path. When the problem is solved in detail, it turns out that a perfect conductor when placed in a magnetic field cannot produce a permanent eddy current screen: the field will penetrate about 1 cm in an hour. \(^5\)

\(^4\) Diamagnetism, the magnetization \( M \), and the magnetic susceptibility are defined in Chapter 14. The magnitude of the apparent diamagnetic susceptibility of bulk superconductors is very much larger than in typical diamagnetic substances. In (1), \( M \) is the magnetization equivalent to the superconducting currents in the specimen.

Figure 4  (a) Magnetization versus applied magnetic field for a bulk superconductor exhibiting a complete Meissner effect (perfect diamagnetism). A superconductor with this behavior is called a type I superconductor. Above the critical field $H_c$ the specimen is a normal conductor and the magnetization is too small to be seen on this scale. Note that minus $4\pi M$ is plotted on the vertical scale: the negative value of $M$ corresponds to diamagnetism. (b) Superconducting magnetization curve of a type II superconductor. The flux starts to penetrate the specimen at a field $H_{c1}$ lower than the thermodynamic critical field $H_c$. The specimen is in a vortex state between $H_{c1}$ and $H_{c2}$, and it has superconducting electrical properties up to $H_{c2}$. Above $H_{c2}$ the specimen is a normal conductor in every respect, except for possible surface effects. For given $H_c$ the area under the magnetization curve is the same for a type II superconductor as for a type I. (CGS units in all parts of this figure.)

The magnetization curve expected for a superconductor under the conditions of the Meissner-Ochsenfeld experiment is sketched in Fig. 4a. This applies quantitatively to a specimen in the form of a long solid cylinder placed in a longitudinal magnetic field. Pure specimens of many materials exhibit this behavior; they are called type I superconductors or, formerly, soft superconductors. The values of $H_c$ are always too low for type I superconductors to have any useful technical application in coils for superconducting magnets.

Other materials exhibit a magnetization curve of the form of Fig. 4b and are known as type II superconductors. They tend to be alloys (as in Fig. 5a) or transition metals with high values of the electrical resistivity in the normal state: that is, the electronic mean free path in the normal state is short. We shall see later why the mean free path is involved in the "magnetization" of superconductors.

Type II superconductors have superconducting electrical properties up to a field denoted by $H_{c2}$. Between the lower critical field $H_{c1}$ and the upper critical field $H_{c2}$ the flux density $B \neq 0$ and the Meissner effect is said to be incomplete. The value of $H_{c2}$ may be 100 times or more higher (Fig. 5b) than the value of the critical field $H_c$ calculated from the thermodynamics of the transition. In the region between $H_{c1}$ and $H_{c2}$ the superconductor is threaded by flux lines and is said to be in the vortex state. A field $H_{c2}$ of 410 kG (41 teslas) has been attained in an alloy of Nb, Al, and Ge at the boiling point of helium, and 540 kG (54 teslas) has been reported for PbMo$_6$S$_8$. 
Figure 5a  Superconducting magnetization curves of annealed polycrystalline lead and lead-indium alloys at 4.2 K. (A) lead; (B) lead–2.08 wt. percent indium; (C) lead–6.23 wt. percent indium; (D) lead–20.4 wt. percent indium. (After Livingston.)

Figure 5b  Stronger magnetic fields than any now contemplated in practical superconducting devices are within the capability of certain Type II materials. These materials cannot be exploited, however, until their critical current density can be raised and until they can be fabricated as finely divided conductors. (Magnetic fields of more than about 20 teslas can be generated only in pulses, and so portions of the curves shown as broken lines were measured in that way.)
Figure 6 Entropy $S$ of aluminum in the normal and superconducting states as a function of the temperature. The entropy is lower in the superconducting state because the electrons are more ordered here than in the normal state. At any temperature below the critical temperature $T_c$ the specimen can be put in the normal state by application of a magnetic field stronger than the critical field.

Commercial solenoids wound with a hard superconductor produce high steady fields over 100 kG. A “hard superconductor” is a type II superconductor with a large magnetic hysteresis, usually induced by mechanical treatment. Such materials have an important medical application in magnetic resonance imaging (MRI).

**Heat Capacity**

In all superconductors the entropy decreases markedly on cooling below the critical temperature $T_c$. Measurements for aluminum are plotted in Fig. 6. The decrease in entropy between the normal state and the superconducting state tells us that the superconducting state is more ordered than the normal state, for the entropy is a measure of the disorder of a system. Some or all of the electrons thermally excited in the normal state are ordered in the superconducting state. The change in entropy is small, in aluminum of the order of $10^{-4} k_B$ per atom. The small entropy change must mean that only a small fraction (of the order of $10^{-4}$) of the conduction electrons participate in the transition to the ordered superconducting state. The free energies of normal and superconducting states are compared in Fig. 7.

The heat capacity of gallium is plotted in Fig. 8: (a) compares the normal and superconducting states; (b) shows that the electronic contribution to the heat capacity in the superconducting state is an exponential form with an argu-
Figure 7  Experimental values of the free energy as a function of temperature for aluminum in the superconducting state and in the normal state. Below the transition temperature $T_c = 1.180$ K the free energy is lower in the superconducting state. The two curves merge at the transition temperature, so that the phase transition is second order (there is no latent heat of transition at $T_c$). The curve $F_S$ is measured in zero magnetic field, and $F_N$ is measured in a magnetic field sufficient to put the specimen in the normal state. (Courtesy of N. E. Phillips.)

Figure 8  (a) The heat capacity of gallium in the normal and superconducting states. The normal state (which is restored by a 200 G field) has electronic, lattice, and (at low temperatures) nuclear quadrupole contributions. In (b) the electronic part $C_{es}$ of the heat capacity in the superconducting state is plotted on a log scale versus $T_c/T$: the exponential dependence on $1/T$ is evident. Here $\gamma = 0.60$ mJ mol$^{-1}$K$^{-2}$. (After N. E. Phillips.)
ment proportional to \(-1/T\), suggestive of excitation of electrons across an energy gap. An energy gap (Fig. 9) is a characteristic, but not universal, feature of the superconducting state. The gap is accounted for by the Bardeen-Cooper-Schrieffer (BCS) theory of superconductivity (see Appendix H).

**Energy Gap**

The energy gap of superconductors is of entirely different origin and nature than the energy gap of insulators. In an insulator the energy gap is caused by the electron-lattice interaction, Chapter 7. This interaction ties the electrons to the lattice. In a superconductor the important interaction is the electron-electron interaction which orders the electrons in k space with respect to the Fermi gas of electrons.

The argument of the exponential factor in the electronic heat capacity of a superconductor is found to be \(-E_g/2k_B T\) and not \(-E_g/k_B T\). This has been learnt from comparison with optical and electron tunneling determinations of the gap \(E_g\). Values of the gap in several superconductors are given in Table 3.

The transition in zero magnetic field from the superconducting state to the normal state is observed to be a second-order phase transition. At a second-order transition there is no latent heat, but there is a discontinuity in the heat capacity, evident in Fig. 8a. Further, the energy gap decreases continuously to zero as the temperature is increased to the transition temperature \(T_c\), as in Fig. 10. A first-order transition would be characterized by a latent heat and by a discontinuity in the energy gap.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Energy gaps in superconductors, at (T = 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E_g(0)) in (10^{-2})eV. (E_g(0)/k_BT_c).</td>
<td>Al</td>
</tr>
<tr>
<td>Sc</td>
<td>Ti</td>
</tr>
<tr>
<td>16.34</td>
<td>3.4</td>
</tr>
<tr>
<td>Y</td>
<td>Zr</td>
</tr>
<tr>
<td>30.5</td>
<td>3.80</td>
</tr>
<tr>
<td>La</td>
<td>Hf</td>
</tr>
<tr>
<td>19.37</td>
<td>14.360</td>
</tr>
</tbody>
</table>


Figure 9 (a) Conduction band in the normal state; (b) energy gap at the Fermi level in the superconducting state. Electrons in excited states above the gap behave as normal electrons in rf fields: they cause resistance; at dc they are shorted out by the superconducting electrons. The gap $E_g$ is exaggerated in the figure: typically $E_g \sim 10^{-4} \epsilon_F$.

Figure 10 Reduced values of the observed energy gap $E_g(T)/E_g(0)$ as a function of the reduced temperature $T/T_c$, after Townsend and Sutton. The solid curve is drawn for the BCS theory.

Microwave and Infrared Properties

The existence of an energy gap means that photons of energy less than the gap energy are not absorbed. Nearly all the photons incident are reflected as for any metal because of the impedance mismatch at the boundary between vacuum and metal, but for a very thin ($\sim 20$ Å) film more photons are transmitted in the superconducting state than in the normal state.

For photon energies less than the energy gap, the resistivity of a superconductor vanishes at absolute zero. At $T \ll T_c$ the resistance in the superconducting state has a sharp threshold at the gap energy. Photons of lower energy see a resistanceless surface. Photons of higher energy see a resistance that
approaches that of the normal state because such photons cause transitions to unoccupied normal energy levels above the gap.

As the temperature is increased not only does the gap decrease in energy, but the resistivity for photons with energy below the gap energy no longer vanishes, except at zero frequency. At zero frequency the superconducting electrons short-circuit any normal electrons that have been thermally excited above the gap. At finite frequencies the inertia of the superconducting electrons prevents them from completely screening the electric field, so that thermally excited normal electrons now can absorb energy (Problem 3).

**Isotope Effect**

It has been observed that the critical temperature of superconductors varies with isotopic mass. In mercury $T_c$ varies from 4.185 K to 4.146 K as the average atomic mass $M$ varies from 199.5 to 203.4 atomic mass units. The transition temperature changes smoothly when we mix different isotopes of the same element. The experimental results within each series of isotopes may be fitted by a relation of the form

$$M^{\alpha}T_c = \text{constant}.$$  \hspace{1cm} (2)

Observed values of $\alpha$ are given in Table 4.

From the dependence of $T_c$ on the isotopic mass we learn that lattice vibrations and hence electron-lattice interactions are deeply involved in superconductivity. This was a fundamental discovery: there is no other reason for the superconducting transition temperature to depend on the number of neutrons in the nucleus.

The original BCS model gave the result $T_c \propto \theta_{\text{Debye}} \propto M^{-1/2}$, so that $\alpha = \frac{1}{2}$ in (2), but the inclusion of coulomb interactions between the electrons changes the relation. Nothing is sacred about $\alpha = \frac{1}{2}$. The absence of an isotope effect in Ru and Zr has been accounted for in terms of the electron band structure of these metals.

**THEORETICAL SURVEY**

A theoretical understanding of the phenomena associated with superconductivity has been reached in several ways. Certain results follow directly from thermodynamics. Many important results can be described by phenomenological equations: the London equations and the Landau-Ginzburg equations (Appendix I). A successful quantum theory of superconductivity was given by Bardeen, Cooper, and Schrieffer, and has provided the basis for subsequent work. Josephson and Anderson discovered the importance of the phase of the superconducting wavefunction.
Table 4  Isotope effect in superconductors

Experimental values of $\alpha$ in $M^0 T_c = \text{constant}$, where $M$ is the isotopic mass.

<table>
<thead>
<tr>
<th>Substance</th>
<th>$\alpha$</th>
<th>Substance</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zn</td>
<td>$0.45 \pm 0.05$</td>
<td>Ru</td>
<td>$0.00 \pm 0.05$</td>
</tr>
<tr>
<td>Cd</td>
<td>$0.32 \pm 0.07$</td>
<td>Os</td>
<td>$0.15 \pm 0.05$</td>
</tr>
<tr>
<td>Sn</td>
<td>$0.47 \pm 0.02$</td>
<td>Mo</td>
<td>$0.33$</td>
</tr>
<tr>
<td>Hg</td>
<td>$0.50 \pm 0.03$</td>
<td>Nb$_3$Sn</td>
<td>$0.08 \pm 0.02$</td>
</tr>
<tr>
<td>Pb</td>
<td>$0.49 \pm 0.02$</td>
<td>Zr</td>
<td>$0.00 \pm 0.05$</td>
</tr>
</tbody>
</table>

**Thermodynamics of the Superconducting Transition**

The transition between the normal and superconducting states is thermodynamically reversible, just as the transition between liquid and vapor phases of a substance is reversible. Thus we may apply thermodynamics to the transition, and we thereby obtain an expression for the entropy difference between normal and superconducting states in terms of the critical field curve $H_c$ versus $T$. This is analogous to the vapor pressure equation for the liquid-gas coexistence curve (TP, Chapter 10).

We treat a type I superconductor with a complete Meissner effect, so that $B = 0$ inside the superconductor. We shall see that the critical field $H_c$ is a quantitative measure of the free energy difference between the superconducting and normal states at constant temperature. The symbol $H_c$ will always refer to a bulk specimen, never to a thin film. For type II superconductors, $H_c$ is understood to be the thermodynamic critical field related to the stabilization free energy.

The stabilization free energy of the superconducting state with respect to the normal state can be determined by calorimetric or magnetic measurements. In the calorimetric method the heat capacity is measured as a function of temperature for the superconductor and for the normal conductor, which means the superconductor in a magnetic field larger than $H_c$. From the difference of the heat capacities we can compute the free energy difference, which is the stabilization free energy of the superconducting state.

In the magnetic method the stabilization free energy is found from the value of the applied magnetic field that will destroy the superconducting state, at constant temperature. The argument follows. Consider the work done (Fig. 11) on a superconductor when it is brought reversibly at constant temperature from a position at infinity (where the applied field is zero) to a position $r$ in the field of a permanent magnet:

$$W = -\int_0^{B_r} M \cdot dB_a ,$$

(3)
Figure 11  (a) A superconductor in which the Meissner effect is complete has $B = 0$, as if the magnetization were $M = -B_a/4\pi$, in CGS units. (b) When the applied field reaches the value $B_{ac}$, the normal state can coexist in equilibrium with the superconducting state. In coexistence the free energy densities are equal: $F_N(T, B_{ac}) = F_S(T, B_{ac})$. 

per unit volume of specimen. This work appears in the energy of the magnetic field. The thermodynamic identity for the process is

$$dF = -M \cdot dB_a ,$$  \hspace{1cm} (4)

as in TP, Chapter 8.

For a superconductor with $M$ related to $B_a$ by (1) we have

\begin{align*}
(CGS) & \quad dF_S = \frac{1}{4\pi} B_a \, dB_a ; \\
(SI) & \quad dF_S = \frac{1}{\mu_0} B_a \, dB_a .
\end{align*}

The increase in the free energy density of the superconductor is

\begin{align*}
(CGS) & \quad F_S(B_a) - F_S(0) = B_a^2/8\pi ; \\
(SI) & \quad F_S(B_a) - F_S(0) = B_a^2/2\mu_0 .
\end{align*}

on being brought from a position where the applied field is zero to a position where the applied field is $B_a$. 
Now consider a normal nonmagnetic metal. If we neglect the small susceptibility of a metal in the normal state, then \( M = 0 \) and the energy of the normal metal is independent of field. At the critical field we have

\[
F_N(B_{ac}) = F_N(0).
\]  

(7)

The results (6) and (7) are all we need to determine the stabilization energy of the superconducting state at absolute zero. At the critical value \( B_{ac} \) of the applied magnetic field the energies are equal in the normal and superconducting states:

(CG) \[
F_N(B_{ac}) = F_S(B_{ac}) = F_S(0) + B_{ac}^2/8\pi,
\]  

(8)

(SI) \[
F_N(B_{ac}) = F_S(B_{ac}) = F_S(0) + B_{ac}^2/2\mu_0.
\]

In SI units \( H_c = B_{ac}/\mu_0 \), whereas in CGS units \( H_c = B_{ac} \).

The specimen is stable in either state when the applied field is equal to the critical field. Now by (7) it follows that

(CG) \[
\Delta F = F_N(0) - F_S(0) = B_{ac}^2/8\pi,
\]  

(9)

where \( \Delta F \) is the stabilization free energy density of the superconducting state. For aluminum, \( B_{ac} \) at absolute zero is 105 gauss, so that at absolute zero \( \Delta F = (105)^2/8\pi = 439 \text{ erg cm}^{-3} \), in excellent agreement with the result of thermal measurements, 430 \text{ erg cm}^{-3}.

At a finite temperature the normal and superconducting phases are in equilibrium when the magnetic field is such that their free energies \( F = U - TS \) are equal. The free energies of the two phases are sketched in Fig. 12 as a function of the magnetic field. Experimental curves of the free energies of the two phases for aluminum are shown in Fig. 7. Because the slopes \( dF/dT \) are equal at the transition temperature, there is no latent heat at \( T_c \).

**London Equation**

We saw that the Meissner effect implies a magnetic susceptibility \( \chi = -1/4\pi \) in CGS in the superconducting state or, in SI, \( \chi = -1 \). This sweeping assumption tends to cut off further discussion, and it does not account for the flux penetration observed in thin films. Can we modify a constitutive equation of electrodynamics (such as Ohm's law) in some way to obtain the Meissner effect?

\(^6\)This is an adequate assumption for type I superconductors. In type II superconductors in high fields the change in spin paramagnetism of the conduction electrons lowers the energy of the normal phase significantly. In some, but not all, type II superconductors the upper critical field is limited by this effect. Clogston has suggested that \( H_{ac}^{\text{max}} = 18,400 T_c \), where \( H_{ac} \) is in gauss and \( T_c \) in K. See A. M. Clogston, Phys. Rev. Lett. 9, 266 (1962); B. Chandrasekhar, Appl. Phys. Lett. 1, 7 (1962).
Figure 12 The free energy density $F_N$ of a nonmagnetic normal metal is approximately independent of the intensity of the applied magnetic field $B_a$. At a temperature $T < T_c$, the metal is a superconductor in zero magnetic field, so that $F_N(T, 0)$ is lower than $F_N(T, 0)$. An applied magnetic field increases $F_N$ by $B^2/2\pi$, in CGS units, so that $F_N(T, B_a) = F_N(T, 0) + B^2/(2\pi)$. If $B_a$ is larger than the critical field $B_{cr}$ the free energy density is lower in the normal state than in the superconducting state, and now the normal state is the stable state. The origin of the vertical scale in the drawing is at $F_N(T, 0)$. The figure equally applies to $U_S$ and $U_N$ at $T = 0$.

effect? We do not want to modify the Maxwell equations themselves. Electrical conduction in the normal state of a metal is described by Ohm’s law $\mathbf{j} = \sigma \mathbf{E}$. We need to modify this drastically to describe conduction and the Meissner effect in the superconducting state. Let us make a postulate and see what happens.

We postulate that in the superconducting state the current density is directly proportional to the vector potential $\mathbf{A}$ of the local magnetic field, where $\mathbf{B} = \text{curl} \mathbf{A}$. The gauge of $\mathbf{A}$ will be specified. In CGS units we write the constant of proportionality as $-c/4\pi \lambda_L^2$ for reasons that will become clear. Here $c$ is the speed of light and $\lambda_L$ is a constant with the dimensions of length. In SI units we write $-1/\mu_0 \lambda_L^2$. Thus

\begin{align*}
\text{(CGS)} \quad & \mathbf{j} = -\frac{c}{4\pi \lambda_L^2} \mathbf{A} ; \\
\text{(SI)} \quad & \mathbf{j} = -\frac{1}{\mu_0 \lambda_L^2} \mathbf{A}.
\end{align*} 

This is the London equation. We express it another way by taking the curl of both sides to obtain

\begin{align*}
\text{(CGS)} \quad & \text{curl } \mathbf{j} = -\frac{c}{4\pi \lambda_L^2} \mathbf{B} ; \\
\text{(SI)} \quad & \text{curl } \mathbf{j} = -\frac{1}{\mu_0 \lambda_L^2} \mathbf{B}.
\end{align*} 

The London equation (10) is understood to be written with the vector potential in the London gauge in which $\text{div } \mathbf{A} = 0$, and $A_r = 0$ on any external
surface through which no external current is fed. The subscript $n$ denotes the component normal to the surface. Thus $\text{div } j = 0$ and $j_n = 0$, the actual physical boundary conditions. The form (10) applies to a simply connected superconductor; additional terms may be present in a ring or cylinder, but (11) holds true independent of geometry.

First we show that the London equation leads to the Meissner effect. By a Maxwell equation we know that

\begin{equation}
\text{(CGS)} \quad \text{curl } B = \frac{4\pi}{c} j ; \quad \text{ (SI) } \text{curl } B = \mu_0 j ; \quad (12)
\end{equation}

under static conditions. We take the curl of both sides to obtain

\begin{equation}
\text{CGS} \quad \text{curl curl } B = -\nabla^2 B = \frac{4\pi}{c} \text{ curl } j ; \\
\text{SI} \quad \text{curl curl } B = -\nabla^2 B = \mu_0 \text{ curl } j ,
\end{equation}

which may be combined with the London equation (11) to give for a superconductor,

\begin{equation}
\nabla^2 B = B/\lambda_L^2 . \quad (13)
\end{equation}

This equation is seen to account for the Meissner effect because it does not allow a solution uniform in space, so that a uniform magnetic field cannot exist in a superconductor. That is, $B(r) = B_0 = \text{constant}$ is not a solution of (13) unless the constant field $B_0$ is identically zero. The result follows because $\nabla^2 B_0$ is always zero, but $B_0/\lambda_L^2$ is not zero unless $B_0$ is zero. Note further that (12) ensures that $j = 0$ in a region where $B = 0$.

In the pure superconducting state the only field allowed is exponentially damped as we go in from an external surface. Let a semi-infinite superconductor occupy the space on the positive side of the $x$ axis, as in Fig. 13. If $B(0)$ is the field at the plane boundary, then the field inside is

\begin{equation}
B(x) = B(0) \exp(-x/\lambda_L) , \quad (14)
\end{equation}
for this is a solution of (13). In this example the magnetic field is assumed to be parallel to the boundary. Thus we see \( \lambda_L \) measures the depth of penetration of the magnetic field; it is known as the London penetration depth. Actual penetration depths are not described precisely by \( \lambda_L \) alone, for the London equation is now known to be somewhat oversimplified. It is shown by comparison of (22) with (11) that

\[
\text{(CGS)} \quad \lambda_L = \left( \frac{mc^2}{4\pi q^2} \right)^{1/2} ;
\]

\[
\lambda_L = \left( \frac{e_0 mc^2}{nq^2} \right)^{1/2} \quad \text{(SI)}
\]

(14a)

for particles of charge \( q \) and mass \( m \) in concentration \( n \). Values are given in Table 5.

An applied magnetic field \( B_\perp \) will penetrate a thin film fairly uniformly if the thickness is much less than \( \lambda_L \); thus in a thin film the Meissner effect is not complete. In a thin film the induced field is much less than \( B_\perp \), and there is little effect of \( B_\perp \) on the energy density of the superconducting state, so that (6) does not apply. It follows that the critical field \( H_c \) of thin films in parallel magnetic fields will be very high.

\[\text{Coherence Length}\]

The London penetration depth \( \lambda_L \) is a fundamental length that characterizes a superconductor. An independent length is the coherence length \( \xi \). The coherence length is a measure of the distance within which the superconducting electron concentration cannot change drastically in a spatially-varying magnetic field.

The London equation is a local equation: it relates the current density at a point \( \mathbf{r} \) to the vector potential at the same point. So long as \( j(\mathbf{r}) \) is given as a constant times \( A(\mathbf{r}) \), the current is required to follow exactly any variation in the vector potential. But the coherence length \( \xi \) is a measure of the range over which we should average \( A \) to obtain \( j \). It is also a measure of the minimum spatial extent of a transition layer between normal and superconductor. The coherence length is best introduced into the theory through the Landau-Ginzburg equations, Appendix I. Now we give a plausibility argument for the energy required to modulate the superconducting electron concentration.

Any spatial variation in the state of an electronic system requires extra kinetic energy. A modulation of an eigenfunction increases the kinetic energy because the modulation will increase the integral of \( d^2\phi/dx^2 \). It is reasonable to restrict the spatial variation of \( j(\mathbf{r}) \) in such a way that the extra energy is less than the stabilization energy of the superconducting state.

We compare the plane wave \( \psi(x) = e^{ikx} \) with the strongly modulated wavefunction:

\[
\varphi(x) = 2^{-1/2} (e^{i(k+q)x} + e^{ikx}) .
\]

(15a)
Table 5 Calculated intrinsic coherence length and London penetration depth, at absolute zero

<table>
<thead>
<tr>
<th>Metal</th>
<th>Intrinsic Pippard coherence length $\xi_0$, in $10^{-6}$ cm</th>
<th>London penetration depth $\lambda_L$, in $10^{-6}$ cm</th>
<th>$\lambda_L/\xi_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sn</td>
<td>23.</td>
<td>3.4</td>
<td>0.16</td>
</tr>
<tr>
<td>Al</td>
<td>160.</td>
<td>1.6</td>
<td>0.010</td>
</tr>
<tr>
<td>Pb</td>
<td>8.3</td>
<td>3.7</td>
<td>0.45</td>
</tr>
<tr>
<td>Cd</td>
<td>76.</td>
<td>11.0</td>
<td>0.14</td>
</tr>
<tr>
<td>Nb</td>
<td>3.8</td>
<td>3.9</td>
<td>1.02</td>
</tr>
</tbody>
</table>

After R. Meservey and B. B. Schwartz.

The probability density associated with the plane wave is uniform in space: $\psi^*\psi = e^{-ikx}e^{ikx} = 1$, whereas $\varphi^*\varphi$ is modulated with the wavevector $q$:

$$
\varphi^*\varphi = \frac{1}{2}(e^{-i(k+q)x} + e^{-ikx})(e^{i(k+q)x} + e^{ikx})
= \frac{1}{2}(2 + e^{iqx} + e^{-iqx}) = 1 + \cos qx .
$$  \hfill (15b)

The kinetic energy of the wave $\psi(x)$ is $\epsilon = \hbar^2k^2/2m$; the kinetic energy of the modulated density distribution is higher, for

$$
\int dx \varphi^* \left( -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \right) \varphi = \frac{1}{2} \left( \frac{\hbar^2}{2m} \right) [(k+q)^2 + k^2] = \frac{\hbar^2}{2m} k^2 + \frac{\hbar^2}{2m} k q ,
$$

where we neglect $q^2$ for $q \ll k$.

The increase of energy required to modulate is $\hbar^2kq/2m$. If this increase exceeds the energy gap $E_g$, superconductivity will be destroyed. The critical value $q_0$ of the modulation wavevector is given by

$$
\frac{\hbar^2}{2m} k_F q_0 = E_g .
$$  \hfill (16a)

We define an intrinsic coherence length $\xi_0$ related to the critical modulation by $\xi_0 = 1/q_0$. We have

$$
\xi_0 = \frac{\hbar^2k_F}{2mE_g} = \hbar v_F/2E_g ,
$$  \hfill (16b)

where $v_F$ is the electron velocity at the Fermi surface. On the BCS theory a similar result is found (QTS, p. 173):

$$
\xi_0 = 2\hbar v_F/\pi E_g .
$$  \hfill (17)

Calculated values of $\xi_0$ from (17) are given in Table 5. The intrinsic coherence length $\xi_0$ is characteristic of a pure superconductor.
In impure materials and in alloys the coherence length $\xi$ is shorter than $\xi_0$. This may be understood qualitatively: in impure material the electron eigenfunctions already have wiggles in them; we can construct a given localized variation of current density with less energy from wavefunctions with wiggles than from smooth wavefunctions.

The coherence length first appeared in the Landau-Ginzburg equations; these equations also follow from the BCS theory. They describe the structure of the transition layer between normal and superconducting phases in contact. The coherence length and the actual penetration depth $\lambda$ depend on the mean free path $\ell$ of the electrons measured in the normal state; the relationships are indicated in Fig. 14. When the superconductor is very impure, with a very small $\ell$, then $\xi \approx (\xi_0 \ell)^{1/2}$ and $\lambda \approx \lambda_L (\xi_0 / \ell)^{1/2}$, so that $\lambda / \xi \approx \lambda_L / \ell$. This is the “dirty superconductor” limit. The ratio $\lambda / \xi$ is denoted by $\kappa$.

**BCS Theory of Superconductivity**

The basis of a quantum theory of superconductivity was laid by the classic 1957 papers of Bardeen, Cooper, and Schrieffer. There is a “BCS theory of superconductivity” with a very wide range of applicability, from He$^3$ atoms in their condensed phase, to type I and type II metallic superconductors, and to high-temperature superconductors based on planes of cuprate ions. Further, there is a “BCS wavefunction” composed of particle pairs $k \uparrow$ and $-k \downarrow$, which, when treated by the BCS theory, gives the familiar electronic superconductivity observed in metals and exhibits the energy gaps of Table 3. This pairing is known as s-wave pairing. There are other forms of particle pairing possible with the BCS theory, but we do not have to consider other than the BCS wavefunction here at this time. In this chapter we treat the specific accomplishments of BCS theory with a BCS wavefunction, which include:

1. An attractive interaction between electrons can lead to a ground state separated from excited states by an energy gap. The critical field, the thermal properties, and most of the electromagnetic properties are consequences of the energy gap.

2. The electron-lattice-electron interaction leads to an energy gap of the observed magnitude. The indirect interaction proceeds when one electron interacts with the lattice and deforms it; a second electron sees the deformed lattice and adjusts itself to take advantage of the deformation to lower its energy. Thus the second electron interacts with the first electron via the lattice deformation.

3. The penetration depth and the coherence length emerge as natural consequences of the BCS theory. The London equation is obtained for magnetic fields that vary slowly in space. Thus the central phenomenon in superconductivity, the Meissner effect, is obtained in a natural way.

---

4. The criterion for the transition temperature of an element or alloy involves the electron density of orbitals \( D(\epsilon_F) \) of one spin at the Fermi level and the electron-lattice interaction \( U \), which can be estimated from the electrical resistivity because the resistivity at room temperature is a measure of the electron-phonon interaction. For \( UD(\epsilon_F) \ll 1 \) the BCS theory predicts

\[
T_c = 1.14\theta \exp[-1/UD(\epsilon_F)] ,
\]

where \( \theta \) is the Debye temperature and \( U \) is an attractive interaction. The result for \( T_c \) is satisfied at least qualitatively by the experimental data. There is an interesting apparent paradox: the higher the resistivity at room temperature the higher is \( U \), and thus the more likely it is that a metal will be a superconductor when cooled.

5. Magnetic flux through a superconducting ring is quantized and the effective unit of charge is \( 2e \) rather than \( e \). The BCS ground state involves pairs of electrons; thus flux quantization in terms of the pair charge \( 2e \) is a consequence of the theory.

**BCS Ground State**

The filled Fermi sea is the ground state of a Fermi gas of noninteracting electrons. This state allows arbitrarily small excitations—we can form an excited state by taking an electron from the Fermi surface and raising it just above the Fermi surface. The BCS theory shows that with an appropriate attractive interaction between electrons the new ground state is superconducting and is separated by a finite energy \( E_\text{F} \) from its lowest excited state.

The formation of the BCS ground state is suggested by Fig. 15. The BCS state in (b) contains admixtures of one-electron orbitals from above the Fermi energy \( \epsilon_F \). At first sight the BCS state appears to have a higher energy than the Fermi state: the comparison of (b) with (a) shows that the kinetic energy of the BCS state is higher than that of the Fermi state. But the attractive potential
energy of the BCS state, although not represented in the figure, acts to lower the total energy of the BCS state with respect to the Fermi state.

When the BCS ground state of a many-electron system is described in terms of the occupancy of one-particle orbitals, those near $\varepsilon_F$ are filled somewhat like a Fermi-Dirac distribution for some finite temperature.

The central feature of the BCS state is that the one-particle orbitals are occupied in pairs: if an orbital with wavevector $k$ and spin up is occupied, then the orbital with wavevector $-k$ and spin down is also occupied. If $k_\uparrow$ is vacant, then $-k_\downarrow$ is also vacant. The pairs are called Cooper pairs, treated in Appendix H. They have spin zero and have many attributes of bosons.

The boson condensation temperature ($TP$, Chapter 7) calculated for metallic concentrations is of the order of the Fermi temperature ($10^4 - 10^5$ K). The superconducting transition temperature is much lower and takes place when the electron pairs break up into two fermions. The model of a superconductor as composed of noninteracting bosons cannot be taken absolutely literally, for there are about $10^6$ electrons in the volume occupied by a single Cooper pair.

**Flux Quantization in a Superconducting Ring**

We prove that the total magnetic flux that passes through a superconducting ring may assume only quantized values, integral multiples of the flux quantum $2\pi\hbar c/\hbar q$, where by experiment $q = 2e$, the charge of an electron pair. Flux quantization is a beautiful example of a long-range quantum effect in which the coherence of the superconducting state extends over a ring or solenoid.

Let us first consider the electromagnetic field as an example of a similar boson field. The electric field intensity $E(r)$ acts qualitatively as a probability field amplitude. When the total number of photons is large, the energy density may be written as

$$E^*(r)E(r)/4\pi \equiv n(r)\hbar \omega,$$

where $n(r)$ is the number density of photons of frequency $\omega$. Then we may write the electric field in a semiclassical approximation as

$$E(r) \equiv (4\pi\hbar \omega)^{1/2} n(r)^{1/2} e^{i\theta(r)} \quad E^*(r) \equiv (4\pi\hbar \omega)^{1/2} n(r)^{1/2} e^{-i\theta(r)},$$

where $\theta(r)$ is the phase of the field. A similar probability amplitude describes Cooper pairs.

The arguments that follow apply to a boson gas with a large number of bosons in the same orbital. We then can treat the boson probability amplitude as a classical quantity, just as the electromagnetic field is used for photons. Both amplitude and phase are then meaningful and observable. The arguments do not apply to a metal in the normal state because an electron in the normal state acts as a single unpaired fermion that cannot be treated classically.

We first show that a charged boson gas obeys the London equation. Let
ψ(r) be the particle probability amplitude. We suppose that the pair concentration \( n = \psi^*\psi = \text{constant} \). At absolute zero \( n \) is one-half of the concentration of electrons in the conduction band, for \( n \) refers to pairs. Then we may write
\[
\psi = n^{1/2} e^{i\theta(r)}; \quad \psi^* = n^{1/2} e^{-i\theta(r)}.
\] (19)
The phase \( \theta(r) \) is important for what follows. In SI units, set \( c = 1 \) in the equations that follow.

The velocity of a particle is, from the Hamilton equations of mechanics,
\begin{equation}
\mathbf{v} = \frac{1}{m} \left( \mathbf{p} - \frac{q}{c} \mathbf{A} \right) = \frac{1}{m} \left( -i\hbar \nabla - \frac{q}{c} \mathbf{A} \right).
\end{equation}

(CGS)

The particle flux is given by
\[
\psi^* \mathbf{v} \psi = \frac{n}{m} \left( \hbar \nabla \theta - \frac{q}{c} \mathbf{A} \right),
\] (20)
so that the electric current density is
\[
\mathbf{j} = q \psi^* \mathbf{v} \psi = \frac{nq}{m} \left( \hbar \nabla \theta - \frac{q}{c} \mathbf{A} \right).
\] (21)

We may take the curl of both sides to obtain the London equation:
\begin{equation}
\text{curl} \mathbf{j} = -\frac{nq^2}{mc} \mathbf{B},
\end{equation}

(CGS)

with use of the fact that the curl of the gradient of a scalar is identically zero. The constant that multiplies \( \mathbf{B} \) agrees with (14a). We recall that the Meissner effect is a consequence of the London equation, which we have here derived.

Quantization of the magnetic flux through a ring is a dramatic consequence of Eq. (21). Let us take a closed path \( \mathbf{C} \) through the interior of the supercon-
ducting material, well away from the surface (Fig. 16). The Meissner effect tells us that \( B \) and \( j \) are zero in the interior. Now (21) is zero if
\[ \hbar c \nabla \theta = qA \quad . \tag{23} \]
We form
\[ \oint_C \nabla \theta \cdot dl = \theta_2 - \theta_1 \]
for the change of phase on going once around the ring.

The probability amplitude \( \psi \) is measurable in the classical approximation, so that \( \psi \) must be single-valued and
\[ \theta_2 - \theta_1 = 2\pi s \quad , \tag{24} \]
where \( s \) is an integer. By the Stokes theorem,
\[ \oint_C A \cdot dl = \int_C (\text{curl} \ A) \cdot d\sigma = \int_C B \cdot d\sigma = \Phi \quad , \tag{25} \]
where \( d\sigma \) is an element of area on a surface bounded by the curve \( C \), and \( \Phi \) is the magnetic flux through \( C \). From (23), (24), and (25) we have
\[ 2\pi \hbar c s = q\Phi \quad , \]
for
\[ \Phi = (2\pi \hbar c/q)s \quad . \tag{26} \]
Thus the flux through the ring is quantized in integral multiples of \( 2\pi \hbar c/q \).

By experiment \( q = -2e \) as appropriate for electron pairs, so that the quantum of flux in a superconductor is
\[
\begin{align*}
\text{(CGS)} & \quad \Phi_0 = 2\pi \hbar c/2e = 2.0678 \times 10^{-7} \text{ gauss cm}^2 \quad ; \\
\text{(SI)} & \quad \Phi_0 = 2\pi \hbar /2e = 2.0678 \times 10^{-15} \text{ tesla m}^2 . \tag{27}
\end{align*}
\]
This unit of flux is called a fluxoid or fluxon.

The flux through the ring is the sum of the flux \( \Phi_{\text{ext}} \) from external sources and the flux \( \Phi_{\text{sc}} \) from the superconducting currents which flow in the surface of the ring: \( \Phi = \Phi_{\text{ext}} + \Phi_{\text{sc}} \). The flux \( \Phi \) is quantized. There is normally no quanti-
zation condition on the flux from external sources, so that $\Phi_{sc}$ must adjust itself appropriately in order that $\Phi$ assume a quantized value.

Suppose a magnetic monopole of strength $g$ is situated just below the center of a superconducting ring. The magnetic flux through the ring is $(g/r^2)(2\pi r^2) = 2\pi g$, and by (27) this must equal an integral multiple of $\pi\hbar c/e$. Thus the minimum permissible value of $g$ in CGS is $\hbar c/2e$, the famous Dirac result.

**Duration of Persistent Currents**

Consider a persistent current that flows in a ring of a type I superconductor of wire of length $L$ and cross-sectional area $A$. The persistent current maintains a flux through the ring of some integral number of fluxoids (27). A fluxoid cannot leak out of the ring and thereby reduce the persistent current unless by a thermal fluctuation a minimum volume of the superconducting ring is momentarily in the normal state.

The probability per unit time that a fluxoid will leak out is the product

$$P = \text{(attempt frequency)}(\text{activation barrier factor}) \quad (28)$$

The activation barrier factor is $\exp(-\Delta F/k_B T)$, where the free energy of the barrier is

$$\Delta F \approx (\text{minimum volume})(\text{excess free energy density of normal state}) \quad (29)$$

The minimum volume of the ring that must turn normal to allow a fluxoid to escape is of the order of $R\xi^2$, where $\xi$ is the coherence length of the superconductor and $R$ the wire thickness. The excess free energy density of the normal state is $H_c^2/8\pi$, whence the barrier free energy is

$$\Delta F \approx R\xi^2 H_c^2/8\pi \quad (29)$$

Let the wire thickness be $10^{-4}$ cm, the coherence length $= 10^{-4}$ cm, and $H_c = 10^3$ G; then $\Delta F \approx 10^{-7}$ erg. As we approach the transition temperature from below, $\Delta F$ will decrease toward zero, but the value given is a fair estimate between absolute zero and 0.8 $T_c$. Thus the activation barrier factor is

$$\exp(-\Delta F/k_B T) \approx \exp(-10^8) \approx 10^{-(4.34 \times 10^7)}$$

The characteristic frequency with which the minimum volume can attempt to change its state must be of order of $E_g/\hbar$. If $E_g = 10^{-15}$ erg, the attempt frequency is $\approx 10^{-15}/10^{-27} = 10^{12}$ s$^{-1}$. The leakage probability (28) becomes

$$P = 10^{12} 10^{-4.34 \times 10^7} s^{-1} \approx 10^{-4.34 \times 10^7} s^{-1}$$

The reciprocal of this is a measure of the time required for a fluxoid to leak out, $T = 1/P = 10^{4.34 \times 10^7}$ s.

The age of the universe is only $10^{18}$ s, so that a fluxoid will never leak out
in the age of the universe, under our assumed condition. Accordingly, the current is maintained.

There are two circumstances in which the activation energy is much lower and a fluxoid can be observed to leak out of a ring—either very close to the critical temperature, where \( H_c \) is very small, or when the material of the ring is a type II superconductor and already has fluxoids embedded in it. These special situations are discussed in the literature under the subject of fluctuations in superconductors.

**Type II Superconductors**

There is no difference in the mechanism of superconductivity in type I and type II superconductors. Both types have similar thermal properties at the superconductor-normal transition in zero magnetic field. But the Meissner effect is entirely different (Fig. 5).

A good type I superconductor excludes a magnetic field until superconductivity is destroyed suddenly, and then the field penetrates completely. A good type II superconductor excludes the field completely up to a field \( H_{c1} \). Above \( H_{c1} \) the field is partially excluded, but the specimen remains electrically superconducting. At a much higher field, \( H_{c2} \), the flux penetrates completely and superconductivity vanishes. (An outer surface layer of the specimen may remain superconducting up to a still higher field \( H_{c3} \).)

An important difference in a type I and a type II superconductor is in the mean free path of the conduction electrons in the normal state. If the coherence length \( \xi \) is longer than the penetration depth \( \lambda \), the superconductor will be type I. Most pure metals are type I, with \( \kappa < 1 \) (see Table 5 on p. 353).

But, when the mean free path is short, the coherence length is short and the penetration depth is great (Fig. 14). This is the situation when \( \kappa = \lambda / \xi > 1 \), and the superconductor will be type II.

We can change some metals from type I to type II by a modest addition of an alloying element. In Figure 5 the addition of 2 wt. percent of indium changes lead from type I to type II, although the transition temperature is scarcely changed at all. Nothing fundamental has been done to the electronic structure of lead by this amount of alloying, but the magnetic behavior as a superconductor has changed drastically.

The theory of type II superconductors was developed by Ginzburg, Landau, Abrikosov, and Gorkov. Later Kunzler and co-workers observed that Nb\(_3\)Sn wires can carry large supercurrents in fields approaching 100 kG; this led to the commercial development of strong field superconducting magnets.

Consider the interface between a region in the superconducting state and a region in the normal state. The interface has a surface energy that may be positive or negative and that decreases as the applied magnetic field is in-
creased. A superconductor is type I if the surface energy is always positive as the magnetic field is increased, and type II if the surface energy becomes negative as the magnetic field is increased. The sign of the surface energy has no importance for the transition temperature.

The free energy of a bulk superconductor is increased when the magnetic field is expelled. However, a parallel field can penetrate a very thin film nearly uniformly (Fig. 17), only a part of the flux is expelled, and the energy of the superconducting film will increase only slowly as the external magnetic field is increased. This causes a large increase in the field intensity required for the destruction of superconductivity. The film has the usual energy gap and will be resistanceless. A thin film is not a type II superconductor, but the film results show that under suitable conditions superconductivity can exist in high magnetic fields.

**Vortex State.** The results for thin films suggest the question: Are there stable configurations of a superconductor in a magnetic field with regions (in the form of thin rods or plates) in the normal state, each normal region surrounded by a superconducting region? In such a mixed state, called the vortex state, the external magnetic field will penetrate the thin normal regions uniformly, and the field will also penetrate somewhat into the surrounding superconducting material, as in Fig. 18.

The term vortex state describes the circulation of superconducting currents in vortices throughout the bulk specimen, as in Fig. 19 below. There is no chemical or crystallographic difference between the normal and the superconducting regions in the vortex state. The vortex state is stable when the penetra-
Figure 18  Variation of the magnetic field and energy gap parameter $\Delta(x)$ at the interface of superconducting and normal regions, for type I and type II superconductors. The energy gap parameter is a measure of the stabilization energy density of the superconducting state.

tion of the applied field into the superconducting material causes the surface energy to become negative. A type II superconductor is characterized by a vortex state stable over a certain range of magnetic field strength; namely, between $H_{c1}$ and $H_{c2}$.

Estimation of $H_{c1}$ and $H_{c2}$. What is the condition for the onset of the vortex state as the applied magnetic field is increased? We estimate $H_{c1}$ from the penetration depth $\lambda$. The field in the normal core of the fluxoid will be $H_{c1}$ when the applied field is $H_{c1}$.

The field will extend out from the normal core a distance $\lambda$ into the superconducting environment. The flux thus associated with a single core is $\pi \lambda^2 H_{c1}$, and this must be equal to the flux quantum $\Phi_0$ defined by (27). Thus

$$H_{c1} = \Phi_0 / \pi \lambda^2 .$$

(30)

This is the field for nucleation of a single fluxoid.
Figure 19 Flux lattice in NbSe$_2$ at 1,000 gauss at 0.2K, as viewed with a scanning tunneling microscope. The photo shows the density of states at the Fermi level, as in Figure 23. The vortex cores have a high density of states and are shaded white; the superconducting regions are dark, with no states at the Fermi level. The amplitude and spatial extent of these states is determined by a potential well formed by $\Delta(\alpha)$ as in Figure 18 for a Type II superconductor. The potential well confines the core state wavefunctions in the image here. The star shape is a finer feature, a result special to NbSe$_2$ of the sixfold disturbance of the charge density at the Fermi surface. Photo courtesy of H. F. Hess, AT&T Bell Laboratories.

At $H_{c2}$ the fluxoids are packed together as tightly as possible, consistent with the preservation of the superconducting state. This means as densely as the coherence length $\xi$ will allow. The external field penetrates the specimen almost uniformly, with small ripples on the scale of the fluxoid lattice. Each core is responsible for carrying a flux of the order of $\pi\xi^2 H_{c2}$, which also is quantized to $\Phi_0$. Thus

$$H_{c2} = \frac{\Phi_0}{\pi\xi^2}$$

(31)

gives the upper critical field. The larger the ratio $\lambda/\xi$, the larger is the ratio of $H_{c2}$ to $H_{c1}$.

It remains to find a relation between these critical fields and the thermodynamic critical field $H_c$ that measures the stabilization energy density of the superconducting state, which is known by (9) to be $H_c^2/8\pi$. In a type II superconductor we can determine $H_c$ only indirectly by calorimetric measurement of
the stabilization energy. To estimate $H_{c1}$ in terms of $H_c$, consider the stability of the vortex state at absolute zero in the impure limit $\xi < \lambda$; here $k > 1$ and the coherence length is short in comparison with the penetration depth.

We estimate in the vortex state the stabilization energy of a fluxoid core viewed as a normal metal cylinder which carries an average magnetic field $B_a$. The radius is of the order of the coherence length, the thickness of the boundary between N and S phases. The energy of the normal core referred to the energy of a pure superconductor is given by the product of the stabilization energy times the area of the core:

\[
\langle \text{CGS} \rangle \quad f_{\text{core}} \approx \frac{1}{8\pi} H_c^2 \times \pi \xi^2 ,
\]

per unit length. But there is also a decrease in magnetic energy because of the penetration of the applied field $B_a$ into the superconducting material around the core:

\[
\langle \text{CGS} \rangle \quad f_{\text{mag}} \approx -\frac{1}{8\pi} B_a^2 \times \pi \lambda^2 .
\]

For a single fluxoid we add these two contributions to obtain

\[
\langle \text{CGS} \rangle \quad f = f_{\text{core}} + f_{\text{mag}} \approx \frac{1}{8\pi} \left( H_c^2 \xi^2 - B_a^2 \lambda^2 \right) .
\]

The core is stable if $f < 0$. The threshold field for a stable fluxoid is at $f = 0$, or, with $H_{c1}$ written for $B_a$,

\[
H_{c1}/H_c = \frac{\xi}{\lambda} .
\]

The threshold field divides the region of positive surface energy from the region of negative surface energy.

We can combine (30) and (35) to obtain a relation for $H_c$:

\[
\pi \xi \lambda H_c = \Phi_0 .
\]

We can combine (30), (31), and (35) to obtain

\[
(H_{c1} H_{c2})^{1/2} = H_c ,
\]

and

\[
H_{c2} \approx (\lambda/\xi) H_c = \kappa H_c .
\]

**Single Particle Tunneling**

Consider two metals separated by an insulator, as in Fig. 20. The insulator normally acts as a barrier to the flow of conduction electrons from one metal to the other. If the barrier is sufficiently thin (less than 10 or 20 Å) there is a significant probability that an electron which impinges on the barrier will pass from one metal to the other: this is called tunneling. In many experiments the
insulating layer is simply a thin oxide layer formed on one of two evaporated metal films, as in Fig. 21.

When both metals are normal conductors, the current-voltage relation of the sandwich or tunneling junction is ohmic at low voltages, with the current directly proportional to the applied voltage. Giaever (1960) discovered that if one of the metals becomes superconducting the current-voltage characteristic changes from the straight line of Fig. 22a to the curve shown in Fig. 22b.

Figure 23a contrasts the electron density of orbitals in the superconductor with that in the normal metal. In the superconductor there is an energy gap centered at the Fermi level. At absolute zero no current can flow until the applied voltage is $V = E_g/2e = \Delta/e$. 
Figure 23 The density of orbitals and the current-voltage characteristic for a tunneling junction. In (a) the energy is plotted on the vertical scale and the density of orbitals on the horizontal scale. One metal is in the normal state and one in the superconducting state. (b) I versus V; the dashes indicate the expected break at $T = 0$. (After Giaever and Megerle.)

The gap $E_g$ corresponds to the break-up of a pair of electrons in the superconducting state, with the formation of two electrons, or an electron and a hole, in the normal state. The current starts when $eV = \Delta$. At finite temperatures there is a small current flow even at low voltages, because of electrons in the superconductor that are thermally excited across the energy gap.

**Josephson Superconductor Tunneling**

Under suitable conditions we observe remarkable effects associated with the tunneling of superconducting electron pairs from a superconductor through a layer of an insulator into another superconductor. Such a junction is called a weak link. The effects of pair tunneling include:

**Dc Josephson effect.** A dc current flows across the junction in the absence of any electric or magnetic field.

**Ac Josephson effect.** A dc voltage applied across the junction causes rf current oscillations across the junction. This effect has been utilized in a precision determination of the value of $\hbar/e$. Further, an rf voltage applied with the dc voltage can then cause a dc current across the junction.

**Macroscopic long-range quantum interference.** A dc magnetic field applied through a superconducting circuit containing two junctions causes the maximum supercurrent to show interference effects as a function of magnetic field intensity. This effect can be utilized in sensitive magnetometers.

**Dc Josephson Effect.** Our discussion of Josephson junction phenomena follows the discussion of flux quantization. Let $\psi_1$ be the probability amplitude of electron pairs on one side of a junction, and let $\psi_2$ be the amplitude on the other side. For simplicity, let both superconductors be identical. For the present we suppose that they are both at zero potential.

The time-dependent Schrödinger equation $i\hbar \frac{\partial \psi}{\partial t} = \hat{H} \psi$ applied to the two amplitudes gives
\[
\frac{i\hbar}{\partial t} \frac{\partial \psi_1}{\partial t} = \hbar T \psi_2 ; \quad \frac{i\hbar}{\partial t} \frac{\partial \psi_2}{\partial t} = \hbar T \psi_1 .
\] (38)

Here \(\hbar T\) represents the effect of the electron-pair coupling or transfer interaction across the insulator; \(T\) has the dimensions of a rate or frequency. It is a measure of the leakage of \(\psi_1\) into the region 2, and of \(\psi_2\) into the region 1. If the insulator is very thick, \(T\) is zero and there is no pair tunneling.

Let \(\psi_1 = n_1^{1/2} e^{i\theta_1}\) and \(\psi_2 = n_2^{1/2} e^{i\theta_2}\). Then

\[
\frac{\partial \psi_1}{\partial t} = \frac{1}{2} n_1^{-1/2} e^{i\theta_1} \frac{\partial n_1}{\partial t} + i n_1 \frac{\partial \theta_1}{\partial t} = -i T \psi_2 ;
\] (39)

\[
\frac{\partial \psi_2}{\partial t} = \frac{1}{2} n_2^{-1/2} e^{i\theta_2} \frac{\partial n_2}{\partial t} + i n_2 \frac{\partial \theta_2}{\partial t} = -i T \psi_1 .
\] (40)

We multiply (39) by \(n_1^{1/2} e^{-i\theta_1}\) to obtain, with \(\delta = \theta_2 - \theta_1\),

\[
\frac{1}{2} \frac{\partial n_1}{\partial t} + i n_1 \frac{\partial \theta_1}{\partial t} = -i T(n_1 n_2)^{1/2} e^{i\delta} .
\] (41)

We multiply (40) by \(n_2^{1/2} e^{-i\theta_2}\) to obtain

\[
\frac{1}{2} \frac{\partial n_2}{\partial t} + i n_2 \frac{\partial \theta_2}{\partial t} = -i T(n_1 n_2)^{1/2} e^{-i\delta} .
\] (42)

Now equate the real and imaginary parts of (41) and similarly of (42):

\[
\frac{\partial n_1}{\partial t} = 2 T(n_1 n_2)^{1/2} \sin \delta ; \quad \frac{\partial n_2}{\partial t} = -2 T(n_1 n_2)^{1/2} \sin \delta ;
\] (43)

\[
\frac{\partial \theta_1}{\partial t} = -T \left( \frac{n_2}{n_1} \right)^{1/2} \cos \delta ; \quad \frac{\partial \theta_2}{\partial t} = -T \left( \frac{n_1}{n_2} \right)^{1/2} \cos \delta .
\] (44)

If \(n_1 = n_2\) as for identical superconductors 1 and 2, we have from (44) that

\[
\frac{\partial \theta_1}{\partial t} = \frac{\partial \theta_2}{\partial t} ; \quad \frac{\partial}{\partial t} (\theta_2 - \theta_1) = 0 .
\] (45)

From (43) we see that

\[
\frac{\partial n_2}{\partial t} = -\frac{\partial n_1}{\partial t} .\] (46)

The current flow from (1) to (2) is proportional to \(\partial n_2/\partial t\) or, the same thing, \(-\partial n_1/\partial t\). We therefore conclude from (43) that the current \(J\) of superconductor pairs across the junction depends on the phase difference \(\delta\) as

\[
J = J_0 \sin \delta = J_0 \sin (\theta_2 - \theta_1) .
\] (47)

where \(J_0\) is proportional to the transfer interaction \(T\). The current \(J_0\) is the
maximum zero-voltage current that can be passed by the junction. With no applied voltage a dc current will flow across the junction (Fig. 24), with a value between $J_0$ and $-J_0$ according to the value of the phase difference $\theta_2 - \theta_1$. This is the dc Josephson effect.

**Ac Josephson Effect.** Let a dc voltage $V$ be applied across the junction. We can do this because the junction is an insulator. An electron pair experiences a potential energy difference $qV$ on passing across the junction, where $q = -2e$. We can say that a pair on one side is at potential energy $-eV$ and a pair on the other side is at $eV$. The equations of motion that replace (38) are

$$i\hbar \frac{\partial \psi_1}{\partial t} = \hbar T \psi_2 - eV \psi_1; \quad i\hbar \frac{\partial \psi_2}{\partial t} = \hbar T \psi_1 + eV \psi_2.$$  \hspace{1cm} (48)

We proceed as above to find in place of (41) the equation

$$\frac{1}{2} \frac{\partial n_1}{\partial t} + in_1 \frac{\partial \theta_1}{\partial t} = ieV n_1 \hbar^{-1} - iT(n_1 n_2)^{1/2} e^{i\delta}. \hspace{1cm} (49)$$

This equation breaks up into the real part

$$\frac{\partial n_1}{\partial t} = 2T(n_1 n_2)^{1/2} \sin \delta,$$

exactly as without the voltage $V$, and the imaginary part

$$\frac{\partial \theta_1}{\partial t} = (eV/\hbar) - T(n_2/n_1)^{1/2} \cos \delta,$$

which differs from (44) by the term $eV/\hbar$.

Further, by extension of (42),

$$\frac{1}{2} \frac{\partial n_2}{\partial t} + in_2 \frac{\partial \theta_2}{\partial t} = -i eV n_2 \hbar^{-1} - iT(n_1 n_2)^{1/2} e^{-i\delta}.$$  \hspace{1cm} (52)
whence
\[
\frac{\partial n_2}{\partial t} = -2T(n_1n_2)^{1/2} \sin \delta; \quad (53)
\]
\[
\frac{\partial \delta}{\partial t} = -(eV/\hbar) - T(n_1n_2)^{1/2} \cos \delta. \quad (54)
\]

From (51) and (54) with \( n_1 = n_2 \), we have
\[
\frac{\partial (\delta_2 - \delta_1)}{\partial t} = \frac{\partial \delta}{\partial t} = -2eV/\hbar. \quad (55)
\]

We see by integration of (55) that with a dc voltage across the junction the relative phase of the probability amplitudes vary as
\[
\delta(t) = \delta(0) - (2eVt/\hbar). \quad (56)
\]

The superconducting current is given by (47) with (56) for the phase:
\[
J = J_0 \sin [\delta(0) - (2eVt/\hbar)]. \quad (57)
\]

The current oscillates with frequency
\[
\omega = 2eV/\hbar. \quad (58)
\]

This is the ac Josephson effect. A dc voltage of 1 \( \mu V \) produces a frequency of 483.6 MHz. The relation (58) says that a photon of energy \( \hbar \omega = 2eV \) is emitted or absorbed when an electron pair crosses the barrier. By measuring the voltage and the frequency it is possible to obtain a very precise value\(^8\) of \( e/\hbar \).

**Macroscopic Quantum Interference.** We saw in (24) and (26) that the phase difference \( \theta_2 - \theta_1 \) around a closed circuit which encompasses a total magnetic flux \( \Phi \) is given by
\[
\theta_2 - \theta_1 = (2e/\hbar c)\Phi. \quad (59)
\]

The flux is the sum of that due to external fields and that due to currents in the circuit itself.

We consider two Josephson junctions in parallel, as in Fig. 25. No voltage is applied. Let the phase difference between points 1 and 2 taken on a path through junction \( a \) be \( \delta_a \). When taken on a path through junction \( b \), the phase

difference is $\delta_b$. In the absence of a magnetic field these two phases must be equal.

Now let the flux $\Phi$ pass through the interior of the circuit. We do this with a straight solenoid normal to the plane of the paper and lying inside the circuit. By (59), $\delta_b - \delta_a = (2e/\hbar c)\Phi$, or

$$\delta_b = \delta_0 + \frac{e}{\hbar c} \Phi \quad ; \quad \delta_a = \delta_0 - \frac{e}{\hbar c} \Phi .$$  \hspace{1cm} (60)

The total current is the sum of $J_a$ and $J_b$. The current through each junction is of the form (47), so that

$$J_{\text{Total}} = J_0 \left\{ \sin \left( \delta_0 + \frac{e}{\hbar c} \Phi \right) + \sin \left( \delta_0 - \frac{e}{\hbar c} \Phi \right) \right\} = 2J_0 \sin \delta_0 \cos \frac{e\Phi}{\hbar c} .$$

The current varies with $\Phi$ and has maxima when

$$e\Phi/\hbar c = s\pi , \quad s = \text{integer} .$$  \hspace{1cm} (61)

The periodicity of the current is shown in Fig. 26. The short period variation is produced by interference from the two junctions, as predicted by (61). The longer period variation is a diffraction effect and arises from the finite dimensions of each junction—this causes $\Phi$ to depend on the particular path of integration (Problem 6). The diffraction effect was not anticipated in early work that resolved only single electron tunneling.
HIGH-TEMPERATURE SUPERCONDUCTORS

High $T_c$ or HTS, denotes superconductivity in materials, chiefly oxides, with high transition temperatures, accompanied by high critical currents and magnetic fields. By 1988 the long-standing 23 K ceiling of $T_c$ in intermetallic compounds had been elevated to 125 K in bulk superconducting oxides; these passed the standard tests for superconductivity—the Meissner effect, ac Josephson effect, persistent currents of long duration, and substantially zero dc resistivity. By 1994 HTS showed promise in pre-commercial applications, as in thin film devices, and wires were being fabricated. Levitated vehicles and long distance power transmission have not arrived.

Memorable steps in the advance include:

- BaPb$_{0.75}$Bi$_{0.25}$O$_3$ \( T_c = 12 \text{ K} \) [BPBO]
- La$_{1.85}$Ba$_{0.15}$CuO$_4$ \( T_c = 36 \text{ K} \) [LBCO]
- YBa$_2$Cu$_3$O$_7$ \( T_c = 90 \text{ K} \) [YBCO]
- Tl$_2$Ba$_2$Ca$_2$Cu$_3$O$_{10}$ \( T_c = 120 \text{ K} \) [TBCO]

A related mercury compound may have $T_c = 140$ K. The material with most research behind it is YBCO.

The crystal structures of these materials, except for BPBO, are oxygen-defect modifications of the perovskite structure of Figure 13.10, with about one-third of the oxygen positions vacant. The primitive cell is developed from that of a tetragonal perovskite tripled along the $c$ axis. Per formula unit of YBa$_2$Cu$_3$O$_7$, the positive ion valencies based on Y$^{+3}$, Ba$^{+2}$, Cu$^{+2}$ are $3 \times 1 + 2 \times 2 + 2 \times 3 = 13$ and the negative (oxygen) valency is $-2 \times 7 = -14$. The quasitetragonal or orthorhombic cell dimensions of YBCO are, in nm, $a = 0.388$, $b = 0.384$, $c = 1.163$, with cell volume 0.173 nm$^3$; if there is one excess charge carrier in this volume the carrier concentration in $n_0 = 5.77 \times 10^{21}$ cm$^{-3}$. The excess carriers typically are holes embedded in doped antiferromagnetic insulators.

Parallel sheets of CuO$_2$ are a structural feature of all HTS with $T_c > 40$ K; these are known as the cuprate HTS. In YBCO shown in Figure 27 the CuO$_2$ planes are parallel to the plane of the $ab$ axes in the orthorhombic primitive cell. Two of these planes, those through the interior of the cell, are the dominant conducting pathways. The oxygen content per cell can be changed reversibly from 7.0 to 6.0 atoms, simply by pumping oxygen in and out of the parallel chains of CuO that run along the $b$ axis of the figure. At the composition YBa$_2$Cu$_3$O$_6$ the crystal is an insulator, with antiferromagnetic order of the Cu spins. An increase in oxygen above O$_{6.5}$ makes the crystal metallic and non-magnetic; the crystal is superconducting above O$_{6.64}$. A charge-neutral formula for YBa$_2$Cu$_3$O$_7$ can be written as YBa$_2$(Cu$^{2+})_2$(Cu$^{3+}$)(O$^{2-})_7$ or as YBa$_2$(Cu$^{2+})_3$(O$^{2-})_6$(O$^-$).
In the TlBaCaCu compounds, $T_c$ increases with the number of sheets of CuO$_2$ stacked continuously in a crystal cell; thus for the series (CuO$_2$)$_n$BaO$_2$TlO$_2$, (CuO$_2$)$_n$CaBaO$_2$TlO$_2$, (CuO$_2$)$_n$Ca$_2$Ba$_2$O$_2$Tl$_2$O$_2$ the $T_c$'s are 82, 109, 122 K. Isostructural compounds where Bi replaces Tl and Sr replaces Ba have $T_c = 10, 90, 110$ K for the one, two, and three layer compounds.

All these layer compounds are strongly anisotropic in electrical properties measured parallel or perpendicular to the layers. It is unrealistic in any approximation to neglect the anisotropy of the normal or superconducting properties.

**Critical Fields and Critical Currents**

Critical field and current values in HTS may be quite uncertain, because in anisotropic media the measurements are difficult and perhaps subject to large reevaluation for giant flux creep time effects. We give reasonably accepted values of the critical parameters for comparison with conventional materials.

High $T_c$ suggests a high stabilization energy and high energy gap for the superconducting state, from which follow high $H_c$ (from Eq. (9)) and short coherence lengths $\xi_0$ (from Eq. (17)). Taken in combination with a normal penetration depth from Eq. (14a), these results lead to extreme type II behavior; high $\kappa$ in Eq. (37b) and very high $H_{c2}$ in Eqs. (31) and (I.25). At the high temperatures accessible with HTS materials, thermally activated creep of
fluxons may limit tolerable current values; thus the achievement of high $T_c$ does not assure that a given material will have useful high current applications.

In the low temperature limit the London penetration depths are $\lambda_{ab} \approx 140$ nm and $\lambda_c \approx 700$ nm, from magnetization measurements; the coherence length $\xi_{ab} \approx 1.5$ nm from critical field measurements. The Ginzburg-Landau parameter $\kappa_{ab} = \lambda_{ab}/\xi_{ab} \approx 100$, which amply satisfies the criterion $\kappa > 1$ for a type II superconductor. The coherence length normal to the CuO$_2$ planes is $\xi_c \approx 0.2 - 0.6$ nm, of the order of the interplanar spacing itself. In thin films the critical current density at 4 K is $j_{ab} = 1.2 \times 10^7$ A/cm$^2$ in the $ab$ plane, and $j_c = 4.2 \times 10^5$ A/cm$^2$ in the $c$ direction, suggestive of a quasi-2D superconductor with planes connected by Josephson tunneling. Critical currents in ceramic powders are much smaller, typically $10^4$ A/cm$^2$. From $j_{ab}$ we may infer $H_{c2}(ab) \approx 10^7$ G = 1000 T; values up to 100 T have been measured in YBCO.

In narrow bridges (50 nm) much higher critical current densities have been observed, being limited by the depairing of Cooper pairs. A bridge is narrow if no fluxons form within the volume of the bridge, and the GL parameter $\psi$ may be taken as constant within the bridge. Then (1.10) becomes

$$-\alpha + \beta \psi + \frac{1}{2} m v^2 = 0,$$

(62)

where the last term is the kinetic energy of the Cooper pairs, each of mass $m$. The current density is

$$j = 2e|\psi|^2 v = (2e|\psi|^2/m)^{1/2}(2\alpha - 2\beta|\psi|^2)^{1/2},$$

(63)

which is a maximum with respect to $f = |\psi|^2/|\psi_0|^2$ when $f_c = \sqrt{(2/3)}$. Here $\psi_0$ is the GL parameter at zero current. Thus the maximum (depairing) current density is

$$j_c = (4/3\sqrt{3})e k m^2|\psi_0|^2;$$

(64)

the critical depairing velocity $v_c \approx k m \xi$. For $\lambda \approx 100$ nm and $\xi \approx 1$ nm, we have $j_c \approx 1.3 \times 10^9$ A/cm$^2$ in good agreement with experiment.

**Hall Number**

The Hall number is defined as the inverse Hall constant $1/R_{Hk}$ normalized to the formula volume. If only holes or electrons conduct, as in simple metals, the Hall number gives the carrier concentration per formula volume. In YBCO the number is p-type and proportional to temperature for $H||c$, for $T$ somewhat above $T_c$, and is of the order of unity at 300 K, giving a carrier concentration $\approx 7 \times 10^{21}$ cm$^{-3}$.

**Fullerenes**

These are stable, cage-like molecules that constitute the third form of pure carbon; the other two are diamond and graphite. The archetype fullerene is
C₆₀; each molecule has the form of a truncated icosahedron with 20 hexagonal faces and 12 pentagonal faces, like a soccer ball. This is one of the 13 Archimedean solids characterized by having all their angles equal and all their faces regular polygons. C₆₀ crystallizes in a face-centered cubic structure, as in Figure 28. Alkali-fullerene compounds such as K₂C₆₀ are superconducting; this one has $T_C = 19.2$ K. The K atoms occupy the octahedral sites in the cubic cell. RbCs₂C₆₀ has $T_c = 33$ K.

**SUMMARY**

*(In CGS Units)*

- A superconductor exhibits infinite conductivity.
- A bulk specimen of metal in the superconducting state exhibits perfect diamagnetism, with the magnetic induction $B = 0$. This is the Meissner effect. The external magnetic field will penetrate the surface of the specimen over a distance determined by the penetration depth $\lambda$.
- There are two types of superconductors, I and II. In a bulk specimen of type I superconductor the superconducting state is destroyed and the normal state is restored by application of an external magnetic field in excess of a critical value $H_c$. A type II superconductor has two critical fields $H_{c1} < H_c < H_{c2}$; a
vortex state exists in the range between \( H_{c1} \) and \( H_{c2} \). The stabilization energy density of the pure superconducting state is \( H^2/8\pi \) in both type I and II superconductors.

- In the superconducting state an energy gap, \( E_g = 4k_BT_\circ \), separates superconducting electrons below from normal electrons above the gap. The gap is detected in experiments on heat capacity, infrared absorption, and tunneling.

- The London equation

\[
j = -\frac{c}{4\pi\lambda_L^2} A \quad \text{or} \quad \text{curl } j = -\frac{c}{4\pi\lambda_L^2} B
\]

leads to the Meissner effect through the penetration equation \( \nabla^2 B = B/\lambda_L^2 \), where \( \lambda_L = (m^2/4\pi\hbar v_F)^{1/2} \) is the London penetration depth.

- Three important lengths enter the theory of superconductivity: the London penetration depth \( \lambda_L \); the intrinsic coherence length \( \xi_0 \); and the normal electron mean free path \( \ell \).

- In the London equation \( A \) or \( B \) should be a weighted average over the coherence length \( \xi \). The intrinsic coherence length \( \xi_0 = 2\hbar v_F/\pi E_g \).

- The BCS theory accounts for a superconducting state formed from pairs of electrons \( k \uparrow \) and \( -k \downarrow \). These pairs act as bosons.

- Type II superconductors have \( \xi < \lambda \). The critical fields are related by \( H_{c1} = (\xi/\lambda) H_c \) and \( H_{c2} = (\lambda/\xi) H_c \). The Ginzburg-Landau parameter \( \kappa \) is defined as \( \lambda/\xi \). Values of \( H_{c2} \) are as high as 500 kG = 50 T.

**Problems**

1. **Magnetic field penetration in a plate.** The penetration equation may be written as \( \lambda^2 \nabla^2 B = B \), where \( \lambda \) is the penetration depth. (a) Show that \( B(x) \) inside a superconducting plate perpendicular to the \( x \) axis and of thickness \( \delta \) is given by

\[
B(x) = B_a \frac{\cosh (x/\lambda)}{\cosh (\delta/2\lambda)}
\]

where \( B_a \) is the field outside the plate and parallel to it; here \( x = 0 \) is at the center of the plate. (b) The effective magnetization \( M(x) \) in the plate is defined by \( E(x) = B_a = 4\pi M(x) \). Show that, in CGS, \( 4\pi M(x) = -B_a(1/8\lambda^2)(\delta^2 - 4x^2) \), for \( \delta \ll \lambda \). In SI we replace the 4\( \pi \) by \( \mu_0 \).

2. **Critical field of thin films.** (a) Using the result of Problem 1b, show that the free energy density at \( T = 0 \) K within a superconducting film of thickness \( \delta \) in an external magnetic field \( B_a \) is given by, for \( \delta \ll \lambda \),

\[
F_S(x, B_a) = U_S(0) + (\delta^2 - 4x^2)B_a^2/64\pi\lambda^2
\]

(CGS)
In SI the factor $\pi$ is replaced by $\frac{1}{2}\mu_0$. We neglect a kinetic energy contribution to the problem. (b) Show that the magnetic contribution to $F_S$ when averaged over the thickness of the film is $B_0^2(\delta\lambda)^2/96\pi$. (c) Show that the critical field of the thin film is proportional to $(\lambda/\delta)H_c$, where $H_c$ is the bulk critical field, if we consider only the magnetic contribution to $U_S$.

3. **Two-fluid model of a superconductor.** On the two-fluid model of a superconductor we assume that at temperatures $0 < T < T_c$ the current density may be written as the sum of the contributions of normal and superconducting electrons: $j = j_N + j_s$, where $j_N = \sigma_0 E$ and $j_s$ is given by the London equation. Here $\sigma_0$ is an ordinary normal conductivity, decreased by the reduction in the number of normal electrons at temperature $T$ as compared to the normal state. Neglect inertial effects on both $j_N$ and $j_s$. (a) Show from the Maxwell equations that the dispersion relation connecting wavevector $k$ and frequency $\omega$ for electromagnetic waves in the superconductor is

\[
(CGS) \quad k^2\epsilon^2 = 4\pi\sigma_0 \omega i - c^2 \lambda_L^{-2} + \omega^2 ; \quad \text{or}
\]

\[
(SI) \quad k^2\epsilon^2 = (\sigma_0/\epsilon_0)\omega i - c^2 \lambda_F^{-2} + \omega^2 ,
\]

where $\lambda_n^2$ is given by (14a) with $n$ replaced by $n_s$. Recall that curl $\nabla \times B = -\nabla \times B$.

(b) If $\tau$ is the relaxation time of the normal electrons and $n_N$ is their concentration, show by use of the expression $\sigma_0 = n_N e^2/\tau m$ that at frequencies $\omega \ll 1/\tau$ the dispersion relation does not involve the normal electrons in an important way, so that the motion of the electrons is described by the London equation alone. The supercurrent short-circuits the normal electrons. The London equation itself only holds true if $\hbar\omega$ is small in comparison with the energy gap. Note: The frequencies of interest are such that $\omega \ll \omega_p$, where $\omega_p$ is the plasma frequency.

4. **Structure of a vortex.** (a) Find a solution to the London equation that has cylindrical symmetry and applies outside a line core. In cylindrical polar coordinates, we want a solution of

\[
B - \lambda^2 \nabla^2 B = 0
\]

that is singular at the origin and for which the total flux is the flux quantum:

\[
2\pi\int_0^\infty d\rho \, \rho B(\rho) = \Phi_0.
\]

The equation is in fact valid only outside the normal core of radius $\xi$. (b) Show that the solution has the limits

\[
B(\rho) = (\Phi_0/2\pi\lambda^2)\ln(\lambda/\rho) , \quad (\xi \ll \rho \ll \lambda)
\]

\[
B(\rho) = (\Phi_0/2\pi\lambda^2)(\pi\lambda/2\rho)^{1/2} \exp(-\rho/\lambda) , \quad (\rho \gg \lambda)
\]

*This problem is somewhat difficult.
5. **London penetration depth.** (a) Take the time derivative of the London equation (10) to show that \( \partial j / \partial t = (e^2 / 4 \pi \lambda_L^2) E \). (b) If \( mv / dt = qE \), as for free carriers of charge \( q \) and mass \( m \), show that \( \lambda_L^2 = me^2 / 4 \pi q^2 \).

*6. **Diffraction effect of Josephson junction.** Consider a junction of rectangular cross section with a magnetic field \( B \) applied in the plane of the junction, normal to an edge of width \( w \). Let the thickness of the junction be \( T \). Assume for convenience that the phase difference of the two superconductors is \( \pi / 2 \) when \( B = 0 \). Show that the dc current in the presence of the magnetic field is

\[
J = J_0 \frac{\sin(wTB e / hc)}{(wTB e / hc)}
\]

7. **Meissner effect in sphere.** Consider a sphere of a type I superconductor with critical field \( H_c \). (a) Show that in the Meissner regime the effective magnetization \( M \) within the sphere is given by \( -8 \pi M / 3 = B_o \), the uniform applied magnetic field. (b) Show that the magnetic field at the surface of the sphere in the equatorial plane is \( 3B_o / 2 \). (It follows that the applied field at which the Meissner effect starts to break down is \( 2H_c / 3 \).) Reminder: The demagnetization field of a uniformly magnetized sphere is \( -4 \pi M / 3 \).

**References**


**TYPE II SUPERCONDUCTIVITY**


**HIGH TEMPERATURE SUPERCONDUCTORS**


**JOSEPHSON EFFECTS**


# Dielectrics and Ferroelectrics

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maxwell equations</td>
<td>381</td>
</tr>
<tr>
<td>Polarization</td>
<td>381</td>
</tr>
<tr>
<td><strong>MACROSCOPIC ELECTRIC FIELD</strong></td>
<td>382</td>
</tr>
<tr>
<td>Depolarization field, $E_1$</td>
<td>384</td>
</tr>
<tr>
<td><strong>LOCAL ELECTRIC FIELD AT AN ATOM</strong></td>
<td>386</td>
</tr>
<tr>
<td>Lorentz field, $E_5$</td>
<td>388</td>
</tr>
<tr>
<td>Field of dipoles inside cavity, $E_3$</td>
<td>388</td>
</tr>
<tr>
<td><strong>DIELECTRIC CONSTANT AND POLARIZABILITY</strong></td>
<td>388</td>
</tr>
<tr>
<td>Electronic polarizability</td>
<td>390</td>
</tr>
<tr>
<td><strong>STRUCTURAL PHASE TRANSITIONS</strong></td>
<td>393</td>
</tr>
<tr>
<td><strong>FERROELECTRIC CRYSTALS</strong></td>
<td>393</td>
</tr>
<tr>
<td>Classification of ferroelectric crystals</td>
<td>394</td>
</tr>
<tr>
<td><strong>DISPLACITIVE TRANSITIONS</strong></td>
<td>398</td>
</tr>
<tr>
<td>Soft optical phonons</td>
<td>400</td>
</tr>
<tr>
<td>Landau theory of the phase transition</td>
<td>400</td>
</tr>
<tr>
<td>Second-order transition</td>
<td>402</td>
</tr>
<tr>
<td>First-order transition</td>
<td>402</td>
</tr>
<tr>
<td>Antiferroelectricity</td>
<td>404</td>
</tr>
<tr>
<td>Ferroelectric domains</td>
<td>406</td>
</tr>
<tr>
<td>Piezoelectricity</td>
<td>408</td>
</tr>
<tr>
<td>Ferroelasticity</td>
<td>410</td>
</tr>
<tr>
<td>Optical ceramics</td>
<td>410</td>
</tr>
<tr>
<td><strong>SUMMARY</strong></td>
<td>410</td>
</tr>
<tr>
<td><strong>PROBLEMS</strong></td>
<td>411</td>
</tr>
<tr>
<td>1. Atomic hydrogen</td>
<td>411</td>
</tr>
<tr>
<td>2. Conducting sphere</td>
<td>412</td>
</tr>
<tr>
<td>3. Effect of air gap</td>
<td>412</td>
</tr>
<tr>
<td>4. Interfacial polarization</td>
<td>412</td>
</tr>
<tr>
<td>5. Polarization of sphere</td>
<td>412</td>
</tr>
<tr>
<td>6. Ferroelectric criterion</td>
<td>412</td>
</tr>
<tr>
<td>7. Saturation polarization</td>
<td>412</td>
</tr>
<tr>
<td>8. Dielectric constant below transition</td>
<td>413</td>
</tr>
<tr>
<td>9. Soft modes</td>
<td>413</td>
</tr>
<tr>
<td>10. Ferroelectric linear array</td>
<td>414</td>
</tr>
</tbody>
</table>

**NOTATION:**

$$\varepsilon_0 = \frac{1}{4\pi c^2};$$  
(CGS)  

$$D = E + 4\pi P = \varepsilon E = (1 + 4\pi \chi)E;$$  

$$\alpha = \frac{\rho}{E_{\text{local}}};$$  

(SI)  

$$D = \varepsilon_0 E + P = \varepsilon_0 E = (1 + \chi)\varepsilon_0 E;$$  

$$\alpha_{\text{SI}} = 4\pi \varepsilon_0 \alpha_{\text{CGS}}.$$  

$$\varepsilon_{\text{CGS}} = \varepsilon_{\text{SI}};$$  

$$4\pi \chi_{\text{CGS}} = \chi_{\text{SI}};$$
Figure 1 The permanent dipole moment of a molecule of water has the magnitude $1.9 \times 10^{-18}$ esu-cm and is directed from O$^-$ ion toward the midpoint of the line connecting the H$^+$ ions. (To convert to SI units, multiply by $\frac{1}{4} \times 10^{11}$.)

Figure 2 Electrostatic potential and field components in CGS at position $r$, $\theta$ for a dipole $p$ directed along the $z$ axis. For $\theta = 0$, we have $E_x = E_y = 0$ and $E_z = 2p/r^3$; for $\theta = \pi/2$ we have $E_x = E_y = 0$ and $E_z = -p/r^3$. To convert to SI, replace $p$ by $p/4\pi\varepsilon_0$. (After E. M. Purcell.)
CHAPTER 13: DIELECTRICS AND FERROELECTRICS

First we relate the applied electric field to the internal electric field in a dielectric crystal. The study of the electric field within dielectric matter arises when we ask:

- What is the relation in the material between the dielectric polarization \( \mathbf{P} \) and the macroscopic electric field \( \mathbf{E} \) in the Maxwell equations?
- What is the relation between the dielectric polarization and the local electric field which acts at the site of an atom in the lattice? The local field determines the dipole moment of the atom.

**Maxwell Equations**

\[
\begin{align*}
\text{(CGS)} & \quad \text{curl } \mathbf{H} = \frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \frac{\partial}{\partial t} (\mathbf{E} + 4\pi \mathbf{P}) \; , \\
\text{curl } \mathbf{E} & = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} \; , \\
\text{div } \mathbf{E} & = 4\pi \rho \; , \\
\text{div } \mathbf{B} & = 0 \; , \\
\text{(SI)} & \quad \text{curl } \mathbf{H} = \mathbf{J} + \frac{\partial}{\partial t} (\varepsilon_0 \mathbf{E} + \mathbf{P}) \; , \\
\text{curl } \mathbf{E} & = -\frac{\partial \mathbf{B}}{\partial t} \; , \\
\text{div } \varepsilon_0 \mathbf{E} & = \rho \; , \\
\text{div } \mathbf{B} & = 0 \; .
\end{align*}
\]

**Polarization**

The polarization \( \mathbf{P} \) is defined as the dipole moment per unit volume, averaged over the volume of a cell. The total dipole moment is defined as

\[
\mathbf{p} = \sum q_n \mathbf{r}_n \; ,
\]

where \( \mathbf{r}_n \) is the position vector of the charge \( q_n \). The value of the sum will be independent of the origin chosen for the position vectors, provided that the system is neutral. The dipole moment of a water molecule is shown in Fig. 1.

The electric field at a point \( \mathbf{r} \) from a dipole moment \( \mathbf{p} \) is given by a standard result of elementary electrostatics:

\[
\begin{align*}
\text{(CGS)} & \quad \mathbf{E}(\mathbf{r}) = \frac{3(\mathbf{p} \cdot \mathbf{r}) \mathbf{r} - \mathbf{r}^2 \mathbf{p}}{r^5} \; , \\
\text{(SI)} & \quad \mathbf{E}(\mathbf{r}) = \frac{3(\mathbf{p} \cdot \mathbf{r}) \mathbf{r} - \mathbf{r}^2 \mathbf{p}}{4\pi \varepsilon_0 r^5} \; .
\end{align*}
\]

The lines of force of a dipole pointing along the \( z \) axis are shown in Fig. 2.
MACROSCOPIC ELECTRIC FIELD

One contribution to the electric field inside a body is that of the applied electric field, defined as

\[ E_0 = \text{field produced by fixed charges external to the body} \]  \hspace{1cm} (3)

The other contribution to the electric field is the sum of the fields of all charges that constitute the body. If the body is neutral, the contribution to the average field may be expressed in terms of the sum of the fields of atomic dipoles.

We define the average electric field \( E(r_0) \) as the average field over the volume of the crystal cell that contains the lattice point \( r_0 \):

\[ E(r_0) = \frac{1}{V_c} \int dV \, e(r) \]  \hspace{1cm} (4)

where \( e(r) \) is the microscopic electric field at the point \( r \). The field \( E \) is a much smoother quantity than the microscopic field \( e \). We could well have written the dipole field (2) as \( e(r) \) because it is a microscopic unsmoothed field.

We call \( E \) the macroscopic electric field. It is adequate for all problems in the electrodynamics of crystals provided that we know the connection between \( E \), the polarization \( P \), and the current density \( j \), and provided that the wavelengths of interest are long in comparison with the lattice spacing.\(^1\)

To find the contribution of the polarization to the macroscopic field, we can simplify the sum over all the dipoles in the specimen. By a famous theorem of electrostatics\(^2\) the macroscopic electric field caused by a uniform polarization is equal to the electric field in vacuum of a fictitious surface charge density

\(^1\)A detailed derivation of the Maxwell equations for the macroscopic fields \( E \) and \( B \), starting from the Maxwell equations in terms of the microscopic fields \( e \) and \( h \), is given by E. M. Purcell, Electricity and magnetism, 2nd ed., McGraw-Hill, 1965.

\(^2\)The electrostatic potential in CGS units of a dipole \( p \) is \( \phi(r) = p \cdot \text{grad}(1/r) \). For a volume distribution of polarization \( P \) we have

\[ \phi(r) = \int dV \left( P \cdot \text{grad} \frac{1}{r} \right) \]

which by a vector identity becomes

\[ \phi(r) = \int dV \left( -\frac{1}{r} \, \text{div} \, P + \text{div} \frac{P}{r^2} \right) \]

If \( P \) is constant, then \( \text{div} \, P = 0 \) and by the Gauss theorem we have

\[ \phi(r) = \int dS \frac{P_n}{r} = \int dS \frac{\sigma}{r} \]

where \( dS \) is an element of area on the surface of the body. This completes the proof.
\[ \sigma = \hat{n} \cdot \mathbf{P} \] on the surface of the body. Here \( \hat{n} \) is the unit normal to the surface, drawn outward from the polarized matter.

We apply the result to a thin dielectric slab (Fig. 3a) with a uniform volume polarization \( \mathbf{P} \). The electric field \( \mathbf{E}_1(r) \) produced by the polarization is equal to the field produced by the fictitious surface charge density \( \sigma = \hat{n} \cdot \mathbf{P} \) on the surface of the slab. On the upper boundary the unit vector \( \hat{n} \) is directed upward and on the lower boundary \( \hat{n} \) is directed downward. The upper boundary bears the fictitious charge \( \sigma = \hat{n} \cdot \mathbf{P} = +P \) per unit area, and the lower boundary bears \( -P \) per unit area.

The electric field \( \mathbf{E}_1 \) due to these charges has a simple form at any point between the plates, but comfortably removed from their edges. By Gauss's law

\[
\text{(CGS)} \quad E_1 = -4\pi|\sigma| = -4\pi P ;
\]

\[
\text{(SI)} \quad E_1 = -\frac{|\sigma|}{\varepsilon_0} = -\frac{P}{\varepsilon_0} .
\]

(4a)

We add \( \mathbf{E}_1 \) to the applied field \( \mathbf{E}_0 \) to obtain the total macroscopic field inside the slab, with \( \hat{z} \) the unit vector normal to the plane of the slab:

\[
\text{(CGS)} \quad \mathbf{E} = \mathbf{E}_0 + \mathbf{E}_1 = \mathbf{E}_0 - 4\pi P \hat{z} ;
\]

\[
\text{(SI)} \quad \mathbf{E} = \mathbf{E}_0 + \mathbf{E}_1 = \mathbf{E}_0 - \frac{P}{\varepsilon_0} \hat{z} .
\]

(5)

We define

\[
\mathbf{E}_1 = \text{field of the surface charge density } \hat{n} \cdot \mathbf{P} \text{ on the boundary} .
\]

(6)

This field is smoothly varying in space inside and outside the body and satisfies the Maxwell equations as written for the macroscopic field \( \mathbf{E} \). The reason \( \mathbf{E}_1 \) is a smooth function when viewed on an atomic scale is that we have replaced the discrete lattice of dipoles \( \mathbf{p}_j \) with the smoothed polarization \( \mathbf{P} \).
Depolarization Field, $E_1$

The geometry in many of our problems is such that the polarization is uniform within the body, and then the only contributions to the macroscopic field are from $E_0$ and $E_1$:

$$E = E_0 + E_1$$ (7)

Here $E_0$ is the applied field and $E_1$ is the field due to the uniform polarization.

The field $E_1$ is called the depolarization field, for within the body it tends to oppose the applied field $E_0$ as in Fig. 4. Specimens in the shape of ellipsoids, a class that includes spheres, cylinders, and discs as limiting forms, have an advantageous property: a uniform polarization produces a uniform depolarization field inside the body. This is a famous mathematical result demonstrated in classic texts on electricity and magnetism.\(^3\)

If $P_x, P_y, P_z$ are the components of the polarization $P$ referred to the principal axes of an ellipsoid, then the components of the depolarization field are written

\[
\begin{align*}
E_{1x} &= -N_x P_x ; & E_{1y} &= -N_y P_y ; & E_{1z} &= -N_z P_z ; \\
E_{1x} &= -\frac{N_x P_x}{\epsilon_0} ; & E_{1y} &= -\frac{N_y P_y}{\epsilon_0} ; & E_{1z} &= -\frac{N_z P_z}{\epsilon_0}.
\end{align*}
\] (8)

Here $N_x, N_y, N_z$ are the depolarization factors; their values depend on the ratios of the principal axes of the ellipsoid. The $N$'s are positive and satisfy the sum rule $N_x + N_y + N_z = 4\pi$ in CGS, and $N_x + N_y + N_z = 1$ in SI.

Values of $N$ parallel to the figure axis of ellipsoids of revolution are plotted in Fig. 5, additional cases have been calculated by Osborn\(^4\) and by Stoner. In limiting cases $N$ has the values:

<table>
<thead>
<tr>
<th>Shape</th>
<th>Axis</th>
<th>$N$ (CGS)</th>
<th>$N$ (SI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sphere</td>
<td>any</td>
<td>4$\pi$/3</td>
<td>1/3</td>
</tr>
<tr>
<td>Thin slab</td>
<td>normal</td>
<td>4$\pi$</td>
<td>1</td>
</tr>
<tr>
<td>Thin slab</td>
<td>in plane</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Long circular cylinder</td>
<td>longitudinal</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Long circular cylinder</td>
<td>transverse</td>
<td>2$\pi$</td>
<td>1/2</td>
</tr>
</tbody>
</table>

We can reduce the depolarization field to zero in two ways, either by working with a long fine specimen or by making an electrical connection between electrodes deposited on the opposite surfaces of a thin slab.

---


A uniform applied field $E_0$ will induce uniform polarization in an ellipsoid. We introduce the dielectric susceptibility $\chi$ such that the relations

\begin{align}
\text{(CGS)} \quad P &= \chi E ; \\
\text{(SI)} \quad P &= \varepsilon_0 \chi E ,
\end{align}

connect the macroscopic field $E$ inside the ellipsoid with the polarization $P$. Here $\chi_{\text{SI}} = 4\pi \chi_{\text{CGS}}$.

If $E_0$ is uniform and parallel to a principal axis of the ellipsoid, then

\begin{align}
\text{(CGS)} \quad E &= E_0 + E_1 = E_0 - NP ; \\
\text{(SI)} \quad E &= E_0 - \frac{NP}{\varepsilon_0} ,
\end{align}

by (8), whence

\begin{align}
\text{(CGS)} \quad P &= \chi (E_0 - NP) ; \\
&= \frac{\chi}{1 + N \chi} E_0 ; \\
\text{(SI)} \quad P &= \chi (\varepsilon_0 E_0 - NP) ; \\
&= \frac{\chi \varepsilon_0}{1 + N \chi} E_0 .
\end{align}

The value of the polarization depends on the depolarization factor $N$. 

---

**Figure 4** The depolarization field $E_1$ is opposite to $P$. The fictitious surface charges are indicated; the field of these charges is $E_1$ within the ellipsoid.

**Figure 5** Depolarization factor $N$ parallel to the figure axis of ellipsoids of revolution, as a function of the axial ratio $c/a$. 

![Diagram showing the depolarization field and the relationship between parameters](image-url)
LOCAL ELECTRIC FIELD AT AN ATOM

The value of the local electric field that acts at the site of an atom is significantly different from the value of the macroscopic electric field. We can convince ourselves of this by consideration of the local field at a site with a cubic arrangement of neighbors in a crystal of spherical shape. The macroscopic electric field in a sphere is

\[(\text{CGS})\]

\[E = E_0 + E_1 = E_0 - \frac{4\pi}{3} \rho ; \quad (12)\]

\[(\text{SI})\]

\[E = E_0 + E_1 = E_0 - \frac{1}{3\varepsilon_0} \rho , \quad \text{by (10).}\]

But consider the field that acts on the atom at the center of the sphere (this atom is not unrepresentative). If all dipoles are parallel to the z axis and have magnitude \(p\), the z component of the field at the center due to all other dipoles is, from (2),

\[(\text{CGS})\]

\[E_{\text{dipole}} = p \sum_i \frac{3x_i^2 - r_i^2}{r_i^5} = p \sum_i \frac{2x_i^2 - x_i^2 - y_i^2}{r_i^5} . \quad (13)\]

In SI we replace \(p\) by \(p/4\pi\varepsilon_0\). The \(x, y, z\) directions are equivalent because of the symmetry of the lattice and of the sphere; thus

\[\sum_i \frac{x_i^2}{r_i^5} = \sum_i \frac{x_i^2}{r_i^5} = \sum_i \frac{y_i^2}{r_i^5} ,\]

whence \(E_{\text{dipole}} = 0\).

The correct local field is just equal to the applied field, \(E_{\text{local}} = E_0\), for an atom site with a cubic environment in a spherical specimen. Thus the local field is not the same as the macroscopic average field \(E\).

We now develop an expression for the local field at a general lattice site, not necessarily of cubic symmetry. The local field at an atom is the sum of the electric field \(E_0\) from external sources and of the field from the dipoles within the specimen. It is convenient to decompose the dipole field so that part of the summation over dipoles may be replaced by integration.

We write

\[E_{\text{local}} = E_0 + E_1 + E_2 + E_3 . \quad (14)\]

\(^5\text{Atom sites in a cubic crystal do not necessarily have cubic symmetry; thus the O}^2^-\text{ sites in the barium titanate structure of Fig. 10 do not have a cubic environment. However, the Na}^+\text{ and Cl}^-\text{ sites in the NaCl structure and the Cs}^+\text{ and Cl}^-\text{ sites in the CsCl structure have cubic symmetry.}\)
Figure 6 The internal electric field on an atom in a crystal is the sum of the external applied field $E_0$ and of the field due to the other atoms in the crystal. The standard method of summing the dipole fields of the other atoms is first to sum individually over a moderate number of neighboring atoms inside an imaginary sphere concentric with the reference atom: this defines the field $E_3$, which vanishes at a reference site with cubic symmetry. The atoms outside the sphere can be treated as a uniformly polarized dielectric. Their contribution to the field at the reference point is $E_1 + E_2$, where $E_1$ is the depolarization field associated with the outer boundary and $E_2$ is the field associated with the surface of the spherical cavity.

Here

$E_0 = \text{field produced by fixed charges external to the body;}$

$E_1 = \text{depolarization field, from a surface charge density } \hat{n} \cdot P \text{ on the outer surface of the specimen;}$

$E_2 = \text{Lorentz cavity field: field from polarization charges on inside of a spherical cavity cut (as a mathematical fiction) out of the specimen with the reference atom as center, as in Fig. 6; } E_1 + E_2 \text{ is the field due to uniform polarization of the body in which a hole has been created;}$

$E_3 = \text{field of atoms inside cavity.}$

The contribution $E_1 + E_2 + E_3$ to the local field is the total field at one atom caused by the dipole moments of all the other atoms in the specimen:

\[
E_1 + E_2 + E_3 = \sum_i \frac{3(p_i \cdot r_i)r_i - r_i^3 p_i}{r_i^5},
\]

(CGS) \hspace{1cm} (15)

and in SI we replace $p_i$ by $p_i/4\pi\varepsilon_0$.

Dipoles at distances greater than perhaps ten lattice constants from the reference site make a smoothly varying contribution to this sum, a contribution which may be replaced by two surface integrals. One surface integral is taken over the outer surface of the ellipsoidal specimen and defines $E_1$, as in Eq. (6). The second surface integral defines $E_2$ and may be taken over any interior surface that is a suitable distance (say 50 Å) from the reference site. We count in $E_3$ any dipoles not included in the volume bounded by the inner and outer surfaces. It is convenient to let the interior surface be spherical.
**Lorentz Field, \( E_2 \)**

The field \( E_2 \) due to the polarization charges on the surface of the fictitious cavity was calculated by Lorentz. If \( \theta \) is the polar angle (Fig. 7) referred to the polarization direction, the surface charge density on the surface of the cavity is \(-P \cos \theta\). The electric field at the center of the spherical cavity of radius \( a \) is

\[
\text{(CGS)} \quad E_2 = \int_0^\pi (a^{-2})(2\pi a \sin \theta)(a \, d\theta)(P \cos \theta)(\cos \theta) = \frac{4\pi}{3} \frac{1}{\epsilon_0} P ; \quad (16)
\]

\[
\text{(SI)} \quad E_2 = \frac{1}{3\epsilon_0} P .
\]

This is the negative of the depolarization field \( E_1 \) in a polarized sphere, so that \( E_1 + E_2 = 0 \) for a sphere.

**Field of Dipoles Inside Cavity, \( E_3 \)**

The field \( E_3 \) due to the dipoles within the spherical cavity is the only term that depends on the crystal structure. We showed for a reference site with cubic surroundings in a sphere that \( E_3 = 0 \) if all the atoms may be replaced by point dipoles parallel to each other.

The total local field at a cubic site is, from (14) and (16),

\[
\text{(CGS)} \quad E_{\text{local}} = E_0 + E_1 + \frac{4\pi}{3} P = E + \frac{4\pi}{3} P ; \quad (17)
\]

\[
\text{(SI)} \quad E_{\text{local}} = E + \frac{1}{3\epsilon_0} P .
\]

This is the Lorentz relation: the field acting at an atom in a cubic site is the macroscopic field \( E \) of Eq. (7) plus \( 4\pi P/3 \) or \( P/3\epsilon_0 \) from the polarization of the other atoms in the specimen. Experimental data for cubic ionic crystals support the Lorentz relation.

**DIELECTRIC CONSTANT AND POLARIZABILITY**

The dielectric constant \( \epsilon \) of an isotropic or cubic medium relative to vacuum is defined in terms of the macroscopic field \( E \):

\[
\text{(CGS)} \quad \epsilon = \frac{E + 4\pi P}{E} = 1 + 4\pi \chi ; \quad (18)
\]

\[
\text{(SI)} \quad \epsilon = \frac{\epsilon_0 E + P}{\epsilon_0 E} = 1 + \chi .
\]

Remember that \( \chi_{\text{SI}} = 4\pi \chi_{\text{CGS}} \), by definition, but \( \epsilon_{\text{SI}} = \epsilon_{\text{CGS}} \).
The susceptibility $\chi$ is related to the dielectric constant by

\[
\frac{\chi}{\epsilon} = \frac{P}{E} = \frac{\epsilon - 1}{4\pi} ; \quad \frac{\chi}{\epsilon_0} = \frac{P}{\epsilon_0 E} = \epsilon - 1 .
\]

(19)

In a noncubic crystal the dielectric response is described by the components of the susceptibility tensor or of the dielectric constant tensor:

\[
\begin{align*}
\text{(CGS)} & \quad P_{\mu} = \chi_{\mu\nu}E_{\nu} ; \quad \epsilon_{\mu\nu} = \delta_{\mu\nu} + 4\pi\chi_{\mu\nu} . \\
\text{(SI)} & \quad P_{\mu} = \chi_{\mu\nu}\epsilon_0 E_{\nu} ; \quad \epsilon_{\mu\nu} = \delta_{\mu\nu} + \chi_{\mu\nu} . 
\end{align*}
\]

(20)

The polarizability $\alpha$ of an atom is defined in terms of the local electric field at the atom:

\[
p = \alpha E_{\text{local}} ,
\]

(21)

where $p$ is the dipole moment. This definition applies in CGS and in SI, but $\alpha_{\text{SI}} = 4\pi\epsilon_0\alpha_{\text{CGS}}$. The polarizability is an atomic property, but the dielectric constant will depend on the manner in which the atoms are assembled to form a crystal. For a non-spherical atom $\alpha$ will be a tensor.

The polarization of a crystal may be expressed approximately as the product of the polarizabilities of the atoms times the local electric field:

\[
P = \sum_j N_j p_j = \sum_j N_j \alpha_j E_{\text{loc}(j)} ,
\]

(22)

where $N_j$ is the concentration and $\alpha_j$ the polarizability of atoms $j$, and $E_{\text{loc}(j)}$ is the local field at atom sites $j$.

We want to relate the dielectric constant to the polarizabilities; the result will depend on the relation that holds between the macroscopic electric field and the local electric field. We give the derivation in CGS units and state the result in both systems of units.
If the local field is given by the Lorentz relation (17), then

\[
(P = (\Sigma N_j \alpha_j) \left( E + \frac{4\pi}{3} P \right);
\]

and we solve for \( P \) to find the susceptibility

\[
\chi = \frac{P}{E} = \frac{\Sigma N_j \alpha_j}{1 - \frac{4\pi}{3} \Sigma N_j \alpha_j}.
\] (23)

By definition \( \epsilon = 1 + 4\pi\chi \) in CGS; we may rearrange (23) to obtain

\[
\frac{\epsilon - 1}{\epsilon + 2} = \frac{4\pi}{3} \Sigma N_j \alpha_j; \quad (S1) \quad \frac{\epsilon - 1}{\epsilon + 2} = \frac{1}{3\epsilon_0} \Sigma N_j \alpha_j, \]

the Clausius-Mossotti relation. This relates the dielectric constant to the electronic polarizability, but only for crystal structures for which the Lorentz local field (17) obtains.

**Electronic Polarizability**

The total polarizability may usually be separated into three parts: electronic, ionic, and dipolar, as in Fig. 8. The electronic contribution arises from the displacement of the electron shell relative to a nucleus. The ionic contribution comes from the displacement of a charged ion with respect to other ions. The dipolar polarizability arises from molecules with a permanent electric dipole moment that can change orientation in an applied electric field.

In heterogeneous materials there is usually also an interfacial polarization arising from the accumulation of charge at structural interfaces. This is of little fundamental interest, but it is of considerable practical interest because commercial insulating materials are usually heterogeneous.\(^6\)

The dielectric constant at optical frequencies arises almost entirely from the electronic polarizability. The dipolar and ionic contributions are small at high frequencies because of the inertia of the molecules and ions. In the optical range (24) reduces to

\[
\frac{n^2 - 1}{n^2 + 2} = \frac{4\pi}{3} \Sigma N_j \alpha_j(\text{electronic}); \]

here we have used the relation \( n^2 = \epsilon \), where \( n \) is the refractive index.

By applying (25) to large numbers of crystals we determine in Table 1 empirical values of the electronic polarizabilities that are reasonably consistent with the observed values of the refractive index. The scheme is not entirely

Figure 8  Frequency dependence of the several contributions to the polarizability.

### Table 1  Electronic polarizabilities of ions, in $10^{-24}$ cm$^3$

<table>
<thead>
<tr>
<th></th>
<th>He</th>
<th>Li$^+$</th>
<th>Be$^{2+}$</th>
<th>B$^{3+}$</th>
<th>C$^{4+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pauling JS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O$^{2-}$</td>
<td>0.201</td>
<td>0.029</td>
<td>0.008</td>
<td>0.003</td>
<td>0.0013</td>
</tr>
<tr>
<td>F$^-$</td>
<td>3.88</td>
<td>0.104</td>
<td>0.390</td>
<td>0.290</td>
<td>0.0165</td>
</tr>
<tr>
<td>Ne</td>
<td>0.858</td>
<td>0.179</td>
<td>0.094</td>
<td>0.052</td>
<td>0.185</td>
</tr>
<tr>
<td>Na$^+$</td>
<td>0.290</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg$^{2+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al$^{3+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Si$^{4+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S$^{2-}$</td>
<td>10.2</td>
<td>2.947</td>
<td>1.133</td>
<td>(1.1)</td>
<td>(0.19)</td>
</tr>
<tr>
<td>Cl$^-$</td>
<td>1.62</td>
<td>0.83</td>
<td>0.47</td>
<td>0.286</td>
<td>0.37</td>
</tr>
<tr>
<td>Ar</td>
<td>2.46</td>
<td>1.40</td>
<td>0.86</td>
<td>0.55</td>
<td>0.73</td>
</tr>
<tr>
<td>K$^+$</td>
<td>2.46</td>
<td>1.40</td>
<td>0.86</td>
<td>0.55</td>
<td>0.73</td>
</tr>
<tr>
<td>Ca$^{2+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sc$^{3+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ti$^{4+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Se$^{2-}$</td>
<td>10.5</td>
<td>4.77</td>
<td>2.46</td>
<td>1.40</td>
<td>0.86</td>
</tr>
<tr>
<td>Br$^-$</td>
<td>4.091</td>
<td>1.679</td>
<td>1.133</td>
<td>(1.1)</td>
<td>(0.19)</td>
</tr>
<tr>
<td>Kr</td>
<td>1.679</td>
<td>1.133</td>
<td>(1.1)</td>
<td>(0.19)</td>
<td></td>
</tr>
<tr>
<td>Rb$^+$</td>
<td>1.679</td>
<td>1.133</td>
<td>(1.1)</td>
<td>(0.19)</td>
<td></td>
</tr>
<tr>
<td>Sr$^{2+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y$^{3+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zr$^{4+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Te$^{2-}$</td>
<td>14.0</td>
<td>7.10</td>
<td>3.99</td>
<td>2.42</td>
<td>1.55</td>
</tr>
<tr>
<td>I$^-$</td>
<td>6.116</td>
<td>2.743</td>
<td>(2.5)</td>
<td>(0.19)</td>
<td></td>
</tr>
<tr>
<td>Xe</td>
<td>2.743</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cs$^+$</td>
<td>2.743</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ba$^{2+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>La$^{3+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ce$^{4+}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values from L. Pauling, Proc. R. Soc. London A114, 181 (1927); S. S. Jaswal and T. P. Sharma, J. Phys. Chem. Solids 34, 509 (1973); and J. Tessman, A. Kahn, and W. Shockley, Phys. Rev. 92, 890 (1953). The TKS polarizabilities are at the frequency of the D lines of sodium. The values are in CGS; to convert to SI, multiply by $\frac{1}{2} \times 10^{-15}$. 
self-consistent, because the electronic polarizability of an ion depends somewhat on the environment in which it is placed. The negative ions are highly polarizable because they are large.

**Classical Theory of Electronic Polarizability.** An electron bound harmonically to an atom will show resonance absorption at a frequency \( \omega_0 = (\beta/m)^{1/2} \), where \( \beta \) is the force constant. The displacement \( x \) of the electron occasioned by the application of a field \( E_{\text{loc}} \) is given by

\[
-eE_{\text{loc}} = \beta x = m\omega_0^2 x ,
\]

so that the static electronic polarizability is

\[
\alpha(\text{electronic}) = \frac{p}{E_{\text{loc}}} = -ex/E_{\text{loc}} = e^2/m\omega_0^2 .
\]

The electronic polarizability will depend on frequency, and it is shown in the following example that for frequency \( \omega \)

(CGS) \[
\alpha(\text{electronic}) = \frac{e^2/m}{\omega_0^2 - \omega^2} ;
\]

but in the visible region the frequency dependence (dispersion) is not usually very important in most transparent materials.

**EXAMPLE: Frequency dependence.** Find the frequency dependence of the electronic polarizability of an electron having the resonance frequency \( \omega_0 \), treating the system as a simple harmonic oscillator.

The equation of motion in the local electric field \( E_{\text{loc}} \sin \omega t \) is

\[
m \frac{d^2x}{dt^2} + m\omega_0^2 x = -eE_{\text{loc}} \sin \omega t ,
\]

so that, for \( x = x_0 \sin \omega t \),

\[
m(-\omega^2 + \omega_0^2)x_0 = -eE_{\text{loc}} .
\]

The dipole moment has the amplitude

\[
p_0 = -ex_0 = \frac{e^2E_{\text{loc}}}{m(\omega_0^2 - \omega^2)} ,
\]

from which (28) follows.

In quantum theory the expression corresponding to (28) is

(CGS) \[
\alpha(\text{electronic}) = \frac{e^2}{m} \sum_j \frac{f_{ij}}{\omega_j^2 - \omega^2} ,
\]

where \( f_{ij} \) is called the oscillator strength of the electric dipole transition between the atomic states \( i \) and \( j \).
STRUCTURAL PHASE TRANSITIONS

It is not uncommon for crystals to transform from one crystal structure to another as the temperature or pressure is varied. The stable structure $A$ at absolute zero generally has the lowest accessible internal energy of all the possible structures. Even this selection of a structure $A$ can be varied with application of pressure, because a low atomic volume will favor closest-packed or even metallic structures. Hydrogen and xenon, for example, become metallic under extreme pressure.

Some other structure $B$ may have a softer or lower frequency phonon spectrum than $A$. As the temperature is increased the phonons in $B$ will be more highly excited (higher thermal average occupancies) than the phonons in $A$. Because the entropy increases with the occupancy, the entropy of $B$ will become higher than the entropy of $A$ as the temperature is increased.

It is thereby possible for the stable structure to transform from $A$ to $B$ as the temperature is increased. The stable structure at a temperature $T$ is determined by the minimum of the free energy $F = U - TS$. There will be a transition from $A$ to $B$ if a temperature $T_c$ exists (below the melting point) such that $F_A(T_c) = F_B(T_c)$.

Often several structures have nearly the same internal energy at absolute zero. The phonon dispersion relations for the structures may, however, be rather different. The phonon energies are sensitive to the number and arrangement of nearby atoms; these are the quantities that change as the structure is changed.

Some structural phase transitions have only small effects on the macroscopic physical properties of the material. However, if the transition is influenced by an applied stress, the crystal may yield mechanically quite easily near the transition temperature because the relative proportions in the two phases will change under stress. Some other structural phase transitions may have spectacular effects on the macroscopic electrical properties.

Ferroelectric transitions are a subgroup of structural phase transitions, a subgroup marked by the appearance of a spontaneous dielectric polarization in the crystal. Ferroelectrics are of theoretical and technical interest because they often have unusually high and unusually temperature-dependent values of the dielectric constant, the piezoelectric effect, the pyroelectric effect, and optical effects, including optical frequency doubling.

FERROELECTRIC CRYSTALS

A ferroelectric crystal exhibits an electric dipole moment even in the absence of an external electric field. In the ferroelectric state the center of positive charge of the crystal does not coincide with the center of negative charge.

The plot of polarizations versus electric field for the ferroelectric state shows a hysteresis loop. A crystal in a normal dielectric state usually does not
show significant hysteresis when the electric field is increased and then reversed, both slowly.

In some crystals the ferroelectric dipole moment is not changed by an electric field of the maximum intensity which it is possible to apply before causing electrical breakdown. In these crystals we are often able to observe a change in the spontaneous moment when the temperature is changed (Fig. 9). Such crystals are called pyroelectric.

Lithium niobate, LiNbO$_3$, is pyroelectric at room temperature. It has a high transition temperature ($T_c=1480$ K) and a high saturation polarization (50 $\mu$C/cm$^2$). It can be “poled,” which means given a remanent polarization, by an electric field applied over 1400 K.

Ferroelectricity usually disappears above a certain temperature called the transition temperature. Above the transition the crystal is said to be in a paraelectric state. The term paraelectric suggests an analogy with paramagnetism; similarly, there is usually a rapid drop in the dielectric constant as the temperature increases.

**Classification of Ferroelectric Crystals**

We list in Table 2 some of the crystals commonly considered to be ferroelectric, along with the transition temperature or Curie point $T_c$ at which the crystal changes from the low temperature polarized state to the high temperature unpolarized state. Thermal motion tends to destroy the ferroelectric order. Some ferroelectric crystals have no Curie point because they melt before leaving the ferroelectric phase. The table also includes values of the spontaneous polarization $P_s$.

Ferroelectric crystals may be classified into two main groups, order-disorder or displacive. If in the paraelectric phase the atomic displacements are oscillations about a nonpolar site, then after a displacive transition the oscillations are about a polar site.

If in the paraelectric phase the displacements are about some double-well or multi-well configuration of sites, then in an order-disorder transition the displacements are about an ordered subset of these wells.

There has been a tendency recently to define the character of the transition in terms of the dynamics of the lowest frequency (“soft”) optical phonon modes. If a soft mode can propagate in the crystal at the transition, then the transition is displacive. If the soft mode is only diffusive (non-propagating) there is really not a phonon at all, but only a large amplitude hopping motion between the wells of the order-disorder system. Many ferroelectrics have soft modes that fall between these two extremes.

The order-disorder class of ferroelectrics includes crystals with hydrogen bonds in which the motion of the protons is related to the ferroelectric properties, as in potassium dihydrogen phosphate (KH$_2$PO$_4$) and isomorphous salts.
Figure 9 The temperature variation of (a) the dielectric constant $\varepsilon$, (b) the pyroelectric coefficient $dP/dT$, and (c) the specific heat $c_p$, of PbTiO$_3$. (After Remeika and Glass.)
Table 2  Ferroelectric crystals

To obtain \( P_s \) in the CGS unit of esu cm\(^{-2} \), multiply the value given in \( \mu C \) cm\(^{-2} \) by \( 3 \times 10^9 \).

<table>
<thead>
<tr>
<th></th>
<th>( T_c ) in K</th>
<th>( P_s ) in ( \mu C ) cm(^{-2} ), at ( T ) K</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KDP type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{KH}_2\text{PO}_4 )</td>
<td>123</td>
<td>4.75</td>
</tr>
<tr>
<td>( \text{KD}_2\text{PO}_4 )</td>
<td>213</td>
<td>4.83</td>
</tr>
<tr>
<td>( \text{RbH}_2\text{PO}_4 )</td>
<td>147</td>
<td>5.6</td>
</tr>
<tr>
<td>( \text{KH}_2\text{AsO}_4 )</td>
<td>97</td>
<td>5.0</td>
</tr>
<tr>
<td>GeTe</td>
<td>670</td>
<td>—</td>
</tr>
<tr>
<td><strong>TGS type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tri-glycine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sulfate</td>
<td>322</td>
<td>2.8</td>
</tr>
<tr>
<td>Tri-glycine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>selenate</td>
<td>295</td>
<td>3.2</td>
</tr>
<tr>
<td><strong>Perovskites</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{BaTiO}_3 )</td>
<td>408</td>
<td>26.0</td>
</tr>
<tr>
<td>( \text{KNbO}_3 )</td>
<td>708</td>
<td>30.0</td>
</tr>
<tr>
<td>( \text{PbTiO}_3 )</td>
<td>765</td>
<td>&gt; 50</td>
</tr>
<tr>
<td>( \text{LiTaO}_3 )</td>
<td>938</td>
<td>50</td>
</tr>
<tr>
<td>( \text{LiNbO}_3 )</td>
<td>1480</td>
<td>71</td>
</tr>
</tbody>
</table>

* A compilation of data on ferroelectric and antiferroelectric materials is given by E. C. Subarao, Ferroelectrics 5, 267 (1973).

The behavior of crystals in which the hydrogen has been replaced by deuterium is interesting:

\[
\begin{array}{cccc}
\text{KH}_2\text{PO}_4 & \text{KD}_2\text{PO}_4 & \text{KH}_2\text{AsO}_4 & \text{KD}_2\text{AsO}_4 \\
123 \text{ K} & 213 \text{ K} & 96 \text{ K} & 162 \text{ K} \\
\end{array}
\]

The substitution of deuterons for protons nearly doubles \( T_c \), although the fractional change in the molecular weight of the compound is less than 2 percent. This extraordinarily large isotope shift is believed to be a quantum effect involving the mass-dependence of the de Broglie wavelength. Neutron diffraction data show that above the Curie temperature the proton distribution along the hydrogen bond is symmetrically elongated. Below the Curie temperature the distribution is more concentrated and asymmetric with respect to neighboring ions, so that one end of the hydrogen bond is preferred by the proton over the other end.

The displacive class of ferroelectrics includes ionic crystal structures closely related to the perovskite and ilmenite structures. The simplest ferroelectric crystal is GeTe with the sodium chloride structure. We shall devote ourselves primarily to crystals with the perovskite structure, Fig. 10.
Figure 10  (a) The crystal structure of barium titanate. The prototype crystal is calcium titanate (perovskite). The structure is cubic, with Ba\(^{2+}\) ions at the cube corners, O\(^{2-}\) ions at the face centers, and a Ti\(^{4+}\) ion at the body center. (b) Below the Curie temperature the structure is slightly deformed, with Ba\(^{2+}\) and Ti\(^{4+}\) ions displaced relative to the O\(^{2-}\) ions, thereby developing a dipole moment. The upper and lower oxygen ions may move downward slightly.

Figure 11  Spontaneous polarization projected on cube edge of barium titanate, as a function of temperature. (After W. J. Merz.)

Consider the order of magnitude of the ferroelectric effects in barium titanate: the observed saturation polarization \(P_s\) at room temperature (Fig. 11) is \(8 \times 10^4\) esu cm\(^{-2}\). The volume of a cell is \((4 \times 10^{-8})^3 = 64 \times 10^{-24}\) cm\(^3\), so that the dipole moment of a cell is

\[
\text{(CGS)} \quad p \equiv (8 \times 10^4 \text{ esu cm}^{-2})(64 \times 10^{-24} \text{ cm}^3) \approx 5 \times 10^{-18} \text{ esu cm} ;
\]

\[
\text{(SI)} \quad p \equiv (3 \times 10^{-1} \text{ C m}^{-2})(64 \times 10^{-30} \text{ m}^3) \approx 2 \times 10^{-29} \text{ C m} .
\]
If the positive ions Ba\(^{2+}\) and Ti\(^{4+}\) were moved by \(\delta = 0.1\ \text{Å}\) with respect to the negative O\(^{2-}\) ions, the dipole moment of a cell would be \(6e\delta \approx 3 \times 10^{-18}\) esu cm. In LiNbO\(_3\) the displacements are considerably larger, being 0.9 Å and 0.5 Å for the lithium and niobium ions respectively.

**DISPLACIVE TRANSITIONS**

Two viewpoints contribute to an understanding of a ferroelectric displacive transition and by extension to displacive transitions in general. We may speak of a polarization catastrophe in which for some critical condition the polarization or some Fourier component of the polarization becomes very large. Equally, we may speak of the condensation of a transverse optical phonon. Here the word condensation is to be understood in the Bose-Einstein sense (TP, p. 199) of a time-independent displacement of finite amplitude. This can occur when the corresponding TO phonon frequency vanishes at some point in the Brillouin zone. LO phonons always have higher frequencies than the TO phonons of the same wavevector, so we are not concerned with LO phonon condensation.

In a polarization catastrophe the local electric field caused by the ionic displacement is larger than the elastic restoring force, thereby giving an asymmetrical shift in the positions of the ions. Higher order restoring forces will limit the shift to a finite displacement.

The occurrence of ferroelectricity (and antiferroelectricity) in many perovskite-structure crystals suggests that this structure is favorably disposed to a displacive transition. Local field calculations make clear the reason for the favored position of this structure: the O\(^{2-}\) ions do not have cubic surroundings, and the local field factors turn out to be unusually large.

We give first the simple form of the catastrophe theory, supposing that the local field at all atoms is equal to \(E + 4\pi P/3\) in CGS or \(E + P/3\epsilon_0\) in SI. The theory given now leads to a second-order transition; the physical ideas can be carried over to a first-order transition. In a second-order transition there is no latent heat; the order parameter (in this instance, the polarization) is not discontinuous at the transition temperature. In a first-order transition there is a latent heat; the order parameter changes discontinuously at the transition temperature.

We rewrite (24) for the dielectric constant in the form

\[
\epsilon = \frac{1 + \frac{8\pi}{3} \sum N_i \alpha_i}{1 - \frac{4\pi}{3} \sum N_i \alpha_i},
\]

(CGS)
where $\alpha_i$ is the electronic plus ionic polarizability of an ion of type $i$ and $N_i$ is the number of ions $i$ per unit volume. The dielectric constant becomes infinite and permits a finite polarization in zero applied field when

\[
\Sigma N_i \alpha_i = 3/4 \pi .
\]  

(31)

This is the condition for a polarization catastrophe.

The value of $\varepsilon$ in (30) is sensitive to small departures of $\Sigma N_i \alpha_i$ from the critical value $3/4 \pi$. If we write

\[
(4\pi/3) \Sigma N_i \alpha_i = 1 - 3s ,
\]

(32)

where $s \ll 1$, the dielectric constant in (30) becomes

\[
\varepsilon = \frac{1}{s} .
\]

(33)

Suppose near the critical temperature $s$ varies linearly with temperature:

\[
s = (T - T_c)/\xi ,
\]

(34)

where $\xi$ is a constant. Such a variation of $s$ or $\Sigma N_i \alpha_i$ might come from normal thermal expansion of the lattice. The dielectric constant has the form

\[
\varepsilon = \frac{\xi}{T - T_c} ,
\]

(35)

close to the observed temperature variation in the paraelectric state, Fig. 12.
Soft Optical Phonons

The Lyddane-Sachs-Teller relation (Chapter 10) is

$$\omega_{\gamma}^2/\omega_{\chi}^2 = \varepsilon(\infty)/\varepsilon(0).$$

(36)

The static dielectric constant increases when the transverse optical phonon frequency decreases. When the static dielectric constant $\varepsilon(0)$ has a high value, such as 100 to 10,000, we find that $\omega_{\gamma}$ has a low value.

When $\omega_{\gamma} = 0$ the crystal is unstable because there is no effective restoring force. The ferroelectric BaTiO$_3$ at 24°C has a TO mode at 12 cm$^{-1}$, a low frequency for an optical mode.

If the transition to a ferroelectric state is first order, we do not find $\omega_{\gamma} = 0$ or $\varepsilon(0) = \infty$ at the transition. The LST relation suggests only that $\varepsilon(0)$ extrapolates to a singularity at a temperature $T_0$ below $T_c$. In disordered (“dirty”) ferroelectrics the static dielectric constant can be larger than suggested by the LST relations.\(^7\)

The association of a high static dielectric constant with a low frequency optical mode is supported by experiments on strontium titanate, SrTiO$_3$. According to the LST relation, if the reciprocal of the static dielectric constant has a temperature dependence $1/\varepsilon(0) \propto (T - T_0)$, then the square of the optical mode frequency will have a similar temperature dependence: $\omega_{\gamma}^2 \propto (T - T_0)$, if $\omega_{\gamma}$ is independent of temperature. The result for $\omega_{\gamma}^2$ is very well confirmed by Fig. 13. Measurements of $\omega_{\gamma}$ versus $T$ for another ferroelectric crystal, SbSI, are shown in Fig. 14.

Landau Theory of the Phase Transition

A ferroelectric with a first-order phase transition between the ferroelectric and the paraelectric state is distinguished by a discontinuous change of the saturation polarization at the transition temperature. The transition between the normal and superconducting states is a second-order transition, as is the transition between the ferromagnetic and paramagnetic states. In these transitions the degree of order goes to zero without a discontinuous change as the temperature is increased.

We can obtain a consistent formal thermodynamic theory of the behavior of a ferroelectric crystal by considering the form of the expansion of the energy as a function of the polarization $P$. We assume that the Landau\(^8\) free energy density $\hat{F}$ in one dimension may be expanded formally as

$$\hat{F}(P;T,E) = -EP + g_0 + \frac{1}{2}g_2P^2 + \frac{1}{4}g_4P^4 + \frac{1}{6}g_6P^6 + \cdots,$$  \hspace{1cm} (37)

where the coefficients $g_n$ depend on the temperature.

\(^7\)G. Burns and E. Burstein, Ferroelectrics 7, 297 (1974).

\(^8\)In TP, see pp. 69 and 298 for a discussion of the Landau function.
Figure 13  Plot of the square of the frequency of the zero wavevector transverse optic mode against temperature, for SrTiO$_3$, as observed in neutron diffraction experiments by Cowley. The broken line is the reciprocal of the dielectric constant from the measurements of Mitsui and Westphal.

Figure 14  Decrease of a transverse phonon frequency as the Curie temperature is approached from below, in the ferroelectric crystal antimony sulphoiodide, SbSI. (After Raman scattering experiments by C. H. Perry and D. K. Agrawal.)

The series does not contain terms in odd powers of $P$ if the unpolarized crystal has a center of inversion symmetry, but crystals are known in which odd powers are important. Power series expansions of the free energy do not always exist, for nonanalytic terms are known to occur, especially when very near a transition. For example, the transition in KH$_2$PO$_4$ appears to have a logarithmic singularity in the heat capacity at the transition, which is not classifiable as either first or second order.
The value of $P$ in thermal equilibrium is given by the minimum of $\hat{F}$ as a function of $P$; the value of $\hat{F}$ at this minimum defines the Helmholtz free energy $F(T,E)$. The equilibrium polarization in an applied electric field $E$ satisfies the extremum condition

$$\frac{\partial \hat{F}}{\partial P} = 0 = -E + g_2 P + g_4 P^3 + g_6 P^5 + \cdots .$$

(38)

In this section we assume that the specimen is a long rod with the external applied field $E$ parallel to the long axis.

To obtain a ferroelectric state we must suppose that the coefficient of the term in $P^2$ in (37) passes through zero at some temperature $T_0$:

$$g_2 = \gamma (T - T_0) ,$$

(39)

where $\gamma$ is taken as a positive constant and $T_0$ may be equal to or lower than the transition temperature. A small positive value of $g_2$ means that the lattice is "soft" and is close to instability. A negative value of $g_2$ means that the unpolarized lattice is unstable. The variation of $g_2$ with temperature is accounted for by thermal expansion and other effects of anharmonic lattice interactions.

**Second-Order Transition**

If $g_4$ in (37) is positive, nothing new is added by the term in $g_6$, and this may then be neglected. The polarization for zero applied electric field is found from (38):

$$\gamma (T - T_0) P_s + g_4 P_s^3 = 0 ,$$

(40)

so that either $P_s = 0$ or $P_s^2 = (\gamma / g_4) (T_0 - T)$. For $T \ge T_0$ the only real root of (40) is at $P_s = 0$, because $\gamma$ and $g_4$ are positive. Thus $T_0$ is the Curie temperature. For $T < T_0$ the minimum of the Landau free energy in zero applied field is at

$$|P_s| = (\gamma / g_4)^{1/2} (T_0 - T)^{1/2} ,$$

(41)

as plotted in Fig. 15. The phase transition is a second-order transition because the polarization goes continuously to zero at the transition temperature. The transition in LiTaO$_3$ is an example (Fig. 16) of a second-order transition.

**First-Order Transition**

The transition is first order if $g_4$ in (37) is negative. We must now retain $g_6$ and take it positive in order to restrain $\hat{F}$ from going to minus infinity (Fig. 17). The equilibrium condition for $E = 0$ is given by (38):

$$\gamma (T - T_0) P_s - |g_4| P_s^3 + g_6 P_s^5 = 0 ,$$

(42)

so that either $P_s = 0$ or

$$\gamma (T - T_0) - |g_4| P_s^2 + g_6 P_s^4 = 0 .$$

(43)
Figure 15  Spontaneous polarization versus temperature, for a second-order phase transition.

Figure 16  Temperature variation of the polar-axis static dielectric constant of LiTaO₃. (After Glass.)
At the transition temperature $T_c$ the free energies of the paraelectric and ferroelectric phases will be equal. That is, the value of $F$ for $P_s = 0$ will be equal to the value of $F$ at the minimum given by (43). In Fig. 18 we show the characteristic variation with temperature of $P_s$ for a first-order phase transition; contrast this with the variation shown in Fig. 15 for a second-order phase transition. The transition in BaTiO$_3$ is first order.

The dielectric constant is calculated from the equilibrium polarization in an applied electric field $E$ and is found from (36). In equilibrium at temperatures over the transition the terms in $P^4$ and $P^6$ may be neglected; thus $E = \gamma(T - T_0)P$, or

(CGS) \[ \epsilon(T > T_c) = 1 + 4\pi P/E = 1 + 4\pi \gamma(T - T_0). \]  

of the form of (36). The result applies whether the transition is of the first or second order, but if second order we have $T_0 = T_c$; if first order, then $T_0 < T_c$. Equation (39) defines $T_0$, but $T_c$ is the transition temperature.

**Antiferroelectricity**

A ferroelectric displacement is not the only type of instability that may develop in a dielectric crystal. Other deformations occur, as in Fig. 19. These deformations, even if they do not give a spontaneous polarization, may be accompanied by changes in the dielectric constant. One type of deformation is called antiferroelectric and has neighboring lines of ions displaced in opposite senses. The perovskite structure appears to be susceptible to many types of deformation, often with little difference in energy between them. The phase
Figure 18 Calculated values of the spontaneous polarization as a function of temperature, with parameters as for barium titanate. (After W. Cochran.)

Figure 19 Schematic representation of fundamental types of structural phase transitions from a centrosymmetric prototype. (After Lines and Glass.)
Table 3  Antiferroelectric crystals

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Transition temperature to antiferroelectric state, in K</th>
</tr>
</thead>
<tbody>
<tr>
<td>WO₃</td>
<td>1010</td>
</tr>
<tr>
<td>NaNbO₃</td>
<td>793, 911</td>
</tr>
<tr>
<td>PbZrO₃</td>
<td>506</td>
</tr>
<tr>
<td>PbHfO₃</td>
<td>488</td>
</tr>
<tr>
<td>NH₄H₂PO₄</td>
<td>148</td>
</tr>
<tr>
<td>ND₄D₂PO₄</td>
<td>242</td>
</tr>
<tr>
<td>NH₄H₂AsO₄</td>
<td>216</td>
</tr>
<tr>
<td>ND₄D₂AsO₄</td>
<td>304</td>
</tr>
<tr>
<td>(NH₄)₂H₃IO₆</td>
<td>254</td>
</tr>
</tbody>
</table>

From a compilation by Walter J. Merz.

diagrams of mixed perovskite systems, such as the PbZrO₃–PbTiO₃ system, show transitions between para-, ferro-, and antiferroelectric states (Fig. 20).

Ordered antiferroelectric arrangements of permanent electric dipole moments occur at low temperatures in ammonium salts and in hydrogen halides. Several crystals believed to have an ordered nonpolar state are listed in Table 3.

**Ferroelectric Domains**

Consider a ferroelectric crystal (such as barium titanate in the tetragonal phase) in which the spontaneous polarization may be either up or down the c axis of the crystal. A ferroelectric crystal generally consists of regions called domains within each of which the polarization is in the same direction, but in adjacent domains the polarization is in different directions. In Fig. 21 the polarization is in opposite directions. The net polarization depends on the difference in the volumes of the upward- and downward-directed domains. The crystal as a whole will appear to be unpolarized, as measured by the charge on electrodes covering the ends, when the volumes of domains in opposite senses are equal. The total dipole moment of the crystal may be changed by the movement of the walls between domains or by the nucleation of new domains.

Figure 22 is a series of photomicrographs of a single crystal of barium titanate in an electric field normal to the plane of the photographs and parallel to the tetragonal axis. The closed curves are boundaries between domains polarized into and out of the plane of the photographs.

The domain boundaries change size and shape when the intensity of the electric field is altered. The motion of domain walls in ferroelectrics is not
Figure 20 Ferroelectric $F$, antiferroelectric $A$, and paraelectric $P$ phases of the lead zirconate–lead titanate solid solution system. The subscript $T$ denotes a tetragonal phase; $C$ a cubic phase; $R$ a rhombohedral phase, of which there are high temperature (HT) and low temperature (LT) forms. Near the rhombohedral–tetragonal phase boundaries one finds very high piezoelectric coupling coefficients. (After Jaffe.)

Figure 21 (a) Schematic drawing of atomic displacements on either side of a boundary between domains polarized in opposite directions in a ferroelectric crystal; (b) view of a domain structure, showing 180° boundaries between domains polarized in opposite directions.
Figure 22. Ferroelectric domains on the face of a single crystal of barium titanate. The face is normal to the tetragonal or c axis. The net polarization of the crystal as judged by domain volumes is increased markedly as the electric field intensity parallel to the axis is increased from 550 volts/cm to 980 V/cm. The domain boundaries are made visible by etching the crystal in a weak acid solution. (R. C. Miller.)

simple: it is known that in an electric field a $180^\circ$ wall in BaTiO$_3$ appears to move by the repeated nucleation of steps by thermal fluctuations along the parent wall. A "$180^\circ$ wall" is the boundary between regions having opposite polarization directions.

**Piezoelectricity**

All crystals in a ferroelectric state are also piezoelectric: a stress $Z$ applied to the crystal will change the electric polarization (Fig. 23). Similarly, an electric field $E$ applied to the crystal will cause the crystal to become strained. In schematic one-dimensional notation, the piezoelectric equations are

$$P = Zd + E\chi; \quad e = Zs + Ed,$$

where $P$ is the polarization, $Z$ the stress, $d$ the piezoelectric strain constant, $E$ the electric field, $\chi$ the dielectric susceptibility, $e$ the elastic strain, and $s$ the elastic compliance constant. To obtain (45) in SI, replace $\chi$ by $\varepsilon_0\chi$. These relations exhibit the development of polarization by an applied stress and the development of elastic strain by an applied electric field.

A crystal may be piezoelectric without being ferroelectric: a schematic example of such a structure is given in Fig. 24. Quartz is piezoelectric, but not ferroelectric; barium titanate is both. For order of magnitude, in quartz

---

Figure 23 (a) Unstressed ferroelectric crystal and (b) stressed ferroelectric crystal. The stress changes the polarization by $\Delta P$, the induced piezoelectric polarization.

Figure 24 (a) The unstressed crystal has a threefold symmetry axis. The arrows represent dipole moments, each set of three arrows represents a planar group of ions denoted by $A_2^+B_3^-$, with a $B_2^-$ ion at each vertex. The sum of the three dipole moments at each vertex is zero. (b) The crystal when stressed develops a polarization in the direction indicated. The sum of the dipole moments about each vertex is no longer zero.

$$d \approx 10^{-7} \text{ cm/statvolt and in barium titanate } d \approx 10^{-5} \text{ cm/statvolt.}$$

The general definition\(^{10}\) of the piezoelectric strain constants is

$$d_{ik} = (\partial e_k/\partial E_i)_{Z},$$

where $i = x, y, z$ and $k = xx, yy, zz, yz, zx, xy$. To convert to cm/stat V from values of $d_{ik}$ given in m/V, multiply by $3 \times 10^4$.

The lead zirconate–lead titanate system (called the PZT system), Fig. 20, is widely used in polycrystalline (ceramic) form with compositions of very high piezoelectric coupling. The synthetic polymer polyvinylidene fluoride (PVF$_2$) is five times more strongly piezoelectric than crystalline quartz. Thin stretched films of PVF$_2$ are flexible and are easy to handle as ultrasonic transducers. Being flexible, the films are easy to handle in medical applications to monitor blood pressure and respiration.

The response of piezoelectric crystals in transducer applications is characterized by the electromechanical coupling factor $k$, whose square is defined by

$$k^2 = \frac{\text{mechanical energy stored}}{\text{electric energy stored}}. \quad (47)$$

The result can be expressed in terms of the material constants used in (45). The relation is given in works on the applications of piezoelectric crystals.

**Ferroelasticity**

A crystal is ferroelastic if it has two or more stable orientation states in the absence of mechanical stress and if it can be reproducibly transformed from one to another of these states by application of mechanical stress. Realizations are few: gadolinium molybdate Gd$_2$(MoO$_4$)$_3$ has a combined ferroelectric-ferroelastic transition; lead phosphate Pb$_3$(PO$_4$)$_2$ has a pure ferroelastic transition; several crystals have combined ferroelectric-ferroelastic transitions.

**Optical Ceramics**

The addition of lanthanum to the PZT system gives a reasonably transparent product with useful electro-optical properties. The PLZT system has been used as the basis of optical memories.

**SUMMARY**

(In CGS Units)

- The electric field averaged over the volume of the specimen defines the macroscopic electric field $\mathbf{E}$ of the Maxwell equations.

- The electric field that acts at the site $r_j$ of an atom $j$ is the local electric field, $\mathbf{E}_{\text{loc}}$. It is a sum over all charges, grouped in terms as $\mathbf{E}_{\text{loc}}(r_j) = \mathbf{E}_0 + \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3(r_j)$, where only $\mathbf{E}_3$ varies rapidly within a cell. Here:
  - $\mathbf{E}_0 =$ external electric field;
  - $\mathbf{E}_1 =$ depolarization field associated with the boundary of the specimen;
  - $\mathbf{E}_2 =$ field from polarization outside a sphere centered about $r_j$;
  - $\mathbf{E}_3(r_j) =$ field at $r_j$ due to all atoms inside the sphere.

- The macroscopic field $\mathbf{E}$ of the Maxwell equations is equal to $\mathbf{E}_0 + \mathbf{E}_1$, which, in general, is not equal to $\mathbf{E}_{\text{loc}}(r_j)$.

- The depolarization field in an ellipsoid is $E_{\mu\nu} = -N_{\mu\nu}P_{\nu}$, where $N_{\mu\nu}$ is the depolarization tensor; the polarization $\mathbf{P}$ is the dipole moment per unit volume. In a sphere $N = 4\pi/3$. 

• The Lorentz field is $E_2 = 4\pi P/3$.

• The polarizability $\alpha$ of an atom is defined in terms of the local electric field as $p = \alpha E_{loc}$.

• The dielectric susceptibility $\chi$ and dielectric constant $\varepsilon$ are defined in terms of the macroscopic electric field $E$ as $D = E + 4\pi P = \varepsilon E = (1 + 4\pi\chi)E$, or $\chi = P/E$. In SI, we have $\chi = p/\varepsilon_0 E$.

• At atom at a site with cubic symmetry has $E_{loc} = E + (4\pi/3)P$ and satisfies the Clausius-Mossotti relation (24).

**Problems**

1. Polarizability of atomic hydrogen. Consider a semiclassical model of the ground state of the hydrogen atom in an electric field normal to the plane of the orbit (Fig. 25), and show that for this model $\alpha = a_\text{H}^3$, where $a_\text{H}$ is the radius of the unperturbed orbit. Note: If the applied field is in the $x$ direction, then the $x$ component of the field of the nucleus at the displaced position of the electron orbit must be equal to the applied field. The correct quantum-mechanical result is larger than this by the factor $\frac{\alpha}{2}$. (We are speaking of $a_0$ in the expansion $\alpha = a_0 + a_1 E + \cdots$.) We assume $x \ll a_\text{H}$. One can also calculate $a_1$ on this model.

![Figure 25](image.png)
2. Polarizability of conducting sphere. Show that the polarizability \( \gamma \) of a conducting metallic sphere of radius \( a \) is \( \alpha = a^3 \). This result is most easily obtained by noting that \( E = 0 \) inside the sphere and then using the depolarization factor \( 4\pi/3 \) for a sphere (Fig. 26). The result gives values of \( \alpha \) of the order of magnitude of the observed polarizabilities of atoms. A lattice of \( N \) conducting spheres per unit volume has dielectric constant \( \epsilon = 1 + 4\pi Na^3 \), for \( Na^3 \ll 1 \). The suggested proportionality of \( \alpha \) to the cube of the ionic radius is satisfied quite well for alkali and halogen ions. To do the problem in SI, use \( \frac{4}{3} \) as the depolarization factor.

3. Effect of air gap. Discuss the effect of an air gap (Fig. 27) between condenser plates and dielectric on the measurement of high dielectric constants. What is the highest apparent dielectric constant possible if the air gap thickness is \( 10^{-3} \) of the total thickness? The presence of air gaps can seriously distort the measurement of high dielectric constants.

4. Interfacial polarization. Show that a parallel-plate capacitor made up of two parallel layers of material—one layer with dielectric constant \( \epsilon \), zero conductivity, and thickness \( d \), and the other layer with \( \epsilon = 0 \) for convenience, finite conductivity \( \sigma \), and thickness \( qd \)—behaves as if the space between the condenser plates were filled with a homogeneous dielectric with dielectric constant

\[
\text{(CGS)} \quad \epsilon_{\text{eff}} = \frac{\epsilon (1 + q)}{1 - (i\omega q/4\pi \sigma)},
\]

where \( \omega \) is the angular frequency. Values of \( \epsilon_{\text{eff}} \) as high as \( 10^4 \) or \( 10^5 \) caused largely by this Maxwell-Wagner mechanism are sometimes found, but the high values are always accompanied by large ac losses.

5. Polarization of sphere. A sphere of dielectric constant \( \epsilon \) is placed in a uniform external electric field \( E_0 \). (a) What is the volume average electric field \( E \) in the sphere? (b) Show that the polarization in the sphere is \( P = \chi E_0 [1 + (4\pi \chi/3)] \), where \( \chi = (\epsilon - 1)/4\pi \). Hint: You do not need to calculate \( E_{\text{loc}} \) in this problem; in fact it is confusing to do so, because \( \epsilon \) and \( \chi \) are defined so that \( P = \chi E \). We require \( E_0 \) to be unchanged by insertion of the sphere. We can produce a fixed \( E_0 \) by placing positive charges on one thin plate of an insulator and negative charges on an opposite plate. If the plates are always far from the sphere, the field of the plates will remain unchanged when the sphere is inserted between them. The results above are in CGS.

6. Ferroelectric criterion for atoms. Consider a system of two neutral atoms separated by a fixed distance \( a \), each atom having a polarizability \( \alpha \). Find the relation between \( a \) and \( \alpha \) for such a system to be ferroelectric. Hint: The dipolar field is strongest along the axis of the dipole.

7. Saturation polarization at Curie point. In a first-order transition the equilibrium condition (43) with \( T \) set equal \( T_c \) gives one equation for the polarization \( P_s(T_c) \). A
Figure 26  The total field inside a conducting sphere is zero. If a field $E_0$ is applied externally, then the field $E_1$ due to surface charges on the sphere must just cancel $E_0$, so that $E_0 + E_1 = 0$ within the sphere. But $E_1$ can be simulated by the depolarization field $-4\pi P/3$ of a uniformly polarized sphere of polarization $P$. Relate $P$ to $E_0$ and calculate the dipole moment $p$ of the sphere. In SI the depolarization field is $-P/3\varepsilon_0$.

Figure 27  An air gap of thickness $qd$ is in series in a capacitor with a dielectric slab of thickness $d$.

further condition at the Curie point is that $\hat{F}(P_s, T_c) = \hat{F}(0, T_c)$. (a) Combining these two conditions, show that $P_s(T_c) = 3|g_4|/16g_6$. (b) Using this result, show that $T_c = T_0 + 3g_4^2/16\gamma g_6$.

8. Dielectric constant below transition temperature. In terms of the parameters in the Landau free energy expansion, show that for a second-order phase transition the dielectric constant below the transition temperature is

$$\varepsilon = 1 + 4\pi\Delta P/E = 1 + 2\pi/\gamma(T_c - T)$$

This result may be compared with (44) above the transition.

9. Soft modes and lattice transformations. Sketch a monatomic linear lattice of lattice constant $a$. (a) Add to each of six atoms a vector to indicate the direction of the displacement at a given time caused by a longitudinal phonon with wavevector at the zone boundary. (b) Sketch the crystal structure that results if this zone boundary
phonon becomes unstable ($\omega \to 0$) as the crystal is cooled through $T_c$. (c) Sketch on one graph the essential aspects of the longitudinal phonon dispersion relation for the monatomic lattice at $T$ well above $T_c$ and at $T = T_c$. Add to the graph the same information for phonons in the new structure at $T$ well below $T_c$.

10. **Ferroelectric linear array.** Consider a line of atoms of polarizability $\alpha$ and separation $a$. Show that the array can polarize spontaneously if $\alpha \geq \frac{a^3}{4\Sigma n^{-3}}$, where the sum is over all positive integers and is given in tables as 1.202.

---

**References**


*Ferroelectrics. A Journal.*


14

Diamagnetism and Paramagnetism

LANGEVIN DIAMAGNETISM EQUATION 417

QUANTUM THEORY OF DIAMAGNETISM OF MONONUCLEAR SYSTEMS 419

PARAMAGNETISM 420

QUANTUM THEORY OF PARAMAGNETISM 420
  Rare earth ions 423
  Hund rules 424
  Iron group ions 425
  Crystal field splitting 426
  Quenching of the orbital angular momentum 426
  Spectroscopic splitting factor 429
  Van Vleck temperature-independent paramagnetism 430

COOLING BY ISENTROPIC DEMAGNETIZATION 431
  Nuclear demagnetization 432

PARAMAGNETIC SUSCEPTIBILITY OF CONDUCTION ELECTRONS 433

SUMMARY 436

PROBLEMS 436

1. Diamagnetic susceptibility of atomic hydrogen 436
2. Hund rules 437
3. Triplet excited states 437
4. Heat capacity from internal degrees of freedom 438
5. Pauli spin susceptibility 438
6. Conduction electron ferromagnetism 438
7. Two-level system 440
8. Paramagnetism of \( S = 1 \) system 440

REFERENCES 440

NOTATION: In the problems treated in this chapter the magnetic field \( B \) is always closely equal to the applied field \( B_a \), so that we write \( B \) for \( B_a \) in most instances.
Figure 1 Characteristic magnetic susceptibilities of diamagnetic and paramagnetic substances.
Magnetism is inseparable from quantum mechanics, for a strictly classical system in thermal equilibrium can display no magnetic moment, even in a magnetic field. The magnetic moment of a free atom has three principal sources: the spin with which electrons are endowed; their orbital angular momentum about the nucleus, and the change in the orbital moment induced by an applied magnetic field.

The first two effects give paramagnetic contributions to the magnetization, and the third gives a diamagnetic contribution. In the ground 1s state of the hydrogen atom the orbital moment is zero, and the magnetic moment is that of the electron spin along with a small induced diamagnetic moment. In the 1s² state of helium the spin and orbital moments are both zero, and there is only an induced moment. Atoms with filled electron shells have zero spin and zero orbital moment: these moments are associated with unfilled shells.

The magnetization $M$ is defined as the magnetic moment per unit volume. The magnetic susceptibility per unit volume is defined as

$$(\text{CGS}) \quad \chi = \frac{M}{B} ; \quad (\text{SI}) \quad \chi = \frac{\mu_0 M}{B} . \tag{1}$$

where $B$ is the macroscopic magnetic field intensity. In both systems of units $\chi$ is dimensionless. We shall sometimes for convenience refer to $M/B$ as the susceptibility without specifying the system of units.

Quite frequently a susceptibility is defined referred to unit mass or to a mole of the substance. The molar susceptibility is written as $\chi_M$; the magnetic moment per gram is sometimes written as $\sigma$. Substances with a negative magnetic susceptibility are called diamagnetic. Substances with a positive susceptibility are called paramagnetic, as in Fig. 1.

Ordered arrays of magnetic moments are discussed in Chapter 15, the arrays may be ferromagnetic, ferrimagnetic, antiferromagnetic, helical, or more complex in form. Nuclear magnetic moments give rise to nuclear paramagnetism. Magnetic moments of nuclei are of the order of $10^{-3}$ times smaller than the magnetic moment of the electron.

**LANGEVIN DIAMAGNETISM EQUATION**

Diamagnetism is associated with the tendency of electrical charges partially to shield the interior of a body from an applied magnetic field. In electromagnetism we are familiar with Lenz's law: when the flux through an electrical circuit is changed, an induced current is set up in such a direction as to oppose the flux change.
In a superconductor or in an electron orbit within an atom, the induced current persists as long as the field is present. The magnetic field of the induced current is opposite to the applied field, and the magnetic moment associated with the current is a diamagnetic moment. Even in a normal metal there is a diamagnetic contribution from the conduction electrons, and this diamagnetism is not destroyed by collisions of the electrons.

The usual treatment of the diamagnetism of atoms and ions employs the Larmor theorem: in a magnetic field the motion of the electrons around a central nucleus is, to the first order in \( B \), the same as a possible motion in the absence of \( B \) except for the superposition of a precession of the electrons with angular frequency

\[
(\text{CGS}) \quad \omega = \frac{eB}{2mc} ; \\
(\text{SI}) \quad \omega = \frac{eB}{2m} .
\]

(2)

If the field is applied slowly, the motion in the rotating reference system will be the same as the original motion in the rest system before the application of the field.

If the average electron current around the nucleus is zero initially, the application of the magnetic field will cause a finite current around the nucleus. The current is equivalent to a magnetic moment opposite to the applied field. It is assumed that the Larmor frequency (2) is much lower than the frequency of the original motion in the central field. This condition is not satisfied in free carrier cyclotron resonance, and the cyclotron frequency is twice the frequency (2).

The Larmor precession of \( Z \) electrons is equivalent to an electric current

\[
(\text{SI}) \quad I = (\text{charge})(\text{revolutions per unit time}) = (-Ze)\left(\frac{1}{2m} \cdot \frac{eB}{2m}\right) .
\]

(3)

The magnetic moment \( \mu \) of a current loop is given by the product (current) \( \times \) (area of the loop). The area of the loop of radius \( \rho \) is \( \pi \rho^2 \). We have

\[
(\text{SI}) \quad \mu = \frac{Ze^2B}{4m} \langle \rho^2 \rangle ; \\
(\text{CGS}) \quad \mu = \frac{Ze^2B}{4mc^2} \langle \rho^2 \rangle .
\]

(4)

Here \( \langle \rho^2 \rangle = \langle x^2 \rangle + \langle y^2 \rangle \) is the mean square of the perpendicular distance of the electron from the field axis through the nucleus. The mean square distance of the electrons from the nucleus is \( \langle r^2 \rangle = \langle x^2 \rangle + \langle y^2 \rangle + \langle z^2 \rangle \). For a spherically symmetrical distribution of charge we have \( \langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle \), so that \( \langle r^2 \rangle = \frac{3}{2} \langle \rho^2 \rangle \).

From (4) the diamagnetic susceptibility per unit volume is, if \( N \) is the number of atoms per unit volume,

\[
(\text{CGS}) \quad \chi = \frac{N\mu}{B} = \frac{NZe^2}{6mc^2} \langle r^2 \rangle ;
\]

(5)
\[ \chi = \frac{\mu_0 N \mu}{B} = \frac{\mu_0 N Ze^2}{6m} \langle r^2 \rangle \]  

This is the classical Langevin result.

The problem of calculating the diamagnetic susceptibility of an isolated atom is reduced to the calculation of \( \langle r^2 \rangle \) for the electron distribution within the atom. The distribution can be calculated by quantum mechanics.

Experimental values for neutral atoms are most easily obtained for the inert gases. Typical experimental values of the molar susceptibilities are the following:

<table>
<thead>
<tr>
<th></th>
<th>He</th>
<th>Ne</th>
<th>Ar</th>
<th>Kr</th>
<th>Xe</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi_m ) in CGS in ( 10^{-6} ) cm(^3)/mole</td>
<td>-1.9</td>
<td>-7.2</td>
<td>-19.4</td>
<td>-28.0</td>
<td>-43.0</td>
</tr>
</tbody>
</table>

In dielectric solids the diamagnetic contribution of the ion cores is described roughly by the Langevin result. The contribution of conduction electrons is more complicated, as is evident from the de Haas-van Alphen effect discussed in Chapter 9.

**QUANTUM THEORY OF DIAMAGNETISM OF MONONUCLEAR SYSTEMS**

From (G. 18) the effect of a magnetic field is to add to the hamiltonian the terms

\[ \mathcal{H}' = \frac{ie\hbar}{2mc} (\nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla) + \frac{e^2}{2mc^2} A^2 ; \]  

for an atomic electron these terms may usually be treated as a small perturbation. If the magnetic field is uniform and in the \( z \) direction, we may write

\[ A_x = -\frac{1}{2} y B, \quad A_y = \frac{1}{2} x B, \quad A_z = 0, \]

and (6) becomes

\[ \mathcal{H}' = \frac{ie\hbar B}{2mc} \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) + \frac{e^2 B^2}{8mc^2} (x^2 + y^2) . \]

The first term on the right is proportional to the orbital angular momentum component \( L_z \) if \( r \) is measured from the nucleus. In mononuclear systems this term gives rise only to paramagnetism. The second term gives for a spherically symmetric system a contribution

\[ E' = \frac{e^2 B^2}{12mc^2} \langle r^2 \rangle , \]
by first-order perturbation theory. The associated magnetic moment is diamagnetic:

\[ \mu = -\frac{\partial E'}{\partial B} = -\frac{e^2 \langle r^2 \rangle}{6mc^2} B , \]  

in agreement with the classical result.

**PARAMAGNETISM**

Electronic paramagnetism (positive contribution to \( \chi \)) is found in:

1. Atoms, molecules, and lattice defects possessing an odd number of electrons, as here the total spin of the system cannot be zero. Examples: free sodium atoms; gaseous nitric oxide (NO); organic free radicals such as triphenylmethyl, C(C₆H₅)₃; F centers in alkali halides.

2. Free atoms and ions with a partly filled inner shell: transition elements; ions isoelectronic with transition elements; rare earth and actinide elements. Examples: Mn²⁺, Gd³⁺, U⁴⁺. Paramagnetism is exhibited by many of these ions even when incorporated into solids, but not invariably.

3. A few compounds with an even number of electrons, including molecular oxygen and organic biradicals.


**QUANTUM THEORY OF PARAMAGNETISM**

The magnetic moment of an atom or ion in free space is given by

\[ \mu = \gamma \hbar J = -g\mu_B J , \]  

where the total angular momentum \( \hbar J \) is the sum of the orbital \( \hbar L \) and spin \( \hbar S \) angular momenta.

The constant \( \gamma \) is the ratio of the magnetic moment to the angular momentum; \( \gamma \) is called the gyromagnetic ratio or magnetogyric ratio. For electronic systems a quantity \( g \) called the \( g \) factor or the spectroscopic splitting factor is defined by

\[ g\mu_B \equiv -\gamma \hbar . \]  

For an electron spin \( g = 2.0023 \), usually taken as 2.00. For a free atom the \( g \) factor is given by the Landé equation

\[ g = 1 + \frac{J(J + 1) + S(S + 1) - L(L + 1)}{2J(J + 1)} . \]
Figure 2. Energy level splitting for one electron in a magnetic field $B$ directed along the positive $z$ axis. For an electron the magnetic moment $\mu$ is opposite in sign to the spin $S$, so that $\mu = -g \mu_B S$. In the low energy state the magnetic moment is parallel to the magnetic field.

Figure 3. Fractional populations of a two-level system in thermal equilibrium at temperature $T$ in a magnetic field $B$. The magnetic moment is proportional to the difference between the two curves.

The Bohr magneton $\mu_B$ is defined as $e\hbar/2mc$ in CGS and $e\hbar/2m$ in SI. It is closely equal to the spin magnetic moment of a free electron.

The energy levels of the system in a magnetic field are

$$U = -\mu \cdot B = m_j g \mu_B B ,$$

(14)

where $m_j$ is the azimuthal quantum number and has the values $J, J - 1, \ldots, -J$. For a single spin with no orbital moment we have $m_j = \pm \frac{1}{2}$ and $g = 2$, whence $U = \pm \mu_B B$. This splitting is shown in Fig. 2.

If a system has only two levels the equilibrium populations are, with $\tau \equiv k_B T$,

$$\frac{N_1}{N} = \frac{\exp(\mu B/\tau)}{\exp(\mu B/\tau) + \exp(-\mu B/\tau)} ;$$

(15)

$$\frac{N_2}{N} = \frac{\exp(-\mu B/\tau)}{\exp(\mu B/\tau) + \exp(-\mu B/\tau)} ;$$

(16)

here $N_1$, $N_2$ are the populations of the lower and upper levels, and $N = N_1 + N_2$ is the total number of atoms. The fractional populations are plotted in Fig. 3.

The projection of the magnetic moment of the upper state along the field direction is $-\mu$ and of the lower state is $\mu$. The resultant magnetization for $N$ atoms per unit volume is, with $x = \mu B/k_B T$,

$$M = (N_1 - N_2)\mu = N\mu \cdot \frac{e^x - e^{-x}}{e^x + e^{-x}} = N\mu \tanh x .$$

(17)

For $x \ll 1$, $\tanh x \equiv x$, and we have

$$M \equiv N\mu(\mu B/k_B T) .$$

(18)

In a magnetic field an atom with angular momentum quantum number $J$ has $2J + 1$ equally spaced energy levels. The magnetization (Fig. 4) is given by

$$M = NgJ\mu_B B J(x) , \quad (x = gJ\mu_B B/k_B T) ,$$

(19)
Figure 4  Plot of magnetic moment versus $B/T$ for spherical samples of (I) potassium chromium alum, (II) ferric ammonium alum, and (III) gadolinium sulfate octahydrate. Over 99.5% magnetic saturation is achieved at 1.3 K and about 50,000 gauss. (5T). After W. E. Henry.

where the Brillouin function $B_J$ is defined by

$$B_J(x) = \frac{2J + 1}{2J} \cthn \left( \frac{(2J + 1)x}{2J} \right) - \frac{1}{2J} \cthn \left( \frac{x}{2J} \right).$$

(20)

Equation (17) is a special case of (20) for $J = \frac{1}{2}$.

For $x \ll 1$, we have

$$\cthn x = \frac{1}{x} + \frac{x}{3} - \frac{x^3}{45} + \cdots,$$

(21)

and the susceptibility is

$$\frac{M}{B} \approx \frac{NJ(J + 1)g^2\mu_B^2}{3k_BT} = \frac{Np^2\mu_B^2}{3k_BT} = \frac{C}{T}.$$

(22)

Here $p$ is the effective number of Bohr magnetons, defined as

$$p = g[J(J + 1)]^{1/2}.$$

(23)
The constant $C$ is known as the **Curie constant**. The form (19) is known as the Curie-Brillouin law, and (22) is known as the **Curie law**. Results for the paramagnetic ions in a gadolinium salt are shown in Fig. 5.

**Rare Earth Ions**

The ions of the rare-earth elements have closely similar chemical properties, and their chemical separation in tolerably pure form was accomplished only long after their discovery. Their magnetic properties are fascinating: the ions exhibit a systematic variety and intelligible complexity. The chemical properties of the trivalent ions are similar because the outermost electron shells are identically in the $5s^25p^6$ configuration, like neutral xenon. In lanthanum, just before the rare earth group begins, the $4f$ shell is empty; at cerium there is one $4f$ electron, and the number of $4f$ electrons increases steadily through the group until we have $4f^{13}$ at ytterbium and the filled shell $4f^{14}$ at lutecium. The radii of the trivalent ions contract fairly smoothly as we go through the group from $1.11$ Å at cerium to $0.94$ Å at ytterbium. This is the famous "lanthanide contraction." What distinguishes the magnetic behavior of one ion species from another is the number of $4f$ electrons compacted in the inner shell with a radius
of perhaps 0.3 Å. Even in the metals the 4f core retains its integrity and its atomic properties: no other group of elements in the periodic table is as interesting.

The preceding discussion of paramagnetism applies to atoms that have a \((2J + 1)\)-fold degenerate ground state, the degeneracy being lifted by a magnetic field. The influence of all higher energy states of the system is neglected. These assumptions appear to be satisfied by a number of rare-earth ions, Table 1. The calculated magneton numbers are obtained with \(g\) values from the Landé result (13) and the ground-state level assignment predicted by the Hund theory of spectral terms. The discrepancy between the experimental magneton numbers and those calculated on these assumptions is quite marked for \(\text{Eu}^{3+}\) and \(\text{Sm}^{3+}\) ions. For these ions it is necessary to consider the influence of the high states of the \(L - S\) multiplet, as the intervals between successive states of the multiplet are not large compared to \(k_B T\) at room temperature. A multiplet is the set of levels of different \(J\) values arising out of a given \(L\) and \(S\). The levels of a multiplet are split by the spin-orbit interaction.

**Hund Rules**

The Hund rules as applied to electrons in a given shell of an atom affirm that electrons will occupy orbitals in such a way that the ground state is characterized by the following:

1. The maximum value of the total spin \(S\) allowed by the exclusion principle;
2. The maximum value of the orbital angular momentum \(L\) consistent with this value of \(S\);
3. The value of the total angular momentum \(J\) is equal to \(|L - S|\) when the shell is less than half full and to \(L + S\) when the shell is more than half full. When the shell is just half full, the application of the first rule gives \(L = 0\), so that \(J = S\).

The first Hund rule has its origin in the exclusion principle and the coulomb repulsion between electrons. The exclusion principle prevents two electrons of the same spin from being at the same place at the same time. Thus electrons of the same spin are kept apart, further apart than electrons of opposite spin. Because of the coulomb interaction the energy of electrons of the same spin is lower—the average potential energy is less positive for parallel spin than for antiparallel spin. A good example is the ion \(\text{Mn}^{2+}\). This ion has five electrons in the \(3d\) shell, which is therefore half-filled. The spins can all be parallel if each electron enters a different orbital, and there are exactly five different orbitals available, characterized by the orbital quantum numbers \(m_L = 2, 1, 0, -1, -2\). These will be occupied by one electron each. We expect \(S = \frac{5}{2}\), and because \(\sum m_L = 0\) the only possible value of \(L\) is 0, as observed.
Table 1 Effective magneton numbers \( p \) for trivalent lanthanide group ions

(Near room temperature)

<table>
<thead>
<tr>
<th>Ion</th>
<th>Configuration</th>
<th>Basic level</th>
<th>( p(\text{calc}) )</th>
<th>( p(\text{exp}) ), approximate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ce(^{3+})</td>
<td>( 4f^{15}5s^2p^6 )</td>
<td>( ^2F_{5/2} )</td>
<td>2.54</td>
<td>2.4</td>
</tr>
<tr>
<td>Pr(^{3+})</td>
<td>( 4f^{13}5s^2p^6 )</td>
<td>( ^3H_4 )</td>
<td>3.58</td>
<td>3.5</td>
</tr>
<tr>
<td>Nd(^{3+})</td>
<td>( 4f^{15}5s^2p^6 )</td>
<td>( ^4I_{9/2} )</td>
<td>3.62</td>
<td>3.5</td>
</tr>
<tr>
<td>Pm(^{3+})</td>
<td>( 4f^{15}5s^2p^6 )</td>
<td>( ^5I_4 )</td>
<td>2.68</td>
<td>—</td>
</tr>
<tr>
<td>Sm(^{3+})</td>
<td>( 4f^{13}5s^2p^6 )</td>
<td>( ^6H_{5/2} )</td>
<td>0.84</td>
<td>1.5</td>
</tr>
<tr>
<td>Eu(^{3+})</td>
<td>( 4f^{13}5s^2p^6 )</td>
<td>( ^7F_0 )</td>
<td>0</td>
<td>3.4</td>
</tr>
<tr>
<td>Gd(^{3+})</td>
<td>( 4f^{13}5s^2p^6 )</td>
<td>( ^6S_{7/2} )</td>
<td>7.94</td>
<td>8.0</td>
</tr>
<tr>
<td>T(_{1}^{3+})</td>
<td>( 4f^{13}5s^2p^6 )</td>
<td>( ^7F_6 )</td>
<td>9.72</td>
<td>9.5</td>
</tr>
<tr>
<td>T(_{2}^{3+})</td>
<td>( 4f^{13}5s^2p^6 )</td>
<td>( ^6H_{15/2} )</td>
<td>10.63</td>
<td>10.6</td>
</tr>
<tr>
<td>Ho(^{3+})</td>
<td>( 4f^{11}5s^2p^6 )</td>
<td>( ^5I_8 )</td>
<td>10.60</td>
<td>10.4</td>
</tr>
<tr>
<td>Er(^{3+})</td>
<td>( 4f^{11}5s^2p^6 )</td>
<td>( ^4I_{15/2} )</td>
<td>9.59</td>
<td>9.5</td>
</tr>
<tr>
<td>Tm(^{3+})</td>
<td>( 4f^{13}5s^2p^6 )</td>
<td>( ^3H_6 )</td>
<td>7.57</td>
<td>7.3</td>
</tr>
<tr>
<td>Yb(^{3+})</td>
<td>( 4f^{13}5s^2p^6 )</td>
<td>( ^2F_{7/2} )</td>
<td>4.54</td>
<td>4.5</td>
</tr>
</tbody>
</table>

The second Hund rule is best approached by model calculations. Pauling and Wilson,\(^1\) for example, give a calculation of the spectral terms that arise from the configuration \( p^2 \). The third Hund rule is a consequence of the sign of the spin-orbit interaction: For a single electron the energy is lowest when the spin is antiparallel to the orbital angular momentum. But the low energy pairs \( m_L, m_S \) are progressively used up as we add electrons to the shell; by the exclusion principle when the shell is more than half full the state of lowest energy necessarily has the spin parallel to the orbit.

Consider two examples of the Hund rules: The ion Ce\(^{3+}\) has a single \( f \) electron; an \( f \) electron has \( l = 3 \) and \( s = \frac{1}{2} \). Because the \( f \) shell is less than half full, the \( J \) value by the preceding rule is \( |L - S| = L - \frac{1}{2} = \frac{1}{2} \). The ion Pr\(^{3+}\) has two \( f \) electrons: one of the rules tells us that the spins add to give \( S = 1 \). Both \( f \) electrons cannot have \( m_l \) without violating the Pauli exclusion principle, so that the maximum \( L \) consistent with the Pauli principle is not 6, but 5. The \( J \) value is \( |L - S| = 5 - 1 = 4 \).

**Iron Group Ions**

Table 2 shows that the experimental magneton numbers for salts of the iron transition group of the periodic table are in poor agreement with (18). The values often agree quite well with magneton numbers \( p = 2[S(S + 1)]^{1/2} \) calcu-

---

Table 2  Effective magneton numbers for iron group ions

<table>
<thead>
<tr>
<th>Ion</th>
<th>Configuration</th>
<th>Basic level</th>
<th>$p(\text{calc}) = \frac{g(J(J + 1))^{1/2}}{2(S(S + 1))^{1/2}}$</th>
<th>$p(\text{calc}) = \frac{g(J(J + 1))^{1/2}}{2(S(S + 1))^{1/2}}$</th>
<th>$p(\text{exp})^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ti$^{3+}$, V$^{4+}$</td>
<td>$3d^1$</td>
<td>$^2D_{3/2}$</td>
<td>1.55</td>
<td>1.73</td>
<td>1.8</td>
</tr>
<tr>
<td>V$^{3+}$</td>
<td>$3d^2$</td>
<td>$^3F_{2}$</td>
<td>1.63</td>
<td>2.83</td>
<td>2.8</td>
</tr>
<tr>
<td>Cr$^{3+}$, V$^{2+}$</td>
<td>$3d^3$</td>
<td>$^4F_{3/2}$</td>
<td>0.77</td>
<td>3.87</td>
<td>3.8</td>
</tr>
<tr>
<td>Mn$^{3+}$, Cr$^{2+}$</td>
<td>$3d^4$</td>
<td>$^5D_0$</td>
<td>0</td>
<td>4.90</td>
<td>4.9</td>
</tr>
<tr>
<td>Fe$^{3+}$, Mn$^{2+}$</td>
<td>$3d^5$</td>
<td>$^6S_{5/2}$</td>
<td>5.92</td>
<td>5.92</td>
<td>5.9</td>
</tr>
<tr>
<td>Fe$^{2+}$</td>
<td>$3d^6$</td>
<td>$^5D_4$</td>
<td>6.70</td>
<td>4.90</td>
<td>5.4</td>
</tr>
<tr>
<td>Co$^{2+}$</td>
<td>$3d^7$</td>
<td>$^4F_{9/2}$</td>
<td>6.63</td>
<td>3.87</td>
<td>4.8</td>
</tr>
<tr>
<td>Ni$^{2+}$</td>
<td>$3d^8$</td>
<td>$^3F_{4}$</td>
<td>5.59</td>
<td>2.83</td>
<td>3.2</td>
</tr>
<tr>
<td>Cu$^{2+}$</td>
<td>$3d^9$</td>
<td>$^2D_{5/2}$</td>
<td>3.55</td>
<td>1.73</td>
<td>1.9</td>
</tr>
</tbody>
</table>

*Representative values.

lated as if the orbital moment were not there at all. We say that the orbital moments are quenched.

Crystal Field Splitting

The difference in behavior of the rare earth and the iron group salts is that the 4f shell responsible for paramagnetism in the rare earth ions lies deep inside the ions, within the 5s and 5p shells, whereas in the iron group ions the 3d shell responsible for paramagnetism is the outermost shell. The 3d shell experiences the intense inhomogeneous electric field produced by neighboring ions. This inhomogeneous electric field is called the crystal field. The interaction of the paramagnetic ions with the crystal field has two major effects: the coupling of L and S vectors is largely broken up, so that the states are no longer specified by their J values; further, the $2L + 1$ sublevels belonging to a given L which are degenerate in the free ion may now be split by the crystal field, as in Fig. 6. This splitting diminishes the contribution of the orbital motion to the magnetic moment.

Quenching of the Orbital Angular Momentum

In an electric field directed toward a fixed nucleus, the plane of a classical orbit is fixed in space, so that all the orbital angular momentum components $L_x$, $L_y$, $L_z$ are constant. In quantum theory one angular momentum component, usually taken as $L_z$, and the square of the total orbital angular momentum $L^2$ are constant in a central field. In a noncentral field the plane of the orbit will move about; the angular momentum components are no longer constant and may average to zero. In a crystal $L_z$ will no longer be a constant of the motion, although to a good approximation $L^2$ may continue to be constant. When $L_z$ averages to zero, the orbital angular momentum is blocked to be quenched. The
magnetic moment of a state is given by the average value of the magnetic moment operator $\mu_B(L + 2S)$. In a magnetic field along the $z$ direction the orbital contribution to the magnetic moment is proportional to the quantum expectation value of $L_z$; the orbital magnetic moment is quenched if the mechanical moment $L_z$ is quenched.

When the spin-orbit interaction energy is introduced, the spin may drag some orbital moment along with it. If the sign of the interaction favors parallel orientation of the spin and orbital magnetic moments, the total magnetic moment will be larger than for the spin alone, and the $g$ value will be larger than 2. The experimental results are in agreement with the known variation of sign of the spin-orbit interaction: $g > 2$ when the $3d$ shell is more than half full, $g = 2$ when the shell is half full, and $g < 2$ when the shell is less than half full.

We consider a single electron with orbital quantum number $L = 1$ moving about a nucleus, the whole being placed in an inhomogeneous crystalline electric field. We omit electron spin.

In a crystal of orthorhombic symmetry the charges on neighboring ions will produce an electrostatic potential $\varphi$ about the nucleus of the form

$$e\varphi = Ax^2 + By^2 - (A + B)z^2,$$

(24)

where $A$ and $B$ are constants. This expression is the lowest degree polynomial in $x, y, z$ which is a solution of the Laplace equation $\nabla^2 \varphi = 0$ and compatible with the symmetry of the crystal.
In free space the ground state is three-fold degenerate, with magnetic quantum numbers $m_L = 1, 0, -1$. In a magnetic field these levels are split by energies proportional to the field $B$, and it is this field-proportional splitting which is responsible for the normal paramagnetic susceptibility of the ion. In the crystal the picture may be different. We take as the three wavefunctions associated with the unperturbed ground state of the ion
\[ U_x = xf(r) ; \quad U_y = yf(r) ; \quad U_z = zf(r) \].
(25)

These wavefunctions are orthogonal, and we assume that they are normalized. Each of the $U$'s can be shown to have the property
\[ \mathcal{L}^2 U_i = L(L + 1)U_i = 2U_i \],
(26)
where $\mathcal{L}^2$ is the operator for the square of the orbital angular momentum, in units of $\hbar$. The result (26) confirms that the selected wavefunctions are in fact $p$ functions, having $L = 1$.

We observe now that the $U$'s are diagonal with respect to the perturbation, as by symmetry the nondiagonal elements vanish:
\[ \langle U_x|e\varphi|U_y \rangle = \langle U_x|e\varphi|U_z \rangle = \langle U_y|e\varphi|U_x \rangle = 0 \].
(27)

Consider for example
\[ \langle U_x|e\varphi|U_y \rangle = \int xy[f(r)^2\{Ax^2 + By^2 - (A + B)x^2\}] \, dx \, dy \, dz ; \]
(28)
the integrand is an odd function of $x$ (and also of $y$) and therefore the integral must be zero. The energy levels are then given by the diagonal matrix elements:
\[
\langle U_x|e\varphi|U_x \rangle = \int [f(r)^2\{Ax^4 + By^2x^2 - (A + B)x^2y^2\}] \, dx \, dy \, dz \\
= A(I_1 - I_2) ,
\]
(29)
where
\[ I_1 = \int [f(r)^2x^4] \, dx \, dy \, dz ; \quad I_2 = \int [f(r)^2x^2y^2] \, dx \, dy \, dz \].

In addition,
\[ \langle U_y|e\varphi|U_y \rangle = B(I_1 - I_2) ; \quad \langle U_z|e\varphi|U_z \rangle = -(A + B)(I_1 - I_2) \].

The three eigenstates in the crystal field are $p$ functions with their angular lobes directed along each of the $x$, $y$, $z$ axes, respectively.

The orbital moment of each of the levels is zero, because
\[ \langle U_x|L_x|U_x \rangle = \langle U_y|L_y|U_y \rangle = \langle U_z|L_z|U_z \rangle = 0 \].

This effect is known as quenching. The level still has a definite total angular momentum, since $\mathcal{L}^2$ is diagonal and gives $L = 1$, but the spatial components of the angular momentum are not constants of the motion and their time average is zero in the first approximation. Therefore the components of the orbital magnetic moment also vanish in the same approximation. The role of the crystal
field in the quenching process is to split the originally degenerate levels into nonmagnetic levels separated by energies $\geq \mu H$, so that the magnetic field is a small perturbation in comparison with the crystal field.

At a lattice site of cubic symmetry there is no term in the potential of the form (24), that is, quadratic in the electron coordinates. Now the ground state of an ion with one $p$ electron (or with one hole in a $p$ shell) will be triply degenerate. However, the energy of the ion will be lowered if the ion displaces itself with respect to the surroundings, thereby creating a noncubic potential such as (24). Such a spontaneous displacement is known as the Jahn-Teller effect and is often large and important, particularly with the Mn$^{3+}$ and Cu$^{2+}$ ions$^2$ and with holes in alkali and silver halides.

### Spectroscopic Splitting Factor

We suppose for convenience that the crystal field constants, $A$, $B$ are such that $U_x = x f(r)$ is the orbital wave function of the ground state of the atom in the crystal. For a spin $S = \frac{1}{2}$ there are two possible spin states $S_z = \pm \frac{1}{2}$ represented by the spin functions $\alpha$, $\beta$, which in the absence of a magnetic field are degenerate in the zeroth approximation. The problem is to take into account the spin-orbit interaction energy $\lambda \mathbf{L} \cdot \mathbf{S}$.

If the ground state function is $\psi_0 = U_x \alpha = x f(r) \alpha$ in the zeroth approximation, then in the first approximation, considering the $\lambda \mathbf{L} \cdot \mathbf{S}$ interaction by standard perturbation theory, we have

$$
\psi = [U_x - i(\lambda/2\Delta_1)U_y] \alpha - i(\lambda/2\Delta_2)U_z \beta,
$$

(30)

where $\Delta_1$ is the energy difference between the $U_x$ and $U_y$ states, and $\Delta_2$ is the difference between the $U_x$ and $U_z$ states. The term in $U_z \beta$ actually has only a second order effect on the result and may be discarded. The expectation value of the orbital angular momentum to the first order is given directly by

$$
\langle \psi | L_z | \psi \rangle = -\lambda/\Delta_1,
$$

and the magnetic moment of the state as measured in the $z$ direction is

$$
\mu_B \langle \psi | L_z + 2S_z | \psi \rangle = [- (\lambda/\Delta_1) + 1] \mu_B
$$

As the separation between the levels $S_z = \pm \frac{1}{2}$ in a field $H$ is

$$
\Delta E = g \mu_B H = 2[1 - (\lambda/\Delta_1)] \mu_B H,$$

the $g$ value or spectroscopic splitting factor (12) in the $z$ direction is

$$
g = 2[1 - (\lambda/\Delta_1)].
$$

Van Vleck Temperature-Independent Paramagnetism

We consider an atomic or molecular system which has no magnetic moment in the ground state, by which we mean that the diagonal matrix element of the magnetic moment operator $\mu_z$ is zero.

Suppose that there is a nondiagonal matrix element $\langle s | \mu_z | 0 \rangle$ of the magnetic moment operator, connecting the ground state $0$ with the excited state $s$ of energy $\Delta = E_s - E_0$ above the ground state. Then by standard perturbation theory the wavefunction of the ground state in a weak field ($\mu_z B \ll \Delta$) becomes

$$\psi'_0 = \psi_0 + \frac{B}{\Delta} \langle s | \mu_z | 0 \rangle \psi_s,$$

and the wavefunction of the excited state becomes

$$\psi'_s = \psi_s - \frac{B}{\Delta} \langle 0 | \mu_z | s \rangle \psi_0.$$

The perturbed ground state now has a moment

$$\langle 0' | \mu_z | 0' \rangle = 2B |\langle s | \mu_z | 0 \rangle|^2 / \Delta,$$

and the upper state has a moment

$$\langle s' | \mu_z | s' \rangle = -2B |\langle s | \mu_z | 0 \rangle|^2 / \Delta.$$

There are two interesting cases to consider:

Case (a). $\Delta \ll k_B T$. The surplus population in the ground state over the excited state is approximately equal to $N\Delta / 2k_B T$, so that the resultant magnetization is

$$M = \frac{2B |\langle s | \mu_z | 0 \rangle|^2}{\Delta} \frac{N\Delta}{2k_BT},$$

which gives for the susceptibility

$$\chi = N |\langle s | \mu_z | 0 \rangle|^2 / k_B T.$$

Here $N$ is the number of molecules per unit volume. This contribution is of the usual Curie form, although the mechanism of magnetization here is by polarization of the states of the system, whereas with free spins the mechanism of magnetization is the redistribution of ions among the spin states. We note that the splitting $\Delta$ does not enter in (37).

Case (b). $\Delta \gg k_B T$. Here the population is nearly all in the ground state, so that

$$M = \frac{2NB |\langle s | \mu_z | 0 \rangle|^2}{\Delta},$$

The susceptibility is

$$\chi = \frac{2N |\langle s | \mu_z | 0 \rangle|^2}{\Delta}. $$
independent of temperature. This type of contribution is known as Van Vleck paramagnetism.

**COOLING BY ISENTROPIC DEMAGNETIZATION**

The first method for attaining temperatures much below 1 K was that of isentropic, or adiabatic, demagnetization. By its use temperatures of $10^{-3}$ K and lower have been reached. The method rests on the fact that at a fixed temperature the entropy of a system of magnetic moments is lowered by the application of a magnetic field.

The entropy is a measure of the disorder of a system: the greater the disorder, the higher is the entropy. In the magnetic field the moments will be partly lined up (partly ordered), so that the entropy is lowered by the field. The entropy is also lowered if the temperature is lowered, as more of the moments line up.

If the magnetic field can then be removed without changing the entropy of the spin system, the order of the spin system will look like a lower temperature than the same degree of order in the presence of the field. When the specimen is demagnetized at constant entropy, entropy can flow into the spin system only from the system of lattice vibrations, as in Fig. 7. At the temperatures of interest the entropy of the lattice vibrations is usually negligible; thus the entropy of the spin system will be essentially constant during adiabatic demagnetization of the specimen. Magnetic cooling is a one-shot operation, not cyclic.

We first find an expression for the spin entropy of a system of $N$ ions, each of spin $S$, at a temperature sufficiently high that the spin system is entirely disordered. That is, $T$ is supposed to be much higher than some temperature $\Delta$ which characterizes the energy of the interactions ($E_{\text{int}} = k_B \Delta$) tending to orient the spins preferentially. Some of these interactions are discussed in Chapter 16. The definition of the entropy $\sigma$ of a system of $G$ accessible states is $\sigma = k_B \ln G$. At a temperature so high that all of the $2S + 1$ states of each ion are nearly equally populated, $G$ is the number of ways of arranging $N$ spins in $2S + 1$ states. Thus $G = (2S + 1)^N$, whence the spin entropy $\sigma_S$ is:

$$\sigma_S = k_B \ln (2S + 1)^N = N k_B \ln (2S + 1).$$

This spin entropy is reduced by a magnetic field if the lower levels gain in population when the field separates the $2S + 1$ states in energy.

---

3 The method was suggested by P. Debye, Ann. Physik 81, 1154 (1926); and W. F. Giauque, J. Am. Chem. Soc. 49, 1864 (1927). For many purposes the method has been supplanted by the He$^3$-He$^4$ dilution refrigerator which operates in a continuous cycle. The He$^3$ atoms in solution in liquid He$^4$ play the role of atoms in a gas, and cooling is effected by "vaporization" of He$^3$; see TP, Chapter 12.
Figure 7 During adiabatic demagnetization the total entropy of the specimen is constant. For effective cooling the initial entropy of the lattice should be small in comparison with the entropy of the spin system.

The steps carried out in the cooling process are shown in Fig. 8. The field is applied at temperature $T_1$ with the specimen in good thermal contact with the surroundings, giving the isothermal path $ab$. The specimen is then insulated ($\Delta \sigma = 0$) and the field removed; the specimen follows the constant entropy path $bc$, ending up at temperature $T_2$. The thermal contact at $T_1$ is provided by helium gas, and the thermal contact is broken by removing the gas with a pump.

**Nuclear Demagnetization**

The population of a magnetic sublevel is a function only of $\mu B/k_B T$, hence of $B/T$. The spin-system entropy is a function only of the population distribution; hence the spin entropy is a function only of $B/T$. If $B_\Delta$ is the effective field that corresponds to the local interactions, the final temperature $T_2$ reached in an adiabatic demagnetization experiment is

$$T_2 = T_1 (B_\Delta/B) ,$$

where $B$ is the initial field and $T_1$ the initial temperature.

Because nuclear magnetic moments are weak, nuclear magnetic interactions are much weaker than similar electronic interactions. We expect to reach a temperature 100 times lower with a nuclear paramagnet than with an electron paramagnet. The initial temperature $T_1$ of the nuclear stage in a nuclear spin-cooling experiment must be lower than in an electron spin-cooling experiment. If we start at $B = 50$ kG and $T_1 = 0.01$ K, then $\mu B/k_B T_1 = 0.5$, and the en-
Figure 8  Entropy for a spin $\frac{1}{2}$ system as a function of temperature, assuming an internal random magnetic field $B_\Delta$ of 100 gauss. The specimen is magnetized isothermally along $ab$, and is then insulated thermally. The external magnetic field is turned off along $bc$. In order to keep the figure on a reasonable scale the initial temperature $T_1$ is lower than would be used in practice, and so is the external magnetic field.

entropy decrease on magnetization is over 10 percent of the maximum spin entropy. This is sufficient to overwhelm the lattice and from (41) we estimate a final temperature $T_2 \approx 10^{-7}$ K. The first$^4$ nuclear cooling experiment was carried out on Cu nuclei in the metal, starting from a first stage at about 0.02 K as attained by electronic cooling. The lowest temperature reached was $1.2 \times 10^{-6}$ K.

The results in Fig. 9 fit a line of the form of (41): $T_2 = T_1 (3.1/B)$ with $B$ in gauss, so that $B_\Delta = 3.1$ gauss. This is the effective interaction field of the magnetic moments of the Cu nuclei. The motivation for using nuclei in a metal is that conduction electrons help ensure rapid thermal contact of lattice and nuclei at the temperature of the first stage. The present record$^5$ for a spin temperature is 280 pK, in rhodium.

**PARAMAGNETIC SUSCEPTIBILITY OF CONDUCTION ELECTRONS**

We are going to try to show how on the basis of these statistics the fact that many metals are diamagnetic, or only weakly paramagnetic, can be brought into agreement with the existence of a magnetic moment of the electrons.

W. Pauli, 1927

Classical free electron theory gives an unsatisfactory account of the paramagnetic susceptibility of the conduction electrons. An electron has associated with it a magnetic moment of one Bohr magneton, $\mu_B$. One might expect that


the conduction electrons would make a Curie-type paramagnetic contribution (22) to the magnetization of the metal: \( M = N\mu_B^2B/k_BT \). Instead it is observed that the magnetization of most normal nonferromagnetic metals is independent of temperature.

Pauli showed that the application of the Fermi-Dirac distribution (Chapter 6) would correct the theory as required. We first give a qualitative explanation of the situation. The result (18) tells us that the probability an atom will be lined up parallel to the field \( B \) exceeds the probability of the antiparallel orientation by roughly \( \mu B/k_BT \). For \( N \) atoms per unit volume, this gives a net magnetization \( \approx N\mu^2B/k_BT \), the standard result.

Most conduction electrons in a metal, however, have no possibility of turning over when a field is applied, because most orbitals in the Fermi sea with parallel spin are already occupied. Only the electrons within a range \( k_BT \) of the top of the Fermi distribution have a chance to turn over in the field; thus only the fraction \( T/T_F \) of the total number of electrons contribute to the susceptibility. Hence

\[
M \approx \frac{N\mu^2B}{k_BT} \cdot T \cdot \frac{T}{T_F} = \frac{N\mu^2}{k_BT_F} B,
\]

which is independent of temperature and of the observed order of magnitude.

We now calculate the expression for the paramagnetic susceptibility of a free electron gas at \( T \ll T_F \). We follow the method of calculation suggested by Fig. 10. An alternate derivation is the subject of Problem 5.
The concentration of electrons with magnetic moments parallel to the magnetic field is

\[ N_+ = \frac{1}{2} \int_{-\mu_B}^{\epsilon_f} d\epsilon \: D(\epsilon + \mu B) \equiv \frac{1}{2} \int_0^{\epsilon_f} d\epsilon \: D(\epsilon) + \frac{1}{2} \mu B \: D(\epsilon_F), \]

written for absolute zero. Here \( \frac{1}{2} D(\epsilon + \mu B) \) is the density of orbitals of one spin orientation, with allowance for the downward shift of energy by \(-\mu B\). The approximation is written for \( k_B T \ll \epsilon_F \).

The concentration of electrons with magnetic moments antiparallel to the magnetic field is

\[ N_- = \frac{1}{2} \int_{\mu_B}^{\epsilon_f} d\epsilon \: D(\epsilon - \mu B) \equiv \frac{1}{2} \int_0^{\epsilon_f} d\epsilon \: D(\epsilon) - \frac{1}{2} \mu B \: D(\epsilon_F). \]

The magnetization is given by \( M = \mu(N_+ - N_-) \), so that

\[ M = \mu^2 \: D(\epsilon_F)B = \frac{3N\mu^2}{2k_B T_F} \: B, \]

with \( D(\epsilon_F) = 3N/2\epsilon_F = 3N/2k_B T_F \) from Chapter 6. The result (42) gives the Pauli spin magnetization of the conduction electrons, for \( k_B T \ll \epsilon_F \).

In deriving the paramagnetic susceptibility, we have supposed that the spatial motion of the electrons is not affected by the magnetic field. But the wavefunctions are modified by the magnetic field; Landau has shown that for
free electrons this causes a diamagnetic moment equal to $-\frac{1}{3}$ of the paramagnetic moment. Thus the total magnetization of a free electron gas is

$$M = \frac{N\mu_B^2}{k_B T_F} B.$$  \hspace{1cm} (43)

Before comparing (43) with the experiment we must take account of the diamagnetism of the ionic cores, of band effects, and of electron-electron interactions. In sodium the interaction effects increase the spin susceptibility by perhaps 75 percent.

The magnetic susceptibility is considerably higher for most transition metals (with unfilled inner electron shells) than for the alkali metals (Fig. 11). The high values suggest that the density of orbitals is unusually high for transition metals, in agreement with measurements of the electronic heat capacity. We saw in Chapter 9 how this arises from band theory.

**SUMMARY**

(In CGS Units)

- The diamagnetic susceptibility of $N$ atoms of atomic number $Z$ is $\chi = -Ze^2N\langle r^2 \rangle/6mc^2$, where $\langle r^2 \rangle$ is the mean square atomic radius. (Langevin)

- Atoms with a permanent magnetic moment $\mu$ have a paramagnetic susceptibility $\chi = N\mu^2/3k_BT$, for $\mu B < k_BT$. (Curie-Langevin)

- For a system of spins $S = \frac{1}{2}$, the exact magnetization is $M = N\mu \tanh(\mu B/k_BT)$, where $\mu = \frac{1}{2}g\mu_B$. (Brillouin)

- The ground state of electrons in the same shell have the maximum value of $S$ allowed by the Pauli principle and the maximum $L$ consistent with this $S$. The $J$ value is $L + S$ if the shell is more than half full and $|L - S|$ if the shell is less than half full.

- A cooling process operates by demagnetization of a paramagnetic salt at constant entropy. The final temperature reached is of the order of $(B_\Delta/B)T_{\text{initial}}$, where $B_\Delta$ is the effective local field and $B$ is the initial applied magnetic field.

- The paramagnetic susceptibility of a Fermi gas of conduction electrons is $\chi = 3N\mu^2/2\varepsilon_F$, independent of temperature for $k_BT \ll \varepsilon_F$. (Pauli)

**Problems**

1. **Diamagnetic susceptibility of atomic hydrogen.** The wave function of the hydrogen atom in its ground state (1s) is $\psi = (\pi a_0^2)^{-1/2} \exp(-r/a_0)$, where $a_0 = h^2/mc^2 = 0.529 \times 10^{-8}$ cm. The charge density is $\rho(x, y, z) = -e|\psi|^2$, according to the statistical interpretation of the wave function. Show that for this state $\langle r^2 \rangle = 3a_0^2$, and calculate the molar diamagnetic susceptibility of atomic hydrogen $(-2.36 \times 10^{-6}$ cm$^3$/mole).
2. **Hund rules.** Apply the Hund rules to find the ground state (the basic level in the notation of Table 1) of (a) Eu^{2+}, in the configuration 4f^7 5s^2 p^6; (b) Yb^{3+}; (c) Tb^{3+}. The results for (b) and (c) are in Table 1, but you should give the separate steps in applying the rules.

3. **Triplet excited states.** Some organic molecules have a triplet (S = 1) excited state at an energy k_BΔ above a singlet (S = 0) ground state. (a) Find an expression for the magnetic moment ⟨μ⟩ in a field B. (b) Show that the susceptibility for T ≫ Δ is approximately independent of Δ. (c) With the help of a diagram of energy levels versus field and a rough sketch of entropy versus field, explain how this system might be cooled by adiabatic magnetization (not demagnetization).
4. **Heat capacity from internal degrees of freedom.** (a) Consider a two-level system with an energy splitting $k_B \Delta$ between upper and lower states, the splitting may arise from a magnetic field or in other ways. Show that the heat capacity per system is

$$ C = \left( \frac{\partial U}{\partial T} \right)_\Delta = k_B \frac{(\Delta/T)^2 e^{\Delta/T}}{(1 + e^{\Delta/T})^2} \, . $$

The function is plotted in Fig. 12. Peaks of this type in the heat capacity are often known as Schottky anomalies. The maximum heat capacity is quite high, but for $T \ll \Delta$ and for $T \gg \Delta$ the heat capacity is low. (b) Show that for $T \gg \Delta$ we have $C = k_B (\Delta/2T)^2 + \cdots$. The hyperfine interaction between nuclear and electronic magnetic moments in paramagnetic salts (and in systems having electron spin order) causes splittings with $\Delta \approx 1$ to $100$ mK. These splittings are often detected experimentally by the presence of a term in $1/T^2$ in the heat capacity in the region $T \gg \Delta$. Nuclear electric quadrupole interactions (see Chapter 16) with crystal fields also cause splittings, as in Fig. 13.

5. **Pauli spin susceptibility.** The spin susceptibility of a conduction electron gas at absolute zero may be discussed by another method. Let

$$ N^+ = \frac{1}{2} N (1 + \zeta) \, ; \quad N^- = \frac{1}{2} N (1 - \zeta) $$

be the concentrations of spin-up and spin-down electrons. (a) Show that in a magnetic field $B$ the total energy of the spin-up band in a free electron gas is

$$ E^+ = E_0 (1 + \zeta)^{5/3} - \frac{1}{2} N \mu_B (1 + \zeta) \, , $$

where $E_0 = \frac{3}{4} N \varepsilon_F$, in terms of the Fermi energy $\varepsilon_F$ in zero magnetic field. Find a similar expression for $E^-$. (b) Minimize $E_{\text{total}} = E^+ + E^-$ with respect to $\zeta$ and solve for the equilibrium value of $\zeta$ in the approximation $\zeta \ll 1$. Go on to show that the magnetization is $M = 3N \mu_B^2 / 2 \varepsilon_F$, in agreement with Eq. (42).

6. **Conduction electron ferromagnetism.** We approximate the effect of exchange interactions among the conduction electrons if we assume that electrons with parallel spins interact with each other with energy $-V$, and $V$ is positive, while electrons with antiparallel spins do not interact with each other. (a) Show with the help of Problem 5 that the total energy of the spin-up band is

$$ E^+ = E_0 (1 + \zeta)^{5/3} - \frac{1}{2} V N^2 (1 + \zeta)^2 - \frac{1}{2} N \mu_B (1 + \zeta) \, ; $$

find a similar expression for $E^-$. (b) Minimize the total energy and solve for $\zeta$ in the limit $\zeta \ll 1$. Show that the magnetization is

$$ M = \frac{3N \mu_B^2}{2 \varepsilon_F - \frac{1}{2} V N} B \, , $$

so that the exchange interaction enhances the susceptibility. (c) Show that with $B = 0$ the total energy is unstable at $\zeta = 0$ when $V > 4 \varepsilon_F / 3N$. If this is satisfied a ferromagnetic state ($\zeta \neq 0$) will have a lower energy than the paramagnetic state. Because of the assumption $\zeta \ll 1$, this is a sufficient condition for ferromagnetism, but it may not be a necessary condition. It is known as the Stoner condition.
Figure 12  Heat capacity of a two-level system as a function of $T/\Delta$, where $\Delta$ is the level splitting. The Schottky anomaly is a very useful tool for determining energy level splittings of ions in rare-earth and transition-group metals, compounds, and alloys.

Figure 13  The normal-state heat capacity of gallium at $T < 0.21$ K. The nuclear quadrupole ($C \propto T^{-2}$) and conduction electron ($C \propto T$) contributions dominate the heat capacity at very low temperatures. (After N. E. Phillips.)
7. **Two-level system.** The result of Problem 4 is often seen in another form. (a) If the two energy levels are at $\Delta$ and $-\Delta$, show that the energy and heat capacity are

$$ U = -\Delta \tanh(\Delta/k_B T) \quad ; \quad C = k_B (\Delta/k_B T)^2 \sech^2(\Delta/k_B T). $$

(b) If the system has a random composition such that all values of $\Delta$ are equally likely up to some limit $\Delta_0$, show that the heat capacity is linearly proportional to the temperature, provided $k_B T \ll \Delta_0$. This result was applied to the heat capacity of dilute magnetic alloys by W. Marshall, Phys. Rev. 118, 1519 (1960). It is also used in the theory of glasses, Chapter 17.

8. **Paramagnetism of $S = 1$ system.** (a) Find the magnetization as a function of magnetic field and temperature for a system of spins with $S = 1$, moment $\mu$, and concentration $n$. (b) Show that in the limit $\mu B \ll kT$ the result is $M = (2n\mu^2/3kT)B$.

---

**References**


Ferromagnetism and Antiferromagnetism

FERROMAGNETIC ORDER
- Curie point and the exchange integral 443
- Temperature dependence of the saturation magnetization 446
- Saturation magnetization at absolute zero 448

MAGNONS
- Quantization of spin waves 450
- Thermal excitation of magnons 455

NEUTRON MAGNETIC SCATTERING 456

FERRIMAGNETIC ORDER
- Curie temperature and susceptibility of ferrimagnets 460
- Iron garnets 461

ANTIFERROMAGNETIC ORDER
- Susceptibility below the Néel temperature 462
- Antiferromagnetic magnons 466

FERROMAGNETIC DOMAINS
- Anisotropy energy 468
- Transition region between domains 472
- Solitons 473
- Origin of domains 474
- Coercivity and hysteresis 475

SINGLE DOMAIN PARTICLES
- Geomagnetism and biomagnetism 477
- Magnetic force microscopy 478

MAGNETIC BUBBLE DOMAINS 480

SUMMARY 481

PROBLEMS
1. Magnon dispersion relation 481
2. Heat capacity of magnons 482
3. Néel temperature 482
4. Magnetoelastic coupling 492
5. Coercive force of a small particle 482
6. Saturation magnetization near $T_c$ 483
7. Néel wall 483

REFERENCES 483

NOTATION: (CGS) $B = H + 4\pi M$; (SI) $B = \mu_0(H + M)$. We call $B_a$ the applied magnetic field in both systems of units: in CGS we have $B_a = H_a$ and in SI we have $B_a = \mu_0 H_a$. The susceptibility is $\chi = M/B_a$ in CGS and $\chi = M/H_a = \mu_0 M/B_a$ in SI. One tesla = $10^4$ gauss.
Figure 1 Ordered arrangements of electron spins.
CHAPTER 15: FERROMAGNETISM AND ANTFERROMAGNETISM

FERROMAGNETIC ORDER

A ferromagnet has a spontaneous magnetic moment—a magnetic moment even in zero applied magnetic field. The existence of a spontaneous moment suggests that electron spins and magnetic moments are arranged in a regular manner. The order need not be simple: all of the spin arrangements sketched in Fig. 1 except the simple antiferromagnet have a spontaneous magnetic moment, called the saturation moment.

Curie Point and the Exchange Integral

Consider a paramagnet with a concentration of $N$ ions of spin $S$. Given an internal interaction tending to line up the magnetic moments parallel to each other, we shall have a ferromagnet. Let us postulate such an interaction and call it the exchange field.\(^1\) The orienting effect of the exchange field is opposed by thermal agitation, and at elevated temperatures the spin order is destroyed.

We treat the exchange field as equivalent to a magnetic field $B_E$. The magnitude of the exchange field may be as high as $10^7$ gauss ($10^3$ tesla). We assume that $B_E$ is proportional to the magnetization $M$.

The magnetization is defined as the magnetic moment per unit volume; unless otherwise specified it is understood to be the value in thermal equilibrium at the temperature $T$. If domains (regions magnetized in different directions) are present, the magnetization refers to the value within a domain.

In the mean field approximation we assume each magnetic atom experiences a field proportional to the magnetization:

$$B_E = \lambda M,$$

where $\lambda$ is a constant, independent of temperature. According to (1), each spin sees the average magnetization of all the other spins. In truth, it may see only near neighbors, but our simplification is good for a first look at the problem.

The Curie temperature $T_c$ is the temperature above which the spontaneous magnetization vanishes; it separates the disordered paramagnetic phase at $T > T_c$ from the ordered ferromagnetic phase at $T < T_c$. We can find $T_c$ in terms of $\lambda$.

\(^1\)Also called the molecular field or the Weiss field, after Pierre Weiss who was the first to imagine such a field. The exchange field $B_E$ simulates a real magnetic field in the expressions for the energy $-\mu \cdot B_E$ and the torque $\mu \times B_E$ on a magnetic moment $\mu$. But $B_E$ is not really a magnetic field and therefore does not enter into the Maxwell equations: for example, there is no current density $j$ related to $B_E$ by curl $\mathbf{H} = 4\pi j/c$. The magnitude of $B_E$ is typically $10^4$ larger than the average magnetic field of the magnetic dipoles of the ferromagnet.
Consider the paramagnetic phase: an applied field $\mu_a$ will cause a finite magnetization and this in turn will cause a finite exchange field $B_E$. If $\chi_p$ is the paramagnetic susceptibility,

\[ M = \chi_p(B_a + B_E) \]  

The magnetization is equal to a constant susceptibility times a field only if the fractional alignment is small: this is where the assumption enters that the specimen is in the paramagnetic phase.

The paramagnetic susceptibility is given by the Curie law $\chi_p = C/T$, where $C$ is the Curie constant. Substitute (1) in (2); we find $MT = C(B_a + \lambda M)$ and

\[ \chi = \frac{M}{B_a} = \frac{C}{(T - C\lambda)} . \]  

The susceptibility (3) has a singularity at $T = C\lambda$. At this temperature (and below) there exists a spontaneous magnetization, because if $\chi$ is infinite we can have a finite $M$ for zero $B_a$. From (3) we have the Curie-Weiss law

\[ \chi = \frac{C}{T - T_c} ; \quad T_c = C\lambda . \]  

This expression describes fairly well the observed susceptibility variation in the paramagnetic region above the Curie point.

Detailed calculations\(^2\) predict

\[ \chi \propto \frac{1}{(T - T_c)^{1.33}} \]

at temperatures very close to $T_c$, in general agreement with the experimental data summarized in Table 1. The reciprocal susceptibility of nickel is plotted in Fig. 2.

From (4) and the definition (14.22) of the Curie constant $C$ we may determine the value of the mean field constant $\lambda$ in (1):

\[ \lambda = \frac{T_c}{C} \frac{3k_B T_c}{Ng^2S(S + 1)\mu_B^2} . \]

For iron $T_c = 1000$ K, $g = 2$, and $S = 1$; from (5) we have $\lambda \approx 5000$. With $M_s = 1700$ we have $B_E \approx \lambda M \approx (5000)(1700) \approx 10^7$ G = $10^3$ T. The exchange field in iron is very much stronger than the real magnetic field due to the other magnetic ions in the crystal: a magnetic ion produces a field $\approx \mu_B/\alpha^3$ or about $10^3$ G = 0.1 T at a neighboring lattice point.

\(^2\)Experimentally the susceptibility for $T \gg T_c$ is given quite accurately by $C/(T - \theta)$, where $\theta$ is appreciably greater than the actual transition temperature $T_c$. See the review by C. Domb in *Magnetism*, Vol. 2A, G. T. Rado and H. Suhl, eds., Academic Press, 1965.
Table 1 Critical point exponents for ferromagnets

As $T \to T_c$ from above, the susceptibility $\chi$ becomes proportional to $(T - T_c)^{-\gamma}$; as $T \to T_c$ from below, the magnetization $M_s$ becomes proportional to $(T_c - T)^{\beta}$. In the mean field approximation, $\gamma = 1$ and $\beta = \frac{1}{4}$.

<table>
<thead>
<tr>
<th></th>
<th>$\gamma$</th>
<th>$\beta$</th>
<th>$T_c$, in K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>1.33 ± 0.015</td>
<td>0.34 ± 0.04</td>
<td>1043</td>
</tr>
<tr>
<td>Co</td>
<td>1.21 ± 0.04</td>
<td>—</td>
<td>1388</td>
</tr>
<tr>
<td>Ni</td>
<td>1.35 ± 0.02</td>
<td>0.42 ± 0.07</td>
<td>627.2</td>
</tr>
<tr>
<td>Gd</td>
<td>1.3 ± 0.1</td>
<td>—</td>
<td>292.5</td>
</tr>
<tr>
<td>CrO$_2$</td>
<td>1.63 ± 0.02</td>
<td>—</td>
<td>386.5</td>
</tr>
<tr>
<td>CrBr$_3$</td>
<td>1.215 ± 0.02</td>
<td>0.368 ± 0.005</td>
<td>32.56</td>
</tr>
<tr>
<td>EuS</td>
<td>—</td>
<td>0.33 ± 0.015</td>
<td>16.50</td>
</tr>
</tbody>
</table>

Experimental data collected by H. E. Stanley.

Figure 2 Reciprocal of the susceptibility per gram of nickel in the neighborhood of the Curie temperature (355°C). The density is $\rho$. The dashed line is a linear extrapolation from high temperatures. (After P. Weiss and R. Forrer.)
The exchange field gives an approximate representation of the quantum-mechanical exchange interaction. On certain assumptions it can be shown\(^3\) that the energy of interaction of atoms \(i, j\) bearing electron spins \(S_i, S_j\) contains a term

\[
U = -2JS_i \cdot S_j ,
\]

where \(J\) is the exchange integral and is related to the overlap of the charge distributions of the atoms \(i, j\). Equation (6) is called the Heisenberg model.

The charge distribution of a system of two spins depends on whether the spins are parallel or antiparallel,\(^4\) for the Pauli principle excludes two electrons of the same spin from being at the same place at the same time. It does not exclude two electrons of opposite spin. Thus the electrostatic energy of a system will depend on the relative orientation of the spins: the difference in energy defines the exchange energy.

The exchange energy of two electrons may be written in the form \(-2JS_1 \cdot s_2\) as in (6), just as if there were a direct coupling between the directions of the two spins. For many purposes in ferromagnetism it is a good approximation to treat the spins as classical angular momentum vectors.

We can establish an approximate connection between the exchange integral \(J\) and the Curie temperature \(T_c\). Suppose that the atom under consideration has \(z\) nearest neighbors, each connected with the central atom by the interaction \(J\). For more distant neighbors we take \(J\) as zero. The mean field theory result is

\[
J = \frac{3k_B T_c}{2zS(S+1)} .
\]

Better statistical approximations give somewhat different results. For the sc, bcc, and fcc structures with \(S = \frac{1}{2}\), Rushbrooke and Wood give \(k_B T_c/zJ = 0.28; 0.325;\) and 0.346, respectively, as compared with 0.500 from (7) for all three structures. If iron is represented by the Heisenberg model with \(S = 1\), then the observed Curie temperature corresponds to \(J = 11.9\) meV.

**Temperature Dependence of the Saturation Magnetization**

We can also use the mean field approximation below the Curie temperature to find the magnetization as a function of temperature. We proceed as


\(^4\)If two spins are antiparallel, the wavefunctions of the two electrons must be symmetric, as in the combination \(u(r_1)v(r_2) + u(r_2)v(r_1)\). If the two spins are parallel, the Pauli principle requires that the orbital part of the wavefunction be antisymmetric, as in \(u(r_1)v(r_2) - u(r_2)v(r_1)\), for here if we interchange the coordinates \(r_1, r_2\) the wavefunction changes sign. If we set the positions equal so that \(r_1 = r_2\), then the antisymmetric function vanishes: for parallel spins there is zero probability of finding the two electrons at the same position.
Figure 3  Graphical solution of Eq. (9) for the reduced magnetization \( m \) as a function of temperature. The reduced magnetization is defined as \( m = M/N\mu \). The left-hand side of Eq. (9) is plotted as a straight line \( m \) with unit slope. The right-hand side is \( \tanh(m/t) \) and is plotted vs. \( m \) for three different values of the reduced temperature \( t = k_B T/N\mu^2 \lambda = T/T_c \). The three curves correspond to the temperatures \( 2T_c, T_c, \) and \( 0.5T_c \). The curve for \( t = 2 \) intersects the straight line \( m \) only at \( m = 0 \), as appropriate for the paramagnetic region (there is no external applied magnetic field). The curve for \( t = 1 \) (or \( T = T_c \)) is tangent to the straight line \( m \) at the origin; this temperature marks the onset of ferromagnetism. The curve for \( t = 0.5 \) is in the ferromagnetic region and intersects the straight line \( m \) at about \( m = 0.94N\mu \). As \( t \to 0 \) the intercept moves up to \( m = 1 \), so that all magnetic moments are lined up at absolute zero.

before, but instead of the Curie law we use the complete Brillouin expression for the magnetization. For spin \( \frac{1}{2} \) this is \( M = N\mu \tanh(\mu B/k_B T) \).

If we omit the applied magnetic field and replace \( B \) by the molecular field \( B_E = \lambda M \), then

\[
M = N\mu \tanh(\mu\lambda M/k_B T) .
\]

We shall see that solutions of this equation with nonzero \( M \) exist in the temperature range between 0 and \( T_c \).

To solve (8) we write it in terms of the reduced magnetization \( m = M/N\mu \) and the reduced temperature \( t = k_B T/N\mu^2 \lambda \), whence

\[
m = \tanh(m/t) .
\]

We then plot the right and left sides of this equation separately as functions of \( m \), as in Fig. 3. The intercept of the two curves gives the value of \( m \) at the temperature of interest. The critical temperature is \( t = 1 \), or \( T_c = N\mu^2 \lambda/k_B \).

The curves of \( M \) versus \( T \) obtained in this way reproduce roughly the features of the experimental results, as shown in Fig. 4 for nickel. As \( T \) increases the magnetization decreases smoothly to zero at \( T = T_c \). This behavior classifies the usual ferromagnetic/paramagnetic transition as a second order transition.
The mean field theory does not give a good description of the variation of $M$ at low temperatures. For $T \ll T_c$ the argument of $\tanh$ in (9) is large, and

$$\tanh \xi \approx 1 - 2e^{-2\xi}.$$ 

To lowest order the magnetization deviation $\Delta M = M(0) - M(T)$ is

$$\Delta M \approx 2N\mu \exp(-2\lambda N\mu^2/k_BT).$$  (10)

The argument of the exponential is equal to $-2T_c/T$. For $T = 0.1T_c$ we have $\Delta M/N\mu \approx 4 \times 10^{-9}$.

The experimental results show a much more rapid dependence of $\Delta M$ on temperature at low temperatures. At $T = 0.1T_c$ we have $\Delta M/M \approx 2 \times 10^{-3}$ from the data of Fig. 5. The leading term in $\Delta M$ is observed from experiment to have the form

$$\frac{\Delta M/M(0)}{M(0)} = AT^{3/2},$$  (11)

where the constant $A$ has the experimental value $(7.5 \pm 0.2) \times 10^{-6}$ deg$^{-3/2}$ for Ni and $(3.4 \pm 0.2) \times 10^{-6}$ deg$^{-3/2}$ for Fe. The result (11) finds a natural explanation in terms of spin wave theory, as discussed below.

**Saturation Magnetization at Absolute Zero**

Table 2 gives representative values of the saturation magnetization $M_s$, the ferromagnetic Curie temperature, and the effective magneton number defined by $M_s(0) = n_B N \mu_B$, where $N$ is the number of formula units per unit volume. Do not confuse $n_B$ with the paramagnetic effective magneton number $p$ defined by (14.23).
Figure 5  Decrease in magnetization of nickel with temperature, after Argyle, Charap, and Pugh. In the plot $\Delta M = 0$ at 4.2 K.

<table>
<thead>
<tr>
<th>Table 2  Ferromagnetic crystals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substance</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Fe</td>
</tr>
<tr>
<td>Co</td>
</tr>
<tr>
<td>Ni</td>
</tr>
<tr>
<td>Gd</td>
</tr>
<tr>
<td>Dy</td>
</tr>
<tr>
<td>MnAs</td>
</tr>
<tr>
<td>MnBi</td>
</tr>
<tr>
<td>MnSb</td>
</tr>
<tr>
<td>CrO$_2$</td>
</tr>
<tr>
<td>MnOFe$_2$O$_3$</td>
</tr>
<tr>
<td>FeOFe$_2$O$_3$</td>
</tr>
<tr>
<td>NiOFe$_2$O$_3$</td>
</tr>
<tr>
<td>CuOFe$_2$O$_3$</td>
</tr>
<tr>
<td>MgOFe$_2$O$_3$</td>
</tr>
<tr>
<td>EuO</td>
</tr>
<tr>
<td>Y$_3$Fe$<em>3$O$</em>{12}$</td>
</tr>
</tbody>
</table>
Observed values of $n_B$ are often nonintegral. There are many possible causes. One is the spin-orbit interaction which adds or subtracts some orbital magnetic moment. Another cause in ferromagnetic metals is the conduction electron magnetization induced locally about a paramagnetic ion core. A third cause is suggested by the drawing in Fig. 1 of the spin arrangement in a ferromagnet: if there is one atom of spin projection $-S$ for every two atoms $+S$, the average spin is $\frac{1}{2}S$.

Are there in fact any simple ferromagnetic insulators, with all ionic spins parallel in the ground state? The few simple ferromagnets known at present include CrBr$_3$, EuO, and EuS.

A band or itinerant electron model\textsuperscript{5} accounts for the ferromagnetism of the transition metals Fe, Co, Ni. The approach is indicated in Figs. 6 and 7. The relationship of $4s$ and $3d$ bands is shown in Fig. 6 for copper, which is not ferromagnetic. If we remove one electron from copper, we obtain nickel which has the possibility of a hole in the $3d$ band. In the band structure of nickel shown in Fig. 7a for $T > T_c$ we have taken $2 \times 0.27 = 0.54$ of an electron away from the $3d$ band and $0.46$ away from the $4s$ band, as compared with copper.

The band structure of nickel at absolute zero is shown in Fig. 7b. Nickel is ferromagnetic, and at absolute zero $n_B = 0.60$ Bohr magnetons per atom. After allowance for the magnetic moment contribution of orbital electronic motion, nickel has an excess of $0.54$ electron per atom having spin preferentially oriented in one direction. The exchange enhancement of the susceptibility of metals was the subject of Problem 14.6. The properties of amorphous ferromagnets are treated in Chapter 17.

**MAGNONS**

A magnon is a quantized spin wave. We use a classical argument, just as we did for phonons, we find the magnon dispersion relation for $\omega$ versus $k$. We then quantize the magnon energy and interpret the quantization in terms of spin reversal.

The ground state of a simple ferromagnet has all spins parallel, as in Fig. 8a. Consider $N$ spins each of magnitude $S$ on a line or a ring, with nearest neighbor spins coupled by the Heisenberg interaction:

$$U = -2J \sum_{p=1}^{N} S_p \cdot S_{p+1}. \quad (12)$$

Here $J$ is the exchange integral and $\hbar S_p$ is the angular momentum of the spin at site $p$. If we treat the spins $S_p$ as classical vectors, then in the ground state $S_p \cdot S_{p+1} = S^2$ and the exchange energy of the system is $U_0 = -2NJS^2$.

\textsuperscript{5}C. Herring in Rado and Suhl, Vol. IV (1965).
Figure 6a  Schematic relationship of 4s and 3d bands in metallic copper. The 3d band holds 10 electrons per atom and is filled in copper. The 4s band can hold two electrons per atom; it is shown half-filled, as copper has one valence electron outside the filled 3d shell.

Figure 6b  The filled 3d band of copper shown as two separate sub-bands of opposite electron spin orientation, each band holding five electrons. With both sub-bands filled as shown, the net spin (and hence the net magnetization) of the d band is zero.

Figure 7a  Band relationships in nickel above the Curie temperature. The net magnetic moment is zero, as there are equal numbers of holes in both 3d ↓ and 3d ↑ bands.

Figure 7b  Schematic relationship of bands in nickel at absolute zero. The energies of the 3d ↑ and 3d ↓ sub-bands are separated by an exchange interaction. The 3d ↑ band is filled; the 3d ↓ band contains 4.46 electrons and 0.54 hole. The 4s band is usually thought to contain approximately equal numbers of electrons in both spin directions, and so we have not troubled to divide it into sub-bands. The net magnetic moment of 0.54 μB per atom arises from the excess population of the 3d ↑ band over the 3d ↓ band. It is often convenient to speak of the magnetization as arising from the 0.54 hole in the 3d ↓ band.
What is the energy of the first excited state? Consider ... excited state with one particular spin reversed, as in Fig. 8b. We see from (12) that this increases the energy by $8JS^2$, so that $U_1 = U_0 + 8JS^2$.

We can form an excitation of much lower energy if we let all the spins share the reversal, as in Fig. 8c. The elementary excitations of a spin system have a wavelike form and are called magnons (Fig. 9). These are analogous to lattice vibrations or phonons. Spin waves are oscillations in the relative orientations of spins on a lattice; lattice vibrations are oscillations in the relative positions of atoms on a lattice.

We now give a classical derivation of the magnon dispersion relation. The terms in (12) which involve the $p$th spin are

$$-2JS_p \cdot (S_{p-1} + S_{p+1}) .$$

(13)

We write the magnetic moment at site $p$ as $\mu_p = -g\mu_BS_p$. Then (13) becomes

$$-\mu_p \cdot [(-2J/g\mu_B)(S_{p-1} + S_{p+1})] ,$$

(14)

which is of the form $-\mu_p \cdot B_p$, where the effective magnetic field or exchange field that acts on the $p$th spin is

$$B_p = (-2J/g\mu_B)(S_{p-1} + S_{p+1}) .$$

(15)

From mechanics the rate of change of the angular momentum $\hbar S_p$ is equal to the torque $\mu_p \times B_p$ which acts on the spin: $\hbar \frac{dS_p}{dt} = \mu_p \times B_p$, or

$$dS_p/dt = (-g\mu_B/\hbar) S_p \times B_p = (2J/\hbar)(S_p \times S_{p-1} + S_p \times S_{p+1}) .$$

(16)

In Cartesian components

$$dS^x_p/dt = (2J/\hbar)[S^y_p(S^z_{p-1} + S^z_{p+1}) - S^z_p(S^y_{p-1} + S^y_{p+1})] ,$$

(17)

and similarly for $dS^y_p/dt$ and $dS^z_p/dt$. These equations involve products of spin components and are nonlinear.

If the amplitude of the excitation is small (if $S_p^x, S_p^y \ll S$), we may obtain an approximate set of linear equations by taking all $S_p^z = S$ and by neglecting terms in the product of $S^x$ and $S^y$ which appear in the equation for $dS^z_p/dt$. The linearized equations are

$$dS^x_p/dt = (2JS/\hbar)(2S^y_p - S^y_{p-1} - S^y_{p+1}) ;$$

(18a)

$$dS^y_p/dt = -(2JS/\hbar)(2S^x_p - S^x_{p-1} - S^x_{p+1}) ;$$

(18b)

$$dS^z_p/dt = 0 .$$

(19)

By analogy with phonon problems we look for traveling wave solutions of (18) of the form

$$S^x_p = u \exp [i(pka - \omega t)] ; \quad S^y_p = v \exp [i(pka - \omega t)] ,$$

(20)

where $u$, $v$ are constants, $p$ is an integer, and $a$ is the lattice constant. On substitution into (18) we have
Figure 8 (a) Classical picture of the ground state of a simple ferromagnet; all spins are parallel. (b) A possible excitation, one spin is reversed. (c) The low-lying elementary excitations are spin waves. The ends of the spin vectors precess on the surfaces of cones, with successive spins advanced in phase by a constant angle.

Figure 9 A spin wave on a line of spins. (a) The spins viewed in perspective. (b) Spins viewed from above, showing one wavelength. The wave is drawn through the ends of the spin vectors.

\[-i \omega u = (2JS/\hbar)(2 - e^{-ika} - e^{ika})v = (4JS/\hbar)(1 - \cos ka)v ; \]
\[-i \omega v = -(2JS/\hbar)(2 - e^{-ika} - e^{ika})u = -(4JS/\hbar)(1 - \cos ka)u . \]

These equations have a solution for \( u \) and \( v \) if the determinant of the coefficients is equal to zero:

\[
\begin{vmatrix}
  i \omega & (4JS/\hbar)(1 - \cos ka) \\
  -(4JS/\hbar)(1 - \cos ka) & i \omega 
\end{vmatrix} = 0 ,
\]

whence

\[\hbar \omega = 4JS(1 - \cos ka) .\]

This result is plotted in Fig. 10. With this solution we find that \( v = -iu \), corresponding to circular precession of each spin about the \( z \) axis. We see this on taking real parts of (20), with \( v \) set equal to \(-iu\). Then

\[S^x_p = u \cos(pka - \omega t) ; \quad S^y_p = u \sin(pka - \omega t) .\]

Equation (22) is the dispersion relation for spin waves in one dimension with nearest-neighbor interactions. Precisely the same result is obtained from the quantum-mechanical solution; see QTS, Chapter 4. At long wavelengths \( ka \ll 1 \), so that \( 1 - \cos ka \approx \frac{1}{2}(ka)^2 \) and

\[\hbar \omega \approx (2JSa^2)k^2 .\]

The frequency is proportional to \( k^2 \); in the same limit the frequency of a phonon is proportional to \( k \).
The dispersion relation for a ferromagnetic cubic lattice with nearest-neighbor interactions

$$\hbar \omega = 2JS[z - \sum_\delta \cos(k \cdot \delta)] ,$$  \hfill (24)

where the summation is over the $z$ vectors denoted by $\delta$ which join the central atom to its nearest neighbors. For $ka \ll 1$,

$$\hbar \omega = (2JSd^2)k^2$$  \hfill (25)

for all three cubic lattices, where $a$ is the lattice constant.

The coefficient of $k^2$ often may be determined accurately by neutron scattering or by spin wave resonance in thin films, Chapter 16. By neutron scattering G. Shirane and coworkers find, in the equation $\hbar \omega = Dk^2$, the values 281, 500, and 364 meV Å$^2$ for $D$ at 295 K in Fe, Co, Ni, respectively.

**Quantization of Spin Waves.** The quantization of spin waves proceeds exactly as for photons and phonons. The energy of a mode of frequency $\omega_k$ with $n_k$ magnons is given by

$$\epsilon_k = (n_k + \frac{1}{2})\hbar \omega_k .$$  \hfill (26)

The excitation of a magnon corresponds to the reversal of one spin $\frac{1}{2}$. 
Thermal Excitation of Magnons

In thermal equilibrium the average value of the number of magnons excited in the mode $k$ is given by the Planck distribution:

$$
\langle n_k \rangle = \frac{1}{\exp(\hbar\omega/k_B T) - 1}.
$$

(27)

The total number of magnons excited at a temperature $T$ is

$$
\sum_k n_k = \int d\omega \ D(\omega) \langle n(\omega) \rangle,
$$

(28)

where $D(\omega)$ is the number of magnon modes per unit frequency range. The integral is taken over the allowed range of $k$, which is the first Brillouin zone. At sufficiently low temperatures we may carry the integral between 0 and $\infty$ because $\langle n(\omega) \rangle \to 0$ exponentially as $\omega \to \infty$.

Magnons have a single polarization for each value of $k$. In three dimensions the number of modes of wavevector less than $k$ is $(1/2\pi)^3(4\pi k^3/3)$ per unit volume, whence the number of magnons $D(\omega)d\omega$ with frequency in $d\omega$ at $\omega$ is $(1/2\pi)^3(4\pi k^3)(dk/d\omega) d\omega$. In the approximation (25),

$$
\frac{d\omega}{dk} = J S a^2 k = 2 \left( \frac{2 J S a^2}{\hbar} \right)^{1/2} \omega^{1/2}.
$$

Thus the density of modes for magnons is

$$
D(\omega) = \frac{1}{4\pi^2} \left( \frac{\hbar}{2 J S a^2} \right)^{3/2} \omega^{1/2},
$$

(29)

so that the total number of magnons is, from (28),

$$
\sum_k n_k = \frac{1}{4\pi^2} \left( \frac{\hbar}{2 J S a^2} \right)^{3/2} \int_0^\infty d\omega \ \omega^{1/2} e^{\hbar\omega/\omega - 1} = \frac{1}{4\pi^2} \left( \frac{k_B T}{2 J S a^2} \right)^{3/2} \int_0^\infty dx \frac{x^{1/2}}{e^x - 1}.
$$

The definite integral is found in tables and has the value $(0.0587)(4\pi^2)$.

The number $N$ of atoms per unit volume is $Q/a^3$, where $Q = 1, 2, 4$ for sc, bcc, fcc lattices, respectively. Now $(\Sigma n_k)/N$ is equal to the fractional change of magnetization $\Delta M/M(0)$, whence

$$
\frac{\Delta M}{M(0)} = \frac{0.0587}{SQ} \left( \frac{k_B T}{2 JS} \right)^{3/2}.
$$

(30)

*The argument is exactly as for phonons or photons. The Planck distribution follows for any problem where the energy levels are identical with those of a harmonic oscillator or collection of harmonic oscillators.*
This result is the Bloch $T^{3/2}$ law and has been confirmed experimentally. In neutron scattering experiments spin waves have been observed up to temperatures near the Curie temperature and even above the Curie temperature.

NEUTRON MAGNETIC SCATTERING

An x-ray photon sees the spatial distribution of electronic charge, whether the charge density is magnetized or unmagnetized. A neutron sees two aspects of a crystal: the distribution of nuclei and the distribution of electronic magnetization. The neutron diffraction pattern for iron is shown in Fig. 11.

The magnetic moment of the neutron interacts with the magnetic moment of the electron. The cross section for the neutron-electron interaction is of the same order of magnitude as for the neutron-nuclear interaction. Diffraction of neutrons by a magnetic crystal allows the determination of the distribution, direction, and order of the magnetic moments.

The magnetic moments associated with particular components of alloys may be investigated by neutron diffraction. Results for the Fe-Co binary alloy system (which is ferromagnetic) are shown in Fig. 12. The magnetic moment on the cobalt atom does not appear to be affected by alloying, but that on the iron atom increases to about $3\mu_B$ as the cobalt concentration increases. The magnetizations are shown in Fig. 13.

A neutron can be inelastically scattered by the magnetic structure, with the creation or annihilation of a magnon (Fig. 14); such events make possible

![Graph showing neutron diffraction pattern](image-url)

**Figure 11** Neutron diffraction pattern for iron. (After C. G. Shull, E. O. Wollan, and W. C. Koehler.)
Figure 12  Moments attributable to 3d electrons in Fe-Co alloys as a function of composition, after M. F. Collins and J. B. Forsyth.

Figure 13  Average atomic moments of binary alloys of the elements in the iron group. (After Bozorth.)

Figure 14  Scattering of a neutron by an ordered magnetic structure, with creation of a magnon.
the experimental determination of magnon spectra. If the incident neutron has wavevector \(\mathbf{k}_n\) and is scattered to \(\mathbf{k}'_n\) with the creation of a magnon of wavevector \(\mathbf{k}\), then by conservation of crystal momentum \(\mathbf{k}_n = \mathbf{k}'_n + \mathbf{k} + \mathbf{G}\), where \(\mathbf{G}\) is a reciprocal lattice vector. By conservation of energy

\[
\frac{\hbar^2 k_n^2}{2M_n} = \frac{\hbar^2 k_n'^2}{2M_n} + \hbar \omega_k ,
\]

where \(\hbar \omega_k\) is the energy of the magnon created in the process. The observed magnon spectrum for MnPt₃ is shown in Fig. 15.

**FERRIMAGNETIC ORDER**

In many ferromagnetic crystals the saturation magnetization at \(T = 0\) K does not correspond to parallel alignment of the magnetic moments of the constituent paramagnetic ions, even in crystals for which the individual paramagnetic ions have their normal magnetic moments.

The most familiar example is magnetite, Fe₃O₄ or FeO · Fe₂O₃. From Table 14.2 we see that ferric (Fe³⁺) ions are in a state with spin \(S = \frac{5}{2}\) and zero orbital moment. Thus each ion should contribute \(5\mu_B\) to the saturation moment. The ferrous (Fe²⁺) ions have a spin of 2 and should contribute \(4\mu_B\), apart from any residual orbital moment contribution. Thus the effective number of Bohr magnetons per Fe₃O₄ formula unit should be about \(2 \times 5 + 4 = 14\) if all spins were parallel.

The observed value (Table 2) is 4.1. The discrepancy is accounted for if the moments of the Fe³⁺ ions are antiparallel to each other: then the observed
moment arises only from the Fe$^{2+}$ ion, as in Fig. 16. Neutron diffraction results agree with this model.

A systematic discussion of the consequences of this type of spin order was given by L. Néel with reference to an important class of magnetic oxides known as ferrites. The usual chemical formula of a ferrite is MO·Fe$_2$O$_3$, where M is a divalent cation, often Zn, Cd, Fe, Ni, Cu, Co, or Mg. The term ferrimagnetic was coined originally to describe the ferrite-type ferromagnetic spin order such as Fig. 16, and by extension the term covers almost any compound in which some ions have a moment antiparallel to other ions. Many ferrimagnets are poor conductors of electricity, a quality exploited in applications such as rf transformer cores.

The cubic ferrites have the spinel crystal structure shown in Fig. 17. There are eight occupied tetrahedral (or A) sites and 16 occupied octahedral (or B) sites in a unit cube. The lattice constant is about 8 Å. A remarkable feature of the spinels is that all exchange integrals $J_{AA}$, $J_{AB}$, and $J_{BB}$ are negative and favor antiparallel alignment of the spins connected by the interaction. But the AB interaction is the strongest, so that the A spins are parallel to each other and the B spins are parallel to each other, just in order that the A spins may be antiparallel to the B spins. If $J$ in $U = -2JS_i \cdot S_j$ is positive, we say that the exchange integral is ferromagnetic; if $J$ is negative, the exchange integral is antiferromagnetic.

We now prove that three antiferromagnetic interactions can result in ferromagnetism. The mean exchange fields acting on the A and B spin lattices may be written

$$B_A = -\lambda M_A - \mu M_B ; \quad B_B = -\mu M_A - \nu M_B ;$$

(32)

taking all mean field constants $\lambda$, $\mu$, $\nu$ to be positive. The minus sign then corresponds to an antiparallel interaction. The interaction energy density is

$$U = -\frac{1}{2}(B_A^2 \cdot M_A + B_B^2 \cdot M_B) = \frac{1}{2} \lambda M_A^2 + \mu M_A^2 + \frac{1}{2} \nu M_B^2 ;$$

(33)
this is lower when $M_A$ is antiparallel to $M_B$ than when $M_A$ is parallel to $M_B$. The energy when antiparallel should be compared with zero, because a possible solution is $M_A = M_B = 0$. Thus when

$$\mu M_A M_B > \frac{1}{2} (\lambda M_A^2 + \nu M_B^2) ,$$

(34)

the ground state will have $M_A$ directed oppositely to $M_B$. (Under certain conditions there may be noncollinear spin arrays of still lower energy.)

**Curie Temperature and Susceptibility of Ferrimagnets**

We define separate Curie constants $C_A$ and $C_B$ for the ions on the $A$ and $B$ sites. For simplicity, let all interactions be zero except for an antiparallel interaction between the $A$ and $B$ sites: $B_A = -\mu M_B$; $B_B = -\mu M_A$, where $\mu$ is positive. The same constant $\mu$ is involved in both expressions because of the form of (33).

We have in the mean field approximation

(CGS) $M_A T = C_A (B_a - \mu M_B) ; \quad M_B T = C_B (B_a - \mu M_A) ,

(35)

where $B_a$ is the applied field. These equations have a nonzero solution for $M_A$ and $M_B$ in zero applied field if

$$\begin{vmatrix} T & \mu C_A \\ \mu C_B & T \end{vmatrix} = 0 ,$$

(36)

so that the ferrimagnetic Curie temperature is given by $T_c = \mu (C_A C_B)^{1/2}$. 
We solve (35) for $M_A$ and $M_B$ to obtain the susceptibility at $T > T_c$:

$$\chi = \frac{M_A + M_B}{B_a} = \frac{(C_A + C_B)T - 2\mu C_A C_B}{T^2 - T_c^2},$$

a result more complicated than (4). Experimental values for Fe$_3$O$_4$ are plotted in Fig. 18. The curvature of the plot of $1/\chi$ versus $T$ is a characteristic feature of a ferrimagnet. We consider below the antiferromagnetic limit $C_A = C_B$.

**Iron Garnets.** The iron garnets are cubic ferrimagnetic insulators with the general formula $M_3$Fe$_5$O$_{12}$, where $M$ is a trivalent metal ion and the Fe is the trivalent ferric ion ($S = \frac{3}{2}$, $L = 0$). An example is yttrium iron garnet $Y_3$Fe$_5$O$_{12}$, known as YIG. Here Y$^{3+}$ is diamagnetic.

The net magnetization of YIG is due to the resultant of two oppositely magnetized lattices of Fe$^{3+}$ ions. At absolute zero each ferric ion contributes $\pm 5\mu_B$ to the magnetization, but in each formula unit the three Fe$^{3+}$ ions on sites denoted as $d$ sites are magnetized in one sense and the two Fe$^{3+}$ ions on $a$ sites are magnetized in the opposite sense, giving a resultant of $5\mu_B$ per formula unit in good agreement with the measurements of Geller et al.

The mean field at an $a$ site due to the ions on the $d$ sites is $B_a = -(1.5 \times 10^4)M_d$. The observed Curie temperature 559 K of YIG is due to the $a$-$d$ interaction. The only magnetic ions in YIG are the ferric ions. Because these are in an $L = 0$ state with a spherical charge distribution their interaction with lattice deformations and phonons is weak. As a result YIG is characterized by very narrow linewidths in ferromagnetic resonance experiments.

In the rare-earth iron garnets the ions M$^{3+}$ are paramagnetic trivalent rare-earth ions. Magnetization curves are given in Fig. 19. The rare-earth ions occupy sites labeled $c$; the magnetization $M_c$ of the ions on the $c$ lattice is opposite to the net magnetization of the ferric ions on the $a + d$ sites. At low temperatures the combined moments of the three rare-earth ions in a formula
Figure 19  Experimental values of the saturation magnetization versus temperature of various iron garnets, after R. Fauthenet. The formula unit is $M_3Fe_5O_{12}$, where $M$ is a trivalent metal ion. The temperature at which the magnetization crosses zero is called the compensation temperature; here the magnetization of the $M$ sublattice is equal and opposite to the net magnetization of the ferric ion sublattices. Per formula unit there are 3 Fe$^{3+}$ ions on tetrahedral sites $d$; 2 Fe$^{3+}$ ions on octahedral sites $c$; and 3 $M^{3+}$ ions on sites denoted by $c$. The ferric ions contribute $(3 - 2)5\mu_B = 5\mu_B$ per formula unit. The ferric ion coupling is strong and determines the Curie temperature. If the $M^{3+}$ ions are rare-earth ions they are magnetized opposite to the resultant of the Fe$^{3+}$ ions. The $M^{3+}$ contribution drops rapidly with increasing temperature because the $M$-Fe coupling is weak. Measurements on single crystal specimens are reported by Geller et al., Phys. Rev. 137, 1034 (1965).

unit may dominate the net moment of the Fe$^{3+}$ ions, but because of the weak $c$-$a$ and $c$-$d$ coupling the rare-earth lattice loses its magnetization rapidly with increasing temperature. The total moment can pass through zero and then increase again as the Fe$^{3+}$ moment starts to be dominant.

**ANTIFERROMAGNETIC ORDER**

A classical example of magnetic structure determination by neutrons is shown in Fig. 20 for MnO, which has the NaCl structure. At 80 K there are extra neutron reflections not present at 293 K. The reflections at 80 K may be classified in terms of a cubic unit cell of lattice constant 8.85 Å. At 293 K the reflections correspond to an fcc unit cell of lattice constant 4.43 Å.
Figure 20 Neutron diffraction patterns for MnO below and above the spin-ordering temperature of 120 K, after C. G. Shull, W. A. Strauser, and E. O. Wollan. The reflection indices are based on an 8.85 Å cell at 80 K and on a 4.43 Å cell at 293 K. At the higher temperature the Mn$^{2+}$ ions are still magnetic, but they are no longer ordered.

But the lattice constant determined by x-ray reflection is 4.43 Å at both temperatures, 80 K and 293 K. We conclude that the chemical unit cell has the 4.43 Å lattice parameter, but that at 80 K the electronic magnetic moments of the Mn$^{2+}$ ions are ordered in some nonferromagnetic arrangement. If the ordering were ferromagnetic, the chemical and magnetic cells would give the same reflections.

The spin arrangement shown in Fig. 21 is consistent with the neutron diffraction results and with magnetic measurements. The spins in a single [111] plane are parallel, but adjacent [111] planes are antiparallel. Thus MnO is an antiferromagnet, as in Fig. 22.

In an antiferromagnet the spins are ordered in an antiparallel arrangement with zero net moment at temperatures below the ordering or Néel temperature (Table 3). The susceptibility of an antiferromagnet is not infinite at $T = T_N$, but has a weak cusp, as in Fig. 23.

An antiferromagnet is a special case of a ferrimagnet for which both sublattices $A$ and $B$ have equal saturation magnetizations. Thus $C_A = C_B$ in (37), and
Figure 21  Ordered arrangements of spins of the Mn$^{2+}$ ions in manganese oxide, MnO, as determined by neutron diffraction. The O$^{2-}$ ions are not shown.

Figure 22  Spin ordering in ferromagnets ($J > 0$) and antiferromagnets ($J < 0$).

Figure 23  Temperature dependence of the magnetic susceptibility in paramagnets, ferromagnets, and antiferromagnets. Below the Néel temperature of an antiferromagnet the spins have antiparallel orientations, the susceptibility attains its maximum value at $T_N$, where there is a well-defined kink in the curve of $\chi$ versus $T$. The transition is also marked by peaks in the heat capacity and the thermal expansion coefficient.
the Néel temperature in the mean field approximation is given by
\[ T_N = \mu C , \]  
where \( C \) refers to a single sublattice. The susceptibility in the paramagnetic region \( T > T_N \) is obtained from (37):
\[ \chi = \frac{2CT - 2\mu C^2}{T^2 - (\mu C)^2} = \frac{2C}{T + \mu C} = \frac{2C}{T + T_N} . \]

The experimental results at \( T > T_N \) are of the form
\[(\text{CGS})\]
\[ \chi = \frac{2C}{T + \theta} . \]

Experimental values of \( \theta/T_N \) listed in Table 3 often differ substantially from the value unity expected from (39). Values of \( \theta/T_N \) of the observed magnitude may be obtained when next-nearest-neighbor interactions are provided for, and when possible sublattice arrangements are considered. If a mean field constant \(-\epsilon\) is introduced to describe interactions within a sublattice, then \( \theta/T_N = (\mu + \epsilon)/(\mu - \epsilon) \).

**Susceptibility Below the Néel Temperature**

There are two situations: with the applied magnetic field perpendicular to the axis of the spins; and with the field parallel to the axis of the spins. At and above the Néel temperature the susceptibility is nearly independent of the direction of the field relative to the spin axis.
For $B_a$ perpendicular to the axis of the spins we can calculate the susceptibility by elementary considerations. The energy density in the presence of the field is, with $M = |M_A| = |M_B|,$

$$U = \mu M_A M_B - B_a(M_A + M_B) \equiv -\mu M^2(1 - \frac{1}{2}(2\varphi)^2) - 2B_aM\varphi,$$  \hspace{1cm} (41)

where $2\varphi$ is the angle the spins make with each other (Fig. 24a). The energy is a minimum when

$$dU/d\varphi = 0 = 4\mu M^2\varphi - 2B_aM; \quad \varphi = B_a/2\mu M,$$  \hspace{1cm} (42)

so that

$$\chi \perp = 2M\varphi/B_a = 1/\mu.$$  \hspace{1cm} (43)

In the parallel orientation (Fig. 24b) the magnetic energy is not changed if the spin systems $A$ and $B$ make equal angles with the field. Thus the susceptibility at $T = 0$ K is zero:

$$\chi \parallel (0) = 0.$$  \hspace{1cm} (44)

The parallel susceptibility increases smoothly with temperature up to $T_N.$ Measurements on MnF$_2$ are shown in Fig. 25. In very strong fields the spin systems will turn discontinuously from the parallel orientation to the perpendicular orientation where the energy is lower.

**Antiferromagnetic Magnons**

We obtain the dispersion relation of magnons in a one-dimensional antiferromagnet by making the appropriate substitutions in the treatment (16)–(22) of the ferromagnetic line. Let spins with even indices $2p$ compose sublattice $A,$ that with spins up ($S^z = S$); and let spins with odd indices $2p + 1$ compose sublattice $B,$ that with spins down ($S^z = -S$).

We consider only nearest-neighbor interactions, with $J$ negative. Then (16) written for $A$ becomes, with a careful look at (17),

$$dS^x_{2p}/dt = (2JS/\hbar)(-2S^y_{2p} - S^y_{2p-1} - S^y_{2p+1});$$  \hspace{1cm} (45a)

$$dS^y_{2p}/dt = -(2JS/\hbar)(-2S^z_{2p} - S^z_{2p-1} - S^z_{2p+1}).$$  \hspace{1cm} (45b)

The corresponding equations for a spin on $B$ are

$$dS^y_{2p+1}/dt = (2JS/\hbar)(2S^y_{2p+1} + S^y_{2p} + S^y_{2p+2});$$  \hspace{1cm} (46a)

$$dS^z_{2p+1}/dt = -(2JS/\hbar)(2S^x_{2p+1} + S^x_{2p} + S^x_{2p+2}).$$  \hspace{1cm} (46b)

We form $S^+ = S^x + is^y; \text{ then}$

$$dS^x_{2p}/dt = (2iJS/\hbar)(2S^y_{2p} + S^y_{2p-1} + S^y_{2p+1});$$  \hspace{1cm} (47)

$$dS^y_{2p+1}/dt = -(2iJS/\hbar)(2S^x_{2p+1} + S^x_{2p} + S^x_{2p+2}).$$  \hspace{1cm} (48)
Figure 24 Calculation of (a) perpendicular and (b) parallel susceptibilities at 0 K, in the mean field approximation.

Figure 25 Magnetic susceptibility of manganese fluoride, MnF$_2$, parallel and perpendicular to the tetragonal axis. (After S. Foner.)

We look for solutions of the form

$$S_{2p}^+ = u \exp[i2pka - iwt]; \quad S_{2p+1}^+ = v \exp[i(2p + 1)ka - iwt],$$

so that (47) and (48) become, with $\omega_{ex} = -4JS/\hbar = 4J/|S|\hbar$,

$$\omega u = \frac{1}{2} \omega_{ex} (2u + ve^{-ik_a} + ve^{ik_a}) ;$$

$$-\omega v = \frac{1}{2} \omega_{ex} (2v + ue^{-ik_a} + ue^{ik_a}).$$

Equations (50) have a solution if

$$\begin{vmatrix} \omega_{ex} - \omega & \omega_{ex}\cos ka \\ \omega_{ex}\cos ka & \omega_{ex} + \omega \end{vmatrix} = 0 ;$$

thus

$$\omega^2 = \omega_{ex}^2 (1 - \cos^2 ka) ; \quad \omega = \omega_{ex} |\sin ka| .$$

The dispersion relation for magnons in an antiferromagnet is quite different from (22) for magnons in a ferromagnet. For $ka \ll 1$ we see that (52) is linear in $k$: $\omega \approx \omega_{ex} |ka|$. The magnon spectrum of RbMnF$_3$ is shown in Fig. 26, as determined by inelastic neutron scattering experiments. There is a large region in which the magnon frequency is linear in the wavevector.

A physical discussion of the difference between the dispersion relations for ferromagnetic and antiferromagnetic magnons is given by F. Keffer, H. Kaplan, and Y. Yafet, Am. J. Phys. 21, 250 (1953).
Well-resolved magnons have been observed in MnF$_2$ at specimen temperatures up to 0.93 of the Néel temperature. Thus even at high temperatures the magnon approximation is useful. Further details concerning antiferromagnetic magnons are given in *QTS*, Chapter 4.

**FERROMAGNETIC DOMAINS**

At temperatures well below the Curie point the electronic magnetic moments of a ferromagnet are essentially parallel when regarded on a microscopic scale. Yet, looking at a specimen as a whole, the magnetic moment may be very much less than the saturation moment, and the application of an external magnetic field may be required to saturate the specimen. The behavior observed in polycrystalline specimens is similar to that in single crystals.

Actual specimens are composed of small regions called domains, within each of which the local magnetization is saturated. The directions of magnetization of different domains need not be parallel. An arrangement of domains with approximately zero resultant magnetic moment is shown in Fig. 27. Domains form also in antiferromagnetics, ferroelectrics, antiferroelectrics, ferroelastics, superconductors, and sometimes in metals under conditions of a strong de Haas-van Alphen effect. The increase in the gross magnetic moment of a ferromagnetic specimen in an applied magnetic field takes place by two independent processes:

- In weak applied fields the volume of domains (Fig. 28) favorably oriented with respect to the field increases at the expense of unfavorably oriented domains;
- In strong applied fields the domain magnetization rotates toward the direction of the field.
Figure 27  Ferromagnetic domain pattern on a single crystal platelet of nickel. The domain boundaries are made visible by the Bitter magnetic powder pattern technique. The direction of magnetization within a domain is determined by observing growth or contraction of the domain in a magnetic field. (After R. W. De Blois.)

Figure 28  Representative magnetization curve, showing the dominant magnetization processes in the different regions of the curve.
Technical terms defined by the hysteresis loop are shown in Fig. 29. The coercivity is usually defined as the reverse field $H_c$ that reduces the induction $B$ to zero, starting from saturation. In high coercivity materials the coercivity $H_{ci}$ is defined as the reverse field that reduces the magnetization $M$ to zero.

The domain structure of ferromagnetic materials enters their applications. In a transformer core we want low losses in taking the core around a cycle: this means that we want low coercivity, which goes with a high permeability. By making pure, homogeneous, and well-oriented material we facilitate domain boundary displacement and thereby attain high permeability. Relative permeability values as high as $4 \times 10^6$ have been reported.

In a permanent magnet we want a high coercivity, which we may achieve by suppressing boundary displacement, which is best done by eliminating domain boundaries. This is accomplished in very fine particles or crystallites that consist of a single domain, without boundaries. By controlled precipitation of a second metallurgical phase the specimen may be heterogeneous on a very fine scale. Characteristically there is a critical radius below which single domain attributes are found.

**Anisotropy Energy**

There is an energy in a ferromagnetic crystal which directs the magnetization along certain crystallographic axes called directions of easy magnetization.
This energy is called the magnetocrystalline or anisotropy energy. It does not come about from the pure isotropic exchange interaction considered thus far.

Cobalt is a hexagonal crystal. The hexagonal axis is the direction of easy magnetization at room temperature, as shown in Fig. 30. One origin of the anisotropy energy is illustrated by Fig. 31. The magnetization of the crystal sees the crystal lattice through orbital overlap of the electrons: the spin interacts with the orbital motion by means of the spin-orbit coupling.

In cobalt the anisotropy energy density is given by

$$U_K = K_1 \sin^2 \theta + K'_2 \sin^4 \theta,$$

where $\theta$ is the angle the magnetization makes with the hexagonal axis. At room temperature $K_1 = 4.1 \times 10^6 \text{ erg/cm}^3$; $K'_2 = 1.0 \times 10^6 \text{ erg/cm}^3$.

Iron is a cubic crystal, and the cube edges are the directions of easy magnetization. To represent the anisotropy energy of iron magnetized in an arbitrary direction with direction cosines $\alpha_1, \alpha_2, \alpha_3$ referred to the cube edges, we
are guided by cubic symmetry. The expression for the anisotropy energy must be an even power of each \( \alpha_i \), provided opposite ends of a crystal axis are equivalent magnetically, and it must be invariant under interchanges of the \( \alpha_i \) among themselves. The lowest order combination satisfying the symmetry requirements is \( \alpha_1^2 + \alpha_2^2 + \alpha_3^2 \), but this is identically equal to unity and does not describe anisotropy effects. The next combination is of the fourth degree: \( \alpha_1^2 \alpha_2^2 + \alpha_1^2 \alpha_3^2 + \alpha_2^2 \alpha_3^2 \), and then of the sixth degree: \( \alpha_1^2 \alpha_2^2 \alpha_3^2 \). Thus

\[
U_K = K_1(\alpha_1^2 \alpha_2^2 + \alpha_2^2 \alpha_3^2 + \alpha_3^2 \alpha_1^2) + K_2\alpha_1^2 \alpha_2^2 \alpha_3^2 .
\]  
(54)

At room temperature \( K_1 = 4.2 \times 10^5 \) erg/cm\(^3\) and \( K_2 = 1.5 \times 10^5 \) erg/cm\(^3\).

**Transition Region Between Domains**

A Bloch wall in a crystal is the transition layer that separates adjacent regions (domains) magnetized in different directions. The entire change in spin direction between domains does not occur in one discontinuous jump across a single atomic plane, but takes place in a gradual way over many atomic planes (Fig. 32). The exchange energy is lower when the change is distributed over many spins.

This behavior may be understood by interpreting the Heisenberg equation (6) classically. We replace \( \cos \varphi \) by \( 1 - \frac{1}{2}\varphi^2 \); then \( w_{\text{ex}} = JS^2\varphi^2 \) is the exchange energy between two spins making a small angle \( \varphi \) with each other. Here \( J \) is the exchange integral and \( S \) is the spin quantum number; \( w_{\text{ex}} \) is referred to the energy for parallel spins.

If a total change of \( \pi \) occurs in \( N \) equal steps, the angle between neighboring spins is \( \pi/N \), and the exchange energy per pair of neighboring atoms is \( w_{\text{ex}} = JS^2(\pi/N)^2 \). The total exchange energy of a line of \( N + 1 \) atoms is

\[
Nw_{\text{ex}} = JS^2\pi^2/N .
\]  
(55)

The wall would thicken without limit were it not for the anisotropy energy, which acts to limit the width of the transition layer. The spins contained within the wall are largely directed away from the axes of easy magnetization, so there is an anisotropy energy associated with the wall, roughly proportional to the wall thickness.

Consider a wall parallel to the cube face of a simple cubic lattice and separating domains magnetized in opposite directions. We wish to determine the number \( N \) of atomic planes contained within the wall. The energy per unit area of wall is the sum of contributions from exchange and anisotropy energies: \( \sigma_w = \sigma_{\text{ex}} + \sigma_{\text{anis}} \).

The exchange energy is given approximately by (55) for each line of atoms normal to the plane of the wall. There are \( 1/a^2 \) such lines per unit area, where \( a \) is the lattice constant. Thus \( \sigma_{\text{ex}} = \pi^2JS^2/Na^2 \).
The anisotropy energy is of the order of the anisotropy constant times the thickness $Na$, or $\sigma_{\text{anis}} \approx KNa$; therefore

$$\sigma_w = (\pi^2 JS^2/Na^2) + KNa.$$  \hspace{1cm} (56)

This is a minimum with respect to $N$ when

$$\partial\sigma_w/\partial N = 0 = -(\pi^2 JS^2/N^2a^2) + Ka;$$  \hspace{1cm} (57)

or

$$N = (\pi^2 JS^2/Ka^3)^{1/2}. \hspace{1cm} (58)$$

For order of magnitude, $N = 300$ in iron.

The total wall energy per unit area on our model is

$$\sigma_w = 2\pi(KJS^2/a)^{1/2}; \hspace{1cm} (59)$$

in iron $\sigma_w \approx 1$ erg/cm$^2$. Accurate calculation for a $180^\circ$ wall in a (100) plane gives $\sigma_w = 2(2KJS^2/a)^{1/2}$.

**Solitons**

Bloch walls are an example of a fascinating class of physical and mathematical problems known as solitons. A rough description of a classical soliton is a solitary wave that shows great stability in collision with other solitary waves. Bloch walls in the absence of applied magnetic fields are static solitary waves, but they move uniformly in an applied field and their collisions can be studied,
at least theoretically. It is somewhat surprising that this can be done because
the equations of motion are nonlinear.8

**Origin of Domains**

Landau and Lifshitz showed that domain structure is a natural conse-
quence of the various contributions to the energy—exchange, anisotropy, and
magnetic—of a ferromagnetic body.

Direct evidence of domain structure is furnished by photomicrographs of
domain boundaries obtained by the technique of magnetic powder patterns and
by optical studies using Faraday rotation. The powder pattern method de-
veloped by F. Bitter consists in placing a drop of a colloidal suspension of finely
divided ferromagnetic material, such as magnetite, on the surface of the ferro-
magnetic crystal. The colloid particles in the suspension concentrate strongly
about the boundaries between domains where strong local magnetic fields exist
which attract the magnetic particles. The discovery of transparent ferromag-
netic compounds has encouraged the use also of optical rotation for domain
studies.

We may understand the origin of domains by considering the structures
shown in Fig. 33, each representing a cross section through a ferromagnetic
single crystal. In (a) we have a single domain; as a consequence of the magnetic
“poles” formed on the surfaces of the crystal this configuration will have a high
value of the magnetic energy \((1/8\pi) \int B^2 \, dV\). The magnetic energy density for
the configuration shown will be of the order of \(M_s^2 = 10^6 \text{ erg/cm}^3\); here \(M_s\)
denotes the saturation magnetization, and the units are CGS.

In (b) the magnetic energy is reduced by roughly one-half by dividing the
crystal into two domains magnetized in opposite directions. In (c) with \(N\) do-
 mains the magnetic energy is reduced to approximately \(1/N\) of the magnetic
energy of (a), because of the reduced spatial extension of the field.

In domain arrangements such as (d) and (e) the magnetic energy is zero.
Here the boundaries of the triangular prism domains near the end faces of the
crystal make equal angles (45°) with the magnetization in the rectangular do-
 mains and with the magnetization in the domains of closure. The component of
magnetization normal to the boundary is continuous across the boundary and
there is no magnetic field associated with the magnetization. The flux circuit is
completed within the crystal—thus giving rise to the term domains of closure
for surface domains that complete the flux circuit, as in Fig. 34.

Domain structures are often more complicated than our simple examples,
but domain structure always has its origin in the possibility of lowering the
energy of a system by going from a saturated configuration with high magnetic
energy to a domain configuration with a lower energy.

---

8Recent monographs on soliton theory include R. K. Bullough, *Solitons*, Springer, 1960, G. L.
Lamb, *Elements of soliton theory*, Wiley, 1960. For a short introductory article, see R. K. Bul-
Coercivity and Hysteresis

The coercivity is the magnetic field $H_c$ required to reduce the magnetization or the induction $B$ to zero (Fig. 29). The value of the coercivity ranges over seven orders of magnitude; it is the most sensitive property of ferromagnetic materials which is subject to control. The coercivity may vary from 600 G in a loudspeaker permanent magnet (Alnico V) and 10,000 G in a special high stability magnet (SmCo$_5$) to 0.5 G in a commercial power transformer (Fe-Si 4 wt. pet.) and 0.002 G in a pulse transformer (Supermalloy). Low coercivity is desired in a transformer, for this means low hysteresis loss per cycle of operation. Materials with low coercivity are called soft; those with high coercivity are called hard, although there is not necessarily a 1:1 relationship of magnetic hardness with mechanical hardness (Chapter 20).

The coercivity decreases as the impurity content decreases and also as internal strains are removed by annealing (slow cooling). Amorphous ferromagnetic alloys may have low coercivity, low hysteresis losses, and high permeability, as treated in Chapter 17. Alloys that contain a precipitated phase may have a high coercivity, as in Alnico V (Fig. 35).
Soft magnetic materials are used to concentrate and shape magnetic flux, as in motors, generators, transformers, and sensors. Classes of soft materials include electrical steels (usually alloyed with several percent of silicon to increase electrical resistivity and to decrease anisotropy); various alloys of Fe-Co-Mn, starting with permalloys of composition near Ni$_{78}$Fe$_{22}$, which have near zero anisotropy energy and near zero magnetostriction; NiZn and MnZn ferrites; and metallic glasses produced by rapid solidification methods (Chapter 17). A commercial metallic glass (METGLAS 2605S-2) with composition Fe$_{70}$B$_{13}$Si$_{9}$ has a hysteresis loss per cycle much lower than the best grain oriented silicon steel.

The high coercivity of materials composed of very small grains or fine powders is well understood. A sufficiently small particle, with diameter less than $10^{-5}$ or $10^{-6}$ cm, is always magnetized to saturation as a single domain because the formation of a flux-closure configuration is energetically unfavorable. In a single domain particle it is not possible for magnetization reversal to take place by means of the process of boundary displacement, which usually requires relatively weak fields. Instead the magnetization of the particle must rotate as a whole, a process that may require large fields depending on the anisotropy energy of the material and the anisotropy of the shape of the particle.

The coercivity of fine iron particles is expected theoretically to be about 500 gauss on the basis of rotation opposed by the crystalline anisotropy energy, and this is of the order of the observed value. Higher coercivities have been reported for elongated iron particles, the rotation here being opposed by the shape anisotropy of the demagnetization energy.
Rare earth metals in alloys with Mn, Fe, Co, and Ni have very large crystal anisotropies and correspondingly large coercivities, of the order of $2K/M$. These alloys are exceptionally good permanent magnets. For example, the hexagonal compound SmCo$_5$ has an anisotropy energy $1.1 \times 10^8$ erg cm$^{-3}$, equivalent to a coercivity $2K/M$ of 290 kG (29 T). Magnets of Nd$_2$Fe$_{14}$B have energy products as high as 50 MGOe, exceeding all other commercially available magnets.

Friedberg and Paul$^9$ developed a theory of the pinning of domain walls in single phase polycrystalline materials that gives a good account of experimental results over a wide range of materials. In particular their theory gives a good account of the high coercivity of SmCo$_5$, predicting $H_c = 9$ kG where 10 kG is observed.

**SINGLE DOMAIN PARTICLES**

The dominant industrial and commercial applications of ferromagnetism are in magnetic recording devices, where the magnetic material is in the form of single domain particles or regions. The total value of the production of magnetic devices for recording may be comparable with the total value of semiconductor device production and greatly exceeds the value of superconducting device production, the latter being held back by low critical temperatures, as compared with magnetic Curie temperatures. The magnetic recording devices or memories typically are in the form of hard disks in computers and tape cassettes in video and audio recorders.

An ideal single domain particle is a fine particle, usually elongated, that has its magnetic moment directed toward one end or the other of the particle. The alternative orientations may be labeled as N or S; + or −; in digital recording, as 0 or 1. To have digital properties a ferromagnetic particle should be fine enough, typically 10–100 nm, so that only one domain is within the particle.$^{10}$ If the fine particle is elongated (acicular) or has uniaxial crystal symmetry, only two values of the magnetic moment of the single domain are permitted, which is what one wants for digital properties. The first successful recording material was acicular $\tau$-Fe$_2$O$_3$ with length-to-width ratio of about 5:1, coercivity near 200 Oe and a length $<$1μm; chromium dioxide CrO$_2$ is the basis of a better material,$^{11}$ in a form highly acicular (20:1) with coercivity near 500 Oe.

Effective elongation can be attained with spheres by making a chain, like a string of beads. An ensemble of such chains or of elongated single domain particles is said to exhibit superparamagnetism if the magnetic moment of a

---

unit is constant. If $\mu$ is the magnetic moment in a magnetic field $B$, then the net magnetization of the ensemble will follow the Curie-Brillouin-Langevin law of (14.19), (14.20) if the particles are embedded in a liquid so that they are each free to rotate as a whole. If the particles are frozen in a solid, there will be a remanent magnetization (Figure 29) after removal of an applied field.

**Geomagnetism and Biomagnetism**

Single domain ferromagnetic properties are of special geological interest in sedimentary rocks because the rocks through their remanent magnetization carry a memory of the direction of the earth's magnetic field at the time that they were laid down, and thus of the geographical location of the rocks at that epoch. The magnetic record is perhaps the most important basis of the theory of the drift of continents. Annually, layers of sediment are deposited in stream beds, layers that may bear some magnetic particles in single domain form. This record persists over at least 500 million years of geological time and can tell us where on the surface of the earth the deposit was laid down at a given time. Lava flows also record magnetic field directions.

The change in magnetization from layer to layer gives a superb historical record of the drift of the continental plates on the earth's surface. The paleomagnetic record is one basis of the branch of geology called plate tectonics. The original interpretation of the record was made more difficult, or more exciting, by the associated discovery (Brunhes, 1906) that the magnetic field of the earth itself can show reversals in direction, an effect contained within the standard dynamo theory of the earth's magnetism. Reversals have taken place once every $1 \times 10^4$ to $25 \times 10^6$ years. When a reversal occurs, it is relatively sudden.

Fine single domain particles, often of magnetite Fe$_3$O$_4$, are even of importance in biology. A direction-seeking effect known as magnetotaxis$^{12}$ often controls, possibly sometimes along with an astronomical guide system, the motion of bacteria, the migration of birds, and the movements of homing pigeons and bees. The effect is due to the interaction of a single domain particle (or cluster of such particles, Figure 36) in the organism with the external magnetic field of the earth.

**Magnetic Force Microscopy**

The success of the scanning tunneling microscope (STM) stimulated the development of related scanning probe devices, of which the scanning magnetic force microscope is one of the most effective. A sharp tip of a magnetic material, such as nickel, is mounted on a cantilever lever (Figure 37). Ideally,

$^{12}$Taxis in biology means the movement of a cell or organism toward or away from some external field of force.
Figure 36 Thin section of a cell of a magnetotactic bacterium showing a chain of 50 nm particles of Fe₃O₄. Drawing by Marta Puebla from a photograph by R. B. Frankel and others.

Figure 37 Basic concept of magnetic force microscopy. A magnetic tip attached to a flexible cantilever is used to detect the magnetic field produced by the regions of alternating magnetization in the plane of the sample. (After Gruetter, Mamin, and Bugar, 1992.)

but not yet, the tip is a single domain particle. Forces from the magnetic sample act on the tip and cause a change, such as a deflection, in the cantilever status, and an image is formed by scanning the sample relative to the tip. The magnetic force microscope (MFM) is the only magnetic imaging technique that can provide high resolution (10–100 nm) with little surface preparation. One can, for example, observe and image the magnetic flux that exits from the surface at the intersection of a Bloch wall with the surface (Figure 32). An important application is to the study of magnetic recording media—Figure 38 shows the magnetic signal from a test pattern of 2 μm bits magnetized in the plane of a Co-alloy disk; the parallel component of the field sensed by the sensor tip is what the photo shows.
MAGNETIC BUBBLE DOMAINS

Magnetic recording discs and tapes offer the lowest cost form of mass data storage or memory. Their drawback is that in order to write or read data they require mechanical motion of the head with respect to the magnetic material. The mechanical motion involves an electrical power drive, slow access times, and periodic maintenance.

Electrical methods of entry and read-out using semiconductor memories can avoid these difficulties. However, semiconductor memories are not truly permanent—they lose their data if all power is removed.

Magnetic bubble domains in a ferrite or garnet film can be used to record data, and a sequence of bubbles can be shifted by electromagnetic signals. Bubble memories are nonvolatile, which means that they retain data without need for electrical power.

Small cylindrical magnetic domains (bubble domains) can be stabilized in a uniaxial magnetic material by a bias magnetic field normal to the plane of the specimen and parallel to the easy axis of magnetization. The formation of bubbles is illustrated in Fig. 39. Magnetic bubbles are used for high density memory storage, for example in portable personal computers.

Below a critical value of the bias field a circular domain as in (b) is unstable with respect to a serpentine domain pattern, as in (a). At still higher fields the bubble is unstable with respect to a saturated configuration. The energetics are treated by Boeck.\(^\text{13}\)

\(^{13}\)A. H. Boeck, Bell Syst. Tech. J. 46, 1901 (1967).
SUMMARY
(In CGS Units)

- The susceptibility of a ferromagnet above the Curie temperature has the form \( \chi = C/(T - T_c) \) in the mean field approximation.

- In the mean field approximation the effective magnetic field seen by a magnetic moment in a ferromagnet is \( B_a + \lambda M \), when \( \lambda = T_c/C \) and \( B_a \) is the applied magnetic field.

- The elementary excitations in a ferromagnet are magnons. Their dispersion relation for \( ka \ll 1 \) has the form \( \hbar \omega \approx Jk^2a^2 \) in zero external magnetic field. The thermal excitation of magnons leads at low temperatures to a heat capacity and to a fractional magnetization change both proportional to \( T^{3/2} \).

- In an antiferromagnet two spin lattices are equal, but antiparallel. In a ferromagnet two lattices are antiparallel, but the magnetic moment of one is larger than the magnetic moment of the other.

- In an antiferromagnet the susceptibility above the Néel temperature has the form \( \chi = 2C/(T + \theta) \).

- The magnon dispersion relation in an antiferromagnet has the form \( \hbar \omega \approx Jka \). The thermal excitation of magnons leads at low temperatures to a term in \( T^3 \) in the heat capacity, in addition to the phonon term in \( T^3 \).

- A Bloch wall separates domains magnetized in different directions. The thickness of a wall is \( \approx (J/Na^2)^{1/2} \) lattice constants, and the energy per unit area is \( \approx (KJ/\alpha)^{1/2} \), where \( K \) is the anisotropy energy density.

Problems

1. Magnon dispersion relation. Derive the magnon dispersion relation (24) for a spin \( S \) on a simple cubic lattice, \( z = 6 \). Hint: Show first that (18a) is replaced by

\[
\frac{dS_p^\gamma}{dt} = (2JS/\hbar)(6S_p^\gamma - \sum_\delta S_{p+\delta}^\gamma),
\]
where the central atom is at \( \rho \) and the six nearest neighbors are connected to it by six vectors \( \delta \). Look for solutions of the equations for \( dS_\rho^\nu /dt \) and \( dS_\nu^\rho /dt \) of the form \( \exp(ik \cdot \rho - i\omega t) \).

2. Heat capacity of magnons. Use the approximate magnon dispersion relation \( \omega = AK^2 \) to find the leading term in the heat capacity of a three-dimensional ferromagnet at low temperatures \( k_B T \ll J \). The result is \( 0.113 \, k_B(k_B T/hA)^{3/2} \), per unit volume. The zeta function that enters the result may be estimated numerically, it is tabulated in Jahnke-Emde.

3. Néel temperature. Taking the effective fields on the two-sublattice model of an antiferromagnetic as

\[
B_A = B_a - \mu M_B - \epsilon M_A ; \quad B_B = B_a - \mu M_A - \epsilon M_B,
\]

show that

\[
\frac{\theta}{T_N} = \frac{\mu + \epsilon}{\mu - \epsilon}.
\]

4. Magnetoelastic coupling. In a cubic crystal the elastic energy density in terms of the usual strain components \( e_{ij} \) is

\[
U_{el} = \frac{1}{2} C_{11}(e_{xx}^2 + e_{yy}^2 + e_{zz}^2) + \frac{1}{3} C_{12}(e_{xy}^2 + e_{yz}^2 + e_{zx}^2) + C_{12}(e_{xy}e_{yz} + e_{xz}e_{zx} + e_{zx}e_{xy}) \;,
\]

and the leading term in the magnetic anisotropy energy density is, from (54),

\[
U_K = K_1(\alpha_1^2e_{xx} + \alpha_2^2e_{yy} + \alpha_3^2e_{zz} + \alpha_4^2\alpha_5^2).
\]

Coupling between elastic strain and magnetization direction may be taken formally into account by including in the total energy density a term

\[
U_c = B_1(\alpha_1^2e_{xx} + \alpha_2^2e_{yy} + \alpha_3^2e_{zz}) + B_2(\alpha_1\alpha_2e_{xy} + \alpha_2\alpha_3e_{yz} + \alpha_3\alpha_1e_{zx})
\]

arising from the strain dependence of \( U_K \); here \( B_1 \) and \( B_2 \) are called magnetoelastic coupling constants. Show that the total energy is a minimum when

\[
e_{ii} = \frac{B_1(C_{12} - \alpha_1^2(C_{11} + 2C_{12})]}{[(C_{11} - C_{12})(C_{11} + 2C_{12})]} \; , \quad e_{ij} = -\frac{B_2\alpha_i\alpha_j}{C_{4i}} \quad (i \neq j). \]

This explains the origin of magnetostriction, the change of length on magnetization.

5. Coercive force of a small particle. (a) Consider a small spherical single-domain particle of a uniaxial ferromagnet. Show that the reverse field along the axis required to reverse the magnetization is \( B_a = 2K/M_s \), in CGS units. The coercive force of single-domain particles is observed to be of this magnitude. Take \( U_K = KS^2\theta \) as the anisotropy energy density and \( U_M = -B_aM \cos \theta \) as the interaction energy density with the external field; here \( \theta \) is the angle between \( B_a \) and \( M \). Hint: Expand the energies for small angles about \( \theta = \pi \), and find the value of \( B_a \) for which \( U_K + U_M \) does not have a minimum near \( \theta = \pi \). (b) Show that the magnetic energy of a satu-
Figure 40 A Bloch wall and a Néel wall in a thin film. The magnetization in the Bloch wall is normal to the plane of the film and adds to the wall energy a demagnetization energy $-M_z^2 \delta d$ per unit length of wall, where $\delta$ is the wall thickness and $d$ the film thickness. In the Néel wall the magnetization is parallel to the surface; the addition to the wall energy is negligible when $d \ll \delta$. The addition to the Néel wall energy when $d \gg \delta$ is the subject of Problem 7. (After S. Miedema.)

A: A demagnetized sphere of diameter $d$ is $\approx M_z^2 d^3$. An arrangement with appreciably less magnetic energy has a single wall in an equatorial plane. The domain wall energy will be $\pi \sigma_w d^2/4$, where $\sigma_w$ is the wall energy per unit area. *Estimate* for cobalt the critical radius below which the particles are stable as single domains, taking the value of $J S^2/a$ as for iron.

6. *Saturation magnetization near $T_c$.* Show that in the mean field approximation the saturation magnetization just below the Curie temperature has the dominant temperature dependence $(T_c - T)^{1/2}$. Assume the spin is $\frac{1}{2}$. The result is the same as that for a second-order transition in a ferroelectric crystal, as discussed in Chapter 13. The experimental data for ferromagnets (Table 1) suggest that the exponent is closer to 0.33.

7. *Néel wall.* The direction of magnetization change in a domain wall goes from that of the Bloch wall to that of a Néel wall (Fig. 40) in thin films of material of negligible crystalline anisotropy energy, such as Permalloy. The intercept of the Bloch wall with the surface of the film creates a surface region of high demagnetization energy. The Néel wall avoids this intercept contribution, but at the expense of a demagnetization contribution throughout the volume of the wall. The Néel wall becomes energetically favorable when the film becomes sufficiently thin. Consider, however, the energetics of the Néel wall in bulk material of negligible crystalline anisotropy energy. There is now a demagnetization contribution to the wall energy density. By a qualitative argument similar to (56), show that $\sigma_w \approx (\pi^2 J S^2/Na^2) + (2\pi M_z^2 Na)$. Find $N$ for which $\sigma_w$ is a minimum. Estimate the order of magnitude of $\sigma_w$ for typical values of $J$, $M_z$, and $a$.

**References**

S. V. Vonsovskii, Magnetism, Halsted, 1975, 2 vol.
Journal of Magnetism and Magnetic Materials.
16

Magnetic Resonance

NUCLEAR MAGNETIC RESONANCE  487
   Equations of motion  490

LINE WIDTH  494
   Motional narrowing  494

HYPERFINE SPLITTING  497
   Examples: paramagnetic point defects  499
   Knight shift  500

NUCLEAR QUADRUPOLE RESONANCE  503

FERROMAGNETIC RESONANCE  503
   Shape effects in FMR  504
   Spin wave resonance  506

ANTIFERROMAGNETIC RESONANCE  506

ELECTRON PARAMAGNETIC RESONANCE  510
   Exchange narrowing  510
   Zero-field splitting  510

PRINCIPLE OF MASER ACTION  510
   Three-level maser  512
   Ruby laser  512

SUMMARY  514

PROBLEMS

1. Equivalent electrical circuit  515
2. Rotating coordinate system  515
3. Hyperfine effects on ESR in metals  515
4. FMR in the anisotropy field  515
5. Exchange frequency resonance  516
6. Rf saturation  516

REFERENCES  516

NOTATION: In this chapter the symbols $B_a$ and $B_0$ refer to the applied field, and $B_I$ is the applied field plus the demagnetizing field. In particular we write $B_a = B_0 z$. For CCS readers it may be simpler to read $H$ for $B$ whenever it occurs in this chapter.
Figure 1  Electron spin resonance absorption in MnSO$_4$ at 298 K at 2.75 GHz, after Zavoisky.
In this chapter we discuss dynamical magnetic effects associated with the spin angular momentum of nuclei and of electrons. The principal phenomena are often identified in the literature by their initial letters, such as

NMR: nuclear magnetic resonance
NQR: nuclear quadrupole resonance
EPR or ESR: electron paramagnetic resonance (Fig. 1)
FMR: ferromagnetic resonance
SWR: spin wave resonance (ferromagnetic films)
AFMR: antiferromagnetic resonance
CESR: conduction electron spin resonance

The information that can be obtained about solids by resonance studies may be categorized:

- Electronic structure of single defects, as revealed by the fine structure of the absorption.
- Motion of the spin or of the surroundings, as revealed by changes in the line width.
- Internal magnetic fields sampled by the spin, as revealed by the position of the resonance line (chemical shift; Knight shift).
- Collective spin excitations.

It is best to discuss NMR as a basis for a brief account of the other resonance experiments. A great impact of NMR has been in organic chemistry and biochemistry, where NMR provides a powerful tool for the identification and the structure determination of complex molecules. This success is due to the extremely high resolution attainable in diamagnetic liquids. A major medical application of NMR is magnetic resonance imaging (MRI), which allows the resolution in 3D of abnormal growths, configurations, and reactions in the whole body.

NUCLEAR MAGNETIC RESONANCE

We consider a nucleus that possesses a magnetic moment $\mu$ and an angular momentum $\hbar I$. The two quantities are parallel, and we may write

$$\mu = \gamma \hbar I ;$$

(1)

the magnetogyric ratio $\gamma$ is constant. By convention $I$ denotes the nuclear angular momentum measured in units of $\hbar$. 
The energy of interaction with the applied magnetic field is
\[ U = -\mu \cdot B_0 ; \]  
if \( B_0 = B_0 \hat{z} \), then
\[ U = -\mu_z B_0 = -\gamma \hbar B_0 I_z . \]  
The allowed values of \( I_z \) are \( m_I = I, I - 1, \ldots, -I \), and \( U = -m_I \gamma \hbar B_0 \).

In a magnetic field a nucleus with \( I = \frac{1}{2} \) has two energy levels corresponding to \( m_I = \pm \frac{1}{2} \), as in Fig. 2. If \( \hbar \omega_0 \) denotes the energy difference between the two levels, then \( \hbar \omega_0 = \gamma \hbar B_0 \) or
\[ \omega_0 = \gamma B_0 . \]  
This is the fundamental condition for magnetic resonance absorption.

For the proton\(^1\) \( \gamma = 2.675 \times 10^4 \text{ s}^{-1} \text{ gauss}^{-1} = 2.675 \times 10^8 \text{ s}^{-1} \text{ tesla}^{-1} \), so that
\[ \nu(\text{MHz}) = 4.258 B_0(\text{kilogauss}) = 42.58 B_0(\text{tesla}) , \]  
where \( \nu \) is the frequency. One tesla is precisely \( 10^4 \) gauss. Magnetic data for selected nuclei are given in Table 1. For the electron spin,
\[ \nu(\text{GHz}) = 2.80 B_0(\text{kilogauss}) = 28.0 B_0(\text{tesla}) . \]  

\(^1\) The magnetic moment \( \mu_p \) of the proton is \( 1.4106 \times 10^{-23} \text{ erg G}^{-1} \) or \( 1.4106 \times 10^{-26} \text{ J T}^{-1} \), and \( \gamma = 2\mu_p / \hbar \). The nuclear magneton \( \mu_n \) is defined as \( e\hbar / 2M_p c \) and is equal to \( 5.0509 \times 10^{-24} \text{ erg G}^{-1} \) or \( 5.0509 \times 10^{-27} \text{ J T}^{-1} \); thus \( \mu_p = 2.793 \) nuclear magnetons.
Table 1  Nuclear magnetic resonance data

For every element the most abundant magnetic isotope is shown. After Varian Associates NMR Table.

<table>
<thead>
<tr>
<th>Element</th>
<th>Isotope</th>
<th>Spin</th>
<th>Natural Abundance</th>
<th>Nuclear Magnetic Moment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li⁷</td>
<td>1/2</td>
<td>3/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Na²³</td>
<td>3/2</td>
<td>5/2</td>
<td>100.</td>
<td>2.216</td>
</tr>
<tr>
<td>K⁹⁹</td>
<td>3/2</td>
<td>7/2</td>
<td>93.08</td>
<td>1.391</td>
</tr>
<tr>
<td>Rb⁸⁵</td>
<td>5/2</td>
<td>3/2</td>
<td>72.8</td>
<td>1.438</td>
</tr>
<tr>
<td>Cs¹³³</td>
<td>7/2</td>
<td>3/2</td>
<td>100.</td>
<td>1.166</td>
</tr>
<tr>
<td>Fr</td>
<td>1/2</td>
<td>5/2</td>
<td>99.98</td>
<td>2.792</td>
</tr>
<tr>
<td>Ce¹⁴¹⁵</td>
<td>7/2</td>
<td>5/2</td>
<td>100.</td>
<td>1.166</td>
</tr>
<tr>
<td>Th</td>
<td>1/2</td>
<td>5/2</td>
<td>99.98</td>
<td>2.792</td>
</tr>
</tbody>
</table>

Most abundant isotope with nonzero nuclear spin

Nuclear spin, in units of $\hbar$

Natural abundance of isotope, in percent

Nuclear magnetic moment, in units of $e\hbar/2M_p$c

<table>
<thead>
<tr>
<th>Element</th>
<th>Isotope</th>
<th>Spin</th>
<th>Natural Abundance</th>
<th>Nuclear Magnetic Moment</th>
</tr>
</thead>
<tbody>
<tr>
<td>B¹¹</td>
<td>3/2</td>
<td>1/2</td>
<td>81.17</td>
<td>0.072</td>
</tr>
<tr>
<td>C¹³</td>
<td>3/2</td>
<td>1/2</td>
<td>99.64</td>
<td>0.040</td>
</tr>
<tr>
<td>N¹⁴</td>
<td>5/2</td>
<td>1/2</td>
<td>100.</td>
<td>0.257</td>
</tr>
<tr>
<td>O¹⁷</td>
<td>1/2</td>
<td>1/2</td>
<td>-1.893</td>
<td>2.627</td>
</tr>
<tr>
<td>F¹⁹</td>
<td>1/2</td>
<td>3/2</td>
<td>-1.217</td>
<td>-0.662</td>
</tr>
<tr>
<td>Al²⁷</td>
<td>5/2</td>
<td>1/2</td>
<td>50.57</td>
<td>11.55</td>
</tr>
<tr>
<td>Si²⁹</td>
<td>3/2</td>
<td>1/2</td>
<td>70.4</td>
<td>0.811</td>
</tr>
<tr>
<td>P³¹</td>
<td>3/2</td>
<td>1/2</td>
<td>60.7</td>
<td>0.811</td>
</tr>
<tr>
<td>S³³</td>
<td>3/2</td>
<td>1/2</td>
<td>75.4</td>
<td>0.811</td>
</tr>
<tr>
<td>Cl³⁵</td>
<td>3/2</td>
<td>1/2</td>
<td>50.57</td>
<td>11.55</td>
</tr>
<tr>
<td>Ar</td>
<td>3/2</td>
<td>1/2</td>
<td>75.4</td>
<td>0.811</td>
</tr>
<tr>
<td>K⁹⁹</td>
<td>3/2</td>
<td>7/2</td>
<td>93.08</td>
<td>1.391</td>
</tr>
<tr>
<td>Ca⁶⁵</td>
<td>7/2</td>
<td>5/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Sc⁶⁵</td>
<td>7/2</td>
<td>5/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Ti⁴⁷</td>
<td>7/2</td>
<td>5/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>V⁵¹</td>
<td>7/2</td>
<td>5/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Cr⁵³</td>
<td>5/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Mn⁵⁵</td>
<td>7/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Fe⁵⁷</td>
<td>5/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Co⁵⁹</td>
<td>7/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Ni⁶¹</td>
<td>5/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Cu⁶³</td>
<td>7/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Zn⁶⁷</td>
<td>3/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Ga⁹⁹</td>
<td>9/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Ge⁷³</td>
<td>9/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>As⁷⁵</td>
<td>7/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Se⁷⁷</td>
<td>7/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Br⁸⁶</td>
<td>3/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Kr⁸³</td>
<td>3/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Rb⁸⁵</td>
<td>9/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Sr⁸⁷</td>
<td>9/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Y⁸⁹</td>
<td>5/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Zr⁹¹</td>
<td>9/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Nb⁹³</td>
<td>5/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Mo⁹⁵</td>
<td>5/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Tc</td>
<td>5/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Ru¹⁰¹</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Rh¹⁰³</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Pd¹⁰⁵</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Ag¹⁰⁷</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Cd¹¹¹</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>In¹¹⁵</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Sn¹¹⁹</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Tl¹²⁵</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Pb²⁰⁷</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Bi²⁰⁹</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Po</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>At</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
<tr>
<td>Rn</td>
<td>1/2</td>
<td>1/2</td>
<td>92.57</td>
<td>3.256</td>
</tr>
</tbody>
</table>

16  Magnetic Resonance
Equations of Motion

The rate of change of angular momentum of a system is equal to the torque that acts on the system. The torque on a magnetic moment $\mu$ in a magnetic field $B$ is $\mu \times B$, so that we have the gyroscopic equation

$$\dot{h} dI/dt = \mu \times B_a; \quad (5)$$

or

$$d\mu/dt = \gamma \mu \times B_a. \quad (6)$$

The nuclear magnetization $M$ is the sum $\Sigma \mu_i$ over all the nuclei in a unit volume. If only a single isotope is important, we consider only a single value of $\gamma$, so that

$$dM/dt = \gamma M \times B_a. \quad (7)$$

We place the nuclei in a static field $B_a = B_0\hat{z}$. In thermal equilibrium at temperature $T$ the magnetization will be along $\hat{z}$:

$$M_x = 0; \quad M_y = 0; \quad M_z = M_0 = \chi_0 B_0 = C B_0 / T, \quad (8)$$

where the Curie constant $C = N \mu^2/3k_B$.

The magnetization of a system of spins with $I = \frac{1}{2}$ is related to the population difference $N_1 - N_2$ of the lower and upper levels in Fig. 2: $M_z = (N_1 - N_2)\mu$, where the $N$'s refer to a unit volume. The population ratio in thermal equilibrium is just given by the Boltzmann factor for the energy difference $2\mu B_0$:

$$(N_2/N_1)_0 = \exp(-2\mu B_0/k_BT). \quad (9)$$

The equilibrium magnetization is $M_0 = N\mu \tanh(\mu B/k_BT)$.

When the magnetization component $M_z$ is not in thermal equilibrium, we suppose that it approaches equilibrium at a rate proportional to the departure from the equilibrium value $M_0$:

$$\frac{dM_z}{dt} = \frac{M_0 - M_z}{T_1}. \quad (10)$$

In the standard notation $T_1$ is called the longitudinal relaxation time or the spin-lattice relaxation time.

If at $t = 0$ an unmagnetized specimen is placed in a magnetic field $B_0\hat{z}$, the magnetization will increase from the initial value $M_z = 0$ to a final value $M_z = M_0$. Before and just after the specimen is placed in the field, the population $N_1$ will be equal to $N_2$, as appropriate to thermal equilibrium in zero magnetic field. It is necessary to reverse some spins to establish the new equilibrium distribution in the field $B_0$. On integrating (10):

$$\int_0^{M_z} \frac{dM_z}{M_0 - M_z} = \frac{1}{T_1} \int_0^t dt, \quad (11)$$
Figure 3  At time \( t = 0 \) an unmagnetized specimen \( M_z(0) = 0 \) is placed in a static magnetic field \( B_0 \). The magnetization increases with time and approaches the new equilibrium value \( M_0 = \chi_0 B_0 \). This experiment defines the longitudinal relaxation time \( T_1 \). The magnetic energy density \(-\mathbf{M} \cdot \mathbf{B}\) decreases as part of the spin population moves into the lower level. The asymptotic value at \( t \gg T_1 \) is \(-M_0 B_0\). The energy flows from the spin system to the system of lattice vibrations; thus \( T_1 \) is also called the spin-lattice relaxation time.

or

\[
\log \frac{M_0}{M_0 - M_z} = \frac{t}{T_1} ; \quad M_z(t) = M_0 \left[ 1 - \exp\left(-t/T_1\right) \right],
\]

(12)
as in Fig. 3. The magnetic energy \(-\mathbf{M} \cdot \mathbf{B}\) decreases as \( M_z \) approaches its new equilibrium value.

Typical processes whereby the magnetization approaches equilibrium are indicated in Fig. 4. The dominant spin-lattice interaction of paramagnetic ions in crystals is by the phonon modulation of the crystalline electric field. Relaxation proceeds by three principal processes (Fig. 4b): direct (emission or absorption of a phonon); Raman (scattering of a phonon); and Orbach (intervention of a third state). A thorough experimental analysis of spin-lattice relaxation in several rare-earth salts at helium temperatures has been given by Scott and Jeffries;\(^2\) they discuss the evidence for the three processes.

Taking account of (10), the \( z \) component of the equation of motion (7) becomes

\[
\frac{dM_z}{dt} = \gamma (\mathbf{M} \times \mathbf{B}_a)_z + \frac{M_0 - M_z}{T_1},
\]

(13a)

where \((M_0 - M_z)/T_1\) is an extra term in the equation of motion, arising from the spin-lattice interactions not included in the magnetic field \( \mathbf{B}_a \). That is, besides precessing about the magnetic field, \( \mathbf{M} \) will relax to the equilibrium value \( M_0 \).

If in a static field \( B_0 \mathbf{z} \) the transverse magnetization component \( M_x \) is not zero, then \( M_x \) will decay to zero, and similarly for \( M_y \). The decay occurs be-

Figure 4a  Some important processes that contribute to longitudinal magnetization relaxation in an insulator and in a metal. For the insulator we show a phonon scattered inelastically by the spin system. The spin system moves to a lower energy state, and the emitted phonon has higher energy by $\hbar \omega_0$ than the absorbed phonon. For the metal we show a similar inelastic scattering process in which a conduction electron is scattered.

Figure 4b  Spin relaxation from $2 \rightarrow 1$ by phonon emission, phonon scattering, and a two-stage phonon process. The temperature dependence of the longitudinal relaxation time $T_1$ is shown for the several processes.

cause in thermal equilibrium the transverse components are zero. We can provide for transverse relaxation:

$$dM_x/dt = \gamma (M \times B_a)_x - M_x/T_2 ;$$  \hspace{1cm} (13b)$$

$$dM_y/dt = \gamma (M \times B_a)_y - M_y/T_2 ,$$  \hspace{1cm} (13c)$$

where $T_2$ is called the transverse relaxation time.

The magnetic energy $-M \cdot B_a$ does not change as $M_x$ or $M_y$ changes, provided that $B_a$ is along $\hat{z}$. No energy need flow out of the spin system during relaxation of $M_x$ or $M_y$, so that the conditions that determine $T_2$ may be less strict than for $T_1$. Sometimes the two times are nearly equal, and sometimes $T_1 \gg T_2$, depending on local conditions.

The time $T_2$ is a measure of the time during which the individual moments that contribute to $M_x$, $M_y$ remain in phase with each other. Different local magnetic fields at the different spins will cause them to precess at different frequencies. If initially the spins have a common phase, the phases will become random in the course of time and the values of $M_x$, $M_y$ will become zero. We can think of $T_2$ as a dephasing time.

The set of equations (13) are called the Bloch equations. They are not symmetrical in $x$, $y$, and $z$ because we have biased the system with a static magnetic field along $\hat{z}$. In experiments an rf magnetic field is usually applied.
along the \( \hat{x} \) or \( \hat{y} \) axes. Our main interest is in the behavior of the magnetization in the combined rf and static fields, as in Fig. 5. The Bloch equations are plausible, but not exact; they do not describe all spin phenomena, particularly not those in solids.

We determine the frequency of free precession of the spin system in a static field \( B_0 = B_0 \hat{z} \) and with \( M_z = M_0 \). The Bloch equations reduce to

\[
\frac{dM_x}{dt} = \gamma B_0 M_y - \frac{M_x}{T_2} ; \quad \frac{dM_y}{dt} = -\gamma B_0 M_x - \frac{M_y}{T_2} ; \quad \frac{dM_z}{dt} = 0 .
\]  

We look for damped oscillatory solutions of the form

\[
M_x = m \exp(-t/T') \cos \omega t ; \quad M_y = -m \exp(-t/T') \sin \omega t .
\]  

On substitution in (14) we have for the left-hand equation

\[
-\omega \sin \omega t - \frac{1}{T'} \cos \omega t = -\gamma B_0 \sin \omega t - \frac{1}{T_2} \cos \omega t ,
\]  

so that the free precession is characterized by

\[
\omega_0 = \gamma B_0 ; \quad T' = T_2 .
\]

The motion (15) is similar to that of a damped harmonic oscillator in two dimensions. The analogy suggests correctly that the spin system will show resonance absorption of energy from a driving field near the frequency \( \omega_0 = \gamma B_0 \), and the frequency width of the response of the system to the driving field will be \( \Delta \omega \approx 1/T_2 \). Figure 6 shows the resonance of protons in water.

The Bloch equations may be solved to give the power absorption from a rotating magnetic field of amplitude \( B_1 \):

\[
B_x = B_1 \cos \omega t ; \quad B_y = -B_1 \sin \omega t .
\]

After a routine calculation one finds that the power absorption is

\[
\mathcal{P}(\omega) = \frac{\omega \gamma M_z T_2}{1 + (\omega_0 - \omega)^2 T_2^2} B_1^2 .
\]

The half-width of the resonance at half-maximum power is

\[
(\Delta \omega)_{1/2} = 1/T_2 .
\]
LINE WIDTH

The magnetic dipolar interaction is usually the most important cause of line broadening in a rigid lattice of magnetic dipoles. The magnetic field $\Delta B$ seen by a magnetic dipole $\mathbf{\mu}_1$ due to a magnetic dipole $\mathbf{\mu}_2$ at a point $r_{12}$ from the first dipole is

\[
\Delta B = \frac{3(\mathbf{\mu}_2 \cdot r_{12})r_{12} - \mathbf{\mu}_2 r_{12}^2}{r_{12}^3},
\]

by a fundamental result of magnetostatics.

The order of magnitude of the interaction is, with $B_i$ written for $\Delta B$,

\[
B_i = \frac{\mu}{r^3}.
\]

The strong dependence on $r$ suggests that close neighbor interactions will be dominant, so that

\[
B_i = \frac{\mu}{a^3},
\]

where $a$ is the separation of nearest neighbors. This result gives us a measure of the width of the spin resonance line, assuming random orientation of the neighbors. For protons at 2 Å separation,

\[
B_i \approx \frac{1.4 \times 10^{-23} \text{ G cm}^3}{8 \times 10^{-24} \text{ cm}^3} \approx 2 \text{ gauss} = 2 \times 10^{-4} \text{ tesla}.
\]

To express (21), (22), and (23) in SI, multiply the right-hand sides by $\mu_0/4\pi$.

Motional Narrowing

The line width decreases for nuclei in rapid relative motion. The effect in solids is illustrated by Fig. 7: diffusion resembles a random walk as atoms jump...
from one crystal site to another. An atom remains in one site for an average time $\tau$ that decreases markedly as the temperature increases.

The motional effects on the line width are even more spectacular in normal liquids, because the molecules are highly mobile. The width of the proton resonance line in water is only $10^{-5}$ of the width expected for water molecules frozen in position.

The effect of nuclear motion on $T_2$ and on the line width is subtle, but can be understood by an elementary argument. We know from the Bloch equations that $T_2$ is a measure of the time in which an individual spin becomes dephased by one radian because of a local perturbation in the magnetic field intensity. Let $(\Delta \omega)_0 = \gamma B_0$ denote the local frequency deviation due to a perturbation $B_0$. The local field may be caused by dipolar interactions with other spins.

If the atoms are in rapid relative motion, the local field $B_0$ seen by a given spin will fluctuate rapidly in time. We suppose that the local field has a value $+B_0$ for an average time $\tau$ and then changes to $-B_0$, as in Fig. 8a. Such a random change could be caused by a change of the angle between $\mu$ and $r$ in (21). In the time $\tau$ the spin will precess by an extra phase angle $\delta \varphi = \pm \gamma B_0 \tau$ relative to the phase angle of the steady precession in the applied field $B_0$.

The motional narrowing effect arises for short $\tau$ such that $\delta \varphi \ll 1$. After $n$ intervals of duration $\tau$ the mean square dephasing angle in the field $B_0$ will be

$$\langle \varphi^2 \rangle = n(\delta \varphi)^2 = n\gamma^2 B_0^2 \tau^2,$$

by analogy with a random walk process: the mean square displacement from the initial position after $n$ steps of length $\ell$ in random directions is $\langle r^2 \rangle = n\ell^2$. 

**Figure 7** Effect of diffusion of nuclei on the Li$^7$ NMR line width in metallic lithium. At low temperatures the width agrees with the theoretical value for a rigid lattice. As the temperature increases the diffusion rate increases and the line width decreases. The abrupt decrease in line width above $T = 230$ K occurs when the diffusion hopping time $\tau$ becomes shorter than $1/\gamma B_0$. Thus the experiment gives a direct measure of the hopping time for an atom to change lattice sites. (After H. S. Gutowsky and B. R. McGarvey.)
Figure 8  Phase of a spin in a constant local field, as compared with dephasing of a spin which after fixed time intervals $\tau$ hops at random among sites having local fields $\pm 1$. The dephasing is measured relative to the phase of a spin in the applied field $B_0$.

The average number of steps necessary to dephase a spin by one radian is $n = 1/\gamma^2 B^2 \tau^2$. (Spins dephased by much more than one radian do not contribute to the absorption signal.) This number of steps takes place in a time

$$T_2 = n\tau = 1/\gamma^2 B^2 \tau,$$

quite different from the rigid lattice result $T_2 \approx 1/\gamma B$. From (26) we obtain as the line width for rapid motion with a characteristic time $\tau$:

$$\Delta \omega = 1/T_2 = (\gamma B_i)^2 \tau,$$

or

$$\Delta \omega = 1/T_2 = (\Delta \omega_0)^2 \tau,$$

where $(\Delta \omega_0)$ is the line width in the rigid lattice.

The argument assumes that $(\Delta \omega_0) \tau \ll 1$, as otherwise $\delta \varphi$ will not be $\ll 1$. Thus $\Delta \omega \ll (\Delta \omega)_0$. The shorter is $\tau$, the narrower is the resonance line! This remarkable effect is known as motional narrowing.\(^3\) The rotational relaxation

\(^3\) The physical ideas are due to N. Bloembergen, E. M. Purcell, and R. V. Pound, Phys. Rev. 73, 679 (1948). The result differs from the theory of optical line width caused by strong collisions between atoms (as in a gas discharge), where a short $\tau$ gives a broad line. In the nuclear spin problem the collisions are weak. In most optical problems the collisions of atoms are strong enough to interrupt the phase of the oscillation. In nuclear resonance the phase may vary smoothly in a collision, although the frequency may vary suddenly from one value to another nearby value.
time of water molecules at room temperature is known from dielectric constant measurements to be of the order of $10^{-10}$ s; if $(\Delta \omega)_o = 10^5$ s$^{-1}$, then $(\Delta \omega)_o \tau = 10^{-5}$ and $\Delta \omega = (\Delta \omega)_o \tau = 1$ s$^{-1}$. Thus the motion narrows the proton resonance line to about $10^{-5}$ of the static width.

**HYPERFINE SPLITTING**

The hyperfine interaction is the magnetic interaction between the magnetic moment of a nucleus and the magnetic moment of an electron. To an observer stationed on the nucleus, the interaction is caused by the magnetic field produced by the nuclear moment of the electron and by the motion of the electron about the nucleus. There is an electron current about the nucleus if the electron is in a state with orbital angular momentum about the nucleus. But even if the electron is in a state of zero orbital angular momentum, there is an electron spin current about the nucleus, and this current gives rise to the contact hyperfine interaction, of particular importance in solids. We can understand the origin of the contact interaction by a qualitative physical argument, given in CGS.

The results of the Dirac theory of the electron suggest that the magnetic moment of $\mu_B = e\hbar/2mc$ of the electron arises from the circulation of an electron with velocity $c$ in a current loop of radius approximately the electron Compton wavelength, $\kappa_e = \hbar/mc \sim 10^{-11}$ cm. The electric current associated with the circulation is

$$ I \sim e \times (\text{turns per unit time}) \sim ec/\kappa_e, \quad (29) $$

and the magnetic field (Fig. 9) produced by the current is

(CGS)$$ B \sim I/\kappa_e c \sim e/\kappa_e^2. \quad (30) $$

The observer on the nucleus has the probability

$$ P = |\psi(0)|^2 \kappa_e^3. \quad (31) $$

of finding himself inside the electron, that is, within a sphere of volume $\kappa_e^3$ about the electron. Here $\psi(0)$ is the value of the electron wavefunction at the nucleus. Thus the average value of the magnetic field seen by the nucleus is

$$ \overline{B} \equiv e|\psi(0)|^2 \kappa_e \approx \mu_B |\psi(0)|^2, \quad (32) $$

where $\mu_B = e\hbar/2mc = \frac{1}{2}e\kappa_e$ is the Bohr magneton.

The contact part of the hyperfine interaction energy is

$$ U = -\mu_I \cdot \overline{B} = -\mu_I \cdot \mu_B |\psi(0)|^2 \approx \gamma h \mu_B |\psi(0)|^2 I \cdot S, \quad (33) $$

where $I$ is the nuclear spin in units of $\hbar$.

The contact interaction in an atom has the form

$$ U = a I \cdot S, \quad (34) $$
Figure 9  Magnetic field B produced by a charge moving in a circular loop. The contact part of the hyperfine interaction with a nuclear magnetic moment arises from the region within or near to the current loop. The field averaged over a spherical shell that encloses the loop gives zero. Thus for an s electron (L = 0) only the contact part contributes to the interaction.

Values of the hyperfine constant $a$ for the ground states of several free atoms are:

<table>
<thead>
<tr>
<th>nucleus</th>
<th>$H^1$</th>
<th>Li$^7$</th>
<th>Na$^{23}$</th>
<th>K$^{39}$</th>
<th>K$^{41}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>$\frac{5}{2}$</td>
<td>$\frac{7}{2}$</td>
<td>$\frac{7}{2}$</td>
</tr>
<tr>
<td>$a$ in gauss</td>
<td>507</td>
<td>144</td>
<td>310</td>
<td>83</td>
<td>85</td>
</tr>
<tr>
<td>$a$ in MHz</td>
<td>1420</td>
<td>402</td>
<td>886</td>
<td>231</td>
<td>127</td>
</tr>
</tbody>
</table>

The value of $a$ in gauss as seen by an electron spin is defined as $a/2\mu_B$.

In a strong magnetic field the energy level scheme of a free atom or ion is dominated by the Zeeman energy splitting of the electron levels; the hyperfine interaction gives an additional splitting that in strong fields is $U' \equiv a m_S m_I$, where $m_S$, $m_I$ are the magnetic quantum numbers.

For the energy level diagram of Fig. 10 the two electronic transitions have the selection rules $\Delta m_S = \pm 1$, $\Delta m_I = 0$; the frequencies are $\omega = \gamma H_0 \pm a/2\hbar$. The nuclear transitions are not marked; they have $\Delta m_S = 0$, so that $\omega_{nuc} = a/2\hbar$. The frequency of the nuclear transition $1 \to 2$ is equal to that of $3 \to 4$. 
The hyperfine interaction in a magnetic atom may split the ground energy level. The splitting in hydrogen is 1420 MHz, and this is the radio frequency line of interstellar atomic hydrogen.

**Examples: Paramagnetic Point Defects**

The hyperfine splitting of the electron spin resonance furnishes valuable structural information about paramagnetic point defects, such as the F centers in alkali halide crystals and the donor atoms in semiconductor crystals.

**F Centers in Alkali Halides.** An F center is a negative ion vacancy with one excess electron bound at the vacancy (Fig. 11). The wavefunction of the trapped electron is shared chiefly among the six alkali ions adjacent to the vacant lattice site, with smaller amplitudes on the 12 halide ions that form the shell of second nearest neighbors. The counting applies to crystals with the NaCl structure. If \( \phi(r) \) is the wavefunction of the valence electron on a single alkali ion, then in the first (or LCAO) approximation

\[
\psi(r) = C \sum_p \phi(r - r_p) ,
\]

where in the NaCl structure the six values of \( r_p \) mark the alkali ion sites that bound the lattice vacancy.

The width of the electron spin resonance line of an F center is determined essentially by the hyperfine interaction of the trapped electron with the nuclear
magnetic moments of the alkali ions adjacent to the vacant lattice site. The observed line width is evidence for the simple picture of the wavefunction of the electron. By line width we mean the width of the envelope of the possible hyperfine structure components.

As an example, consider an F center in KCl. Natural potassium is 93 percent $^{39}$K with nuclear spin $I = \frac{3}{2}$. The total spin of the six potassium nuclei at the F center is $I_{\text{max}} = 6 \times \frac{3}{2} = 9$, so that the number of hyperfine components is $2I_{\text{max}} + 1 = 19$: this is the number of possible values of the quantum number $m_I$. There are $(2I + 1)^6 = 4^6 = 4096$ independent arrangements of the six spins distributed into the 19 components, as in Fig. 12. Often we observe only the envelope of the absorption line of an F center.

**Donor Atoms in Silicon.** Phosphorus is a donor when present in silicon. Each donor atom has five electrons, of which four enter diamagnetically into the covalent bond network of the crystal, and the fifth bound electron acts as a paramagnetic center of spin $S = \frac{1}{2}$. The experimental hyperfine splitting in the strong field limit is shown in Fig. 13.

When the concentration exceeds about $1 \times 10^{18}$ donors cm$^{-3}$, the split line is replaced by a single narrow line. This is a motional narrowing effect (Eq. 28) of the rapid hopping of the donor electrons among many donor atoms. The rapid hopping averages out the hyperfine splitting. The hopping rate increases at the higher concentrations as the overlap of the donor electron wavefunctions is increased, a view supported by conductivity measurements (Chapter 10).

The donor electron wavefunction extends not only over the central donor atom but significantly over some hundreds of silicon atoms. The Si$^{29}$ nuclear spins give additional hyperfine splittings first studied by Feher with a powerful electron nuclear double resonance technique known as ENDOR.

**Knight Shift**

At a fixed frequency the resonance of a nuclear spin is observed at a slightly different magnetic field in a metal than in a diamagnetic solid. The effect is known as the Knight shift or metallic shift and is valuable as a tool for the study of conduction electrons.
Figure 12 The 4096 arrangements of the six nuclear spins of $K^{39}$ as distributed into 19 hyperfine components. Each component will be split further into a very large number of components by virtue of the residual hyperfine interaction with the 12 neighbor Cl nuclei, which may be Cl$^{35}$ (75 percent) or Cl$^{37}$ (25 percent). The envelope of the pattern is approximately gaussian in form.

Figure 13 Electron spin resonance lines of P donor atoms in silicon. At the higher donor concentration near the metal-insulator transition a donor electron can hop from site to site so rapidly that the hyperfine structure is suppressed. (After R. C. Fletcher, W. A. Yager, G. L. Pearson, and F. R. Merritt.)
The interaction energy of a nucleus of spin  \( I \) and magnetogyric ratio \( \gamma_I \) is
\[
U = (-\gamma_I h B_0 + a \langle S_z \rangle) I_z ,
\]
where the first term is the interaction with the applied magnetic field \( B_0 \) and the second is the average hyperfine interaction of the nucleus with the conduction electrons. The average conduction electron spin \( \langle S_z \rangle \) is related to the Pauli spin susceptibility \( \chi_s \) of the conduction electrons: \( M_z = gN\mu_B \langle S_z \rangle = \chi_s B_0 \), whence the interaction may be written as
\[
U = \left(-\gamma_I h + \frac{a\chi_s}{gN\mu_B}\right) B_0 I_z = -\gamma_I h B_0 \left(1 + \frac{\Delta B}{B_0}\right) I_z .
\]
(37)

The Knight shift is defined as
\[
K = -\frac{\Delta B}{B_0} = \frac{a\chi_s}{gN\mu_B \gamma_I h}
\]
(38)
and simulates a fractional change in the magnetogyric ratio. By the definition (34) of the hyperfine contact energy, the Knight shift is given approximately by \( K \approx \chi_s |\psi(0)|^2/N \); that is, by the Pauli spin susceptibility increased in the ratio of the conduction electron concentration at the nucleus to the average conduction electron concentration.

Experimental values are given in Table 2. The value of the hyperfine coupling constant \( a \) is somewhat different in the metal than in the free atom because the wave functions at the nucleus are different. From the Knight shift of metallic Li it is deduced that the value of \( |\psi(0)|^2 \) in the metal is 0.44 of the value in the free atom; a calculated value of the ratio using theoretical wave functions is 0.49.

It is only in rare instances that the absolute value of the spin contribution \( \chi_s \) to the magnetic susceptibility can be determined, usually by very careful conduction electron spin resonance experiments. The Knight shift has been of value in the study of metals, alloys, solid and intermetallic superconductors, and unusual electronic systems such as \( \text{Na}_x\text{WO}_3 \).

**Table 2  Knight shifts in NMR in metallic elements**
(At room temperature)

<table>
<thead>
<tr>
<th>Nucleus</th>
<th>Knight shift in percent</th>
<th>Nucleus</th>
<th>Knight shift in percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li(^{7})</td>
<td>0.0261</td>
<td>Cu(^{63})</td>
<td>0.237</td>
</tr>
<tr>
<td>Na(^{23})</td>
<td>0.112</td>
<td>Rb(^{87})</td>
<td>0.653</td>
</tr>
<tr>
<td>Al(^{27})</td>
<td>0.162</td>
<td>Pd(^{103})</td>
<td>-3.0</td>
</tr>
<tr>
<td>K(^{39})</td>
<td>0.265</td>
<td>Pt(^{195})</td>
<td>-3.533</td>
</tr>
<tr>
<td>V(^{51})</td>
<td>0.580</td>
<td>Au(^{197})</td>
<td>1.4</td>
</tr>
<tr>
<td>Cr(^{53})</td>
<td>0.69</td>
<td>Pb(^{207})</td>
<td>1.47</td>
</tr>
</tbody>
</table>
NUCLEAR QUADRUPOLE RESONANCE

Nuclei of spin $I \geq 1$ have an electric quadrupole moment. The quadrupole moment $Q$ is a measure of the ellipticity of the distribution of charge in the nucleus. The quantity of interest is defined classically by

$$eQ = \frac{1}{2} \int (3z^2 - r^2) \rho(r) d^3x ,$$

where $\rho(r)$ is the charge density. An egg-shaped nucleus has $Q$ positive; a saucer-shaped nucleus has $Q$ negative. The nucleus when placed in a crystal will see the electrostatic field of its environment, as in Fig. 14. If the symmetry of this field is lower than cubic, then the nuclear quadrupole moment will lead to a set of energy levels split by the interaction of the quadrupole moment with the local electric field.

The states that are split are the $2I + 1$ states of a spin $I$. The quadrupole splittings can often be observed directly because an rf magnetic field of the appropriate frequency can cause transitions between the levels. The term nuclear quadrupole resonance refers to observations of nuclear quadrupole splittings in the absence of a static magnetic field. The quadrupole splittings are particularly large in covalently bonded molecules such as Cl$_2$, Br$_2$, and I$_2$; the splittings are of the order $10^7$ or $10^8$ Hz.

FERROMAGNETIC RESONANCE

Spin resonance at microwave frequencies in ferromagnets is similar in principle to nuclear spin resonance. The total electron magnetic moment of the specimen precesses about the direction of the static magnetic field, and energy is absorbed strongly from the rf transverse field when its frequency is equal to the precessional frequency. We may think of the macroscopic vector $S$ representing the total spin of the ferromagnet as quantized in the static magnetic field, with energy levels separated by the usual Zeeman frequencies; the
magnetic selection rule $\Delta m_s = \pm 1$ allows transitions only between adjacent levels.

The unusual features of ferromagnetic resonance include:

- The transverse susceptibility components $\chi'$ and $\chi''$ are very large because the magnetization of a ferromagnet in a given static field is very much larger than the magnetization of electronic or nuclear paramagnets in the same field.
- The shape of the specimen plays an important role. Because the magnetization is large, the demagnetization field is large.
- The strong exchange coupling between the ferromagnetic electrons tends to suppress the dipolar contribution to the line width, so that the ferromagnetic resonance lines can be quite sharp ($<1$ G) under favorable conditions.
- Saturation effects occur at low rf power levels. It is not possible, as it is with nuclear spin systems, to drive a ferromagnetic spin system so hard that the magnetization $M_z$ is reduced to zero or reversed. The ferromagnetic resonance excitation breaks down into spin wave modes before the magnetization vector can be rotated appreciably from its initial direction.

**Shape Effects in FMR**

We treat the effects of specimen shape on the resonance frequency. Consider a specimen of cubic ferromagnetic insulator in the form of an ellipsoid with principal axes parallel to $x$, $y$, $z$ axes of a cartesian coordinate system. The demagnetization factors $N_x$, $N_y$, $N_z$ are identical with the depolarization factors defined in Chapter 13. The components of the internal magnetic field $B_i$ in the ellipsoid are related to the applied field by

$$B_x^i = B_x^0 - N_x M_x ; \quad B_y^i = B_y^0 - N_y M_y ; \quad B_z^i = B_z^0 - N_z M_z .$$

The Lorentz field $(4\pi/3)\mu_b M$ and the exchange field $\lambda M$ do not contribute to the torque because their vector product with $M$ vanished identically. In SI we replace the components of $M$ by $\mu_0 M$, with the appropriate redefinition of the $N$'s.

The components of the spin equation of motion $\dot{M} = \gamma(M \times B^i)$ become, for an applied static field $B_0\hat{z}$,

$$\frac{dM_x}{dt} = \gamma(M_y B_z^i - M_z B_y^i) = \gamma(B_0 + (N_y - N_z)M)M_y ;$$

$$\frac{dM_y}{dt} = \gamma[M(-N_x M_x) - M_z(B_0 - N_z M)] = -\gamma[B_0 + (N_x - N_z)M] M_x .$$

To first order we may set $dM_z/dt = 0$ and $M_z = M$. Solutions of (40) with time dependence $\exp(-i\omega t)$ exist if

$$\begin{vmatrix}
    i\omega & \gamma[B_0 + (N_y - N_z)M] \\
    -\gamma[B_0 + (N_x - N_z)M] & i\omega
\end{vmatrix} = 0 ,$$
so that the ferromagnetic resonance frequency in the applied field \( B_0 \) is

\[
\omega_0^2 = \gamma^2 [B_0 + (N_y - N_z)M][B_0 + (N_x - N_z)M] ;
\]

The frequency \( \omega_0 \) is called the frequency of the uniform mode, in distinction to the frequencies of magnon and other nonuniform modes. In the uniform mode all the moments precess together in phase with the same amplitude.

For a sphere \( N_x = N_y = N_z \), so that \( \omega_0 = \gamma B_0 \). A very sharp resonance line in this geometry is shown in Fig. 15. For a flat plate with \( B_0 \) perpendicular to the plate \( N_x = N_y = 0; N_z = 4\pi \), whence the ferromagnetic resonance frequency is

\[
\omega_0 = \gamma(B_0 - 4\pi M) ;
\]

If \( B_0 \) is parallel to the plane of the plate, the \( xx \) plane, then \( N_x = N_z = 0; N_y = 4\pi \), and

\[
\omega_0 = \gamma[B_0(B_0 + 4\pi M)]^{1/2} ;
\]

\[
\omega_0 = \gamma[B_0(B_0 + \mu_0 M)]^{1/2} .
\]
The experiments determine $\gamma$, which is related to the spectroscopic splitting factor $g$ by $-\gamma = g\mu_B/h$. Values of $g$ for metallic Fe, Co, Ni at room temperature are 2.10, 2.18, and 2.21, respectively.

**Spin Wave Resonance**

Uniform rf magnetic fields can excite long wavelength spin waves in thin ferromagnetic films if the electron spins on the surfaces of the film see different anisotropy fields than the spins within the films. In effect, the surface spins may be pinned by surface anisotropy interactions, as shown in Fig. 16. If the rf field is uniform, it can excite waves with an odd number of half-wavelengths within the thickness of the film. Waves with an even number of half-wavelengths have no net interaction energy with the field.

The condition for spin wave resonance (SWR) with the applied magnetic field normal to the film is obtained from (42) by adding to the right-hand side the exchange contribution to the frequency. The exchange contribution may be written as $DK^2$, where $D$ is the spin wave exchange constant. The assumption $ka < 1$ is valid for the SWR experiments. Thus in an applied field $B_0$ the spin wave resonance frequencies are:

\[
\omega_0 = \gamma(B_0 - 4\pi M) + DK^2 = \gamma(B_0 - 4\pi M) + D(n\pi/L)^2, \tag{44}
\]

where the wavevector for a mode of $n$ half-wavelengths in a film of thickness $L$ is $k = n\pi/L$. An experimental spectrum is shown in Fig. 17.

**ANTIFERROMAGNETIC RESONANCE**

We consider a uniaxial antiferromagnet with spins on two sublattices, 1 and 2. We suppose that the magnetization $M_1$ on sublattice 1 is directed along the $+z$ direction by an anisotropy field $B_Az$; the anisotropy field (Chapter 15) results from an anisotropy energy density $U_K(\theta_1) = K\sin^2 \theta_1$. Here $\theta_1$ is the angle between $M_1$ and the $z$ axis, whence $B_A = 2K/M$, with $M = |M_1| = |M_2|$. The magnetization $M_2$ is directed along the $-z$ direction by an anisotropy field $-B_Az$. If $+z$ is an easy direction of magnetization, so is $-z$. If one sublattice is directed along $+z$, the other will be directed along $-z$.

The exchange interaction between $M_1$ and $M_2$ is treated in the mean field approximation. The exchange fields are

\[
B_1(\text{ex}) = -\lambda M_2, \quad B_2(\text{ex}) = -\lambda M_1, \tag{45}
\]

where $\lambda$ is positive. Here $B_1$ is the field that acts on the spins of sublattice 1, and $B_2$ acts on sublattice 2. In the absence of an external magnetic field the total field acting on $M_1$ is $B_1 = -\lambda M_2 + B_Az$; the total field on $M_2$ is $B_2 = -\lambda M_1 - B_Az$, as in Fig. 18.
Figure 16 Spin wave resonance in a thin film. The plane of the film is normal to the applied magnetic field $B_0$. A cross section of the film is shown here. The internal magnetic field is $B_0 - 4\pi M$. The spins on the surfaces of the film are assumed to be held fixed in direction by surface anisotropy forces. A uniform rf field will excite spin wave modes having an odd number of half-wavelengths. The wave shown is for $n = 3$ half-wavelengths.

Figure 17 Spin wave resonance spectrum in a Permalloy (80Ni20Fe) film at 9 GHz. The order number is the number of half-wavelengths in the thickness of the film. (After R. Weber.)
In what follows we set \( M_1^z = M \); \( M_2^z = -M \). The linearized equations of motion are

\[
\frac{dM_1^y}{dt} = \gamma[M_1^y(\lambda M + B_A) - M(-\lambda M_1^z)];
\]

\[
\frac{dM_1^x}{dt} = \gamma[M_1^x(-\lambda M_1^z) - M_1^y(\lambda M + B_A)]; \tag{46}
\]

\[
\frac{dM_2^y}{dt} = \gamma[M_2^y(-\lambda M - B_A) - (-M)(-\lambda M_1^y)];
\]

\[
\frac{dM_2^x}{dt} = \gamma[(-M)(-\lambda M_1^z) - M_2^y(-\lambda M - B_A)]. \tag{47}
\]

We define \( M_1^+ = M_1^x + iM_1^y; \ M_2^+ = M_2^x + iM_2^y \). Then (46) and (47) become, for time dependence \( \exp(-i\omega t) \),

\[
-i\omega M_1^+ = -i\gamma[M_1^+(B_A + \lambda M) + M_2^+(\lambda M)];
\]

\[
-i\omega M_2^+ = i\gamma[M_2^+(B_A + \lambda M) + M_1^+(\lambda M)].
\]

These equations have a solution if, with \( B_E = \lambda M \),

\[
\begin{vmatrix}
\gamma(B_A + B_E) - \omega & \gamma B_E \\
\gamma B_E & \gamma(B_A + B_E) + \omega
\end{vmatrix} = 0.
\]

Thus the antiferromagnetic resonance frequency is given by

\[
\omega_0^2 = \gamma^2 B_A(B_A + 2B_E). \tag{48}
\]

\( \text{MnF}_2 \) is an extensively studied antiferromagnet. The structure is shown in Fig. 19. The observed variation of \( \omega_0 \) with temperature is shown in Fig. 20. Careful estimates were made by Keffer of \( B_A \) and \( B_E \) for \( \text{MnF}_2 \). He estimated \( B_E = 540 \text{ kG} \) and \( B_A = 8.8 \text{ kG} \) at 0 K, whence \( (2B_A B_E)^{1/2} = 100 \text{ kG} \). The observed value is 93 kG.
Figure 19 Chemical and magnetic structure of MnF$_2$. The arrows indicate the direction and arrangement of the magnetic moments assigned to the manganese atoms.

Figure 20 Antiferromagnetic resonance frequency for MnF$_2$ versus temperature. (After Johnson and Nethercot.)

Richards has made a compilation of AFMR frequencies as extrapolated to 0 K:

<table>
<thead>
<tr>
<th>Crystal</th>
<th>CoF$_2$</th>
<th>NiF$_2$</th>
<th>MnF$_2$</th>
<th>FeF$_2$</th>
<th>MnO</th>
<th>NiO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency in $10^{10}$ Hz</td>
<td>85.5</td>
<td>93.3</td>
<td>26.0</td>
<td>158.</td>
<td>82.8</td>
<td>109</td>
</tr>
</tbody>
</table>
ELECTRON PARAMAGNETIC RESONANCE

Electron spin resonance is a vast area of research. We mention two topics of interest.

Exchange Narrowing

We consider a paramagnet with an exchange interaction \( J \) among nearest-neighbor electron spins. The temperature is assumed to be well above any spin-ordering temperature \( T_c \). Under these conditions the width of the spin resonance line is usually much narrower than expected for the dipole-dipole interaction. The effect is called exchange narrowing; there is a close analogy with motional narrowing. We interpret the exchange frequency \( \omega_{ex} \approx J/\hbar \) as a hopping frequency \( 1/\tau \). Then by generalization of the motional-narrowing result (28) we have for the width of the exchange-narrowed line:

\[
\Delta \omega \approx \frac{(\Delta \omega)^2_{0}}{\omega_{ex}} \quad \text{(49)}
\]

where \( (\Delta \omega)^2_{0} = \gamma^2(B_r^2) \) is the square of the static dipolar width in the absence of exchange.

A useful and striking example of exchange narrowing is the paramagnetic organic crystal known as the \( g \) marker or DPPH, diphenyl picryl hydrazyl, often used for magnetic field calibration. This free radical has a 1.35 G half-width of the resonance line at half-power, only a few percent of the pure dipole width.

Zero-Field Splitting

A number of paramagnetic ions have crystal field splittings of their magnetic ground state energy levels in the range of \( 10^{10} - 10^{11} \) Hz, conveniently accessible by microwave techniques. Much of the pioneer work is due to B. Bleaney and co-workers at Oxford. The \( \text{Mn}^{2+} \) ion is popular and has been studied in many crystals as an additive impurity. A ground state splitting in the range \( 10^{7} - 10^{9} \) Hz is observed, according to the environment.

PRINCIPLE OF MASER ACTION

Crystals can be used as microwave and light amplifiers and as sources of coherent radiation. A maser amplifies microwaves by the stimulated emission of radiation; a laser amplifies light by the same method. The principle, due to Townes, may be understood from the two-level magnetic system of Fig. 21. There are \( n_u \) atoms in the upper state and \( n_l \) atoms in the lower state. We immerse the system in radiation at frequency \( \omega \); the amplitude of the magnetic component of the radiation field is \( B_{rf} \). The probability per atom per unit time
of a transition between the upper and lower states is

\[ P = \left( \frac{\mu B_{\text{rf}}}{\hbar} \right)^2 \frac{1}{\Delta \omega}; \quad (50) \]

here \( \mu \) is the magnetic moment, and \( \Delta \omega \) is the combined width of the two levels. The result (50) is from a standard result of quantum mechanics, called Fermi's golden rule.

The net energy emitted from atoms in both upper and lower states is

\[ \mathcal{P} = \left( \frac{\mu B_{\text{rf}}}{\hbar} \right)^2 \frac{1}{\Delta \omega} \cdot \hbar \omega \cdot (n_u - n_l), \quad (51) \]

per unit time. Here \( \mathcal{P} \) denotes the power out; \( \hbar \omega \) is the energy per photon; and \( n_u - n_l \) is the excess of the number of atoms \( n_u \) initially able to emit a photon over the number of atoms \( n_l \) able to absorb a photon.

In thermal equilibrium \( n_u < n_l \) so there is no net emission of radiation, but in a nonequilibrium condition with \( n_u > n_l \) there will be emission. If we start with \( n_u > n_l \) and reflect the emitted radiation back onto the system, we increase \( B_{\text{rf}} \) and thereby stimulate a higher rate of emission. The enhanced stimulation continues until the population in the upper state decreases and becomes equal to the population in the lower state.

We can build up the intensity of the radiation field by placing the crystal in an electromagnetic cavity. This is like multiple reflection from the walls of the cavity. There will be some power loss in the walls of the cavity: the rate of power loss is

\[
\begin{align*}
\text{(CGS)} & \\
\mathcal{P}_L &= \frac{B_{\text{rf}}^2 V}{8\pi} \cdot \frac{\omega}{Q}; \\
\text{(SI)} & \\
\mathcal{P}_L &= \frac{B_{\text{rf}}^2 V}{2\mu_0} \cdot \frac{\omega}{Q},
\end{align*}
\]

(52)

where \( V \) is the volume and \( Q \) is the \( Q \) factor of the cavity. We understand \( B_{\text{rf}}^2 \) to be a volume average.

The condition for maser action is that the emitted power \( \mathcal{P} \) exceed the power loss \( \mathcal{P}_L \). Both quantities involve \( B_{\text{rf}}^2 \). The maser condition can now be
expressed in terms of the population excess in the upper state:

\[ n_u - n_l > \frac{V \Delta B}{8\pi\mu Q}, \quad \text{(CGS)} \]

\[ n_u - n_l > \frac{V \Delta B}{2\mu_0\mu Q}, \quad \text{(SI)} \]

where \( \mu \) is the magnetic moment. The line width \( \Delta B \) is defined in terms of the combined line width \( \Delta \omega \) of the upper and lower states as \( \mu \Delta B = \hbar \Delta \omega \). The central problem of the maser or laser is to obtain a suitable excess population in the upper state. This is accomplished in various ways in various devices.

**Three-Level Maser**

The three-level maser system (Fig. 22) is a clever solution to the excess population problem. Such a system may derive its energy levels from magnetic ions in a crystal, as Bloembergen showed. Rf power is applied at the pump frequency \( \hbar \omega_p = E_3 - E_1 \) in sufficient intensity to maintain the population of level 3 substantially equal to the population of level 1. Now consider the rate of change of the population \( n_2 \) of level 2 owing to normal thermal relaxation processes. In terms of the indicated transition rates \( P \),

\[ \frac{dn_2}{dt} = -n_2 P(2 \rightarrow 1) - n_2 P(2 \rightarrow 3) + n_2 P(3 \rightarrow 2) + n_1 P(1 \rightarrow 2). \quad (54) \]

In the steady state \( \frac{dn_2}{dt} = 0 \), and by virtue of the saturation rf power we have \( n_3 = n_1 \), whence

\[ \frac{n_2}{n_1} = \frac{P(3 \rightarrow 2) + P(1 \rightarrow 2)}{P(2 \rightarrow 1) + P(2 \rightarrow 3)}. \quad (55) \]

The transition rates are affected by many details of the paramagnetic ion and its environment, but one can hardly fail with this system, for either \( n_2 > n_1 \) and we get maser action between levels 2 and 1, or \( n_2 < n_1 = n_3 \) and we get maser action between levels 3 and 2. The energy levels of the Er\(^{3+} \) ion are used in communication fiber optics amplifiers, as treated at the end of Chapter 17. The ion is optically pumped from level 1 to level 3; there is fast nonradiative decay from level 3 to level 2. The signal at a wavelength of 1.55 \( \mu \)m is amplified by stimulated emission from level 2 to level 1. The wavelength is favorable for long-distance propagation in the optical fiber. The bandwidth is of the order of \( 4 \times 10^{12} \) Hz.

**Ruby Laser**

The same crystal, ruby, used in the microwave maser was also the first crystal to exhibit optical maser action, but a different set of energy levels of Cr\(^{3+} \) are involved (Fig. 23). About 15,000 cm\(^{-1} \) above the ground state there lie a pair of states labeled \(^2E\), spaced 29 cm\(^{-1} \) apart. Above \(^2E\) lie two broad bands of states, labeled \(^4F_1\) and \(^4F_2\). Because the bands are broad they can be
Figure 22. Three-level maser system. Two possible modes of operation are shown, starting from rf saturation of the states 3 and 1 to obtain \( n_3 = n_1 \).

Figure 23. Energy level diagram of \( \text{Cr}^{3+} \) in ruby, as used in laser operation. The initial excitation takes place to the broad bands; they decay to the intermediate levels by the emission of phonons, and the intermediate levels radiate photons as the ion makes the transition to the ground level.

populated efficiently by optical absorption from broadband light sources such as xenon flash lamps.

In operation of a ruby laser both of the broad \( ^4F \) bands are populated by broadband light. Atoms thus excited will decay in \( 10^{-7} \) sec by radiationless processes with the emission of phonons to the states \( ^2E \). Photon emission from the lower of the states \( ^2E \) to the ground state occurs slowly, in about \( 5 \times 10^{-3} \) sec, so that a large excited population can pile up in \( ^2E \). For laser action this population must exceed that in the ground state.

The stored energy in ruby is \( 10^8 \) erg \( \text{cm}^{-3} \) if \( 10^{20} \) \( \text{Cr}^{3+} \) ions \( \text{cm}^{-3} \) are in an excited state. The ruby laser can emit at a very high power level if all this stored energy comes out in a short burst. The overall efficiency of conversion of a ruby laser from input electrical energy to output laser light is about one percent.
Another popular solid state laser is the neodymium glass laser, made of calcium tungstate glass doped with Nd$^{3+}$ ions. This operates as a four level system (Fig. 24). Here it is not necessary to empty out the ground state before laser action can occur.

Semiconductor junction lasers are treated in Chapter 19.

**SUMMARY**  
(In CGS Units)

- The resonance frequency of a free spin is $\omega_0 = \gamma B_0$, where $\gamma = \mu / \hbar I$ is the magnetogyratic ratio.

- The Bloch equations are
  
  \[
  \begin{align*}
  \frac{dM_x}{dt} &= \gamma (M \times B)_x - M_x / T_2 \\
  \frac{dM_y}{dt} &= \gamma (M \times B)_y - M_y / T_2 \\
  \frac{dM_z}{dt} &= \gamma (M \times B)_z + (M_0 - M_z) / T_1 
  \end{align*}
  \]

- The half-width of the resonance at half-power is $(\Delta \omega)_{1/2} = 1/T_2$.
- Saturation effects at high rf power enter when $\gamma^2 B_1^2 T_1 T_2$ exceeds unity.
- The dipolar line width in a rigid lattice is $(\Delta B)_0 = \mu / a^3$.
- If the magnetic moments are ambulatory, with a characteristic time $\tau \ll 1/(\Delta \omega)_0$, the line width is reduced by the factor $(\Delta \omega)_0 / \tau$. In this limit $1/T_1 \approx 1/T_2 \approx (\Delta \omega)_0 / \tau$. With exchange coupling in a paramagnet the line width becomes $= (\Delta \omega)_0^2 \omega_{\text{ex}}$. 


• The ferromagnetic resonance frequency in an ellipsoid of demagnetization factors \(N_x, N_y, N_z\) is \(\omega_0^2 = \gamma^2[B_0 + (N_y - N_z)M][B_0 + (N_x - N_z)M]\).

• The antiferromagnetic resonance frequency is \(\omega_0^2 = \gamma^2B_\Lambda(B_\Lambda + 2B_E)\), in a spherical specimen with zero applied field. Here \(B_\Lambda\) is the anisotropy field and \(B_E\) is the exchange field.

• The condition for maser action is that \(n_u - n_l > V\Delta B / 8\pi\mu Q\).

**Problems**

1. **Equivalent electrical circuit.** Consider an empty coil of inductance \(L_0\) in a series with a resistance \(R_0\); show if the coil is completely filled with a spin system characterized by the susceptibility components \(\chi'(\omega)\) and \(\chi''(\omega)\) that the inductance at frequency \(\omega\) becomes \(L = [1 + 4\pi\chi'(\omega)]L_0\), in series with an effective resistance \(R = 4\pi\omega\chi''(\omega)L_0 + R_0\). In this problem \(\chi = \chi' + i\chi''\) is defined for a linearly polarized rf field. Hint: Consider the impedance of the circuit. (CGS units.)

2. **Rotating coordinate system.** We define the vector \(\mathbf{F}(t) = F_x(t)\hat{x} + F_y(t)\hat{y} + F_z(t)\hat{z}\). Let the coordinate system of the unit vectors \(\hat{x}, \hat{y}, \hat{z}\) rotate with an instantaneous angular velocity \(\Omega\), so that \(d\hat{x}/dt = \Omega_\hat{y}\hat{z} - \Omega_\hat{z}\hat{y}\), etc. (a) Show that \(d\mathbf{F}/dt = (d\mathbf{F}/dt)_R + \Omega \times \mathbf{F}\), where \((d\mathbf{F}/dt)_R\) is the time derivative of \(\mathbf{F}\) as viewed in the rotating frame \(R\). (b) Show that \((7)\) may be written \((d\mathbf{M}/dt)_R = \gamma\mathbf{M} \times (\mathbf{B}_u + \Omega / \gamma)\). This is the equation of motion of \(\mathbf{M}\) in a rotating coordinate system. The transformation to a rotating system is extraordinarily useful; it is exploited widely in the literature. (c) Let \(\Omega = -\gamma B_0\hat{z}\); thus in the rotating frame there is no static magnetic field. Still in the rotating frame, we now apply a dc pulse \(B_1\hat{x}\) for a time \(t\). If the magnetization is initially along \(\hat{z}\), find an expression for the pulse length \(t\) such that the magnetization will be directed along \(-\hat{z}\) at the end of the pulse. (Neglect relaxation effects.) (d) Describe this pulse as viewed from the laboratory frame of reference.

3. **Hyperfine effects on ESR in metals.** We suppose that the electron spin of a conduction electron in a metal sees an effective magnetic field from the hyperfine interaction of the electron spin with the nuclear spin. Let the \(z\) component of the field seen by the conduction electron be written

\[
B_z = \left(\frac{a}{N}\right) \sum_{j=1}^{N} I_j^z,
\]

where \(I_j^z\) is equally likely to be \(\pm \frac{1}{2}\). (a) Show that \(\langle B_z^2 \rangle = (a/2N)^2N\). (b) Show that \(\langle B_z^4 \rangle = 3(a/2N)^4N^2\), for \(N \gg 1\).

4. **FMR in the anisotropy field.** Consider a spherical specimen of a uniaxial ferromagnetic crystal with an anisotropy energy density of the form \(U_K = K \sin^2 \theta\), where \(\theta\) is the angle between the magnetization and the \(z\) axis. We assume that \(K\) is positive. Show that the ferromagnetic resonance frequency in an external magnetic field \(B_0\hat{z}\) is \(\omega_0 = \gamma(B_0 + B_\Lambda)\), where \(B_\Lambda = 2K/M_z\).
5. **Exchange frequency resonance.** Consider a ferrimagnet with two sublattices A and B of magnetizations $M_A$ and $M_B$, where $M_B$ is opposite to $M_A$ when the spin system is at rest. The gyromagnetic ratios are $\gamma_A$, $\gamma_B$ and the molecular fields are $B_A = -\lambda M_B$, $B_B = -\lambda M_A$. Show that there is a resonance at

$$\omega^2_0 = \lambda^2(\gamma_A|M_B| - \gamma_B|M_A|)^2.$$  

This is called the exchange frequency resonance.

6. **Rf saturation.** Given, at equilibrium for temperature $T$, a two-level spin system in a magnetic field $H_0\hat{z}$, with populations $N_1$, $N_2$ and transition rates $W_{12}$, $W_{21}$. We apply an rf signal that gives a transition rate $W_{rf}$. (a) Derive the equation for $dM_z/dt$ and show that in the steady state

$$M_z = M_0/(1 + 2W_{rf}T_1),$$

where $1/T_1 = W_{12} + W_{21}$. It will be helpful to write $N = N_1 + N_2$, $n = N_1 - N_2$, and $n_0 = N(W_{21} - W_{12})/(W_{21} + W_{12})$. We see that as long as $2W_{rf}T_1 \ll 1$ the absorption of energy from the rf field does not substantially alter the population distribution from its thermal equilibrium value. (b) Using the expression for $n$, write down the rate at which energy is absorbed from the rf field. What happens as $W_{rf}$ approaches $1/2T_1$? This effect is called saturation, and its onset may be used to measure $T_1$.

---

**References**


**NUCLEAR MAGNETIC RESONANCE**


V. Jaccarino, “Nuclear resonance in antiferromagnets,” *Magnetism* II A.


**ELECTRON SPIN RESONANCE**


FERRO- AND ANTIFERROMAGNETIC RESONANCE


QUANTUM ELECTRONICS

17
Noncrystalline Solids

DIFFRACTION PATTERN
  Monatomic amorphous materials 521
  Radial distribution function 522
  Structure of vitreous silica, SiO₂ 523

GLASSES
  Viscosity and the hopping rate 527

AMORPHOUS FERROMAGNETS 529

AMORPHOUS SEMICONDUCTORS 530

LOW ENERGY EXCITATIONS IN AMORPHOUS SOLIDS 532
  Heat capacity calculation 533
  Thermal conductivity 534

FIBER OPTICS 535
  Rayleigh attenuation 536

REFERENCES 537
CHAPTER 17: NONCRYSTALLINE SOLIDS

It is generally agreed that the terms amorphous solid, noncrystalline solid, disordered solid, glass, or liquid have no precise structural meaning beyond the description that the structure is "not crystalline on any significant scale." The principal structural order present is imposed by the approximately constant separation of nearest-neighbor atoms or molecules. We exclude from the present discussion disordered crystalline alloys (Chapter 21) where different atoms randomly occupy the sites of a regular crystal lattice.

DIFFRACTION PATTERN

The x-ray or neutron diffraction pattern of an amorphous material such as a liquid or a glass consists of one or more broad diffuse rings, when viewed on the plane normal to the incident x-ray beam. The pattern is different from the diffraction pattern of powdered crystalline material which shows a large number of fairly sharp rings. The result tells us that a liquid does not have a unit of structure that repeats itself identically at periodic intervals in three dimensions.

In a simple monatomic liquid the positions of the atoms show only a short range structure referred to an origin on any one atom. We never find the center of another atom closer than a distance equal to the atomic diameter, but at roughly this distance we expect to find about the number of nearest-neighbor atoms that we find in a crystalline form of the material.

Although the x-ray pattern of a typical amorphous material is distinctly different from that of a typical crystalline material, there is no sharp division between them. For crystalline powder samples of smaller and smaller particle size, the powder pattern lines broaden continuously, and for small enough crystalline particles the pattern becomes similar to the amorphous pattern of a liquid or a glass.

From a typical liquid or glass diffraction pattern, containing three or four diffuse rings, the only quantity which can be determined directly is the radial distribution function. This is obtained from a Fourier analysis of the experimental x-ray scattering curve, and gives directly the average number of atoms to be found at any distance from a given atom. The method of Fourier analysis is equally applicable to a liquid, a glass, or a powdered crystalline material.

It is convenient to begin the analysis of the diffraction pattern with Eq. (2.43). Instead of writing it for the structure factor of the basis, we write the sum for all the atoms in the specimen. Further, instead of specializing the scattering to the reciprocal lattice vectors \( \mathbf{G} \) characteristic of a crystal, we consider arbitrary scattering vectors \( \Delta \mathbf{k} = \mathbf{k}^\prime - \mathbf{k} \), as in Fig. 2.6. We do this because scattering from amorphous materials is not limited to the reciprocal lattice vectors, which in any event cannot here be defined.
Therefore the scattered amplitude from an amorphous material is described by

\[
S(\Delta k) = \sum_m f_m \exp(-i\Delta k \cdot r_m),
\]

with \(f_m\) the atomic form factor of atom, as in Eq. (2.50). The sum runs over all atoms in the specimen.

The scattered intensity at scattering vector \(\Delta k\) is given by

\[
I = S^*S = \sum_m \sum_n f_m f_n \exp[i\Delta k \cdot (r_m - r_n)],
\]

in units referred to the scattering from a single electron. If \(\alpha\) denotes the angle between \(\Delta k\) and \(r_m - r_n\), then

\[
I = \sum_m \sum_n f_m f_n \exp(iKr_{mn} \cos \alpha),
\]

where \(K\) is the magnitude of \(\Delta k\) and \(r_{mn}\) is the magnitude of \(r_m - r_n\).

In an amorphous specimen the vector \(r_m - r_n\) may take on all orientations, so we average the phase factor over a sphere:

\[
\langle \exp(iKr \cos \alpha) \rangle = \frac{1}{4\pi} 2\pi \int_{-\pi}^{\pi} d(cos \alpha) \exp(iKr_{mn} \cos \alpha)
\]

\[
= \sin \frac{K_{mn}}{K_{mn}}.
\]

Thus we have the Debye result for the scattered density:

\[
I = \sum_m \sum_n (f_m f_n \sin Kr_{mn})/K_{mn}.
\]

Monatomic Amorphous Materials

For atoms of only one type, we let \(f_m = f_n = f\) and separate out from the summation (5) the terms with \(n = m\). For a specimen of \(N\) atoms,

\[
I = Nf^2 \left[ 1 + \sum_{n'} (\sin Kr_{mn})/K_{mn} \right].
\]

The sum runs over all atoms \(m\) except the origin atom \(m = n\).

If \(\rho(r)\) is the concentration of atoms at distance \(r\) from a reference atom, we can write (6) as

\[
I = Nf^2 \left[ 1 + \int_0^R dr 4\pi r^2 \rho(r)(\sin Kr)/Kr \right],
\]
where \( R \) is the (very large) radius of the specimen. Let \( p_0 \) denote the concentration; then (7) may be written as

\[
I = Nf^2 \left\{ 1 + \int_0^R dr \, 4\pi r^2 (\rho(r) - p_0) \frac{\sin Kr}{Kr} + (p_0/K) \int_0^R dr \, 4\pi r \sin Kr \right\}.
\]  

(8)

The second integral in (8) gives the scattering from a uniform concentration and may be neglected except in the forward region of very small angles; it reduces to a delta function at the origin as \( R \to \infty \).

**Radial Distribution Function**

It is convenient to introduce the liquid structure factor defined by

\[
S(K) = I/Nf^2.
\]  

(9)

Note that this is not at all the same as \( S(\Delta k) \) in (1). From (8) we have, after dropping the delta function contribution,

\[
S(K) = 1 + \int_0^\infty dr \, 4\pi r^2 (\rho(r) - p_0) \frac{\sin Kr}{Kr}.
\]  

(10)

We define the radial distribution function \( g(r) \) such that

\[
\rho(r) = g(r)p_0.
\]  

(11)

Then (10) becomes

\[
S(K) = 1 + 4\pi p_0 \int_0^\infty dr \, [g(r) - 1] r^2 (\sin Kr)/Kr \]

\[
= 1 + p_0 \int dr \, [g(r) - 1] \exp(iK \cdot r),
\]  

(12)

because \((\sin Kr)/Kr\) is the spherically symmetric or \( s \) term in the expansion of \( \exp(iK \cdot r) \).

By the Fourier integral theorem in three dimensions,

\[
g(r) - 1 = \frac{1}{8\pi^2 p_0} \int dK \, [S(K) - 1] \exp(-iK \cdot r)
\]

\[
= \frac{1}{2\pi^2 p_0 r} \int dK \, [S(K) - 1] K \sin Kr.
\]  

(13)

This result allows us to calculate the radial distribution function \( g(r) \) (also called the two-atom correlation function) from the measured structure factor \( S(K) \).

One of the simplest liquids well suited to x-ray diffraction study is liquid sodium. The plot of the radial distribution \( 4\pi r^2 \rho(r) \) vs. \( r \) is given in Fig. 1, together with the distribution of neighbors in crystalline sodium.
Structure of Vitreous Silica, SiO_2

Vitreous silica (fused quartz) is a simple glass. The x-ray scattering curve is given in Fig. 2. The radial distribution curve $4\pi r^2 \rho(r)$ vs. $r$ is given in Fig. 3. Because there are two kinds of atoms, $\rho(r)$ is actually the superposition of two electron concentration curves, one about a silicon atom as origin and one about an oxygen atom as origin.

The first peak is at 1.62 Å, close to the average Si-O distance found in crystalline silicates. The x-ray workers conclude from the intensity of the first peak that each silicon atom is tetrahedrally surrounded by four oxygen atoms. The relative proportions of Si and O tell us that each O atom is bonded to two Si atoms. From the geometry of a tetrahedron, the O-O distance should be 2.65 Å, compatible with the distance suggested by the shoulder in Fig. 3.

The x-ray results are consistent with the standard model of an oxide glass, due to Zachariasen. Figure 4 illustrates in two dimensions the irregular structure of a glass and the regularly repeating structure of a crystal of identical chemical composition. The x-ray results are completely explained by picturing glassy silica as a random network in which each silicon is tetrahedrally surrounded by four oxygens, each oxygen bonded to two silicons, the two bonds to an oxygen being roughly diametrically opposite. The orientation of one tetrahedral group with respect to a neighboring group about the connecting Si-O-Si
bond can be practically random. There is a definite structural scheme involved: each atom has a definite number of nearest neighbors at a definite distance, but no unit of structure repeats itself identically at regular intervals in three dimensions, and hence the material is not crystalline.

It is not possible to explain the x-ray results by assuming that vitreous silica consists of very small crystals of some crystalline form of quartz, such as cristoballite. Small angle x-ray scattering is not observed, but would be expected from discrete particles with breaks and voids between them. The scheme of bonding in glass must be essentially continuous, at least for the major part of the material, although the scheme of coordination about each atom is the same in vitreous silica and in crystalline cristoballite.

The low thermal conductivity of glasses at room temperature, as discussed presently, also is consistent with the continuous random network model.

A comparison of experimental and calculated results for amorphous germanium is shown in Fig. 5. The calculations are for a random network model and for a microcrystallite model. The latter model gives a very poor agreement. The random network model is supported for amorphous silicon by studies of the band gap and spectroscopic work on the 2p shell.
Figure 3  Radial distribution curve for vitreous SiO$_2$, as the Fourier transform of Fig. 2. The positions of the peaks give the distances of atoms from a silicon or an oxygen. From the areas under the peaks it is possible to calculate the number of neighbors at that distance. The vertical lines indicate the first few average interatomic distances; the heights of the lines are proportional to the peak areas. (After B. E. Warren.)

Figure 4  Schematic two-dimensional analogs illustrating the differences between: (a) the regularly repeating structure of a crystal and (b) continuous random network of a glass. (After Zachariasen.)
GLASSES

A glass has the random structure of the liquid from which it is derived by cooling below the freezing point, without crystallization. Also, a glass has the elastic properties of an isotropic solid.

By general agreement, we say that a liquid on being cooled becomes a glass when the viscosity equals $10^{13}$ poise, where a poise is the CGS unit of viscosity.\(^1\) This defines the glass transition temperature $T_g$. At temperatures above $T_g$ we have a liquid; below $T_g$ we have a glass. The transition is not a thermodynamic phase transition, only a transition for “practical purposes.”

Relatively few liquids can be cooled fast enough in the bulk to form a glass before crystallization intervenes. Molecules of most substances have high

\(^1\)The SI unit of viscosity is 1 Nsm\(^{-2}\), so that 1 poise = 0.1 Nsm\(^{-2}\). It is quite common to find viscosities given in cp or centipoise, being $10^{-2}$ poise.
enough mobility in the liquid so that on cooling a liquid-solid melting transition occurs a long time before the viscosity increases to $10^{13}$ poise or $10^{15}$ cp.

By contrast, liquid water has a viscosity 1.8 cp at the freezing point; the viscosity increases enormously on freezing. We can often make a glass by depositing a jet of atoms of a substrate cooled to a low temperature, a process which will sometimes produce an amorphous layer with glasslike properties. Amorphous ribbons of some metal alloys may be produced in this way in industrial quantities.

The value $10^{13}$ poise used to define $T_g$ is arbitrary, but not unreasonable. If we bond a slab of glass 1 cm thick to two plane parallel vertical surfaces, the glass will flow perceptibly in one year under its own weight when the viscosity drops below $10^{13}$ poise. (For comparison, the viscosity of the mantle of the earth is of the order of $10^{22}$ poise.)

**Viscosity and the Hopping Rate**

The viscosity of a liquid is related to the rate at which molecules undergo thermal rearrangement on a local scale, as by hopping into a vacant neighbor site or by interchange of two neighbor molecules. The physics of the transport process is somewhat different from that of viscosity in the gas phase, but the gas phase result gives a qualitative lower limit to the viscosity of the liquid phase, a limit that applies to nearest-neighbor hopping of atoms.

The gas result (TP 14.34) is

$$\eta = \frac{1}{3} \rho \bar{c} \ell,$$  \hspace{1cm} (14)

where $\eta$ is the viscosity, $\rho$ the density, $\bar{c}$ the mean thermal velocity, and $\ell$ the mean free path. In the liquid $\ell$ is the order of magnitude of the intermolecular separation $a$. With "typical" values $\rho \approx 2$ g cm$^{-3}$; $\bar{c} \approx 10^5$ cm s$^{-1}$; $a \approx 5 \times 10^{-8}$ cm, we have

$$\eta(\text{min}) = 0.3 \times 10^{-2} \text{ poise} = 0.3 \text{ cp}$$  \hspace{1cm} (15)

as an estimate of the lower limit of the viscosity of a liquid. (Tables in chemical handbooks only rarely list values below this.)

We give now a very simple model of the viscosity of a liquid. In order to hop successfully, a molecule must surmount the potential energy barrier presented by its neighbors in the liquid. The preceding estimate of the minimum viscosity applies when this barrier may be neglected. If the barrier is of height $E$, the molecule will have sufficient thermal energy to pass over the barrier only a fraction

$$f = \exp(-E/k_BT)$$  \hspace{1cm} (16)

of the time. Here $E$ is an appropriate free energy and is called the activation
energy for the process that determines the rate of hopping. It is related to the activation energy for self-diffusion treated in Chapter 18.

The viscosity will be increased as the probability of successful hopping is decreased. Thus

$$\eta = \eta(\text{min})/f = \eta(\text{min}) \exp(E/k_B T)$$  \hspace{1cm} (17)

If $\eta = 10^{13}$ poise at the glass transition, the order of magnitude of $f$ must be

$$f \approx 0.3 \times 10^{-15}$$  \hspace{1cm} (18)

at the transition, using (15). The corresponding activation energy is

$$E/k_B T_g = -\ln f = \ln(3 \times 10^{15}) = 35.6$$  \hspace{1cm} (19)

If $T_g = 2000$ K, then $k_B T_g = 2.7 \times 10^{-13}$ erg and $E = 9.6 \times 10^{-12}$ erg $\approx$ 6 eV. This is a high potential energy barrier.

Classes with lower values of $T_g$ will have correspondingly lower values of $E$. (Activation energies obtained in this way are often labeled as $E_{\text{visc}}$.)

Materials that are glass-formers are characterized by activation energies of the order of 1 eV or more; non-glass-formers may have activation energies of the order of 0.01 eV.

When being pressed into molds or drawn into tubes, glass is used in a range of temperatures at which its viscosity is $10^3$ to $10^6$ poises. The working range for vitreous silica begins over 2000°C, so high that the practical usefulness of the material is severely limited. In common glass about 25 percent of NaO is added as a network modifier to SiO$_2$ in order to reduce below 1000°C the temperature needed to make the glass fluid enough for the forming operations needed to make electric lamp bulbs, window glass, and bottles.

**AMORPHOUS FERROMAGNETS**

Amorphous metallic alloys are formed by very rapid quenching (cooling) of a liquid alloy, commonly by directing a molten stream of the alloy onto the surface of a rapidly rotating drum. This process produces a continuous "melt-spun" ribbon of amorphous alloy in industrial quantities.

Ferromagnetic amorphous alloys were developed because amorphous materials have nearly isotropic properties, and isotropic materials should have essentially zero magnetocrystalline anisotropy energy. As discussed in Chapter 15, the absence of directions of hard and easy magnetization should result in low coercivities, low hysteresis losses, and high permeabilities. Because amorphous alloys are also random alloys, their electrical resistivity is high. All these properties have technological value for application as soft magnetic materials. The trade name Metglas is attached to several of these.

The transition metal-metalloid (TM-M) alloys are an important class of
magnetic amorphous alloys. The transition metal component is usually about 80 percent of Fe, Co, or Ni, with the metalloid component B, C, Si, P, or Al. The presence of the metalloids lowers the melting point, making it possible to quench the alloy through the glass transition temperature rapidly enough to stabilize the amorphous phase. For example, the composition Fe_{80}B_{20} (known as Metglas 2605) has \( T_g = 441^\circ\text{C} \), as compared with the melting temperature 1538\(^\circ\text{C}\) of pure iron.

The Curie temperature of this composition in the amorphous phase is 647 K, and the value of the magnetization \( M_s \) at 300 K is 1257, compared with \( T_c = 1043 \) K and \( M_s = 1707 \) for pure iron (Table 15.2). The coercivity is 0.04 G, and the maximum value of the permeability is \( 3 \times 10^5 \). Coercivities as low as 0.006 G have been reported for another composition.\(^2\)

High coercivity materials can be produced by the same melt-spin process if the spin rate or quench rate is decreased to produce a fine-grained crystalline phase, which may be of metastable composition. If the grain size is arranged to match the optimum size for single domains, the coercivity can be quite high. J. L. Croat has reported \( H_{cr} = 7.5 \) kG for the metastable alloy Nd_{0.4}Fe_{0.6} at the optimum melt-spin velocity 5 m s\(^{-1}\). This material may come to have practical application. Figure 6 shows the peak in the coercivity of a Sm-Fe alloy as the spin velocity is varied.

**AMORPHOUS SEMICONDUCTORS**

Amorphous semiconductors can be prepared as thin films by evaporation or sputtering, or in some materials as bulk glasses by supercooling the melt.

What happens to the electron energy band model in a solid without regular crystalline order? The Bloch theorem is not applicable when the structure is not periodic, so that the electron states cannot be described by well-defined \( k \) values. Thus, the momentum selection rule for optical transitions is relaxed; hence all infrared and Raman modes contribute to the absorption spectra. The optical absorption edge is rather featureless. Allowed bands and energy gaps still occur because the form of the density of states vs. energy is determined most strongly by local electron bonding configurations.

Both electrons and holes can carry current in an amorphous semiconductor. The carriers may be scattered strongly by the disordered structure, so that the mean free path may sometimes be of the order of the scale of the disorder. Anderson proposed that the states near band edges may be localized and do not extend through the solid (Fig. 7). Conduction in these states may take place by a thermally-assisted hopping process, for which the Hall effect is anomalous.

and cannot be used to determine the carrier concentration. The quantum Hall effect (Chapter 19) provides experimental evidence for localized states and mobility edges.

Amorphous materials appear to behave almost like intrinsic semiconductors, with the Fermi level lying near to the center of the gap. However, the pinning of the Fermi level is known to arise from the presence of defects, such as dangling bonds and other misfits in the structure, which produce localized states in the gap. The presence of these states, along with the tendency in covalent alloys for local valence requirements to be satisfied, necessitates the use of special techniques for preparation of doped specimens.

Two distinct classes of amorphous semiconductors are widely studied: tetrahedrally-bonded amorphous solids such as silicon and germanium, and the chalcogenide glasses. The latter are multicomponent solids of which one major constituent is a “chalcogen” element—sulfur, selenium, or tellurium.

The tetrahedrally-bonded materials have properties similar to those of their crystalline forms, provided the dangling-bond defects are compensated.
with hydrogen. They can be doped with small amounts of chemical impurities, and their conductivity can be sharply modified by injection of free carriers from a metallic contact. By contrast, the chalcogenide glasses are largely insensitive to chemical impurities and to free carrier injection.

Amorphous hydrogenated silicon is a candidate material for solar cells. Amorphous silicon is a much less expensive material than single crystal silicon. Attempts at using pure amorphous silicon, however, failed because of structural defects (dangling bonds) which were impossible to eliminate. Introduction of hydrogen into amorphous silicon appears to remove the undesirable structure defects. Relatively large proportions of hydrogen are incorporated, of the order of 10 percent or more.

LOW ENERGY EXCITATIONS IN AMORPHOUS SOLIDS

There is a great regularity in the unusual thermal properties of amorphous materials at low temperatures, yet the experimental discovery of these properties was entirely unexpected, completely different from crystalline solids, and not well understood in terms of amorphous structure.

The low temperature heat capacity of pure dielectric crystalline solids is known (Chapter 5) to follow the Debye $T^3$ law, precisely as expected from the excitation of long wavelength phonons. The same behavior was expected in glasses and other amorphous solids—the point was so obvious that it did not encourage experimental investigation.

After early indications by a Berkeley group, Zeller and Pohl discovered that many insulating glasses show an unexpected linear term in the heat capacity below 1 K. Indeed, at 25 mK the observed heat capacity of vitreous silica exceeds the Debye phonon contribution by a factor of 1000.

Anomalous linear terms of comparable magnitude are found in all, or nearly all, amorphous solids. Their presence is believed to be an intrinsic consequence of the amorphous states of matter, but the details of why this is so
remain unclear. Work by Hunklinger and associates gives strong evidence that the anomalous properties arise from two-level systems and not from multi-level oscillator systems; in brief, the evidence is that the systems can be saturated by intense phonon fields, just as a two-level spin system can be saturated by an intense rf magnetic field (Problem 16.6).

**Heat Capacity Calculation**

Consider an amorphous solid with a concentration $N$ of two-level systems at low energies; that is, with a level splitting $\Delta$ much less than the phonon Debye cutoff $k_B\theta$. The partition function of one system is, with $\tau = k_B T$,

$$Z = \exp(\Delta/2\tau) + \exp(-\Delta/2\tau) = 2 \cosh(\Delta/2\tau) .$$

(20)

The thermal average energy is

$$U = -\frac{1}{2}\Delta \tanh(\Delta/2\tau) ,$$

(21)

and the heat capacity of the single system is

$$C_V = k_B (\partial U/\partial \tau) = k_B(\Delta/2\tau)^2 \sech^2(\Delta/2\tau) .$$

(22)

These results are given in detail in TP, pp. 62–63.

Now suppose that $\Delta$ is distributed with uniform probability in the range $\Delta = 0$ to $\Delta = \Delta_0$. The average value of $C_V$ is

$$C_V = \left(\frac{k_B}{4\tau^2}\right) \int_0^{\Delta_0} d\Delta (\Delta^2/\Delta_0) \sech^2(\Delta/2\tau)$$

$$= \left(\frac{2k_B T}{\Delta_0}\right) \int_0^{\Delta_0/2\tau} dx \ x^2 \sech^2 x .$$

(23)

The integral cannot be carried out in closed form.

Two limits are of special interest. For $\tau \ll \Delta_0$, the $\sech^2 x$ term is roughly 1 from $x = 0$ to $x = 1$, and roughly zero for $x > 1$. The value of the integral is roughly 1/3, whence

$$C_V \approx 2k_B T/3\Delta_0 ,$$

(24)

for $T < \Delta_0/k_B$.

For $\tau \gg \Delta_0$, the value of the integral is roughly $1/3(\Delta_0/2k_B T)^3$, so that in this limit

$$C_V \approx \Delta_0^3/12k_B T^2 ,$$

(25)

which approaches zero as $T$ increases.

Thus the interesting region is at low temperatures, for here the two-level system contributes to the heat capacity a term linear in the temperature. This term, originally introduced by W. Marshall for dilute magnetic impurities in metals, has no connection with the usual conduction electron heat capacity which is also proportional to $T$.

The empirical result appears to be that all disordered solids have
Figure 8  Heat capacity of vitreous silica and soda silica glass as a function of temperature. The heat capacity is roughly linear in $T$ below 1 K. The dashed line represents the calculated Debye heat capacity of vitreous silica.

$N \sim 10^{17} \text{ cm}^{-3}$ “new type” low energy excitations uniformly distributed in the energy interval from 0 to 1 K. The anomalous specific heat can now be obtained from (24). For $T = 0.1$ K and $\Delta_0/k_B = 1$ K,

$$C_V \approx \frac{2}{3} N k_B (0.1) \approx 1 \text{ erg cm}^{-3} \text{ K}^{-1}.$$  \hspace{1cm} (26)

For comparison, the phonon contribution at 0.1 K is, from (5.35),

$$C_V \approx 234 N k_B (T/\Theta)^3 \approx (234)(2.3 \times 10^{29})(1.38 \times 10^{-16})(0.1/300)^3$$

$$\approx 2.8 \times 10^{-2} \text{ erg cm}^{-3} \text{ K}^{-1},$$ \hspace{1cm} (27)

much smaller than (26).

The experimental results (Fig. 8) for vitreous SiO$_2$ are represented by

$$C_V = c_1 T + c_3 T^3,$$ \hspace{1cm} (28)

where $c_1 = 12 \text{ erg g}^{-1} \text{ K}^{-2}$ and $c_3 = 18 \text{ erg g}^{-1} \text{ K}^{-4}$.

Thermal Conductivity

The thermal conductivity of glasses is very low. It is limited at room temperature and above by the scale of the disorder of the structure, for this scale determines the mean free path of the dominant thermal phonons. At low temperatures, below 1 K, the conductivity is carried by long wavelength phonons
Figure 9  Short phonon mean free path in a disordered structure. A short wavelength phonon that displaces atom L, as shown, will displace atom R by a much smaller distance, because of the phase cancellation of the upper and lower paths from L to R. The displacement of R is \( \uparrow + \downarrow \sim 0 \), so that the wave incident from L is reflected at R.

and is limited by phonon scattering from the mysterious two-level systems or tunneling states discussed earlier for their contribution to the heat capacity of amorphous solids.

As in Chapter 5, the expression for the thermal conductivity \( K \) has the form

\[
K = \frac{1}{3} cv \ell ,
\]

where \( c \) is the heat capacity per unit volume, \( v \) is an average phonon velocity, and \( \ell \) is the phonon mean free path. For vitreous silica at room temperature,

\[
K \equiv 1.4 \times 10^{-2} \text{ J cm}^{-1} \text{ s}^{-1} \text{ K}^{-1} ;
\]

\[
c \equiv 1.6 \text{ J cm}^{-3} \text{ K}^{-1} ;
\]

\[
\langle v \rangle \equiv 4.2 \times 10^5 \text{ cm s}^{-1} .
\]

Thus the mean free path \( \ell \equiv 6 \times 10^{-8} \text{ cm} \); by reference to Fig. 3 we see that this is of the order of magnitude of the disorder of the structure.

This value of the phonon mean free path is remarkably short. At room temperature and above (that is, above the Debye temperature) most of the phonons have half-wavelengths of the order of the interatomic spacing. It is through phase cancellation processes, as in Fig. 9, that the mean free path is limited to several interatomic spacings. No other structure for fused quartz will give a 6 Å mean free path.

The normal modes of vibration of the glass structure are utterly unlike plane waves. But the modes, as distorted as they may seem, still have quantized amplitudes and may properly be called phonons. The thermal conductivity below room temperature and particularly below 1 K is treated in detail in the book edited by W. A. Phillips that is cited at the end of this chapter.

FIBER OPTICS

Fibers of silica-based lightguides carry a high proportion of the data and information transmitted on the surface of the earth and under the seas, and are the medium of choice for the communications industry. The optical fibers consist of a thin core (\( \approx 10 \mu\text{m} \)) of high-refractive index glass surrounded by a cladding.
The digital data are carried by light, with a minimum attenuation near 0.20 db km\(^{-1}\) at wavelengths near 1.55 \(\mu\)m, which is in the infrared (Figure 10). A range of 100 km corresponds to a loss of 20 db, which is readily supplied by an \(\text{Eu}^{3+}\) laser amplifier.

The optic window of high-purity glasses near this wavelength is limited on the low frequency side by phonon absorption bands and on the high frequency side by Rayleigh scattering and, ultimately, by electronic absorption. In the optic window the losses are determined by the Rayleigh scattering intrinsic to static fluctuations in the local dielectric constant of an inhomogeneous medium, and the attenuation varies as the fourth power of the frequency.

It is fortunate that an excellent source is available for radiation at 1.55 \(\mu\)m. As shown in Figure 16.24, excited (pumped) erbium \(\text{Er}^{3+}\) ions can amplify in an erbium-doped section of fiber.

**Rayleigh attenuation**

The attenuation of light waves in glass is dominated at short wavelengths in the infrared by the same scattering process, called Rayleigh scattering, that is responsible for the blue light of the sky. The extinction coefficient, or attenuation coefficient, \(h\) has the dimension of reciprocal length and for light scattered in a
gas is given after Rayleigh by
\[ h = (2\omega^4/3\pi c^4 N)(n - 1)^2, \]  
where \( n \) is the local refractive index and \( N \) is the number of scattering centers per unit volume. The energy flux as a function of distance has the form \( \exp(-hx) \).

Derivations of (30) are found in good texts on electrodynamics; the structure of the result may be understood by a general argument: The radiant energy scattered from a dipole element \( p \) is proportional to \( (dp^2/dt^2)^2 \), and this accounts for the factor \( \omega^4 \). The local polarizability \( \alpha \) enters as \( \alpha^0 \); if there are \( N \) random scattering centers per unit volume, the scattered energy averaged over these random sources will go as \( N(\langle\Delta\alpha\rangle^0)^2 \), or \( \langle(\Delta n)^2\rangle/N \). Thus we have the essential factors that appear in (30). As applied to a glass, \( \Delta n \) should refer to the variations in polarization around each group of Si—O bonds, and satisfactory numerical estimates of the attenuation may be made in this way.

**Problem**

1. *Metallic Optic Fibres.* In Nature 379, 124 (1996) it is speculated that metallic wires can act as optic fibres, transmitting light with a long delay appropriate to the high refractive index characteristic of metals. Unfortunately the refractive index of a typical metal is dominated by a free-electron term in \( n^2 \), so that the propagation of a light wave is in fact highly damped in a metal. Show that in sodium (p. 331) at room temperature a wave of vacuum wavelength 10 \( \mu \)m will have a damping length of 0.1 \( \mu \)m This may be contrasted with the 100 km damping length found for light in high quality glass fibres.

**References**


Journal of Non-Crystalline Solids.


18

Point Defects

LATTICE VACANCIES 541
DIFFUSION 544
  Metals 547
COLOR CENTERS 548
  F centers 548
  Other centers in alkali halides 548
PROBLEMS 551
  1. Frenkel defects 551
  2. Schottky vacancies 551
  3. F center 552
REFERENCES 552
Figure 1 A plane of a pure alkali halide crystal, showing a vacant positive ion site, a vacant negative ion site, and a coupled pair of vacant sites of opposite sign.
CHAPTER 18: POINT DEFECTS

The common point imperfections in crystals are chemical impurities, vacant lattice sites, and extra atoms not in regular lattice positions. Linear imperfections are treated under dislocations, Chapter 20. The crystal surface is a planar imperfection, with surface electron, phonon, and magnon states, Chapter 19.

Some important properties of crystals are controlled as much by imperfections as by the nature of the host crystal, which may act only as a solvent or matrix or vehicle for the imperfections. The conductivity of some semiconductors is due entirely to trace amounts of chemical impurities. The color and luminescence of many crystals arise from impurities or imperfections. Atomic diffusion may be accelerated enormously by impurities or imperfections. Mechanical and plastic properties are usually controlled by imperfections.

LATTICE VACANCIES

The simplest imperfection is a lattice vacancy, which is a missing atom or ion, also known as a Schottky defect. A lattice vacancy is often indicated in illustrations and in chemical equations by a square (Fig. 1). We create a Schottky defect in a perfect crystal by transferring an atom from a lattice site in the interior to a lattice site on the surface of the crystal. In thermal equilibrium, a certain number of lattice vacancies are always present in an otherwise perfect crystal, because the entropy is increased by the presence of disorder in the structure.

In metals with close-packed structures the proportion of lattice sites vacant at temperatures just below the melting point is of the order of $10^{-3}$ to $10^{-4}$. But in some alloys, in particular the very hard transition metal carbides such as TiC, the proportion of vacant sites of one component can be as high as 50 percent.

The probability that a given site is vacant is proportional to the Boltzmann factor for thermal equilibrium: $P = \exp(-E_V/k_B T)$, where $E_V$ is the energy required to take an atom from a lattice site inside the crystal to a lattice site on the surface. If there are $N$ atoms, the equilibrium number $n$ of vacancies is given by the Boltzmann factor

$$\frac{n}{N-n} = \exp(-E_V/k_B T). \quad (1)$$

If $n \ll N$, then

$$n/N \equiv \exp(-E_V/k_B T). \quad (2)$$

If $E_V \approx 1$ eV and $T \approx 1000$ K, then $n/N \approx e^{-12} \approx 10^{-5}$. 

54
The equilibrium concentration of vacancies decreases as the temperature decreases. The actual concentration of vacancies will be higher than the equilibrium value if the crystal is grown at an elevated temperature and then cooled suddenly, thereby freezing in the vacancies (see the discussion of diffusion below).

In ionic crystals it is usually favorable energetically to form roughly equal numbers of positive and negative ion vacancies. The formation of pairs of vacancies keeps the crystal electrostatically neutral on a local scale. From a statistical calculation we obtain

\[ n = N \exp(-E_v/2k_BT) \]

for the number of pairs, where \( E_v \) is the energy of formation of a pair.

Another vacancy defect is the Frenkel defect (Fig. 2) in which an atom is transferred from a lattice site to an interstitial position, a position not normally occupied by an atom. The calculation of the equilibrium number of Frenkel defects proceeds along the lines of Problem 1. If the number \( n \) of Frenkel defects is much smaller than the number of lattice sites \( N \) and the number of interstitial sites \( N' \), the result is

\[ n = (NN')^{1/2} \exp(-E_l/2k_BT) \]

where \( E_l \) is the energy necessary to remove an atom from a lattice site to an interstitial position.

In pure alkali halides the most common lattice vacancies are Schottky defects; in pure silver halides the most common vacancies are Frenkel defects.

Lattice vacancies are present in alkali halides when these contain additions of divalent elements. If a crystal of KCl is grown with controlled amounts of CaCl₂, the density varies as if a K⁺ lattice vacancy were formed for each Ca²⁺ ion in the crystal. The Ca²⁺ enters the lattice in a normal K⁺ site and the two Cl⁻ ions enter two Cl⁻ sites in the KCl crystal (Fig. 3). Demands of charge neutrality result in a vacant metal on site.

The experimental results show that the addition of CaCl₂ to KCl lowers the density of the crystal. The density would increase if no vacancies were produced, because Ca²⁺ is a heavier and smaller ion than K⁺.

The mechanism of electrical conductivity in alkali and silver halide crystals is usually the motion of ions and not the motion of electrons. This has been established by comparing the transport of charge with the transport of mass as measured by the material plated out on electrodes in contact with the crystal.

The study of ionic conductivity is an important tool in the investigation of lattice defects. Work on alkali and silver halides containing known additions of divalent metal ions shows that at not too high temperatures the ionic conductivity is directly proportional to the amount of divalent addition. This is not because the divalent ions are intrinsically highly mobile, for it is predominantly the monovalent metal ion which deposits at the cathode. The lattice vacancies introduced with the divalent ions are responsible for the enhanced diffusion...
Figure 2. Schottky and Frenkel defects in an ionic crystal. The arrows indicate the displacement of the ions. In a Schottky defect the ion ends up on the surface of the crystal, in a Frenkel defect it is removed to an interstitial position.

Figure 3. Production of a lattice vacancy by the solution of CaCl₂ in KCl: to ensure electrical neutrality a positive ion vacancy is introduced into the lattice with each divalent cation Ca^{++}. The two Cl⁻ ions of CaCl₂ enter normal negative ion sites.

Figure 4. Three basic mechanisms of diffusion: (a) Interchange by rotation about a midpoint. More than two atoms may rotate together. (b) Migration through interstitial sites. (c) Atoms exchange position with vacant lattice sites. (After Seitz.)

(Fig. 4c). The diffusion of a vacancy in one direction is equivalent to the diffusion of an atom in the opposite direction. When lattice defects are generated thermally, their energy of formation gives an extra contribution to the heat capacity of the crystal, as shown in Fig. 5.

An associated pair of vacancies of opposite sign exhibits an electric dipole moment, with contributions to the dielectric constant and dielectric loss due to the motion of pairs of vacancies. The dielectric relaxation time is a measure of
the time required for one of the vacant sites to jump by one atomic position with respect to the other. The dipole moment can change at low frequencies, but not at high. In sodium chloride the relaxation frequency is 1000 s\(^{-1}\) at 85°C.

**DIFFUSION**

When there is a concentration gradient of impurity atoms or vacancies in a solid, there will be a flux of these through the solid. In equilibrium the impurities or vacancies will be distributed uniformly. The net flux \(J_N\) of atoms of one species in a solid is related to the gradient of the concentration \(N\) of this species by a phenomenological relation called **Fick’s law**:

\[
J_N = -D \ \text{grad} \ N .
\] (5)

Here \(J_N\) is the number of atoms crossing unit area in unit time; the constant \(D\) is the **diffusion constant** or **diffusivity** and has the units cm\(^2\)/s or m\(^2\)/s. The minus sign means that diffusion occurs away from regions of high concentration. The form (5) of the law of diffusion is often adequate, but rigorously the gradient of the chemical potential is the driving force for diffusion and not the concentration gradient alone (TP, p. 406).

The diffusion constant is often found to vary with temperature as

\[
D = D_0 \ \exp(-E/k_B T) ;
\] (6)

here \(E\) is the **activation energy** for the process. Experimental results on the diffusion of carbon in alpha iron are shown in Fig. 6. The data are represented by \(E = 0.87\) eV, \(D_0 = 0.020\) cm\(^2\)/s. Representative values of \(D_0\) and \(E\) are given in Table 1.
Figure 6  Diffusion coefficient of carbon in iron, after Wert. The logarithm of $D$ is directly proportional to $1/T$.

Table 1  Diffusion constants and activation energies

<table>
<thead>
<tr>
<th>Host crystal</th>
<th>Atom</th>
<th>$D_0$ cm$^2$s$^{-1}$</th>
<th>$E$ eV</th>
<th>Host crystal</th>
<th>Atom</th>
<th>$D_0$ cm$^2$s$^{-1}$</th>
<th>$E$ eV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu</td>
<td>Cu</td>
<td>0.20</td>
<td>2.04</td>
<td>Si</td>
<td>Al</td>
<td>8.0</td>
<td>3.47</td>
</tr>
<tr>
<td>Cu</td>
<td>Zn</td>
<td>0.34</td>
<td>1.98</td>
<td>Si</td>
<td>Ga</td>
<td>3.6</td>
<td>3.51</td>
</tr>
<tr>
<td>Ag</td>
<td>Ag</td>
<td>0.40</td>
<td>1.91</td>
<td>Si</td>
<td>In</td>
<td>16.0</td>
<td>3.90</td>
</tr>
<tr>
<td>Ag</td>
<td>Cu</td>
<td>1.2</td>
<td>2.00</td>
<td>Si</td>
<td>As</td>
<td>0.32</td>
<td>3.56</td>
</tr>
<tr>
<td>Ag</td>
<td>Au</td>
<td>0.26</td>
<td>1.98</td>
<td>Si</td>
<td>Sb</td>
<td>5.6</td>
<td>3.94</td>
</tr>
<tr>
<td>Ag</td>
<td>Pb</td>
<td>0.22</td>
<td>1.65</td>
<td>Si</td>
<td>Li</td>
<td>$2 \times 10^{-3}$</td>
<td>0.66</td>
</tr>
<tr>
<td>Na</td>
<td>Na</td>
<td>0.24</td>
<td>0.45</td>
<td>Si</td>
<td>Au</td>
<td>$1 \times 10^{-3}$</td>
<td>1.13</td>
</tr>
<tr>
<td>U</td>
<td>U</td>
<td>$2 \times 10^{-3}$</td>
<td>1.20</td>
<td>Ge</td>
<td>Ge</td>
<td>10.0</td>
<td>3.1</td>
</tr>
</tbody>
</table>

To diffuse, an atom must overcome the potential energy barrier presented by its nearest neighbors. We treat the diffusion of impurity atoms between interstitial sites. The same argument will apply to the diffusion of vacant lattice sites.

If the barrier is of height $E$, the atom will have sufficient thermal energy to pass over the barrier a fraction $\exp(-E/k_BT)$ of the time. Quantum tunneling through the barrier is another possible process, but is usually important only for the lightest nuclei, particularly hydrogen.
If $\nu$ is a characteristic atomic vibrational frequency, then the probability $p$ that sometime during unit time the atom will have enough thermal energy to pass over the barrier is

$$p = \nu \exp(-E/k_BT) .$$

(7)

In unit time the atom makes $\nu$ passes at the barrier, with a probability $\exp(-E/k_BT)$ of surmounting the barrier on each try. The quantity $p$ is called the jump frequency.

We consider two parallel planes of impurity atoms in interstitial sites. The planes are separated by lattice constant $a$. There are $S$ impurity atoms on one plane and $(S + a\,dS/dx)$ on the other. The net number of atoms crossing between the planes in unit time is $\approx -pa\,dS/dx$. If $N$ is the total concentration of impurity atoms, then $S = aN$ per unit area of a plane.

The diffusion flux may now be written as

$$J_N = -pa^2(dN/dx) .$$

(8)

On comparison with (5) we have the result

$$D = \nu a^2 \exp(-E/k_BT) ,$$

(9)

of the form (6) with $D_0 = \nu a^2$.

If the impurities are charged, we may find the ionic mobility \( \bar{\mu} \) and the conductivity $\sigma$ from the diffusivity by using the Einstein relation $k_BT\bar{\mu} = qD$ from TP, p. 406:

$$\bar{\mu} = (q\nu a^2/k_BT) \exp(-E/k_BT) ;$$

(10)

$$\sigma = Nq\bar{\mu} = (Nq^2\nu a^2/k_BT) \exp(-E/k_BT) ,$$

(11)

where $N$ is the concentration of impurity ions of charge $q$.

The proportion of vacancies is independent of temperature in the range in which the number of vacancies is determined by the number of divalent metal ions. Then the slope of a plot of $\ln \sigma$ versus $1/k_BT$ gives $E_+$, the barrier activation energy for the jumping of positive ion vacancies (Table 2). Diffusion is very slow at low temperatures. At room temperature the jump frequency is of the order of $1$ s$^{-1}$, and at 100 K it is of the order of $10^{-25}$ s$^{-1}$.

The proportion of vacancies in the temperature range in which the concentration of defects is determined by thermal generation is given by

$$f = \exp(-E_f/2k_BT) ,$$

(12)

where $E_f$ is the energy of formation of a vacancy pair, according to the theory of Schottky or Frenkel defects. Here the slope of a plot of $\ln \sigma$ versus $1/k_BT$ will be $E_+ + \frac{1}{2}E_f$, according to (10) and (12). From measurements in different tem-
Table 2  Activation energy $E_+$ for motion of a positive ion vacancy

Values of the energy of formation of a vacancy pair, $E_f$, are also given. The numbers given in parentheses for the silver salts refer to interstitial silver ions.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>$E_+(eV)$</th>
<th>$E_f(eV)$</th>
<th>Workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>0.86</td>
<td>2.02</td>
<td>Etzel and Maurer</td>
</tr>
<tr>
<td>LiF</td>
<td>0.65</td>
<td>2.68</td>
<td>Haven</td>
</tr>
<tr>
<td>LiCl</td>
<td>0.41</td>
<td>2.12</td>
<td>Haven</td>
</tr>
<tr>
<td>LiBr</td>
<td>0.31</td>
<td>1.80</td>
<td>Haven</td>
</tr>
<tr>
<td>LiI</td>
<td>0.38</td>
<td>1.34</td>
<td>Haven</td>
</tr>
<tr>
<td>KCl</td>
<td>0.89</td>
<td>2.1–2.4</td>
<td>Wagner; Kelting and Witt</td>
</tr>
<tr>
<td>AgCl</td>
<td>0.39(0.10)</td>
<td>1.4$^a$</td>
<td>Tellow</td>
</tr>
<tr>
<td>AgBr</td>
<td>0.25(0.11)</td>
<td>1.1$^a$</td>
<td>Compton</td>
</tr>
</tbody>
</table>

$^a$For Frenkel defect.

Temperature ranges we determine the energy of formation of a vacancy pair $E_f$ and the jump activation energy $E_+$.

The diffusion constant can be measured by radioactive tracer techniques. The diffusion of a known initial distribution of radioactive ions is followed as a function of time or distance. Values of the diffusion constant thus determined may be compared with values from ionic conductivities. The two sets of values do not usually agree within the experimental accuracy, suggesting the presence of a diffusion mechanism that does not involve the transport of charge. For example, the diffusion of pairs of positive and negative ion vacancies does not involve the transport of charge.

Metals

Self-diffusion in monatomic metals most commonly proceeds by lattice vacancies. Self-diffusion means the diffusion of atoms of the metal itself, and not of impurities. The activation energy for self-diffusion in copper is expected to be in the range 2.4 to 2.7 eV for diffusion through vacancies and 5.1 to 6.4 eV for diffusion through interstitial sites. Observed values of the activation energy are 1.7 to 2.1 eV.

Activation energies for diffusion in Li and Na can be determined from measurements of the temperature dependence of the nuclear resonance line width. As discussed in Chapter 16, the resonance line width narrows when the jump frequency of an atom between sites becomes rapid in comparison with the frequency corresponding to the static line width. The values 0.57 eV and 0.45 eV were determined by NMR for Li and Na. Self-diffusion measurements for sodium also give 0.4 eV.
COLOR CENTERS

Pure alkali halide crystals are transparent throughout the visible region of the spectrum. The crystals may be colored in a number of ways:

• by the introduction of chemical impurities;
• by the introduction of an excess of the metal ion (we may heat the crystal in the vapor of the alkali metal and then cool it quickly—an NaCl crystal heated in the presence of sodium vapor becomes yellow; a KCl crystal heated in potassium vapor becomes magenta);
• by x-ray, γ-ray, neutron, and electron bombardment; and
• by electrolysis. A color center is a lattice defect that absorbs visible light. An ordinary lattice vacancy does not color alkali halide crystals, although it affects the absorption in the ultraviolet.

F Centers

The name F center comes from the German word for color, Farbe. We usually produce F centers by heating the crystal in excess alkali vapor or by x-irradiation. The central absorption band (F band) associated with F centers in several alkali halides are shown in Fig. 7, and the quantum energies are listed in Table 3. Experimental properties of F centers have been investigated in detail, originally by Pohl.

The F center has been identified by electron spin resonance as an electron bound at a negative ion vacancy (Fig. 8), in agreement with a model suggested by de Boer. When excess alkali atoms are added to an alkali halide crystal, a corresponding number of negative ion vacancies are created. The valence electron of the alkali atom is not bound to the atom; the electron migrates in the crystal and becomes bound to a vacant negative ion site. A negative ion vacancy in a perfect periodic lattice has the effect of an isolated positive charge: it attracts and binds an electron. We can simulate the electrostatic effect of a negative ion vacancy by adding a positive charge $q$ to the normal charge $-q$ of an occupied negative ion site.

The F center is the simplest trapped-electron center in alkali halide crystals. The optical absorption of an F center arises from an electric dipole transition to a bound excited state of the center.

Other Centers in Alkali Halides

In the $F_A$ center one of the six nearest neighbors of an F center has been replaced by a different alkali ion, Fig. 9. More complex trapped-electron centers are formed by groups of F centers, Fig. 10 and 11. Thus two adjacent F centers form an M center. Three adjacent F centers form an R center. Different centers are distinguished by their optical absorption frequencies.
Table 3  Experimental \( F \) center absorption energies, in eV

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LiCl</td>
<td>3.1</td>
<td>NaBr</td>
</tr>
<tr>
<td>NaCl</td>
<td>2.7</td>
<td>KBr</td>
</tr>
<tr>
<td>KCl</td>
<td>2.2</td>
<td>RbBr</td>
</tr>
<tr>
<td>RbCl</td>
<td>2.0</td>
<td>LiF</td>
</tr>
<tr>
<td>CsCl</td>
<td>2.0</td>
<td>NaF</td>
</tr>
<tr>
<td>LiBr</td>
<td>2.7</td>
<td>KF</td>
</tr>
</tbody>
</table>

Figure 7  The \( F \) bands for several alkali halides: optical absorption versus wavelength for crystals that contain \( F \) centers.

Figure 8  An \( F \) center is a negative ion vacancy with one excess electron bound at the vacancy. The distribution of the excess electron is largely on the positive metal ions adjacent to the vacant lattice site.
Holes may be trapped to form color centers, but hole centers are not usually as simple as electron centers. For example, a hole in the filled $p^6$ shell of a halogen ion leaves the ion in a $p^5$ configuration, whereas an electron added to the filled $p^6$ shell of an alkali ion leaves the ion in a $p^6s$ configuration.

The chemistry of the two centers is different: $p^6s$ acts as a spherically symmetric ion, but $p^5$ acts as an asymmetric ion and, by virtue of the Jahn-Teller effect, will distort its immediate surroundings in the crystal.
The antimorph to the $F$ center is a hole trapped at a positive ion vacancy, but no such center has been identified experimentally in alkali halides; in insulating oxides the $O^-$ (called $V^-$) defect is known. The best-known trapped-hole center is the $V_K$ center, Fig. 12. The $V_K$ center is formed when a hole is trapped by a halogen ion in an alkali halide crystal. Electron spin resonance shows that the center is like a negative halogen molecular ion, such as $\text{Cl}_2^-$ in KCl. The Jahn-Teller trapping of free holes is the most effective form of self-trapping of charge carriers in perfect crystals.

**Problems**

1. **Frenkel defects.** Show that the number $n$ of interstitial atoms in equilibrium with $n$ lattice vacancies in a crystal having $N$ lattice points and $N'$ possible interstitial positions is given by the equation

$$E_i = k_B T \ln \left[ \frac{(N - n)(N' - n)}{n^2} \right],$$

whence, for $n \ll N, N'$, we have $n \approx (NN')^{1/2} \exp(-E_i/2k_BT)$. Here $E_i$ is the energy necessary to remove an atom from a lattice site to an interstitial position.

2. **Schottky vacancies.** Suppose that the energy required to remove a sodium atom from the inside of a sodium crystal to the boundary is 1 eV. Calculate the concentration of Schottky vacancies at 300 K.
3. *F center.* (a) Treat an *F* center as a free electron of mass *m* moving in the field of a point charge *e* in a medium of dielectric constant *ε* = *n*²; what is the 1s-2p energy difference of *F* centers in NaCl? (b) Compare from Table 3 the *F* center excitation energy in NaCl with the 3s-3p energy difference of the free sodium atom.

**References**


Crystal lattice defects. A journal.
19

Surface and Interface Physics

Reconstruction and relaxation 555

SURFACE CRYSTALLOGRAPHY 556
Reflection high-energy electron diffraction 560

SURFACE ELECTRONIC STRUCTURE 560
Work function 560
Thermionic emission 561
Surface states 562
Tangential surface transport 564

MAGNETORESISTANCE IN A TWO-DIMENSIONAL CHANNEL 564
Integral Quantized Hall effect (IQHE) 566
IQHE in real systems 568
Fractional quantized Hall effect (FQHE) 570

p-n JUNCTIONS 570
Rectification 572
Solar cells and photovoltaic detectors 573
Schottky barrier 573

HETEROSTRUCTURES 574
n-N heterojunction 576

SEMICONDUCTOR LASERS 577

LIGHT-EMITTING DIODES 579

SCANNING TUNNELING MICROSCOPY 580

PROBLEMS 581
1. Diffraction from a linear array and square array 581
2. Surface subbands in electric quantum limit 583
3. Shubnikov-de Haas effect in a surface channel 583

REFERENCES 584
Figure 1  Dangling bonds from the (111) surface of a covalently bonded diamond cubic structure. (After M. Prutton, Surface physics, Clarendon, 1975.)
Reconstruction and Relaxation

The surface of a crystalline solid in vacuum is generally defined as the few, approximately three, outermost atomic layers of the solid that differ significantly from the bulk. The surface may be entirely clean or it may have foreign atoms deposited on it or incorporated in it. The bulk of the crystal is called the substrate.

If the surface is clean the top layer may be either reconstructed or, sometimes, unreconstructed. In unreconstructed surfaces the atomic arrangement is in registry with that of the bulk except for an interlayer spacing change (called multilayer relaxation) at the top surface.

The shrinking of the interlayer distance between the first and second layer of atoms with respect to subsequent layers in the bulk is a rather dominant phenomenon. The surface may be thought of as an intermediate between the diatomic molecule and the bulk structure. Because the interatomic distances in diatomic molecules are much smaller than in the bulk, there is a rationale for the surface relaxation. This may be contrasted with reconstruction where the relaxation of atoms yields new surface primitive cells. In relaxation the atoms maintain their structure in the surface plane as it was according to the projection of the bulk cell on the surface; only their distance from the bulk changes.

Experimental values of the top-layer relaxation for metals are given in Table 1. Note the frequency of occurrence of contractions in the interlayer spacing at the surface.

Sometimes in metals, but most often in nonmetals, the atoms in the surface layer form superstructures in which the atoms in the layer are not in registry with the atoms in corresponding layers in the substrate. Surface reconstruction can be a consequence of a rearrangement of broken covalent or ionic bonds at the surface. Under such conditions the atoms at the surface bunch into rows with alternately larger and smaller spacings than in the bulk. That is, for some crystals held together by valence bonds, creation of a surface would leave unsaturated bonds dangling into space (Fig. 1). The energy may then be lowered if neighboring atoms approach each other and form bonds with their otherwise unused valence electrons. Atomic displacements can be as large as 0.5 Å.

Reconstruction does not necessarily require formation of a superstructure. For example, on GaAs (110) surfaces a Ga-As bond rotation occurs that leaves the point group intact. The driving force is electron transfer from Ga to As, which fills the dangling bonds on As and depletes them on Ga.

Surfaces of planes nominally of high indices may be built up of low index
Table 1 Relaxation of topmost interlayer spacing at unreconstructed clean metal surfaces

<table>
<thead>
<tr>
<th>Structure</th>
<th>Change in Interlayer Spacing</th>
</tr>
</thead>
<tbody>
<tr>
<td>hcp(0001)</td>
<td>Re (-5%), Sc (-2%), Ti (-2%), Zr (-1%)</td>
</tr>
<tr>
<td>fcc(111)</td>
<td>Al (+1%), Ag (0%), Cu (-0.7%), Pt (+1%), Rh (0%)</td>
</tr>
<tr>
<td>bcc(110)</td>
<td>Fe (+0.5%), Na (0%), V (-0.3%), W (0%)</td>
</tr>
<tr>
<td>fcc(100)</td>
<td>Al (0%), Cu (-1%), Rh (0%)</td>
</tr>
<tr>
<td>bcc(100)</td>
<td>Fe (-5%), Mo (-9.5%), Ta (-11%), V (-7%), W (-8%)</td>
</tr>
<tr>
<td>fcc(110)</td>
<td>Al (-8.5%), Ag (-8%), Cu (-8.5%), Ni (-8.5%), Pb (-16%), Rh (-3%)</td>
</tr>
<tr>
<td>hcp(10-10)</td>
<td>Re (-17%)</td>
</tr>
<tr>
<td>bcc(211)</td>
<td>Fe (-10%), W (-12%)</td>
</tr>
<tr>
<td>fcc(311)</td>
<td>Al (-13%), Ni (-16%), Cu (-5%)</td>
</tr>
<tr>
<td>bcc(310)</td>
<td>Fe (-16%)</td>
</tr>
<tr>
<td>fcc(331)</td>
<td>Al (-12%)</td>
</tr>
<tr>
<td>fcc(210)</td>
<td>Al (-15.5%)</td>
</tr>
<tr>
<td>bcc(111)</td>
<td>Fe (-17%)</td>
</tr>
<tr>
<td>bcc(210)</td>
<td>Fe (-22%)</td>
</tr>
</tbody>
</table>

planes separated by steps one (or two) atoms in height. Such terrace-step arrangements are important in evaporation and desorption because the attachment energy of atoms is often low at the steps and at kinks in the steps. The chemical activity of such sites may be high. The presence of periodic arrays of steps may be detected by double and triple beams of diffraction in LEED (see below) experiments.

**SURFACE CRYSTALLOGRAPHY**

The surface structure is in general diperiodic. This does not necessarily mean that all its atoms lie in a plane, but rather that the structure is periodic only in two dimensions. The surface structure can be the structure of foreign material deposited on the substrate or it can be the selvage of the pure substrate.

In Chapter 1 we used the term Bravais lattice for the array of equivalent points in two or in three dimensions, that is, for diperiodic or triperiodic structures. In the physics of surfaces it is common to speak of a two-dimensional lattice. Further, the area unit may be called a mesh.

We showed in Fig. 1.9 four of the five nets possible for a diperiodic struc-
ture; the fifth net is the general oblique net, with no special symmetry relation between the mesh basis vectors $a_1$, $a_2$. Thus the five distinct nets are the oblique, square, hexagonal, rectangular, and centered rectangular.

The substrate net parallel to the surface is used as the reference net for the description of the surface. For example, if the surface of a cubic substrate crystal is the (111) surface, the substrate net is hexagonal (Fig. 1.9b), and the surface net is referred to these axes.

The vectors $c_1$, $c_2$ that define the mesh of the surface structure may be expressed in terms of the reference net $a_1$, $a_2$ by a matrix operation $P$:

$$
\begin{pmatrix}
c_1 \\
c_2
\end{pmatrix} = P
\begin{pmatrix}
a_1 \\
a_2
\end{pmatrix} =
\begin{pmatrix}
P_{11} & P_{12} \\
P_{21} & P_{22}
\end{pmatrix}
\begin{pmatrix}
a_1 \\
a_2
\end{pmatrix}.
$$

(1)

Provided that the included angles of the two meshes are equal, the shorthand notation due to E. A. Wood may be used. In this notation, which is widely used, the relation of the mesh $c_1$, $c_2$ to the reference mesh $a_1$, $a_2$ is expressed as

$$
\begin{pmatrix}
c_1 \\
c_2
\end{pmatrix} = \left( \frac{c_1}{a_1} \times \frac{c_2}{a_2} \right) R \alpha,
$$

in terms of the lengths of the mesh basis vectors and the angle $\alpha$ of relative rotation $R$ of the two meshes. If $\alpha = 0$, the angle is omitted. Examples of the Wood notation are given in Fig. 2.

The reciprocal net vectors of the surface mesh may be written as $c_1^*$, $c_2^*$, defined by

$$
c_1 \cdot c_1^* = c_2 \cdot c_2^* = 0 ; \quad c_1 \cdot c_2^* = c_2 \cdot c_1^* = 2\pi \text{ (or 1)}.
$$

Here the $2\pi$ (or 1) indicates that two conventions are in use. The definitions (3) used in Fig. 3 may be compared with the definitions (2.10) and (2.11) for the reciprocal lattice vectors of a triperiodic lattice.

The reciprocal net points of a diperiodic net may be thought of—when we are in three dimensions—as rods. The rods are infinite in extent and normal to the surface plane, where they pass through the reciprocal net points. It may be helpful to think of the rods as generated by a triperiodic lattice which is expanded without limit along one of its axes. Then the reciprocal lattice points along this axis are moved closer together and in the limit form a rod.

The usefulness of the rod concept comes out with the Ewald sphere construction explained in Fig. 2.8. Diffraction occurs everywhere the Ewald sphere intercepts a reciprocal net rod. Each diffracted beam is labelled with the indices $hk$ of the reciprocal net vector

$$
g = h c_1^* + k c_2^*
$$

(4)
forming the beam.

Low energy electron diffraction (LEED) is illustrated by Fig. 4. The elec-
Figure 2  Surface nets of adsorbed atoms. The circles represent atoms in the top layer of the substrate. In (a) the designation fcc(111) means the (111) face of an fcc structure. This face determines a reference net. The lines represent ordered overlays with adatoms at the intersections of two lines. The intersection points represent diperiodic nets (lattices in two dimensions). The designation p(1 × 1) in (a) is a primitive mesh unit for which the basis is identical with the basis of the reference net. In (b) the c(2 × 2) mesh unit is a centered mesh with basis vectors twice as long as those of the reference net. Atomic adsorption on metals takes place most often into those surface sites (hollow sites) that maximize the number of nearest-neighbor atoms on the substrate. (After Van Hove.)
Figure 3  A \((3 \times 1)\) surface structure, (a) real-space; and (b) reciprocal-space diagrams. (After E. A. Wood.)

Figure 4  Ewald sphere construction for diffraction of incident wave \(k\) by a square net, when \(k\) is parallel to one axis of the mesh. The back scattered beams in the plane of the paper are \(k'_4, k'_6, k'_8, k'_7\). Diffracted beams out of the plane of the paper will also occur, such as \((12)\) and \((12)\). The vertical lines are the rods of the reciprocal net.
tron energy is typically in the range 10–1000 eV. This is the arrangement with which Davisson and Germer in 1927 discovered the wave nature of the electron. An experimental pattern is shown in Fig. 5.

**Reflection High-Energy Electron Diffraction.** In the RHEED method a beam of high-energy electrons is directed upon a crystal surface at grazing incidence. By adjustment of the angle of incidence one can arrange the normal component of the incoming wavevector to be very small, which will minimize the penetration of the electron beam and enhance the role of the crystal surface.

The radius $k$ of the Ewald sphere for 100 keV electrons will be $\approx 10^3 \, \text{Å}^{-1}$, which is much longer than the shortest reciprocal lattice vector $2\pi/a \approx 1 \, \text{Å}^{-1}$. It follows that the Ewald sphere will be nearly a flat surface in the central scattering region. The intercept of the rods of the reciprocal net with the nearly flat sphere will be nearly a line when the beam is directed at grazing incidence. The experimental arrangement is shown in Fig. 6.

**SURFACE ELECTRONIC STRUCTURE**

**Work Function**

The work function $W$ of the uniform surface of a metal is defined as the difference in potential energy of an electron between the vacuum level and the Fermi level. The vacuum level is the energy of an electron at rest at a point sufficiently far outside the surface so that the electrostatic image force on the
electron may be neglected—more than 100 Å from the surface. The Fermi level is the electrochemical potential of the electrons in the metal.

Typical values of electron work functions are given in Table 2. The orientation of the exposed crystal face affects the value of the work function because the strength of the electric double layer at the surface depends on the concentration of surface positive ion cores. The double layer exists because the surface ions are in an asymmetrical environment, with vacuum (or an adsorbed foreign atom layer) on one side and the substrate on the other side.

The work function is equal to the threshold energy for photoelectric emission at absolute zero. If $\hbar \omega$ is the energy of an incident photon, then the Einstein equation is $\hbar \omega = W + T$, where $T$ is the kinetic energy of the emitted electron and $W$ is the work function.

**Thermionic Emission**

The rate of emission of thermionic electrons depends exponentially on the work function. The derivation follows.

We first find the electron concentration in vacuum in equilibrium with electrons in a metal at temperature $\tau (= k_B T)$ and chemical potential $\mu$. We treat the electrons in the vacuum as an ideal gas, so that their chemical potential is

$$\mu = \mu_{\text{ext}} + \tau \log(n/n_0) , \tag{5}$$

by *TP*, Chapter 5. Here
Table 2  Electron work functions

(Values obtained by photoemission, except tungsten obtained by field emission.)

<table>
<thead>
<tr>
<th>Element</th>
<th>Surface plane</th>
<th>Work function, in eV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag</td>
<td>(100)</td>
<td>4.64</td>
</tr>
<tr>
<td></td>
<td>(110)</td>
<td>4.52</td>
</tr>
<tr>
<td></td>
<td>(111)</td>
<td>4.74</td>
</tr>
<tr>
<td>Cs</td>
<td>polycrystal</td>
<td>2.14</td>
</tr>
<tr>
<td>Cu</td>
<td>(100)</td>
<td>4.59</td>
</tr>
<tr>
<td></td>
<td>(110)</td>
<td>4.48</td>
</tr>
<tr>
<td></td>
<td>(111)</td>
<td>4.98</td>
</tr>
<tr>
<td>Ce</td>
<td>(111)</td>
<td>4.80</td>
</tr>
<tr>
<td>Ni</td>
<td>(100)</td>
<td>5.22</td>
</tr>
<tr>
<td></td>
<td>(110)</td>
<td>5.04</td>
</tr>
<tr>
<td></td>
<td>(111)</td>
<td>5.35</td>
</tr>
<tr>
<td>W</td>
<td>(100)</td>
<td>4.63</td>
</tr>
<tr>
<td></td>
<td>(110)</td>
<td>5.25</td>
</tr>
<tr>
<td></td>
<td>(111)</td>
<td>4.47</td>
</tr>
</tbody>
</table>

*After H. D. Hagstrom

\[ n_Q = 2\left(\frac{m^*}{2\pi\hbar^2}\right)^{3/2} \]  

(6)

for particles of spin 1/2.

Now \( \mu_{\text{ext}} - \mu = W \), by the definition of the work function \( W \). Thus, from (5),

\[ n = n_Q \exp(-W/\eta) \]  

(7)

The flux that leaves the metal surface when all electrons are drawn off is equal to the flux incident on the surface from outside:

\[ J_n = \frac{1}{2}n\bar{c} = (\pi/2\pi m)^{1/2} n \]  

(8)

by TP(14.95) and (14.121). Here \( \bar{c} \) is the mean speed of the electrons in the vacuum. The electric charge flux is \( eJ_n \) or

\[ J_e = (\pi^2 me/2\pi^2 h^3)\exp(-W/\eta) \]  

(9)

This is called the Richardson-Dushman equation for thermionic emission.

**Surface States**

At the free surface of a semiconductor there often exist surface-bound electronic states with energies in the forbidden gap between the valence and conduction bands of the bulk semiconductor. We can obtain a good impression of the nature of the surface states by considering the wave functions in the weak
binding or two-component approximation of Chapter 7, in one dimension. (The
wave functions in three dimensions will have extra factors \( \exp[i(k_y y + k_z z)] \) in
the \( y,z \) plane of the surface.)

If the vacuum lies in the region \( x > 0 \), the potential energy of an electron
in this region can be set equal to zero:

\[
U(x) = 0, \quad x > 0.
\]  

(10)

In the crystal the potential energy has the usual periodic form:

\[
U(x) = \sum_G U_G \exp(iGx), \quad x < 0.
\]  

(11)

In one dimension \( G = n\pi/a \), where \( n \) is any integer, including zero.

In the vacuum the wave function of a bound surface state must fall off exponentially:

\[
\psi_{\text{out}} = \exp(-sx), \quad x > 0.
\]  

(12)

By the wave equation the energy of the state referred to the vacuum level is

\[
e = -\hbar^2 s^2/2m.
\]  

(13)

Within the crystal the two-component wave function of a bound surface
state will have the form, for \( x < 0 \),

\[
\psi_{\text{in}} = \exp(qx + ikx)[C(k) + C(k - G)\exp(-iGx)]
\]  

(14)

by analogy with (7.49), but with the addition of the factor \( \exp(qx) \) which serves
to bind the electron to the surface.

We now come to an important consideration that restricts the allowed
values of the wavevector \( k \). If the state is bound there can be no current flow in
the \( x \) direction, normal to the surface. This condition is assured in quantum
mechanics if the wave function can be written as a real function of \( x \), a condition
already satisfied by the exterior wave function (12). But (14) can be a real
function only if \( k = \frac{1}{2}G \), so that

\[
\psi_{\text{in}} = \exp(qx)[C(\frac{1}{2}G)\exp(iGx/2) + C(-\frac{1}{2}G)\exp(-iGx/2)]
\]  

(15)

This is real provided \( C^*(\frac{1}{2}G) = C(-\frac{1}{2}G) \). Thus \( k_x \) for a surface state does not
have a continuum of values, but is limited to discrete states associated with
Brillouin zone boundaries.

The state (15) is damped exponentially in the crystal. The constants \( s, q \) are
related by the condition that \( \psi \) and \( d\psi/dx \) are continuous at \( x = 0 \). The binding
energy \( \epsilon \) is determined by solving the two-component secular equation\(^1\) analogous to (7.46). The plot of Fig. 7.12 is helpful in this connection.

Tangential Surface Transport

We have seen that there may exist surface-bound electronic states with energies in the forbidden gap between the valence and conduction bands of the substrate crystal. These states may be occupied or vacant; their existence must affect the statistical mechanics of the problem. This means that the states modify the local equilibrium concentration of electrons and holes, as expressed as a shift of the chemical potential relative to the band edges. Because the chemical potential is independent of position in an equilibrium system, the energy bands must be displaced or bent, as in Fig. 7.

The thickness and carrier concentration in the surface layer may be changed by applying an electric field normal to the surface. The effect of an external field is utilized in the metal-oxide-semiconductor field-effect transistor (MOSFET). This has a metal electrode just outside the semiconductor surface and insulated from it by a layer of oxide (Fig. 8). A voltage, the gate voltage \( V_g \), is applied between the metal and the bulk semiconductor and modulates the conductance between any other electrodes placed in contact with the electrons of the surface space charge. The physics of the surface conductance channel is treated in Problems 2 and 3.

MAGNETORESISTANCE IN A TWO-DIMENSIONAL CHANNEL

The static magnetoconductivity tensor in 3D was found in Problem 6.9. Here we translate that result to a 2D surface conductance channel in the \( xy \) plane, with the static magnetic field in the \( z \) direction, normal to the MOS layer. We assume \( n = N/L^2 \) electrons per unit area. The surface conductance is de-
defined as the volume conductivity times the layer thickness. The surface current density is defined as the current crossing a line of unit length in the surface.

Thus, with (6.43) and (6.65), the surface tensor conductance components become

\[
\sigma_{xx} = \frac{\sigma_0}{1 + (\omega_c \tau)^2} \;
\sigma_{xy} = \frac{\sigma_0 \omega_c \tau}{1 + (\omega_c \tau)^2},
\]

where \( \sigma_0 = ne^2/\tau/m \) and \( \omega_c = eB/mc \) in CGS and \( eB/m \) in SI. The following discussion is written in CGS only, except where ohms are used.

These results apply specifically in the relaxation time approximation used in Chapter 6. When \( \omega_c \tau \gg 1 \), as for strong magnetic field and low temperatures, the surface conductivity components approach the limits

\[
\sigma_{xx} = 0; \quad \sigma_{xy} = nec/B.
\]

The limit for \( \sigma_{xy} \) is a general property of free electrons in crossed electric \( E_y \) and magnetic fields \( B_z \). We establish the result that such electrons drift in the \( x \) direction with velocity \( v_D = eE_y/B_z \). Consider the electrons from a Lorentz frame that moves in the \( x \) direction with this velocity. By electromagnetic theory there is in this frame an electric field \( E'_y = -v_D B_z/c \) that will cancel the applied field \( E_y \) for the above choice of \( v_D \). Viewed in the laboratory frame, all electrons drift in the \( x \) direction with velocity \( v_D \) in addition to any velocity components they had before \( E_y \) was applied.

Thus \( j_x = \sigma_{xy} E_y = nev_D = (neB)E_y \), so that

\[
\sigma_{xy} = nec/B
\]
as in (17). The experiments measure the voltage \( V \) in the \( y \) direction and the current \( I \) in the \( x \) direction (Fig. 9). Here \( I_x = j_x L_y = (neB)(E_y L_y) = (nec/B)V_y \). The Hall resistance is

\[
\rho_H = V_y/I_x = B/nec\ .
\]

(18a)
We see that \( j_x \) can flow with zero \( E_x \), so that the effective conductance can be infinite. Paradoxically, this limit occurs only when \( \sigma_{xx} \) and \( \sigma_{yy} \) are zero. Consider the tensor relations
\[
    j_x = \sigma_{xx} E_x + \sigma_{xy} E_y; \quad j_y = \sigma_{yx} E_x + \sigma_{yy} E_y.
\]
In the Hall effect geometry \( j_y = 0 \), so that \( E_y = (\sigma_{xy}/\sigma_{yy})E_x \), with \( \sigma_{xy} = -\sigma_{yx} \). Thus
\[
    j_x = (\sigma_{xx} + \sigma_{xy}^2/\sigma_{yy})E_x = \sigma(\text{eff}) E_x,
\]
and in the limit \( \sigma_{xx} = \sigma_{yy} = 0 \) the effective conductance is infinite.

**Integral Quantized Hall Effect (IQHE)**

The results of the original measurements\(^2\) under quantum conditions of temperature and magnetic field are shown in Fig. 10. The results are remarkable: at certain values of the gate voltage the voltage drop in the direction of current flow goes essentially to zero, as if the effective conductance were infinite. Further, there are plateaus of the Hall voltage near these same values of gate voltage, and the values of the Hall resistivity \( U_H/I_x \) at these plateaus are accurately equal to \((25,813/\text{integer})\) ohms, where 25,813 is the value of \( h/e^2 \) expressed in ohms.

The IQHE voltage minima \( V_{pp} \) may be explained on a model that is, however, oversimplified. Later we give a general theory. Apply a strong magnetic field such that the separation \( \hbar \omega_c \gg k_B T \). It is meaningful to speak of Landau levels that are completely filled or completely empty. Let the electron surface concentration (proportional to the gate voltage) be adjusted to any of the set of values that cause the Fermi level to fall at a Landau level: from (9.33) and (9.34),
\[
    seB_x/hc = \nu,
\]
where \( s \) is any integer and \( \nu \) is the electron surface concentration.

When the above conditions are satisfied, the electron collision time is greatly enhanced. No elastic collisions are possible from one state to another state in the same Landau level because all possible final states of equal energy are occupied. The Pauli principle prohibits an elastic collision. Inelastic collisions to a vacant Landau level are possible with the absorption of the necessary energy from a phonon, but there are very few thermal phonons of energy greater than the interlevel spacing by virtue of the assumption \( \hbar \omega_c \gg k_B T \).

The quantization of the Hall resistance follows on combining (18a) and (21):
\[
    \rho_H = \hbar/se^2 = 2\pi/sca,
\]
where \( \alpha \) is the fine structure constant \( e^2/\hbar c \equiv 1/137 \), and \( s \) is an integer.

Figure 9 Applied field $E_y$ and drift current $I_x$ in a quantum Hall effect (IQHE) experiment.

Figure 10 In the original IQHE measurements a magnetic field of 180 kG (18 T) points out of the paper. The temperature is 1.5 K. A constant current of 1 $\mu$A is forced to flow between the source and the drain. Voltages $V_{pp}$ and $V_H$ are plotted versus the gate voltage $V_G$, which is proportional to the Fermi level. (After K. von Klitzing, G. Dorda, and M. Pepper.)
**IQHE in Real Systems**

The measurements (Fig. 10) suggest that the above theory of the IQHE is too good. The Hall resistivity is accurately quantized at 25,813/s ohms, whether or not the semiconductor is of very high purity and perfection. The sharp Landau levels (Fig. 11a) are broadened in the real crystal (Fig. 11b), but this does not affect the Hall resistivity. The occurrence of plateaus in the Hall resistance, evident in the $U_H$ curve of Fig. 10, is not expected in ideal systems because partially filled Landau levels will exist for all gate voltages except those for which the Fermi level exactly coincides with a Landau level. Yet the experiments show that a range of $V_e$ values gives the exact Hall resistance.

Laughlin$^3$ interpreted the results for real systems as the expression of the general principle of gauge invariance. The argument is subtle and somewhat reminiscent of the flux quantization in a superconductor in Chapter 12.

In Laughlin’s thought-experiment the 2D electron system is bent to form a cylinder (Fig. 12) whose surface is pierced everywhere by a strong magnetic field $B$ normal to the surface. The current $I$ (former $I_x$) circles the loop. The magnetic field $B$ acts on the charge carriers to produce a Hall voltage $V_H$ (former $V_y$) perpendicular to the current and to $B$—that is, $V_H$ is developed between one edge of the cylinder and the other.

The circulating current $I$ is accompanied by a small magnetic flux $\phi$ that threads the current loop. The aim of the thought-experiment is to find the relation between $I$ and $V_H$. We start with the electromagnetic relation that relates $I$ to the total energy $U$ of a resistanceless system:

$$\frac{\partial U}{\partial t} = -V_x I_x = \frac{I}{c} \frac{\partial \phi}{\partial t}; \quad I = c \frac{\delta U}{\delta \phi}. \quad (23)$$

The value of $I$ can now be found from the variation $\delta U$ of the electronic energy that accompanies a small variation $\delta \phi$ of the flux.

The carrier states divide into two classes:

- Localized states, which are not continuous around the loop.
- Extended states, continuous around the loop.

Localized and extended states cannot coexist at the same energy, according to our present understanding of localization.

The two classes of states respond differently to the application of the flux $\phi$. The localized states are unaffected to first order because they do not enclose any significant part of $\phi$. To a localized state a change in $\phi$ looks like a gauge transformation, which cannot affect the energy of the state.

---

Figure 11  Density of states in a 2D electron gas in a strong magnetic field. (a) Ideal 2D crystal. (b) Real 2D crystal, with impurities and imperfections.

Figure 12  Geometry for Laughlin's thought-experiment. The 2D electron system is wrapped around to form a cylinder. A strong magnetic field $B$ pierces the cylinder everywhere normal to its surface. A current $I$ circles the loop, giving rise to the Hall voltage $V_H$ and a small magnetic flux $\phi$ through the loop.

The extended states enclose $\phi$, and their energy may be changed. However, if the magnetic flux is varied by a flux quantum, $\delta \phi = h c / e$, all extended orbits are identical to those before the flux quantum was added. The argument here is identical to that for the flux quantization in the superconducting ring treated in Chapter 12, but with the $2e$ of the Cooper pair replaced by $e$.

If the Fermi level falls within the localized states of Fig. 11b, all extended states (Landau levels) below the Fermi level will be filled with electrons both before and after the flux change $\delta \phi$. However, during the change an integral number of states, generally one per Landau level, enter the cylinder at one edge and leave it at the opposite edge.

The number must be integral because the system is physically identical before and after the flux change. If the transferred state is transferred while occupied by one electron, it contributes an energy change $e V_H$; if $N$ occupied states are transferred, the energy change is $N e V_H$.

This electron transfer is the only way the degenerate 2D electron system
can change its energy. We can understand the effect by looking at a model system without disorder in the Landau gauge for the vector potential:

$$A = -By \hat{x}.$$  

(24)

An increase $\delta A$ that corresponds to the flux increase $\delta \varphi$ is equivalent to a displacement of an extended state by $\delta A/B$ in the $y$ direction. By the Stokes theorem and the definition of the vector potential we have $\delta \varphi = L_\alpha \delta A$. Thus $\delta \varphi$ causes a motion of the entire electron gas in the $y$ direction.

By $\delta U = NeV_H$ and $\delta \varphi = hc/e$, we have

$$I = c(\Delta U/\Delta \varphi) = cNe^2V_H/hc = (Ne^2/h)V_H,$$  

(25)

so that the Hall resistance is

$$\rho_H = V_H/I = h/Ne^2,$$  

(26)

as in (22).

**Fractional Quantized Hall Effect (FQHE).** A quantized Hall effect has been reported for similar systems at fractional values of the index $s$, by working at lower temperatures and higher magnetic fields. In the extreme quantum limit the lowest Landau level is only partially occupied, and the integral QHE treated above should not occur. It has been observed, however, that the Hall resistance $\rho_H$ is quantized in units of $3h/e^2$ when the occupation of the lowest Landau level is $1/3$ and $2/3$, and $\rho_{xx}$ vanishes for these occupations. Similar breaks have been reported for occupations of $2/5$, $3/5$, $4/5$, and $2/7$.

**p-n JUNCTIONS**

A $p-n$ junction is made from a single crystal modified in two separate regions. Acceptor impurity atoms are incorporated into one part to produce the $p$ region in which the majority carriers are holes. Donor impurity atoms in the other part produce the $n$ region in which the majority carriers are electrons. The interface region may be less than $10^{-4}$ cm thick. Away from the junction region on the $p$ side there are $(-)$ ionized acceptor impurity atoms and an equal concentration of free holes. On the $n$ side there are $(+)$ ionized donor atoms and an equal concentration of free electrons. Thus the majority carriers are holes on the $p$ side and electrons on the $n$ side, Fig. 13.

Holes concentrated on the $p$ side would like to diffuse to fill the crystal uniformly. Electrons would like to diffuse from the $n$ side. But diffusion will upset the local electrical neutrality of the system.

---

A small charge transfer by diffusion leaves behind on the $p$ side an excess of $(-)$ ionized acceptors and on the $n$ side an excess of $(+)$ ionized donors. This charge double layer creates an electric field directed from $n$ to $p$ that inhibits diffusion and thereby maintains the separation of the two carrier types. Because of this double layer the electrostatic potential in the crystal takes a jump in passing through the region of the junction.

In thermal equilibrium the chemical potential of each carrier type is everywhere constant in the crystal, even across the junction. For holes

$$k_B T \ln p(r) + e\varphi(r) = \text{constant} ,$$  \hspace{1cm} (27a)

where $p$ is the hole concentration and $\varphi$ the electrostatic potential. Thus $p$ is low where $\varphi$ is high. For electrons

$$k_B T \ln n(r) - e\varphi(r) = \text{constant} ,$$  \hspace{1cm} (27b)

and $n$ will be low where $\varphi$ is low.

The total chemical potential is constant across the crystal. The effect of the concentration gradient exactly cancels the electrostatic potential, and the net particle flow of each carrier type is zero. However, even in thermal equilibrium there is a small flow of electrons from $n$ to $p$ where the electrons end their lives by recombination with holes. The recombination current $J_{nr}$ is balanced by a current $J_{ng}$ of electrons which are generated thermally in the $p$ region and which are pushed by the built-in field to the $n$ region. Thus in zero external applied electric field

$$J_{nr}(0) + J_{ng}(0) = 0 ,$$  \hspace{1cm} (28)
for otherwise electrons would accumulate indefinitely on one side of the barrier.

**Rectification**

A p-n junction can act as a rectifier. A large current will flow if we apply a voltage across the junction in one direction, but if the voltage is in the opposite direction only a very small current will flow. If an alternating voltage is applied across the junction the current will flow chiefly in one direction—the junction has rectified the current (Fig. 14).

For back voltage bias a negative voltage is applied to the p region and a positive voltage to the n region, thereby increasing the potential difference between the two regions. Now practically no electrons can climb the potential energy hill from the low side of the barrier to the high side. The recombination current is reduced by the Boltzmann factor:

\[ J_{nr}(V \text{ back}) = J_{nr}(0) \exp \left(-\frac{e|V|}{k_B T} \right) . \]  

(29)

The Boltzmann factor controls the number of electrons with enough energy to get over the barrier.

The thermal generation current of electrons is not particularly affected by the back voltage because the generation electrons flow downhill (from p to n) anyway:

\[ J_{ng}(V \text{ back}) = J_{ng}(0) . \]  

(30)

We saw in (28) that \( J_{nr}(0) = -J_{ng}(0) \); thus the generation current dominates the recombination current for a back bias.

When a forward voltage is applied, the recombination current increases because the potential energy barrier is lowered, thereby enabling more electrons to flow from the n side to the p side:

\[ J_{nr}(V \text{ forward}) = J_{nr}(0) \exp \left(e|V|/k_B T \right) . \]  

(31)

Again the generation current is unchanged:

\[ J_{ng}(V \text{ forward}) = J_{ng}(0) . \]  

(32)

The hole current flowing across the junction behaves similarly to the electron current. The applied voltage which lowers the height of the barrier for electrons also lowers it for holes, so that large numbers of electrons flow from the n region under the same voltage conditions that produce large hole currents in the opposite direction.

The electric currents of holes and electrons are additive, so that the total forward electric current is

\[ I = I_s \left[ \exp(eV/k_B T) - 1 \right] , \]  

(33)

where \( I_s \) is the sum of the two generation currents. This equation is well satis-
Figure 14  Rectification characteristic of a p-n junction in germanium, after Shockley. The voltage is plotted vertically and the current horizontally.

fied for p-n junctions in germanium (Fig. 14), but not quite as well in other semiconductors.

**Solar Cells and Photovoltaic Detectors**

Let us shine light on a p-n junction, one without an external bias voltage. Each absorbed photon creates an electron and a hole. When these carriers diffuse to the junction, the built-in electric field of the junction separates them at the energy barrier. The separation of the carriers produces a forward voltage across the barrier: forward, because the electric field of the photoexcited carriers is opposite to the built-in field of the junction.

The appearance of a forward voltage across an illuminated junction is called the photovoltaic effect. An illuminated junction can deliver power to an external circuit. Large area p-n junctions of silicon are used to convert solar photons to electrical energy.

**Schottky Barrier**

When a semiconductor is brought into contact with a metal, there is formed in the semiconductor a barrier layer from which charge carriers are severely depleted. The barrier layer is also called a depletion layer or exhaustion layer.

In Fig. 15 an n-type semiconductor is brought into contact with a metal. The Fermi levels are coincident after the transfer of electrons to the conduction band of the metal. Positively charged donor ions are left behind in this region
that is practically stripped of electrons. Here the Poisson equation is

\[
\text{(CGS)} \quad \text{div } \mathbf{D} = 4\pi ne
\]

where \( n \) is the donor concentration. The electrostatic potential is determined by

\[
\text{(CGS)} \quad \frac{d^2\varphi}{dx^2} = -4\pi ne/\varepsilon
\]

which has a solution of the form

\[
\text{(CGS)} \quad \varphi = -(2\pi ne/\varepsilon)x^2
\]

The origin of \( x \) has been taken for convenience at the right-hand edge of the barrier. The contact is at \(-x_b\), and here the potential energy relative to the right-hand side is \(-e\varphi_0\), whence the thickness of the barrier is

\[
\text{(CGS)} \quad x_b = (e|\varphi_0|/2\pi ne)^{1/2}
\]

With \( \varepsilon = 16 \), \( e\varphi_0 = 0.5 \text{ eV} \), \( n = 10^{16} \text{ cm}^{-3} \), we find \( x_b = 0.3 \mu\text{m} \). This is a somewhat simplified view of the metal-semiconductor contact.

**HETEROSTRUCTURES**

Semiconductor heterostructures are layers of two or more different semiconductors grown coherently with one common crystal structure. Heterostructures offer extra degrees of freedom in the design of semiconductor junction devices, because both the impurity doping and the conduction and valence band offsets at the junction can be controlled. This freedom is the basis of the prediction that most devices that utilize compound semiconductors will in the
future incorporate heterostructures. We treat them here in order to keep ahead of the times.

A heterostructure may be viewed as a single crystal in which the occupancy of the atomic sites changes at the interface. As an example one side of the interface can be Ge and the other side GaAs: both lattice constants are 5.65 Å. One side has the diamond structure and the other side the cubic zinc sulfide structure. Both structures are built up from tetrahedral covalent bonds and fit together coherently as if they were a single crystal. There are a few edge dislocations (Chapter 20) to relieve the strain energy near the interface.

The band gaps, however, are different, and this difference is the source of the real interest in the heterostructure, apart from the technical virtuosity in forming the structure. The band gaps are 0.67 eV for Ge and 1.43 eV for GaAs, at 300 K. The relative alignment of the conduction and valence band edges offers several possibilities, as shown in Fig. 16.

Calculations⁵ suggest that the top of the valence band $E_v$ in Ge should lie about 0.42 eV higher than in GaAs. The bottom of the conduction band $E_c$ in Ge should lie about 0.35 eV lower than in GaAs, so that the offsets are classified as normal in the scheme of Fig. 16.

Band edge offsets act as potential barriers in opposite senses on electrons and holes. Recall that electrons lower their energy by “sinking” on an energy band diagram, whereas holes lower their energy by floating on the same diagram. For the normal alignment both electrons and holes are pushed by the barrier from the wide-gap to the narrow-gap side of the heterostructure.

Other important semiconductor pairs used in heterostructures are AlAs/GaAs, InAs/GaSb, GaP/Si, and ZnSe/GaAs. Good lattice matching in the range

---

Figure 17  (a) Two semiconductors not in contact; the absolute band edge energies are labeled $E_c$ for the conduction band edge and $E_v$ for the valence band edge. An "absolute energy" means referred to infinite distance. The Fermi levels in the two materials are determined by the donor concentrations, as well as by the band structure. (b) The same semiconductors as a heterojunction, so that the two parts are in diffusive equilibrium. This requires that the Fermi level (F.L.) be independent of position, which is accomplished by transfer of electrons from the $N$-side to the $n$-side of the interface. A depletion layer of positively ionized donors is left behind on the $N$-side.

0.1–1.0 percent is often accomplished by use of alloys of different elements, which may also adjust energy gaps to meet specific device needs.

**n-N Heterojunction**

As a practical example, consider two $n$-type semiconductors with a large offset of the two conduction bands, as sketched in Fig. 17a for a semiconductor pair with a normal band line-up. The $n$-type material with the higher conduction band edge is labeled with a capital letter as $N$-type, and the junction shown is called an $n-N$ junction. The electron transport properties across the junction are similar to those across a Schottky barrier. Far from the interface the two semiconductors must be electrically neutral in composition. However, the two Fermi levels, each determined by the doping, must coincide if there is to be zero net electron transport in the absence of an external bias voltage.

These two considerations fix the "far-off" conduction band edge energies relative to the Fermi level, as in Fig. 17b. The combination of a specified band offset (determined by the host material composition) at the interface and the distant band energies (determined by the Fermi level) can be reconciled only if the bands bend near the interface, as in the figure. The necessary band bending is created by space charges consequent to the transfer of electrons from the $N$-side to the lower $n$-side. This transfer leaves behind on the $N$-side a positive donor space charge layer, which through the Poisson equation of electrostatics is the source of the positive second derivative (upward curvature) in the conduction band edge energy on that side.

On the $n$-side there is now a negative space charge because of the excess of electrons on that side. The layer of negative space charge gives a negative
second derivative (α inward curvature) in the conduction band edge energy. On the n-side the band as a whole bends down toward the junction. This differs from the usual p-n junction. The downward bending and the potential step form a potential well for electrons. The well is the basis for the new physical phenomena characteristic of heterostructure physics.

If the doping on the n-side (low $E_g$) is reduced to a negligible value, there will be very few ionized donors on that side in the electron-rich layer. The mobility of these electrons is largely limited only by lattice scattering, which falls off sharply as the temperature is lowered. Low temperature mobilities as high as $2 \times 10^6$ cm$^2$V$^{-1}$s$^{-1}$ have been observed in GaAs/(Al,Ga)As.

If now the thickness of the N-side semiconductor is reduced below the depletion layer thickness on that side, the N material will be entirely depleted of its low mobility electrons. All of the electrical conduction parallel to the interface will be carried by the high-mobility electrons on the n-side, equal in number to the number of ionized N-side donors, but spatially separated from them by the potential step. Such high mobility structures play a large role in solid state studies of 2D electron gases and also in new classes of high speed field effect transistors for computer applications at low temperatures.

**SEMICONDUCTOR LASERS**

Stimulated emission of radiation can occur in direct-gap semiconductors from the radiation emitted when electrons recombine with holes. The electron and hole concentrations created by illumination are larger than their equilibrium concentrations. The recombination times for the excess carriers are much longer than the times for the conduction electrons to reach thermal equilibrium with each other in the conduction band, and for the holes to reach thermal equilibrium with each other in the valence band. This steady state condition for the electron and hole populations is described by separate Fermi levels $\mu_c$ and $\mu_v$ for the two bands, called quasi-Fermi levels.

With $\mu_c$ and $\mu_v$ referred to their band edges, the condition for population inversion is that

$$\mu_c > \mu_v + \epsilon_g .$$

(38)

For laser action the quasi-Fermi levels must be separated by more than the band gap.

Population inversion and laser action can be achieved by forward bias of an ordinary GaAs or InP junction, but almost all practical injection lasers employ the double heterostructure proposed by H. Kroemer (Fig. 18). Here the lasing semiconductor is embedded between two wider-gap semiconductor regions of opposite doping. An example is GaAs embedded in (Al,Ga)As. In such a structure there is a potential barrier that prevents the outflow of electrons to the
Figure 18 Double heterostructure injection laser. Electrons flow from the right into the optically-active layer, where they form a degenerate electron gas. The potential barrier provided by the wide energy gap on the p side prevents the electrons from escaping to the left. Holes flow from the left into the active layer, but cannot escape to the right.

p-type region, and an opposite potential barrier that prevents the outflow of holes to the n-type region.

The value of $\mu_e$ in the optically active layer lines up with $\mu_n$ in the n contact, similarly, $\mu_v$ lines up with $\mu_p$ in the p contact. Inversion can be achieved if we apply a bias voltage larger than the voltage equivalent of the active layer energy gap. The diode wafer provides its own electromagnetic cavity, for the reflectivity at the crystal-air interface is high. Crystals are usually polished to provide two flat parallel surfaces; the radiation is emitted in the plane of the heterojunctions.

Crystals with direct band gaps are required normally for junction lasers. Indirect gaps involve phonons as well as photons; carriers recombine less efficiently because of competing processes, and no laser action has been observed in indirect gap semiconductors.

Gallium arsenide has been widely studied as the optically active layer. It emits in the near infrared at 8383 Å or 1.48 eV; the exact wavelength depends on temperature. The gap is direct (Chapter 8). In a heterojunction the system is very efficient: the ratio of light energy output to dc electrical energy input is near 50 percent, and the differential efficiency for small changes is up to 90 percent.

The wavelength can be adjusted over a wide range in the alloy system.
Ga$_x$In$_{1-x}$P$_y$As$_{1-y}$, so that we can match the laser wavelength to the absorption minimum of optical fibers used as a transmission medium. The combination of double heterostructure lasers with glass fibers forms the basis of the new lightwave communication technology that is gradually replacing transmission of signals over copper lines.

**LIGHT-EMITTING DIODES**

The efficiency of light-emitting diodes is now at the point of exceeding incandescent lamps. Consider a $p$-$n$ junction with a voltage source $V$ splitting the two chemical potentials $\mu_n$ and $\mu_p$ by $eV$, as in Figure 19. Electrons from the $n$ side are injected into the $p$ side, and holes from the $p$ side are injected into the $n$ side. These injected carriers annihilate each other across the junction, thus generating photons if the quantum efficiency is unity.

The generation or recombination process will be much stronger in a direct-gap semiconductor (Figure 8.5a) than in an indirect gap semiconductor (Figure
8.5b). In a direct gap semiconductor such as GaAs, the band-to-band photons are absorbed in a distance $\approx 1 \, \mu m$, which is strong absorption. The direct gap ternary semiconductor $GaAs_{1-x}P_x$ gives light tuned to shorter wavelengths as the composition variable $x$ is increased. This composition was made by Holonyak into one of the first $p$-$n$ diode lasers and into the first visible-spectrum (red) LED. Blue-emitting heterostructures have now been made, such as $In_xGa_{1-x}N - Al_yGa_{1-y}N$.

The performance of LED's has increased markedly over the years, from about 0.1 lumens/watt in 1962 to about 40 lumens/watt in 1994, compared with 15 lumens/watt for a standard white unfiltered incandescent lamp. To quote Craford and Holonyak, "We are entering an entirely new era in lighting (illumination) with an ultimate form of lamp—a direct gap III-V alloy $p$-$n$ heterostructure."

SCANNING TUNNELING MICROSCOPY

The original and still the central application of scanning tunneling microscopy is to the observation of the atomic structure of crystal surfaces. STM is a dream method that sounds much too simple to permit observations on the atomic scale for which it was developed by Binnig and Rohrer. The development was an extraordinary accomplishment, for which they received the 1986 Nobel Prize for Physics. An account of the history is given in their Nobel lecture for that year—accompanied by a quotation from the almost traditional letter of rejection of major work from a scientific publication to which they submitted an early paper.

One takes a sharp metal tip, preferably a tip that has one atom that protrudes a bit more than all the other atoms of the tip, and brings it to within 4 Å to 8 Å of a crystal surface. A voltage applied between the tip and the surface causes a tunneling current of electrons to flow. By tunneling we mean the standard quantum tunneling of a particle through a potential barrier. The transmission of electrons through a barrier is exponentially sensitive to the width of the barrier. The transmission flux in the experimental situation varies by a factor of about 10 for a 1-Å change in the separation of the tip from the surface. This fact is the key to how such a simple device can have a very fine spatial resolution. The greatest part of the tunnel current flows from the surface into the outermost atom of the tip, thus giving resolution on an atomic scale.

The most common way to operate an STM is to apply a voltage between the tip and the surface, and to adjust the height of the tip continuously to keep the tunnel current constant as the tip is scanned across the surface. The signal actually recorded is proportional to the voltage applied to a piezo-ceramic crystal whose displacement moves the tip perpendicular to the surface. The theory shows that for small bias voltage the trajectory of the tip measured in this way
corresponds to a contour of constant local density of states in the surface, taken at the Fermi level. What you are seeing is what the STM “feels”—a map of how the conduction electrons in the surface extend out into the vacuum that separates the surface from the tip.

With STM one can resolve single atoms, as in Figure 1.27, and surface corrugations as small as one thousandth of an atomic diameter. The tip can be used to move atoms around on the surface, strikingly shown here in Figure 20.

Problems

1. Diffraction from a linear array and a square array. The diffraction pattern of a linear structure of lattice constant $a$ is explained$^6$ in Fig. 21. Somewhat similar structures are important in molecular biology: DNA and many proteins are linear helices. (a) A cylindrical film is exposed to the diffraction pattern of Fig. 21b; the axis of the cylinder is coincident with the axis of the linear structure or fiber. Describe the

$^6$Another viewpoint is useful: for a linear lattice the diffraction pattern is described by the single Laue equation $a \cdot \Delta k = 2\pi q$, where $q$ is an integer. The lattice sums which led to the other Laue equations do not occur for a linear lattice. Now $a \cdot \Delta k = \text{const.}$ is the equation of a plane, thus the reciprocal lattice becomes a set of parallel planes normal to the line of atoms.
Figure 21 The diffraction pattern from a single line of lattice constant \( a \) in a monochromatic x-ray beam perpendicular to the line. (a) The condition for constructive interference is \( a \cos \theta = n\lambda \), where \( n \) is an integer. (b) For given \( n \) the diffracted rays of constant \( \lambda \) lie on the surface of a cone.

Figure 22 (a) Backward scattering pattern of 76 eV electrons incident normally on the (110) face of a nickel crystal; a model of the surface is shown in (b). (Courtesy of A. U. MacRae.)
appearance of the diffraction pattern on the film. (b) A flat photographic plate is placed behind the fiber and normal to the incident beam. Sketch roughly the appearance of the diffraction pattern on the plate. (c) A single plane of atoms forms a square lattice of lattice constant a. The plane is normal to the incident x-ray beam. Sketch roughly the appearance of the diffraction pattern on the photographic plate. Hint: The diffraction from a plane of atoms can be inferred from the patterns for two perpendicular lines of atoms. (d) Figure 22 shows the electron diffraction pattern in the backward direction from the nickel atoms on the (110) surface of a nickel crystal. Explain the orientation of the diffraction pattern in relation to the atomic positions of the surface atoms shown in the model. Assume that only the surface atoms are effective in the reflection of low-energy electrons.

2. Surface subbands in electric quantum limit. Consider the contact plane between an insulator and a semiconductor, as in a metal-oxide-semiconductor transistor or MOSFET. With a strong electric field applied across the SiO₂-Si interface, the potential energy of a conduction electron may be approximated by $V(x) = eE$ for $x$ positive and by $V(x) = \infty$ for $x$ negative, where the origin of $x$ is at the interface. The wavefunction is 0 for $x$ negative and may be separated as $\psi(x, y, z) = u(x) \exp[i(k_y y + k_z z)]$, where $u(x)$ satisfies the differential equation

$$-(\hbar^2/2m)u''/dx^2 + V(x)u = eu.$$ 

With the model potential for $V(x)$ the exact eigenfunctions are Airy functions, but we can find a fairly good ground state energy from the variational trial function $x \exp(-ax)$. (a) Show that $\epsilon = (\hbar^2/2m)a^2 + 3eE/2a$. (b) Show that the energy is a minimum when $a = (3eEm/2\hbar^2)^{1/3}$. (c) Show that $\epsilon_{\text{min}} = 1.89(\hbar^2/2m)^{1/3} (3eE/2)^{2/3}$. In the exact solution for the ground state energy the factor 1.89 is replaced by 1.78. As $E$ is increased the extent of the wavefunction in the $x$ direction is decreased. The function $u(x)$ defines a surface conduction channel on the semiconductor side of the interface. The various eigenvalues of $u(x)$ define what are called electric subbands. Because the eigenfunctions are real functions of $x$ the states do not carry current in the $x$ direction, but they do carry a surface channel current in the $y, z$ plane. The dependence of the channel on the electric field $E$ in the $x$ direction makes the device a field effect transistor.

3. Shubnikov-de Haas effect in a surface channel. The effect describes the periodic variation with $1/B$ (or with electron concentration) of the longitudinal electrical resistance of a specimen in a strong perpendicular magnetic field at a low temperature. These are the conditions in which a de Haas-van Alphen effect is observed, as discussed in Chapter 9. The effect can be discussed simply in 2D. (a) Show that the density of free electron orbitals in 2D is $m^2/2\pi\hbar^2$, with omission of spin and valley degeneracy. The surface concentration of electrons can now be written as $\nu = m_e\rho/2\pi\hbar^2$. (b) Show that the S-DH oscillations have the period $\Delta(1/B) = e\hbar/m_e\nu = e/2\pi\hbar cv$. In SI set $c = 1$. (c) According to the view of the inversion layer as a planar capacitor the surface charge concentration $\nu$ varies directly as the gate voltage $V_g$ across the interface. Show that the resistance oscillations are periodic functions of the gate voltage, as in Fig. 23.
Figure 23 Source-drain resistance of an n-channel MOSFET as a function of the gate voltage $V_g$, at transverse magnetic fields of 33.3 and 37.7 kG, at 4 K. (After G. Landwehr et al.)

References

Surface Science. A journal.
20

Dislocations

SHEAR STRENGTH OF SINGLE CRYSTALS 587
Slip 588

DISLOCATIONS 589
Burgers vectors 592
Stress fields of dislocations 593
Low-angle grain boundaries 595
Dislocation densities 598
Dislocation multiplication and slip 600

STRENGTH OF ALLOYS 600

DISLOCATIONS AND CRYSTAL GROWTH 603
Whiskers 604

HARDNESS OF MATERIALS 606

PROBLEMS 606
1. Lines of closest packing 606
2. Dislocation pairs 606
3. Force on dislocation 606

REFERENCES 607
Figure 1  (a) Relative shear of two planes of atoms (shown in section) in a uniformly strained crystal; (b) shear stress as a function of the relative displacement of the planes from their equilibrium position. The heavy broken line drawn at the initial slope defines the shear modulus $G$. 
CHAPTER 20: DISLOCATIONS

This chapter is concerned with the interpretation of the plastic mechanical properties of crystalline solids in terms of the theory of dislocations. Plastic properties are irreversible deformations; elastic properties are reversible. The ease with which pure single crystals deform plastically is striking. This intrinsic weakness of crystals is exhibited in various ways. Pure silver chloride melts at 455°C, yet at room temperature it has a cheeselike consistency and can be rolled into sheets. Pure aluminum crystals are elastic (follow Hooke’s law) only to a strain of about 10^{-5}, after which they deform plastically.

Theoretical estimates of the elastic limit of perfect crystals give values 10^3 or 10^4 higher than the lowest observed values, although a factor 10^2 is more usual. There are few exceptions to the rule that pure crystals are plastic and are not strong: crystals of germanium and silicon are not plastic at room temperature and fail or yield only by fracture. Glass at room temperature fails only by fracture, but it is not crystalline. The fracture of glass is caused by stress concentration at minute cracks.

SHEAR STRENGTH OF SINGLE CRYSTALS

Frenkel gave a simple method of estimating the theoretical shear strength of a perfect crystal. We consider in Fig. 1 the force needed to make a shear displacement of two planes of atoms past each other. For small elastic strains the stress \( \sigma \) is related to the displacement \( x \) by

\[
\sigma = Gx/d .
\]

(1)

Here \( d \) is the interplanar spacing, and \( G \) denotes the appropriate shear modulus. When the displacement is large and has proceeded to the point that atom \( A \) is directly over atom \( B \) in the figure, the two planes of atoms are in a configuration of unstable equilibrium and the stress is zero. As a first approximation we represent the stress-displacement relation by

\[
\sigma = (Ga/2\pi d) \sin (2\pi x/a) ,
\]

(2)

where \( a \) is the interatomic spacing in the direction of shear. This relation is constructed to reduce to (1) for small values of \( x/a \). The critical shear stress \( \sigma_c \) at which the lattice becomes unstable is given by the maximum value of \( \sigma \), or

\[
\sigma_c = Ga/2\pi d .
\]

(3)

If \( a = d \), then \( \sigma_c \approx G/2\pi \); the ideal critical shear stress is of the order of \( \frac{1}{8} \) of the shear modulus.
The observations in Table 1 show the experimental values of the elastic limit are much smaller than (3) would suggest. The theoretical estimate may be improved by consideration of the actual form of the intermolecular forces and by consideration of other configurations of mechanical stability available to the lattice as it is sheared. Mackenzie has shown that these two effects may reduce the theoretical ideal shear strength to about G/30, corresponding to a critical shear strain angle of about 2 degrees. The observed low values of the shear strength can be explained only by the presence of imperfections that can act as sources of mechanical weakness in real crystals. The movement of crystal imperfections called dislocations is responsible for slip at very low applied stresses.

Table 1  Comparison of shear modulus and elastic limit

<table>
<thead>
<tr>
<th></th>
<th>Shear modulus $G$, in dyn/cm²</th>
<th>Elastic limit $\sigma_e$, in dyn/cm²</th>
<th>$G/\sigma_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sn, single crystal</td>
<td>$1.9 \times 10^{11}$</td>
<td>$1.3 \times 10^7$</td>
<td>15,000</td>
</tr>
<tr>
<td>Ag, single crystal</td>
<td>$2.8 \times 10^{11}$</td>
<td>$6 \times 10^6$</td>
<td>45,000</td>
</tr>
<tr>
<td>Al, single crystal</td>
<td>$2.5 \times 10^{11}$</td>
<td>$4 \times 10^6$</td>
<td>60,000</td>
</tr>
<tr>
<td>Al, pure, polycrystal</td>
<td>$2.5 \times 10^{11}$</td>
<td>$2.6 \times 10^8$</td>
<td>900</td>
</tr>
<tr>
<td>Al, commercial drawn</td>
<td>$-2.5 \times 10^{11}$</td>
<td>$9.9 \times 10^8$</td>
<td>250</td>
</tr>
<tr>
<td>Duralumin</td>
<td>$-2.5 \times 10^{11}$</td>
<td>$3.6 \times 10^9$</td>
<td>70</td>
</tr>
<tr>
<td>Fe, soft, polycrystal</td>
<td>$7.7 \times 10^{11}$</td>
<td>$1.5 \times 10^9$</td>
<td>500</td>
</tr>
<tr>
<td>Heat-treated carbon steel</td>
<td>$-8 \times 10^{11}$</td>
<td>$6.5 \times 10^9$</td>
<td>120</td>
</tr>
<tr>
<td>Nickel-chrome steel</td>
<td>$-8 \times 10^{11}$</td>
<td>$1.2 \times 10^{10}$</td>
<td>65</td>
</tr>
</tbody>
</table>

After Mott.

Slip

Plastic deformation in crystals occurs by slip, an example of which is shown in Fig. 2. In slip one part of the crystal slides as a unit across an adjacent part. The surface on which slip takes place is known as the slip plane. The direction of motion is known as the slip direction. The great importance of lattice properties for plastic strain is indicated by the highly anisotropic nature of slip. Displacement takes place along crystallographic planes with a set of small Miller indices, such as the {111} planes in fcc metals and the {110}, {112}, and {123} planes in bcc metals.

The slip direction is in the line of closest atomic packing, {110} in fcc metals and {111} in bcc metals. To maintain the crystal structure after slip, the displacement or slip vector must equal a lattice translation vector. The shortest lattice translation vector expressed in terms of the lattice constant $a$ in a fcc structure is of the form $(a/2)(\hat{x} + \hat{y})$; in a bcc structure it is $(a/2)(\hat{x} + \hat{y} + \hat{z})$. But
in fcc crystals one also observes partial displacements which upset the regular sequence ABCABC . . . of closest-packed planes, to produce a stacking fault such as ABCABABC . . . The result is then a mixture of fcc and hcp stacking.

Deformation by slip is inhomogeneous. Large shear displacements occur on a few widely separated slip planes, while parts of the crystal lying between slip planes remain essentially undeformed. A property of slip is the Schmid law of the critical shear stress: slip takes place along a given slip plane and direction when the corresponding component of shear stress reaches the critical value.

Slip is one mode of plastic deformation. Another mode, twinning, is observed particularly in hcp and bcc structures. During slip a considerable displacement occurs on a few widely separated slip planes. During twinning a partial displacement occurs successively on each of many neighboring crystallographic planes. After twinning, the deformed part of the crystal is a mirror image of the undeformed part. Although both slip and twinning are caused by the motion of dislocations, we shall be concerned primarily with slip.

DISLOCATIONS

The low observed values of the critical shear stress are explained in terms of the motion through the lattice of a line imperfection known as a dislocation. The idea that slip propagates by the motion of dislocations was published in 1934 independently by Taylor, Orowan, and Polanyi; the concept of dislocations was introduced somewhat earlier by Prandtl and Dehlinger. There are several basic types of dislocations. We first describe an edge dislocation. Figure 3 shows a simple cubic crystal in which slip of one atom distance has occurred over the left half of the slip plane but not over the right half. The
boundary between the slipped and unslipped regions is called the dislocation. Its position is marked by the termination of an extra vertical half-plane of atoms crowded into the upper half of the crystal as shown in Fig. 4. Near the dislocation the crystal is highly strained. The simple edge dislocation extends indefinitely in the slip plane in a direction normal to the slip direction. In Fig. 5 we show a photograph of a dislocation in a two-dimensional soap bubble raft obtained by the method of Bragg and Nye.

The mechanism responsible for the mobility of a dislocation is shown in Fig. 6. The motion of an edge dislocation through a crystal is analogous to the passage of a ruck or wrinkle across a rug: the ruck moves more easily than the whole rug. If atoms on one side of the slip plane are moved with respect to those on the other side, atoms at the slip plane will experience repulsive forces from some neighbors and attractive forces from others across the slip plane. These forces cancel to a first approximation. The external stress required to move a dislocation has been calculated and is quite small, probably below $10^5$ dyn/cm², provided that the bonding forces in the crystal are not highly directional. Thus dislocations may make a crystal very plastic. Passage of a dislocation through a crystal is equivalent to a slip displacement of one part of the crystal.

The second simple type of dislocation is the screw dislocation, sketched in Figs. 7 and 8. A screw dislocation marks the boundary between slipped and unslipped parts of the crystal. The boundary parallels the slip direction, instead of lying perpendicular to it as for the edge dislocation. The screw dislocation may be thought of as produced by cutting the crystal partway through with a
Figure 4  Structure of an edge dislocation. The deformation may be thought of as caused by inserting an extra plane of atoms on the upper half of the $y$ axis. Atoms in the upper half-crystal are compressed by the insertion; those in the lower half are extended.

Figure 5  A dislocation in a two-dimensional bubble raft. The dislocation is most easily seen by turning the page by 30° in its plane and sighting at a low angle. (W. M. Lomer, after Bragg and Nye.)

Figure 6  Motion of a dislocation under a shear tending to move the upper surface of the specimen to the right. (D. Hull.)
knife and shearing it parallel to the edge of the cut by one atom spacing. A screw dislocation transforms successive atom planes into the surface of a helix; this accounts for the name of the dislocation.

**Burgers Vectors**

Other dislocation forms may be constructed from segments of edge and screw dislocations. Burgers has shown that the most general form of a linear dislocation pattern in a crystal can be described as shown in Fig. 9. We consider any closed curve not necessarily planar within a crystal, or an open curve terminating on the surface at both ends: (a) Make a cut along any simple surface bounded by the line. (b) Displace the material on one side of this surface by a vector $\mathbf{b}$ relative to the other side; here $\mathbf{b}$ is called the Burgers vector. (c) In regions where $\mathbf{b}$ is not parallel to the cut surface, this relative displacement will either produce a gap or cause the two halves to overlap. In these cases we imagine that we either add material to fill the gap or subtract material to prevent overlap. (d) Rejoin the material on both sides. We leave the strain displacement intact at the time of rewelding, but afterwards we allow the medium
to come to internal equilibrium. The resulting strain pattern is that of the dislocation characterized jointly by the boundary curve and the Burgers vector. The Burgers vector must be equal to a lattice vector in order that the rewelding process will maintain the crystallinity of the material. The Burgers vector of a screw dislocation (Figs. 7 and 8) is parallel to the dislocation line; that of an edge dislocation (Figs. 3 and 4) is perpendicular to the dislocation line and lies in the slip plane.

**Stress Fields of Dislocations**

The stress field of a screw dislocation is particularly simple. Figure 10 shows a shell of material surrounding an axial screw dislocation. The shell of circumference $2\pi r$ has been sheared by an amount $b$ to give a shear strain $\varepsilon = b/2\pi r$. The corresponding shear stress in an elastic continuum is

$$\sigma = Ge = Gb/2\pi r.$$  \hspace{1cm} (4)

This expression cannot hold in the region immediately around the dislocation line, as the strains here are too large for continuum or linear elasticity theory to apply. The elastic energy of the shell is $dE_s = \frac{1}{2}Ge^2 \, dV = (Gb^2/4\pi) \, d\ell/r$ per unit length. The total elastic energy per unit length of a screw dislocation is found on integration to be

$$E_s = \frac{Gb^2}{4\pi} \ln \frac{R}{r_0},$$  \hspace{1cm} (5)
where $R$ and $r_0$ are appropriate upper and lower limits for the variable $r$. A reasonable value of $r_0$ is comparable to the magnitude $b$ of the Burgers vector or to the lattice constant; the value of $R$ cannot exceed the dimensions of the crystal. The value of the ratio $R/r_0$ is not very important because it enters in a logarithm term.

We now show the form of the energy of an edge dislocation. Let $\sigma_{rr}$ and $\sigma_{\theta\theta}$ denote the tensile stresses in the radial and circumferential directions, and let $\sigma_{r\theta}$ denote the shear stress. In an isotropic elastic continuum, $\sigma_{rr}$ and $\sigma_{\theta\theta}$ are proportional to $(\sin \theta)/r$: we need a function that falls off as $1/r$ and that changes sign when $y$ is replaced by $-y$. The shear stress $\sigma_{r\theta}$ is proportional to $(\cos \theta)/r$: considering the plane $y = 0$ we see from Fig. 4 that the shear stress is an odd function of $x$. The constants of proportionality in the stress are proportional to the shear modulus $G$ and to the Burgers vector $b$ of the displacement. The final result, which is derived in books cited in the references, is

$$
\sigma_{rr} = \sigma_{\theta\theta} = \frac{Gb}{2\pi(1 - \nu)} \frac{\sin \theta}{r}, \quad \sigma_{r\theta} = \frac{Gb}{2\pi(1 - \nu)} \frac{\cos \theta}{r},
$$

(6)

where the Poisson ratio $\nu \approx 0.3$ for most crystals. The strain energy of a unit length of edge dislocation is

$$
E_e = \frac{Gb^2}{4\pi(1 - \nu)} \ln \frac{R}{r_0}.
$$

(7)

We want an expression for the shear stress component $\sigma_{xy}$ on planes parallel to the slip plane in Fig. 4. From the stress components $\sigma_{rr}$, $\sigma_{\theta\theta}$, and $\sigma_{r\theta}$ evaluated on the plane a distance $y$ above the slip plane, we find

$$
\sigma_{xy} = \frac{Gb}{2\pi(1 - \nu)} \frac{x(x^2 - y^2)}{(x^2 + y^2)^2}.
$$

(8)
Figure 11 (a) Low-angle grain boundary, after Burgers. (b) Electron micrograph of a low-angle grain boundary in molybdenum. The three dislocations in the image each have the same Burgers vector as in the drawing in Figure 11a. The white circles mark the positions of atomic columns normal to the plane of the paper. Each array of circles defines the position of a dislocation, with four circles on the top of each array and three circles below. Closure failure is indicated by the arrows which define the Burgers vectors. (Courtesy of R. Gronsky.)

It is shown in Problem 3 that the force caused by a resolved uniform shear stress $\sigma$ is $F = b\sigma$ per unit length of dislocation. The force that an edge dislocation at the origin exerts upon a similar one at the location $(y, \theta)$ is

$$F = b\sigma_{xy} = \frac{Gl^2}{2\pi(1 - \nu)} \frac{\sin \theta}{4y}$$

per unit length. Here $F$ is the component of force in the slip direction.

**Low-angle Grain Boundaries**

Burgers suggested that low-angle boundaries between adjoining crystalites or crystal grains consist of arrays of dislocations. A simple example of the Burgers model of a grain boundary is shown in Fig. 11. The boundary occupies a (010) plane in a simple cubic lattice and divides two parts of the crystal that have a [001] axis in common. Such a boundary is called a pure tilt boundary: the misorientation can be described by a small rotation $\theta$ about the common [001] axis of one part of the crystal relative to the other. The tilt boundary is represented as an array of edge dislocations of spacing $D = b/\theta$, where $b$ is the Burgers vector of the dislocations. Experiments have substantiated this model. Figure 12 shows the distribution of dislocations along small-angle grain boundaries, as observed with an electron microscope. Further, Read and Shockley
Figure 12  Electron micrograph of dislocation structures in low-angle grain boundaries in an Al-7 percent Mg solid solution. Notice the lines of small dots on the right. Mag. ×17,000. (R. Goodrich and G. Thomas.)

derived a theory of the interfacial energy as a function of the angle of tilt, with results in excellent agreement with measurements.

Direct verification of the Burgers model is provided by the quantitative x-ray and optical studies of low-angle boundaries in germanium crystals by Vogel and co-workers. By counting etch pits along the intersection of a low-angle grain boundary with an etched germanium surface (Fig. 13), they determined the dislocation spacing $D$. They assumed that each etch pit marked the end of a dislocation. The angle of tilt calculated from the relation $\theta = b/D$ agrees well with the angle measured directly by means of x-rays.

The interpretation of low-angle boundaries as arrays of dislocations is further supported by the fact that pure tilt boundaries move normal to themselves on application of a suitable stress. The motion has been demonstrated in a beautiful experiment, Fig. 14. The specimen is a bicrystal of zinc containing a 2° tilt boundary with dislocations about 30 atomic planes apart. One side of the crystal was clamped, and a force was applied at a point on the opposite side of the boundary. Motion of the boundary took place by cooperative motion of the dislocations in the array, each dislocation moving an equal distance in its own slip plane. The motion was produced by stresses of the order of magnitude of the yield stress for zinc crystals, strong evidence that ordinary deformation results from the motion of dislocations.

Grain boundaries and dislocations offer relatively little resistance to diffusion of atoms in comparison with diffusion in perfect crystals. A dislocation is an
Figure 13 Dislocation etch pits in low-angle boundary on (100) face of germanium, the angle of the boundary is 27.5°. The boundary lies in a (011) plane; the line of the dislocations is [100]. The Burgers vector is the shortest lattice translation vector, or $|b| = a/\sqrt{2} = 4.0$ Å. (F. L. Vogel, Jr.)

Figure 14 Motion of a low-angle grain boundary under stress. The boundary is the straight vertical line, and it is photographed under vertical illumination, thereby making evident the 2° angular change in the cleavage surface of the zinc crystal at the boundary. The irregular horizontal line is a small step in the cleavage surface which serves as a reference mark. The crystal is clamped at the left, at the right it is subject to a force normal to the plane of the page. Top, original position of boundary; bottom, moved back 0.4 mm. (J. Washburn and E. R. Parker.)
open passage for diffusion. Diffusion is greater in plastically deformed material than in annealed crystals. Diffusion along grain boundaries controls the rates of some precipitation reactions in solids: the precipitation of tin from lead-tin solutions at room temperature proceeds about $10^8$ times faster than expected from diffusion in an ideal lattice.

**Dislocation Densities**

The density of dislocations is the number of dislocation lines that intersect a unit area in the crystal. The density ranges from well below $10^2$ dislocations/cm$^2$ in the best germanium and silicon crystals to $10^{11}$ or $10^{12}$ dislocations/cm$^2$ in heavily deformed metal crystals. The methods available for estimating dislocation densities are compared in Table 2. The actual dislocation configurations in cast or annealed (slowly cooled) crystals correspond either to a group of low-angle grain boundaries or to a three-dimensional network of dislocations arranged in cells, as shown in Fig. 15.

Lattice vacancies that precipitate along an existing edge dislocation will eat away a portion of the extra half-plane of atoms and cause the dislocation to climb, which means to move at right angles to the slip direction. If no dislocations are present, the crystal will become supersaturated with lattice vacancies; their precipitation in penny-shaped vacancy plates may be followed by collapse of the plates and formation of dislocation rings that grow with further vacancy precipitation, as in Fig. 16.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Specimen thickness</th>
<th>Width of image</th>
<th>Maximum practical density, per cm$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron microscopy</td>
<td>$&gt;1000$ Å</td>
<td>$\sim 100$ Å</td>
<td>$10^{11}$–$10^{12}$</td>
</tr>
<tr>
<td>X-ray transmission</td>
<td>0.1–1.0 mm</td>
<td>5μm</td>
<td>$10^4$–$10^5$</td>
</tr>
<tr>
<td>X-ray reflection</td>
<td>$&lt;2$μm (min.) – 50μm (max.)</td>
<td>2μm</td>
<td>$10^6$–$10^7$</td>
</tr>
<tr>
<td>Decoration</td>
<td>$\sim 10$μm (depth of focus)</td>
<td>0.5μm</td>
<td>$2 \times 10^7$</td>
</tr>
<tr>
<td>Etch pits</td>
<td>no limit</td>
<td>0.5μm$^b$</td>
<td>$4 \times 10^8$</td>
</tr>
</tbody>
</table>

$^a$W. G. Johnston.

$^b$Limit of resolution of etch pits.
Figure 15  Cell structure of three-dimensional tangles of dislocations in deformed aluminum. (P. R. Swann.)

Figure 16  Electron micrograph of dislocation loops formed by aggregation and collapse of vacancies in Al–5 percent Mg quenched from 550°C. The helical dislocations are formed by vacancy condensation on a screw dislocation. Mag. ×43,000. (A. Eikum and G. Thomas.)
Dislocation Multiplication and Slip

Plastic deformation causes a very great increase in dislocation density, typically from $10^8$ to about $10^{11}$ dislocations/cm$^2$ during deformation. If a dislocation moves completely across its slip plane an offset of one atom spacing is produced, but offsets up to 100 to 1000 atom spacings are observed. This means that dislocations multiply during deformation.

Consider a closed circular dislocation loop of radius $r$ surrounding a slipped area having the radius of the loop. Such a loop will be partly edge, partly screw, and mostly of intermediate character. The strain energy of the loop increases as its circumference, so that the loop will tend to shrink in size. However, the loop will tend to expand if a shear stress is acting that favors slip.

A common feature of all dislocation sources is the bowing of dislocations. A dislocation segment pinned at each end is called a Frank-Read source, and it can lead (Fig. 17) to the generation of a large number of concentric dislocations on a single slip plane (Fig. 18). Related types of dislocation multiplication mechanisms account for slip and for the increased density of dislocations during plastic deformation. Double cross-slip is the most common source.

STRENGTH OF ALLOYS

Pure crystals are very plastic and yield at very low stresses. There appear to be four important ways of increasing the yield strength of an alloy so that it will withstand shear stresses as high as $10^{-2}$ G. They are mechanical blocking of dislocation motion, pinning of dislocations by solute atoms, impeding dislocation motion by short-range order, and increasing the dislocation density so that tangling of dislocations results. All four strengthening mechanisms depend for their success upon impeding dislocation motion. A fifth mechanism, that of removing all dislocations from the crystal, may operate for certain fine hairlike crystals (whiskers) that are discussed in the section on crystal growth.

Mechanical blocking of dislocation motion can be produced most directly by introducing tiny particles of a second phase into a crystal lattice. This process is followed in the hardening of steel, where particles of iron carbide are precipitated into iron, and in hardening aluminum, where particles of Al$_6$Cu are precipitated. The pinning of a dislocation by particles is shown in Fig. 19.

In strengthening by the addition of small particles there are two cases to be considered: either the particle can be deformed with the matrix, which requires that the particle can be traversed by the dislocation, or the particle cannot be traversed by the dislocation. If the particle cannot be cut, the stress necessary to force a dislocation between particles spaced $L$ apart on a slip plane should be approximately

$$\sigma = b/L .$$

(10)

The smaller the spacing $L$, the higher is the yield stress $\sigma$. Before particles precipitate, $L$ is large and the strength is low. Immediately after precipitation is
Figure 17 Frank-Read mechanism for multiplication of dislocations, showing successive stages in the generation of a dislocation loop by the segment $BC$ of a dislocation line. The process can be repeated indefinitely.

Figure 18 A Frank-Read dislocation source in silicon, decorated with copper precipitates and viewed with infrared illumination. Two complete dislocation loops are visible, and the third, innermost loop is near completion. (After W. C. Dash.)
complete and many small particles are present, $L$ is a minimum and the
strength is a maximum. If the alloy is then held at a high temperature, some
particles grow at the expense of others, so that $L$ increases and the strength
drops. Hard intermetallic phases, such as refractory oxides, cannot be cut by
dislocations.

The strength of dilute solid solutions is believed to result from the pinning
of dislocations by solute atoms. The solubility of a foreign atom will be greater
in the neighborhood of a dislocation than elsewhere in a crystal. An atom that
tends to expand the crystal will dissolve preferentially in the expanded region
near an edge dislocation. A small atom will tend to dissolve preferentially in the
contracted region near the dislocation—a dislocation offers both expanded and
contracted regions.

As a result of the affinity of solute atoms for dislocations, each dislocation
will collect a cloud of associated solute atoms during cooling, at a time when the
mobility of solute atoms is high. At still lower temperatures diffusion of solute
atoms effectively ceases, and the solute atom cloud becomes fixed in the crys-
tal. When a dislocation moves, leaving its solute cloud behind, the energy of
the crystal must increase. The increase in energy can only be provided by an
increased stress acting on the dislocation as it pulls away from the solute atom
cloud, and so the presence of the cloud strengthens the crystal.
The passage of a dislocation across a slip plane in pure crystals does not alter the binding energy across the plane after the dislocation is gone. The internal energy of the crystal remains unaffected. The same is true for random solid solutions, because the solution is equally random across a slip plane after slip. Most solid solutions, however, have short-range order. Atoms of different species are not arranged at random on the lattice sites, but tend to have an excess or a deficiency of pairs of unlike atoms. Thus in ordered alloys dislocations tend to move in pairs: the second dislocation reorders the local disorder left by the first dislocation.

The strength of a crystalline material increases with plastic deformation. The phenomenon is called work-hardening or strain-hardening. The strength is believed to increase because of the increased density of dislocations and the greater difficulty of moving a given dislocation across a slip plane that is threaded by many dislocations. Strain-hardening frequently is employed in the strengthening of materials, but its usefulness is limited to low enough temperatures so that annealing does not occur.

An important factor in strain-hardening is the total density of dislocations. In most metals dislocations tend to form cells (Fig. 15) of dislocation-free areas of dimensions of the order of 1 μm. But unless we can get a uniform high density of dislocations we cannot strain-harden a metal to its theoretical strength, because of slip in the dislocation-free areas. A high total density is accomplished by explosive deformation or by special thermal-mechanical treatments, as of martensite in steel.

Each of the mechanisms of strengthening crystals can raise the yield strength to the range of $10^{-3}$ G to $10^{-2}$ G. All mechanisms begin to break down at temperatures where diffusion can occur at an appreciable rate. When diffusion is rapid, precipitated particles dissolve; solute clouds drift along with dislocations as they glide; short-range order repairs itself behind slowly moving dislocations; and dislocation climb and annealing tend to decrease the dislocation density. The resulting time-dependent deformation is called creep. This irreversible motion precedes the elastic limit. The search for alloys for use at very high temperatures is a search for reduced diffusion rates, so that the four strengthening mechanisms will survive to high temperatures. But the central problem of strong alloys is not strength, but ductility, for failure is often by fracture.

**DISLOCATIONS AND CRYSTAL GROWTH**

In some cases the presence of dislocations may be the controlling factor in crystal growth. When crystals are grown in conditions of low supersaturation, of the order of 1 percent, it has been observed that the growth rate is enormously faster than that calculated for an ideal crystal. The actual growth rate is explained in terms of the effect of dislocations on growth.
The theory of growth of ideal crystals predicts that in crystal growth from vapor a supersaturation (pressure/equilibrium vapor pressure) of the order of 10 is required to nucleate new crystals, of the order of 5 to form liquid drops, and of 1.5 to form a two-dimensional monolayer of molecules on the face of a perfect crystal. Volmer and Schultze observed growth of iodine crystals at vapor supersaturations down to less than 1 percent, where the growth rate should have been down by the factor \( \exp(-3000) \) from the rate defined as the minimum observable growth.

The large disagreement expresses the difficulty of nucleating a new monolayer on a completed surface of an ideal crystal. But if a screw dislocation is present (Fig. 20), it is never necessary to nucleate a new layer: the crystal will grow in spiral fashion at the edge of the discontinuity shown. An atom can be bound to a step more strongly than to a plane. The calculated growth rates for this mechanism are in good agreement with observation. We expect that nearly all crystals in nature grown at low supersaturation will contain dislocations, as otherwise they could not have grown. Spiral growth patterns have been observed on a large number of crystals. A beautiful example of the growth pattern from a single screw dislocation is given in Fig. 21.

If the growth rate is independent of direction of the edge in the plane of the surface, the growth pattern is an Archimedes spiral, \( r = a \theta \), where \( a \) is a constant. The limiting minimum radius of curvature near the dislocation is determined by the supersaturation. If the radius of curvature is too small, atoms on the curved edge evaporate until the equilibrium curvature is attained. Away from the origin each part of the step acquires new atoms at a constant rate, so that \( dr/dt = \text{const} \).

**Whiskers**

Fine hairlike crystals or whiskers have been observed to grow under conditions of high supersaturation without the necessity for more than perhaps one dislocation. It may be that these crystals contain a single axial screw dislocation that aids their essentially one-dimensional growth. From the absence of dislocations we would expect these crystal whiskers to have high yield strengths, of the order of the calculated value \( G/30 \) discussed earlier in this chapter. A single axial screw dislocation, if present, could not cause yielding, because in bending the crystal the dislocation is not subjected to a shear stress parallel to its Burgers vector. That is, the stress is not in a direction that can cause slip. Herring and Calt observed whiskers of tin of radius \( \sim 10^{-4} \) cm with elastic properties near those expected from theoretically perfect crystals. They observed yield strains of the order of \( 10^{-2} \), which correspond to shear stresses of order \( 10^{-2} \) G, about 1000 times greater than in bulk tin, confirming the early estimates of the strength of perfect crystals. Theoretical or ideal elastic properties have been observed for a number of materials. A single domain whisker of nickel is shown in Fig. 22.
Figure 20  Development of a spiral step produced by intersection of a screw dislocation with the surface of a crystal as in Fig. 8. Each cube represents a molecule. (F. C. Frank.)

Figure 21  Phase contrast micrograph of a hexagonal spiral growth pattern on a SiC crystal. The step height is 165 Å. (A. R. Verma.)

Figure 22  A nickel whisker of diameter 1000 Å bent in a loop. (R. W. De Blois.)
HARDNESS OF MATERIALS

The hardness of materials is measured in several ways, the simplest test for nonmetals being the scratch test. Substance A is harder than substance B if A will scratch B but B will not scratch A. A standard scale is used for representative minerals, with diamond, the hardest, assigned the value 10 and talc, the softest, assigned the value 1:

10 diamond C  5 apatite $\text{Ca}_5(\text{PO}_4)_3\text{F}$
9 corundum $\text{Al}_2\text{O}_3$  4 fluorite $\text{CaF}_2$
8 topaz $\text{Al}_2\text{Si}_4\text{O}_6\text{F}_2$  3 calcite $\text{CaCO}_3$
7 quartz $\text{SiO}_2$  2 gypsum $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$
6 orthoclase $\text{KAlSi}_3\text{O}_8$  1 talc $3\text{MgO} \cdot 4\text{SiO}_2 \cdot \text{H}_2\text{O}$

There is great current interest in the development of materials of great hardness, for example as films for use as scratch-resistant coatings on lenses. It is widely felt that the scale between diamond and corundum is misleading, because diamond is much, much harder than corundum. It has been suggested that one might assign diamond the hardness 15, with the gap between 9 and 15 to be filled in eventually by synthetic materials, such as compounds of C and B.

Modern scales of hardness, such as the VHN scale, are based on indenter tests in which an indenter is pressed into the surface of the material and the size of the impression is measured. The Vickers Hardness Numbers of selected materials are tabulated below, after conversion by E. R. Weber to units of GPa [GN/m$^2$]:

<table>
<thead>
<tr>
<th>Material</th>
<th>VHN</th>
<th>Material</th>
<th>VHN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond</td>
<td>45.3</td>
<td>BeO</td>
<td>7.01</td>
</tr>
<tr>
<td>SiC</td>
<td>20.0</td>
<td>Steel (quenched)</td>
<td>4.59</td>
</tr>
<tr>
<td>Si$_3$N$_4$</td>
<td>18.5</td>
<td>Cu (annealed)</td>
<td>0.25</td>
</tr>
<tr>
<td>$\text{Al}_2\text{O}_3$</td>
<td>14.0</td>
<td>Al (annealed)</td>
<td>0.12</td>
</tr>
<tr>
<td>B</td>
<td>13.5</td>
<td>Pb</td>
<td>0.032</td>
</tr>
<tr>
<td>WC</td>
<td>11.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data are from the text below by J. C. Anderson and others.

**Problems**

1. **Lines of closest packing.** Show that the lines of closest atomic packing are $\langle 110 \rangle$ in fcc structures and $\langle 111 \rangle$ in bcc structures.

2. **Dislocation pairs.** (a) Find a pair of dislocations equivalent to a row of lattice vacancies; (b) find a pair of dislocations equivalent to a row of interstitial atoms.

3. **Force on dislocation.** Consider a crystal in the form of a cube of side $L$ containing an edge dislocation of Burgers vector $b$. If the crystal is subjected to a shear stress $\sigma$ on the upper and lower faces in the direction of slip, show by considering energy balance that the force acting on the dislocation is $F = b\sigma L$. 
References

ELEMENTARY
S. Amelinckx, Direct observation of dislocations, Suppl. 6 to Solid state physics, Academic, 1964.

ADVANCED
21

Alloys

GENERAL CONSIDERATIONS 611

SUBSTITUTIONAL SOLID SOLUTIONS—
HUME-ROTHEY RULES 614

ORDER-DISORDER TRANSFORMATION 617
Elementary theory of order 618

PHASE DIAGRAMS 622
Eutectics 622

TRANSITION METAL ALLOYS 624
Electrical Conductivity 624

KONDO EFFECT 628

PROBLEMS 630

1. Superlattice lines in Cu$_3$Au 630
2. Configurational heat capacity 630

REFERENCES 630
Figure 1 Ordered (a) and disordered (b) arrangements of AB ions in alloy AB.
CHAPTER 21: ALLOYS

GENERAL CONSIDERATIONS

The theory of the band structure of solids assumes that the crystal has translational invariance. But suppose that the crystal is composed of two elements A and B that occupy at random the regular lattice sites of the structure, in proportions \( x \) and \( 1 - x \) for the composition \( A_x B_{1-x} \). The translational symmetry is no longer perfect.

Will we lose the consequences of band theory, such as the existence of Fermi surfaces and of energy gaps? Will insulators become conductors because the energy gap is gone? We touched on these questions in the discussion of amorphous semiconductors in Chapter 17.

Experiment and theory agree that the consequences of the destruction of perfect translational symmetry are much less serious (nearly always) than we expect at first sight. The viewpoint of the effective screened potential of Chapter 9 is helpful in these matters, first because the effective potentials are relatively weak in comparison with a free ion potential and, second and most important, the differences between the effective potentials of the host and the additive atoms may be very weak in comparison with either alone. Alloys of Si and Ge or of Cu and Ag are classic examples of what we may call the relative ineffectiveness of alloying.

In any event, a low concentration of impurity atoms cannot have much effect on the Fourier components \( U_G \) of the effective potential \( U(r) \) that is responsible for the band gaps and for the form of the Fermi surface. (This statement implies that the \( G \)'s exist, which implies that a regular lattice exists. This is not an important assumption because we know that thermal phonons do not have drastic effects, so that lattice distortions described as frozen-in phonons should not have drastic effects. If the distortions are more serious, as with amorphous solids, the electronic changes can be significant.)

It is true that an impurity atom will introduce Fourier components of \( U(r) \) at wavevectors that are not reciprocal lattice vectors, but at low impurity concentration such components are never large in comparison with the \( U_G \), arguing from the statistics of random potentials. The Fourier components at the reciprocal lattice vectors \( G \) will still be large and will give the band gaps, Fermi surfaces, and sharp x-ray diffraction lines characteristic of a regular lattice.

The consequences of alloying will be particularly small when the impurity element belongs to the same column of the periodic table as the host element it replaces, because the atomic cores will make rather similar contributions to the effective potential.
Figure 2  Resistivity of a disordered binary alloy of copper and gold. The variation of the residual resistivity depends on the composition Cu,Au_{1-x} as x(1 - x), which is known as Nordheim's Rule for a disordered alloy. Here x(1 - x) is a measure of the degree of maximum disorder possible for a given value of x. (Johansson and Linde.)

One measure of the effect of alloying is the residual electrical resistivity, defined as the low temperature limit of the resistivity. Here we must distinguish between disordered and ordered alloys. An alloy is disordered if the A and B atoms are randomly arranged, which occurs for a general value of x in the composition A_xB_{1-x}. For special values of x, such as 1/4, 1/2, and 3/4 for a cubic structure, it is possible for ordered phases to form, phases in which the A and B atoms form an ordered array. The distinction between order and disorder is shown in Fig. 1. The effect of order on the electrical resistivity is shown in Figs. 2 and 3. The residual resistivity increases with disorder, as discussed for amorphous materials in Chapter 17. The effect is shown in Fig. 2 for the Cu-Au alloy system. When the specimen is cooled slowly from a high temperature,
ordered structures are formed at Cu₃Au and CuAu; these structures have a lower residual resistivity by virtue of their order, as in Fig. 3.

Thus we can use the residual electrical resistivity to measure the effect of alloying in a disordered structure. One atomic percent of copper dissolved in silver (which lies in the same column of the periodic table) increases the residual resistivity by 0.077 μohm-cm. This corresponds to a geometrical scattering cross section which is only 3 percent of the naive “projected area” of the impurity atom, so that the scattering effect is very small.

In insulators there is no experimental evidence for a significant reduction of band gap caused by the random potential components. For example, silicon and germanium form homogeneous solid solutions, known as substitutional
alloys, over the entire composition range, but the band edge energies vary continuously with composition from the pure Si gap to the pure Ge gap.

It is widely believed, however, that the density of states near the band edges in amorphous materials is smudged out by the gross absence of translational symmetry. Some of the new states thus formed just inside the gap may not necessarily be current-carrying states because they may not extend throughout the crystal.

SUBSTITUTIONAL SOLID SOLUTIONS—HUME-ROTHERY RULES

We now discuss substitutional solid solutions of one metal A in another metal B of different valence, where A and B occupy, at random, equivalent sites in the structure. Hume-Rothery treated the empirical requirements for the stability of a solid solution of A and B as a single phase system.

One requirement is that the atomic diameters be compatible, which means that they should not differ by more than 15 percent. For example, the diameters are favorable in the Cu (2.55 Å) – Zn (2.65 Å) alloy system: zinc dissolves in copper as an fcc solid solution up to 38 atomic percent zinc. The diameters are less favorable in the Cu (2.55 Å) – Cd (2.97 Å) system, where only 1.7 atomic percent cadmium is soluble in copper. The atomic diameters referred to copper are 1.04 for zinc and 1.165 for cadmium.

Although the atomic diameters may be favorable, solid solutions will not form when there is a strong chemical tendency for A and B to form “intermetallic compounds,” which are compounds of definite chemical proportions. If A is strongly electronegative and B strongly electropositive, compounds such as AB and A2B may precipitate from the solid solution. (This is different from the formation of an ordered alloy phase only by the greater chemical bonding strength of the intermetallic compounds.) Although the atomic diameter ratio is favorable for As in Cu (1.02), only 6 atomic percent As is soluble. The diameter ratio is also favorable for Sb in Mg (1.06), yet the solubility of Sb in Mg is very small.

The electronic structure of alloys can often be described by the average number of conduction electrons (or valence electrons) per atom, denoted by n. In the alloy CuZn the value of n is 1.50, in CuAl, n = 2.00. Changes in electron concentration determine structural changes in many alloy systems.

The phase diagram of the copper-zinc system\(^1\) is shown in Fig. 4. The fcc structure of pure copper (n = 1) persists on the addition of zinc (n = 2) until the electron concentration reaches 1.38. A bcc structure occurs at a minimum elec-

\(^1\)The phases of interest are usually denoted by metallurgists by Greek characters: in the Cu-Zn system we have \(\alpha\) (fcc), \(\beta\) (bcc), \(\gamma\) (complex cubic cell of 52 atoms), \(\epsilon\) (hcp) and \(\eta\) (hcp). \(\epsilon\) and \(\eta\) differ considerably in \(e/a\) ratio. The meaning of the characters depends on the alloy system.
The alpha phase is fcc, beta and beta-prime are bcc, gamma is a complex structure, epsilon and eta are both hcp, but epsilon has a c/a ratio near 1.56 and eta (for pure Zn) has c/a = 1.86. The beta-prime phase is ordered bcc, by which we mean that most of the Cu atoms occupy sites on one sc sublattice and most of the Zn atoms occupy sites on a second sc sublattice that interpenetrates the first sublattice. The beta phase is disordered bcc: any site is equally likely to be occupied by a Cu or Zn atom, almost irrespective of what atoms are in the neighboring sites.

Electron concentration of about 1.48. The gamma phase exists for the approximate range of n between 1.58 and 1.66, and the hcp phase epsilon occurs near 1.75.

The term electron compound denotes an intermediate phase (such as the beta phase of CuZn) whose crystal structure is determined by a fairly well defined electron to atom ratio. For many alloys the ratio is close to the Hume-Rothery rules: 1.50 for the beta phase, 1.62 for the gamma phase, and 1.75 for the epsilon phase. Representative experimental values are collected in Table 1, based on the usual chemical valence of 1 for Cu and Ag; 2 for Zn and Cd; 3 for Al and Ga; 4 for Si, Ge, and Sn.

The Hume-Rothery rules find a simple expression in terms of the band theory of nearly free electrons. The observed limit of the fcc phase occurs close to the electron concentration of 1.36 at which an inscribed Fermi sphere makes contact with the Brillouin zone boundary for the fcc lattice. The observed electron concentration of the bcc phase is close to the concentration 1.48 at which an inscribed Fermi sphere makes contact with the zone boundary for the bcc
lattice. Contact of the Fermi sphere with the zone boundary for the γ phase is at the concentration 1.54. Contact for the hcp phase is at the concentration 1.69 for the ideal c/a ratio.

Why is there a connection between the electron concentrations at which a new phase appears and at which the Fermi surface makes contact with the boundary of the Brillouin zone? We recall that the energy bands split into two at the region of contact on the zone boundary (Chapter 9). If we add more electrons to the alloy at this stage, they will have to be accommodated in the upper band or in states of high energy near the zone corners of the lower band. Both options are possible, and both involve an increase of energy. Therefore it may be energetically favorable for the crystal structure to change to one which can contain a Fermi surface of larger volume (more electrons) before contact is made with the zone boundary. In this way H. Jones made plausible the sequence of structures fcc, bcc, γ, hcp with increasing electron concentration.

Measurements of the lattice parameter of Li-Mg alloys are shown in Fig. 5. In the range shown the structure is bcc. The lattice contracts during the initial stages of the addition of Mg to Li. When the lithium content drops below 50 atomic percent, corresponding to an average electron concentration increasing above 1.5 per atom, the lattice starts to expand. We have seen that for a spherical Fermi surface contact with the zone boundary is established at \( n = 1.48 \) electrons per atom, in a bcc lattice. It appears that the expansion of the lattice arises from the onset of overlap across the zone boundary.

The transformation from fcc to bcc is illustrated by Fig. 6; this shows the number of orbitals per unit energy range as a function of energy, for the fcc and bcc structures. As the number of electrons is increased, a point is reached where it is easier to accommodate additional electrons in the Brillouin zone of the bcc lattice rather than in the Brillouin zone of the fcc lattice. The figure is drawn for copper.
ORDER-DISORDER TRANSFORMATION

The dashed horizontal line in the beta-phase (bcc) region of the phase diagram (Fig. 4) of the Cu-Zn system represents the transition temperature between the ordered (low temperature) and disordered (high temperature) states of the alloy. In the common ordered arrangement of an AB alloy with a bcc structure all the nearest-neighbor atoms of a B atom are A atoms, and vice versa. This arrangement results when the dominant interaction among the atoms is an attraction between A and B atoms. (If the AB interaction is weakly attractive or repulsive, a two-phase system is formed in which some crystallites are largely A and other crystallites are largely B.)
The alloy is completely ordered at absolute zero. It becomes less ordered as the temperature is increased, until a transition temperature is reached above which the structure is disordered. The transition temperature marks the disappearance of long-range order, which is order over many interatomic distances, but some short-range order or correlation among near neighbors may persist above the transition. The long-range order in an AB alloy is shown in Fig. 7a. Long- and short-range order for an alloy of composition $AB_3$ is given in Fig. 7b. The degree of order is defined below.

If an alloy is cooled rapidly from high temperatures to a temperature below the transition, a metastable condition may be produced in which a nonequilibrium disorder is frozen in the structure. The reverse effect occurs when an ordered specimen is disordered at constant temperature by heavy irradiation with nuclear particles. The degree of order may be investigated experimentally by x-ray diffraction. The disordered structure in Fig. 8 has diffraction lines at the same positions as if the lattice points were all occupied by only one type of atom, because the effective scattering power of each plane is equal to the average of the $A$ and $B$ scattering powers. The ordered structure has extra diffraction lines not possessed by the disordered structure. The extra lines are called superstructure lines.

The use of the terms order and disorder in this chapter always refers to regular lattice sites; it is the occupancy that is randomly $A$ or $B$. Do not confuse this usage with that of Chapter 17 on noncrystalline solids where there are no regular lattice sites and the structure itself is random. Both possibilities, however different, occur in nature.

The structure of the ordered CuZn alloy is the cesium chloride structure (Fig. 1.20). The space lattice is simple cubic, and the basis has one Cu atom at 000 and one Zn atom at $\frac{1}{4}\frac{1}{4}\frac{1}{4}$. The diffraction structure factor

$$S(hkl) = f_{Cu} + f_{Zn} \ e^{-i\pi(h+k+l)} . \quad (1)$$

This cannot vanish because $f_{Cu} \neq f_{Zn}$; therefore all reflections of the simple cubic space lattice will occur. In the disordered structure the situation is different: the basis is equally likely to have either Zn or Cu at 000 and either Zn or Cu at $\frac{1}{4}\frac{1}{4}\frac{1}{4}$. Then the average structure factor is

$$\langle S(hkl) \rangle = \langle f \rangle + \langle f \rangle e^{-i\pi(h+k+l)} , \quad (2)$$

where $\langle f \rangle = \frac{1}{2}(f_{Cu} + f_{Zn})$. Equation (2) is exactly the form of the result for the bcc lattice; the reflections vanish when $h + k + l$ is odd. We see that the ordered lattice has reflections (the superstructure lines) not present in the disordered lattice (Fig. 8).

**Elementary Theory of Order**

We give a simple statistical treatment of the dependence of order on temperature for an AB alloy with a bcc structure. The case $A_3^+T$ differs from $AB$, the
Figure 7  (a) Long-range order versus temperature for an $AB$ alloy. The transformation is second order. (b) Long-range and short-range order for an $AB_3$ alloy. The transformation for this composition is first order.

Figure 8  X-ray powder photographs in AuCu$_3$ alloy. (a) Disordered by quenching from $T > T_c$, (b) ordered by annealing at $T < T_c$. (Courtesy of G. M. Gordon.)
former having a first-order transition marked by a latent heat, and the latter having a second-order transition marked by a discontinuity in the heat capacity (Fig. 9). We introduce a measure of the long-range order. We call one simple cubic lattice $a$ and the other $b$: the bcc structure is composed of the two interpenetrating sc lattices, and the nearest neighbors are on one lattice lie on the other lattice. If there are $N$ atoms $A$ and $N$ atoms $B$ in the alloy, the long-range order parameter $P$ is defined so that the number of $A$'s on the lattice $a$ is equal to $\frac{1}{2}(1 + P)N$. The number of $A$'s on lattice $b$ is equal to $\frac{1}{2}(1 - P)N$. When $P = \pm 1$, the order is perfect and each lattice contains only one type of atom. When $P = 0$, each lattice contains equal numbers of $A$ and $B$ atoms and there is no long-range order.

We consider that part of the internal energy associated with the bond energies of $AA$, $AB$, and $BB$ nearest-neighbor pairs. The total bond energy is

$$E = N_{AA}U_{AA} + N_{BB}U_{BB} + N_{AB}U_{AB},$$

where $N_{ij}$ is the number of nearest-neighbor $ij$ bonds, and $U_{ij}$ is the energy of an $ij$ bond.

The probability that an atom $A$ on lattice $a$ will have an $AA$ bond is equal to the probability that an $A$ occupies a particular nearest-neighbor site on $b$, times the number of nearest-neighbor sites, which is 8 for the bcc structure. We assume that the probabilities are independent. Thus, by the preceding expressions for the number of $A$'s on $a$ and $b$,

$$N_{AA} = 8\frac{1}{2}(1 + P)N[\frac{1}{2}(1 - P)] = 2(1 - P^2)N,$$
$$N_{BB} = 8\frac{1}{2}(1 + P)N[\frac{1}{2}(1 - P)] = 2(1 - P^2)N,$$
$$N_{AB} = 8N[\frac{1}{2}(1 + P)]^2 + 8N[\frac{1}{2}(1 - P)]^2 = 4(1 + P^2)N.$$

The energy (3) becomes

$$E = E_0 + 2NP^2U,$$

where

$$E_0 = 2N(U_{AA} + U_{BB} + 2U_{AB}); \quad U = 2U_{AB} - U_{AA} - U_{BB}.$$

We now calculate the entropy of this distribution of atoms. There are $\frac{1}{2}(1 + P)N$ atoms $A$ and $\frac{1}{2}(1 - P)N$ atoms $B$ on lattice $a$; there are $\frac{1}{2}(1 - P)N$ atoms $A$ and $\frac{1}{2}(1 + P)N$ atoms $B$ on lattice $b$. The number of arrangements $G$ of these atoms is

$$G = \left[ \frac{N!}{[\frac{1}{2}(1 + P)N]![\frac{1}{2}(1 - P)N]!} \right]^2.$$

From the definition of the entropy as $S = k_B \ln G$, we have, using Stirling's approximation,

$$S = 2Nk_B \ln 2 - Nk_B[(1 + P) \ln (1 + P) + (1 - P) \ln (1 - P)].$$

This defines the entropy of mixing. For $P = \pm 1$, $S = 0$; for $P = 0$, $S = 2Nk_B \ln 2$. 

Figure 9  Heat capacity versus temperature of CuZn alloy (β-brass).

The equilibrium order is determined by the requirement that the free energy $F = E - TS$ be a minimum with respect to the order parameter $P$. On differentiating $F$ with respect to $P$, we have as the condition for the minimum

$$4NPU + Nk_B T \ln \frac{1 + P}{1 - P} = 0.$$  

The transcendental equation for $P$ may be solved graphically; we find the smoothly decreasing curve shown in Fig. 7a. Near the transition we may expand (9), to find $4NPU + 2Nk_B TP = 0$. At the transition temperature $P = 0$, so that

$$T_c = -2U/k_B .$$  

For a transition to occur, the effective interaction $U$ must be negative.

The short-range order parameter $r$ is a measure of the fraction of the average number $q$ of nearest-neighbor bonds that are $AB$ bonds. When completely disordered, an $AB$ alloy has an average of four $AB$ bonds about each atom $A$. The total possible is eight. We may define

$$r = \frac{1}{4}(q - 4) ,$$

so that $r = 1$ in complete order and $r = 0$ in complete disorder. Observe that $r$ is a measure only of the local order about an atom, whereas the long-range order parameter $P$ refers to the purity of the entire population on a given sublattice. Above the transition temperature $T_c$ the long-range order is rigorously zero, but the short-range order is not.
PHASE DIAGRAMS

There is a large amount of information in a phase diagram even for a binary system, as in Fig. 4. The areas enclosed by curves relate to the equilibrium state in that region of composition and temperature. The curves mark the course of phase transitions as plotted in the $T$-$x$ plane, where $x$ is the composition parameter.

The equilibrium state is the state of minimum free energy of the binary system at given $T$, $x$. Thus the analysis of a phase diagram is the subject of thermodynamics. Several extraordinary results come out of this analysis, in particular the existence of low-melting-point eutectic compositions. Because the analysis has been treated in Chapter 11 of *TP*, we only outline the principal results here.

Two substances will dissolve in each other and form a homogeneous mixture if that is the configuration of lowest free energy accessible to the components. The substances will form a heterogeneous mixture if the combined free energy of the two separate phases side by side is lower than the free energy of the homogeneous mixture. Now we say that the mixture exhibits a solubility gap. In Fig. 4 we see that compositions near Cu$_{0.69}$Zn$_{0.40}$ are in a solubility gap and are mixtures of fcc and bcc phases of different structures and compositions. The phase diagram represents the temperature dependence of the solubility gaps.

When a small fraction of a homogeneous liquid freezes, the composition of the solid that forms is almost always different from that of the liquid. Consider a horizontal section near the composition Cu$_{0.86}$Zn$_{0.20}$ in Fig. 4. Let $x$ denote the weight percent of zinc. At a given temperature, there are three regions:

- $x > x_L$, the equilibrium system is a homogeneous liquid.
- $x_S < x < x_L$, there is a solid phase of composition $x_S$ and a liquid phase of composition $x_L$.
- $x < x_S$, equilibrium system is a homogeneous solid.

The point $x_L$ traces a curve called the liquidus curve, and the point $x_S$ traces the solidus curve.

**Eutectics.** Mixtures with two liquidus branches in their phase diagram are called eutectics, as in Fig. 10 for the Au-Si system. The minimum solidification temperature is called the eutectic temperature; here the composition is the eutectic composition. The solid at this composition consists of two separate phases, as in the microphotograph of Fig. 11.

There are many binary systems in which the liquid phase persists to temperatures below the lower melting temperature of the constituents. Thus Au$_{0.69}$Si$_{0.31}$ solidifies at 370°C as a two-phase heterogeneous mixture, although Au and Si solidify at 1063°C and 1404°C, respectively. One phase of the eutectic is nearly pure gold; the other nearly pure silicon.
Figure 10  Eutectic phase diagram of gold-silicon alloys. The eutectic consists of two branches that come together at $T_e = 370^\circ$C and $x_e = 0.31$ atomic percent Si. (After Kittel and Kroemer, TF.)

Figure 11  Microphotograph of the Pb-Sn eutectic. (Courtesy of J. D. Hunt and K. A. Jackson.)
The Au-Si eutectic is important in semiconductor technology because the eutectic permits low temperature welding of gold contact wires to silicon devices. Lead-tin alloys have a similar eutectic of Pb$_{0.26}$Sn$_{0.74}$ at 183°C. This or nearby compositions are used in solder: nearby if a range of melting temperatures is desired for ease in handling.

**TRANSITION METAL ALLOYS**

When we add copper to nickel the effective magneton number per atom decreases linearly and goes through zero near Cu$_{0.60}$Ni$_{0.40}$, as shown in Fig. 12. At this composition the extra electron from the copper has filled the 3$d$ band, or the spin-up and spin-down 3$d$ sub-bands that were shown in Fig. 15.7b. The situation is shown schematically in Fig. 13.

For simplicity the block drawings represent the density of states as uniform in energy. The actual density is known to be far from uniform; the result of a modern calculation is shown in Fig. 14 for nickel. The width of the 3$d$ band is about 5 eV. At the top, where the magnetic effects are determined, the density of states is particularly high. The average density of states is an order of magnitude higher in the 3$d$ band than in the 4$s$ band. This enhanced density of states ratio gives a rough indication of the expected enhancement of the electronic heat capacity and of the paramagnetic susceptibility in the nonferromagnetic transition metals as compared with the simple monovalent metals.

Figure 15 shows the effect of the addition of small amounts of other elements to nickel. On the band model an alloying metal with $z$ valence electrons outside a filled $d$ shell is expected to decrease the magnetization of nickel by approximately $z$ Bohr magnetons per solute atom. This simple relation holds well for Sn, Al, Zn, and Cu, with $z = 4$, 3, 2, and 1, respectively. For Co, Fe, and Mn the localized moment model of Friedel accounts for effective $z$ values of −1, −2, and −3, respectively.

The average atomic magnetic moments of binary alloys of the elements in the iron group are plotted in Fig. 16 as a function of the concentration of electrons outside the 3$p$ shell. This is called a Slater-Pauling plot. The main sequence of alloys on the right-hand branch follows the rules discussed in connection with Fig. 15. As the electron concentration is decreased a point is reached at which neither of the 3$d$ sub-bands is entirely filled, and the magnetic moment then decreases toward the left-hand side of the plot.

**Electrical Conductivity.** It might be thought that in the transition metals the availability of the 3$d$ band as a path for conduction in parallel with the 4$s$ band would increase the conductivity, but this is not the way it works out. The resistivity of the $s$ electron path is increased by collisions with the $d$ electrons; this is a powerful extra scattering mechanism not available when the $d$ band is filled.
Figure 12  Bohr magneton numbers of nickel-copper alloys.

Figure 13  Distribution of electrons in the alloy 60Cu:40Ni. The extra 0.6 electron provided by the copper has filled the $d$ band entirely and increased slightly the number of electrons in the $s$ band with respect to Fig. 15.7b.

Figure 14  Density of states in nickel. (V. L. Moruzzi, J. F. Janak, and A. R. Williams.)
Figure 15 Saturation magnetization of nickel alloys in Bohr magnetons per atom as a function of the atomic percent of solute element.
We compare the values of the electrical resistivities of Ni, Pd, and Pt in microhm-cm at 18°C with that of the noble metals Cu, Ag, and Au immediately following them in the periodic table:

<table>
<thead>
<tr>
<th></th>
<th>Ni</th>
<th>Pd</th>
<th>Pt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.4</td>
<td>10.8</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>Cu</td>
<td>Ag</td>
<td>Au</td>
</tr>
<tr>
<td></td>
<td>1.7</td>
<td>1.6</td>
<td>2.2</td>
</tr>
</tbody>
</table>

The resistivities of the noble metals are lower than those of the transition metals by a factor of the order of 1/5. This shows the effectiveness of the s-d scattering mechanism.
In dilute solid solutions of a magnetic ion in a nonmagnetic metal crystal (such as Mn in Cu) the exchange coupling between the ion and the conduction electrons has important consequences. The conduction electron gas is magnetized in the vicinity of the magnetic ion, with the spatial dependence shown in Fig. 17. This magnetization causes an indirect exchange interaction between two magnetic ions, because a second ion perceives the magnetization induced by the first ion. The interaction, known as the Friedel or RKKY interaction, also plays a role in the magnetic spin order of the rare-earth metals, where the spins of the $4f$ ion cores are coupled together by the magnetization induced in the conduction electron gas.

A consequence of the magnetic ion-conduction electron interaction is the Kondo effect. A minimum in the electrical resistivity-temperature curve of dilute magnetic alloys at low temperatures has been observed in alloys of Cu, Ag, Au, Mg, Zn with Cr, Mn, and Fe as impurities, among others.

---

Figure 18 A comparison of experimental and theoretical results for the increase of electrical resistivity at low temperatures in dilute alloys of iron in gold. The resistance minimum lies to the right of the figure, for the resistivity increases at high temperatures because of scattering of electrons by thermal phonons. The experiments are due to D. K. C. MacDonald, W. B. Pearson, and I. M. Templeton; the theory is by J. Kondo. An exact solution was given by K. Wilson.

The occurrence of a resistance minimum is connected with the existence of localized magnetic moments on the impurity atoms. Where a resistance minimum is found, there is inevitably a local moment. Kondo showed that the anomalously high scattering probability of magnetic ions at low temperatures is a consequence of the dynamic nature of the scattering by the exchange coupling and of the sharpness of the Fermi surface at low temperatures. The temperature region in which the Kondo effect is important is shown in Fig. 18.

The central result is that the spin-dependent contribution to the resistivity is

$$
\rho_{\text{spin}} = c \rho_M \left[ 1 + \frac{3zJ}{\epsilon_F} \ln T \right] = c \rho_0 - c \rho_1 \ln T ,
$$

(12)

where $J$ is the exchange energy; $z$ the number of nearest neighbors; $c$ the concentration; and $\rho_M$ is a measure of the strength of the exchange scattering.
We see that the spin resistivity increases toward low temperatures if $J$ is negative. If the phonon contribution to the electrical resistivity goes as $T^5$ in the region of interest and if the resistivities are additive, then the total resistivity has the form

$$\rho = aT^5 + c\rho_0 - c\rho_1 \ln T,$$

with a minimum at

$$\frac{d\rho}{dT} = 5aT^4 - \frac{c\rho_1}{T} = 0,$$

whence

$$T_{\text{min}} = \left(\frac{c\rho_1}{5a}\right)^{1/5}.\quad (15)$$

The temperature at which the resistivity is a minimum varies as the one-fifth power of the concentration of the magnetic impurity, in agreement with experiment at least for Fe in Cu.

**Problems**

1. **Superlattice lines in Cu$_2$Au.** Cu$_2$Au alloy (75% Cu, 25% Au) has an ordered state below 400°C, in which the gold atoms occupy the 000 positions and the copper atoms the $\frac{1}{2},0, \frac{1}{2}$, and $0\frac{1}{2}, \frac{1}{2}$ positions in a face-centered cubic lattice. Give the indices of the new x-ray reflections that appear when the alloy goes from the disordered to the ordered state. List all new reflections with indices $\leq 2$.

2. **Configurational heat capacity.** Derive an expression in terms of $P(T)$ for the heat capacity associated with order/disorder effects in an AB alloy. [The entropy (8) is called the configurational entropy or entropy of mixing.]

**References**


APPENDIX A: TEMPERATURE DEPENDENCE OF THE REFLECTION LINES

\[ \ldots I \text{ came to the conclusion that the sharpness of the interference lines would not suffer but that their intensity should diminish with increasing angle of scattering, the more so the higher the temperature.} \]

P. Debye

As the temperature of the crystal is increased, the intensity of the Bragg-reflected beams decreases, but the angular width of the reflected line does not change. Experimental intensities for aluminum are shown in Fig. 1. It is surprising that we can get a sharp x-ray reflection from atoms undergoing large amplitude random thermal motion, with instantaneous nearest-neighbor spacings differing by 10 percent at room temperature. Before the Laue experiment

![Figure 1](image-url)

*Figure 1* The dependence of intensity on temperature for the \((h00)\) x-ray reflections of aluminum. Reflections \((h00)\) with \(h\) odd are forbidden for an fcc structure. (After R. M. Nicklow and R. A. Young.)
was done, but when the proposal was discussed\(^1\) in a coffee house in Munich, the objection was made that the instantaneous positions of the atoms in a crystal at room temperature are far from a regular periodic array, because of the large thermal fluctuation. Therefore, the argument went, one should not expect a well-defined diffracted beam.

But such a beam is found. The reason was given by Debye. Consider the radiation amplitude scattered by a crystal: let the position of the atom nominally at \( r_j \) contain a term \( u(t) \) fluctuating in time: \( r(t) = r_j + u(t) \). We suppose each atom fluctuates independently about its own equilibrium position.\(^2\) Then the thermal average of the structure factor (2.43) contains terms

\[
f_j \exp(-iG \cdot r_j)\langle \exp(-iG \cdot u) \rangle ,
\]

where \( \langle \cdots \rangle \) denotes thermal average. The series expansion of the exponential is

\[
\langle \exp(-iG \cdot u) \rangle = 1 - i\langle G \cdot u \rangle - \frac{1}{2!}\langle (G \cdot u)^2 \rangle + \cdots .
\]

But \( \langle G \cdot u \rangle = 0 \), because \( u \) is a random thermal displacement uncorrelated with the direction of \( G \). Further,

\[
\langle (G \cdot u)^2 \rangle = G^2 \langle u^2 \rangle (\cos^2 \theta) = \frac{1}{3} \langle u^2 \rangle G^2 .
\]

The factor \( \frac{1}{3} \) arises as the geometrical average of \( \cos^2 \theta \) over a sphere.

The function

\[
\exp(-\frac{1}{6} \langle u^2 \rangle G^2) = 1 - \frac{1}{6} \langle u^2 \rangle G^2 + \cdots
\]

has the same series expansion as (2) for the first two terms shown here. For a harmonic oscillator all terms in the series (2) and (3) can be shown to be identical. Then the scattered intensity, which is the square of the amplitude, is

\[
I = I_0 \exp(-\frac{1}{3} \langle u^2 \rangle G^2) ,
\]

where \( I_0 \) is the scattered intensity from the rigid lattice. The exponential factor is the Debye-Waller factor.

Here \( \langle u^2 \rangle \) is the mean square displacement of an atom. The thermal average potential energy \( \langle U \rangle \) of a classical harmonic oscillator in three dimensions is \( \frac{3}{2} k_B T \), whence

\[
\langle U \rangle = \frac{1}{2} C \langle u^2 \rangle = \frac{1}{2} M \omega^2 \langle u^2 \rangle = \frac{3}{2} k_B T ,
\]

where \( C \) is the force constant, \( M \) is the mass of an atom, and \( \omega \) is the frequency of the oscillator. We have used the result \( \omega^2 = C/M \). Thus the scattered intensity is

\[
I(hkl) = I_0 \exp(-k_B T G^2/M \omega^2) ,
\]

\(^1\)P. P. Ewald, private communication.

\(^2\)This is the Einstein model of a solid, it is not a very good model at low temperatures, but it works well at high temperatures. For a general treatment of scattering by thermal fluctuations, see QTS, Chapter 20.
where $hkl$ are the indices of the reciprocal lattice vector $G$. This classical result is a good approximation at high temperatures.

For quantum oscillators $\langle u^2 \rangle$ does not vanish even at $T = 0$; there is zero-point motion. On the independent harmonic oscillator model the zero-point energy is $\frac{1}{2} \hbar \omega$; this is the energy of a three-dimensional quantum harmonic oscillator in its ground state referred to the classical energy of the same oscillator at rest. Half of the oscillator energy is potential energy, so that in the ground state

$$\langle U \rangle = \frac{1}{2} M \omega^2 \langle u^2 \rangle = \frac{3}{4} \hbar \omega; \quad \langle u^2 \rangle = 3 \hbar / 2 M \omega,$$

whence, by (4),

$$I(hkl) = I_0 \exp(-\hbar G^2 / 2M \omega)$$

at absolute zero. If $G = 10^9$ cm$^{-1}$, $\omega = 10^{14}$ s$^{-1}$, and $M = 10^{-22}$ g, the argument of the exponential is approximately 0.1, so that $I/I_0 = 0.9$. At absolute zero 90 percent of the beam is elastically scattered and 10 percent is inelastically scattered.

We see from (6) and from Fig. 1 that the intensity of the diffracted line decreases, but not catastrophically, as the temperature is increased. Reflections of low $G$ are affected less than reflections of high $G$. The intensity we have calculated is that of the coherent diffraction or the elastic scattering in the well-defined Bragg directions. The intensity lost from these directions is the inelastic scattering and appears as a diffuse background. In inelastic scattering the x-ray photon causes the excitation or de-excitation of a lattice vibration, and the photon changes direction and energy.

At a given temperature the Debye-Waller factor of a diffraction line decreases with an increase in the magnitude of the reciprocal lattice vector $G$ associated with the reflection. The larger $|G|$ is, the weaker the reflection at high temperatures. The temperature dependence of the reflected intensity for the $(h00)$ reflections of aluminum is shown in Fig. 1. The theory we have worked out here for x-ray reflection applies equally well to neutron diffraction and to the Mössbauer effect, the recoilless emission of gamma rays by nuclei bound in crystals.

X-rays can be absorbed in a crystal also by the inelastic processes of photo-ionization of electrons and Compton scattering. In the photoeffect the x-ray photon is absorbed and an electron is ejected from an atom. In the Compton effect the photon is scattered inelastically by an electron: the photon loses energy and the electron is ejected from an atom. The depth of penetration$^3$ of the x-ray beam depends on the solid and on the photon energy, but 1 cm is typical. A diffracted beam in Bragg reflection may remove the energy in a much shorter distance, perhaps $10^{-3}$ cm in an ideal crystal.

APPENDIX B: Ewald Calculation of Lattice Sums

The problem is to calculate the electrostatic potential experienced by one ion in the presence of all the other ions in the crystal. We consider a lattice made up of ions with positive or negative charges and shall assume that the ions are spherical.

We compute the total potential \( \varphi = \varphi_1 + \varphi_2 \) at an ion as the sum of two distinct but related potentials. The potential \( \varphi_1 \) is that of a structure with a Gaussian distribution of charge situated at each ion site, with signs the same as those of the real ions. According to the definition of the Madelung constant, the charge distribution on the reference point is not considered to contribute to the potential \( \varphi_1 \) or \( \varphi_2 \) (Fig. 1a). We therefore calculate the potential \( \varphi_1 \) as the difference

\[
\varphi_1 = \varphi_a - \varphi_b
\]

Figure 1  (a) Charge distribution used for computing potential \( \varphi_1 \); the potential \( \varphi_a \) is computed (it includes the dashed curve at the reference point), while \( \varphi_b \) is the potential of the dashed curve alone. (b) Charge distribution for potential \( \varphi_2 \). The reference point is denoted by an X.
of two potentials, \( \varphi_a \) being the potential of a continuous series of Gaussian distributions and \( \varphi_b \) being the potential of the single Gaussian distribution on the reference point.

The potential \( \varphi_2 \) is that of a lattice of point charges with an additional Gaussian distribution of opposite sign superposed upon the point charges (Fig. 1b).

The point of splitting the problem into the two parts \( \varphi_1 \) and \( \varphi_2 \) is that by a suitable choice of the parameter determining the width of each Gaussian peak we can get very good convergence of both parts at the same time. The Gaussian distributions drop out completely on taking the sum of the separate charge distributions giving rise to \( \varphi_1 \) and \( \varphi_2 \), so that the value of the total potential \( \varphi \) is independent of the width parameter, but the rapidity of convergence depends on the value chosen for the parameter.

We calculate first the potential \( \varphi_a \) of a continuous Gaussian distribution. We expand \( \varphi_a \) and the charge density \( \rho \) in Fourier series:

\[
\varphi_a = \sum_C c_C \exp(iG \cdot r) ;
\]

\[
\rho = \sum_C \rho_C \exp(iG \cdot r) ,
\]

where \( G \) is \( 2\pi \) times a vector in the reciprocal lattice. The Poisson equation is

\[
\nabla^2 \varphi_a = -4\pi \rho ,
\]

or

\[
\sum G^2 c_C \exp(iG \cdot r) = 4\pi \sum \rho_C \exp(iG \cdot r) ,
\]

so that

\[
c_C = 4\pi \rho_C / G^2 .
\]

We suppose in finding \( \rho_C \) that there is associated with each lattice point of the Bravais lattice a basis containing ions of charge \( q_i \) at positions \( r_i \) relative to the lattice point. Each ion point is therefore the center of a Gaussian charge distribution of density

\[
\rho(r) = q_i (\eta/\pi)^{3/2} \exp(-\eta r^2) ,
\]

where the factor in front of the exponential ensures that the total charge associated with the ion is \( q_i \); the range parameter \( \eta \) is to be chosen judiciously to ensure rapid convergence of the final result (6), which is in value independent of \( \eta \).

We would normally evaluate \( \rho_C \) by multiplying both sides of (2) by \( \exp(-iG \cdot r) \) and integrating over the volume \( \Delta \) of one cell, so that the charge distribution to be considered is that originating on the ion points within the cell
and also that of the tails of the distributions originating in all other cells. It is easy to see, however, that the integral of the total charge density times \( \exp[-iG \cdot r] \) over a single cell is equal to the integral of the charge density originating in a single cell times \( \exp[-iG \cdot r] \) over all space.

We have therefore

\[
\rho_c \int_{\text{one cell}} \exp(iG \cdot r) \exp(-iG \cdot r) \, dr = \rho_c \Delta \\
= \int \sum_{t} q_t(\eta/\pi)^{3/2} \exp[-\eta(r - r_t)^2] \exp(-iG \cdot r) \, dr .
\]

This expression is readily evaluated:

\[
\rho_c \Delta = \sum_{t} q_t \exp(-iG \cdot r_t) (\eta/\pi)^{3/2} \int_{\text{all space}} \exp[-(iG \cdot \xi + \eta \xi^2)] \, d\xi \\
= \left( \sum_{t} q_t \exp(-iG \cdot r_t) \right) \exp(-G^2/4\eta) = S(G)\exp(-G^2/4\eta) ,
\]

where \( S(G) = \sum_{t} q_t \exp(-iG \cdot r_t) \) is just the structure factor (Chapter 2) in appropriate units. Using (1) and (3),

\[
\varphi = \frac{4\pi}{\Delta} \sum_{G} S(G)G^{-2} \exp(iG \cdot r - G^2/4\eta) .
\]

At the origin \( r = 0 \) we have

\[
\varphi = \frac{4\pi}{\Delta} \sum_{G} S(G)G^{-2} \exp(-G^2/4\eta) .
\]

The potential \( \varphi_b \) at the reference ion point \( i \) due to the central Gaussian distribution is

\[
\varphi_b = \int_{0}^{\infty} (4\pi r^2 \, dr)(\rho/r) = 2q_i(\eta/\pi)^{1/2} ,
\]

and so

\[
\varphi_1(i) = \frac{4\pi}{\Delta} \sum_{G} S(G)G^{-2} \exp(-G^2/4\eta) - 2q_i(\eta/\pi)^{1/2} .
\]

The potential \( \varphi_2 \) is to be evaluated at the reference point, and it differs from zero because other ions have the tails of their Gaussian distributions over-
lapping the reference point. The potential is due to three contributions from each ion point:

\[ q_l \left[ \frac{1}{r_l} - \frac{1}{r_l} \int_0^{r_l} \rho(r) \, dr - \int_{r_l}^{\infty} \frac{\rho(r)}{r} \, dr \right], \]

where the terms are from the point charge, from the part of the Gaussian distribution lying inside a sphere of radius \( r_l \) about the \( l \)th ion point, and from that part lying outside the sphere, respectively. On substituting for \( \rho(r) \) and carrying out elementary manipulations, we have

\[ \varphi_2 = \sum_l \frac{q_l}{r_l} F(\eta^{1/2} r_l), \quad (5) \]

where

\[ F(x) = \left( \frac{2}{\pi^{1/2}} \right) \int_x^{\infty} \exp(-s^2) \, ds. \]

Finally,

\[ \varphi(i) = \frac{4\pi}{\Delta} \sum_G S(G)G^{-2} \exp(-G^2/4\eta) - 2q_i(\eta/\pi)^{1/2} + \sum_l \frac{q_l}{r_l} F(\eta^{1/2} r_l) \quad (6) \]

is the desired total potential of the reference ion \( i \) in the field of all the other ions in the crystal. In the application of the Ewald method the trick is to choose \( \eta \) such that both sums in (6) converge rapidly.

**Ewald-Kornfeld Method for Lattice Sums for Dipole Arrays**

Kornfeld extended the Ewald method to dipolar and quadrupolar arrays. We discuss here the field of a dipole array at a point which is not a lattice point. According to (4) and (5) the potential at a point \( r \) in a lattice of positive unit point charges is

\[ \varphi = (4\pi/\Delta) \sum_G S(G)G^{-2} \exp[iG \cdot r - G^2/4\eta] + \sum_l F(\sqrt{\eta} r_l)/r_l, \quad (7) \]

where \( r_l \) is the distance from \( r \) to the lattice point \( l \).

The first term on the right gives the potential of the charge distribution \( \rho = (\eta/\pi)^{3/2} \exp(-\eta r^2) \) about each lattice point. By a well-known relation in electrostatics we obtain the potential of an array of unit dipoles pointing in the \( z \) direction by taking \(-dV/dz\) of the above potential. The term under discussion contributes

\[ -(4\pi i/\Delta) \sum_G S(G)G_zG^2 \exp[iG \cdot r - G^2/4\eta], \]

and the \( z \) component of the electric field from this term is \( E_z = \partial^2 \varphi / \partial z^2 \), or
\[-(4\pi/\Delta) \sum \frac{S(G)}{(G_z^2/G^2)} \exp[iG \cdot r - G^2/4\eta] \,. \tag{8}\]

The second term on the right of (7) after one differentiation gives

\[-\sum \frac{z_i}{l} \left\{ (F'/(\sqrt{\eta r_i})/r_i^3) + (2/r_i^2)(\eta/\pi)^{1/2} \exp(-\eta r_i^2) \right\} ,
\]

and the z component of this part of the field is

\[\sum \frac{z_i^2}{l} \left\{ (3F/(\sqrt{\eta r_i})/r_i^5) + (6/r_i^4)(\eta/\pi)^{1/2} \exp(-\eta r_i^2) \right\} + (4/r_i^3)(\eta^2/\pi)^{1/2} \exp(-\eta r_i^2) + (2/r_i^2)(\eta/\pi)^{1/2} \exp(-\eta r_i^2) \right\} . \tag{9}\]

The total \(E_z\) is given by the sum of (8) and (9). The effects of any number of lattices may be added.

APPENDIX C: QUANTIZATION OF ELASTIC WAVES: PHONONS

Phonons were introduced in Chapter 4 as quantized elastic waves. How do we quantize an elastic wave? As a simple model of phonons in a crystal, consider the vibrations of a linear lattice of particles connected by springs. We can quantize the particle motion exactly as for a harmonic oscillator or set of coupled harmonic oscillators. To do this we make a transformation from particle coordinates to phonon coordinates, also called wave coordinates because they represent a traveling wave.

Let \(N\) particles of mass \(M\) be connected by springs of force constant \(C\) and length \(a\). To fix the boundary conditions, let the particles form a circular ring. We consider the transverse displacements of the particles out of the plane of the ring. The displacement of particle \(s\) is \(q_s\) and its momentum is \(p_s\). The hamiltonian of the system is

\[H = \sum_{s=1}^{n} \left\{ \frac{1}{2M} p_s^2 + \frac{1}{2} C(q_{s+1} - q_s)^2 \right\} . \tag{1}\]

The hamiltonian of a harmonic oscillator is

\[H = \frac{1}{2M} p^2 + \frac{1}{2} Cx^2 , \tag{2}\]
and the energy eigenvalues are, where \( n = 0, 1, 2, 3, \ldots \),

\[
\epsilon_n = \left(n + \frac{1}{2}\right) \hbar \omega .
\]  

(3)

The eigenvalue problem is also exactly solvable for a chain with the different hamiltonian (1).

To solve (1) we make a Fourier transformation from the coordinates \( p_s, q_s \) to the coordinates \( P_k, Q_k \), which are known as phonon coordinates.

**Phonon Coordinates**

The transformation from the particle coordinates \( q_s \) to the phonon coordinates \( Q_k \) is used in all periodic lattice problems. We let

\[
q_s = N^{-1/2} \sum_k Q_k \exp(iks \alpha) ,
\]  

(4)

consistent with the inverse transformation

\[
Q_k = N^{-1/2} \sum_s q_s \exp(-iks \alpha) .
\]  

(5)

Here the \( N \) values of the wavevector \( k \) allowed by the periodic boundary condition \( q_s = q_{s+N} \) are given

\[
k = 2\pi n / Na ; n = 0, \pm 1, \pm 2, \cdots, \pm \left( \frac{1}{2} N - 1 \right), \frac{1}{2} N .
\]  

(6)

We need the transformation from the particle momentum \( p_s \) to the momentum \( P_k \) that is canonically conjugate to the coordinate \( Q_k \). The transformation is

\[
p_s = N^{-1/2} \sum_k P_k \exp(-iks \alpha) ; P_k = N^{-1/2} \sum_s p_s \exp(iks \alpha) .
\]  

(7)

This is not quite what one would obtain by the naive substitution of \( p \) for \( q \) and \( P \) for \( Q \) in (4) and (5), because \( k \) and \( -k \) have been interchanged between (4) and (7).

We verify that our choice of \( P_k \) and \( Q_k \) satisfies the quantum commutation relation for canonical variables. We form the commutator

\[
[Q_k, P_k] = N^{-1} \left[ \sum_r q_r \exp(-ik'ra), \sum_s p_s \exp(ik's\alpha) \right] 
\]

\[= N^{-1} \sum_r \sum_s [q_r, p_s] \exp[-i(kr - k's)a] .
\]  

(8)
Because the operators $q$, $p$ are conjugate, they satisfy the commutation relation

$$[q_r, p_s] = i\hbar \delta(r, s),$$

where $\delta(r, s)$ is the Kronecker delta symbol.

Thus (8) becomes

$$[Q_k, P_{k'}] = N^{-1} i\hbar \sum_r \exp[-i(k - k')ra] = i\hbar \delta(k, k'),$$

so that $Q_k$, $P_k$ also are conjugate variables. Here we have evaluated the summation as

$$\sum_r \exp[-i(k - k')ra] = \sum_r \exp[-i2\pi n n'/N] = N\delta(n, n') = N\delta(k, k'),$$

where we have used (6) and a standard result for the finite series in (11).

We carry out the transformations (7) and (4) on the hamiltonian (1), and make use of the summation (11):

$$\sum_s p_s^2 = N^{-1} \sum_s \sum_k \sum_{k'} P_k P_{k'} \exp[-i(k + k')sa]$$

$$= \sum_k \sum_{k'} P_k P_{k'} \delta(-k, k') = \sum_k P_k P_{-k};$$

$$\sum_s (q_{s+1} - q_s)^2 = N^{-1} \sum_s \sum_k \sum_{k'} Q_k Q_{k'} \exp(ika)[\exp(ika) - 1]$$

$$\times \exp(ika)[\exp(ika) - 1] = 2 \sum_k Q_k Q_{-k}(1 - \cos ka).$$

Thus the hamiltonian (1) becomes, in phonon coordinates,

$$H = \sum_k \left\{ \frac{1}{2M} P_k P_{-k} + C Q_k Q_{-k}(1 - \cos ka) \right\}.$$

If we introduce the symbol $\omega_k$ defined by

$$\omega_k = (2C/M)^{1/2} (1 - \cos ka)^{1/2},$$

we have the phonon hamiltonian in the form

$$H = \sum_k \left\{ \frac{1}{2M} P_k P_{-k} + \frac{1}{2} M \omega_k^2 Q_k Q_{-k} \right\}. $$
The equation of motion of the phonon coordinate operator $Q_k$ is found by the standard prescription of quantum mechanics:

$$i\hbar  \dot{Q}_k = [Q_k, H] = i\hbar P_{-k}/M,$$

with $H$ given by (14). Further, using the commutator (17),

$$i\hbar  \dot{Q}_k = [Q_k, H] = M^{-1}[P_{-k}, H] = i\hbar \omega_k^2 Q_k,$$

so that

$$\ddot{Q}_k + \omega_k^2 Q_k = 0.$$

This is the equation of motion of a harmonic oscillator with the frequency $\omega_k$.

The energy eigenvalues of a quantum harmonic oscillator are

$$\epsilon_k = \left(n_k + \frac{1}{2}\right)\hbar \omega_k,$$

where the quantum number $n_k = 0, 1, 2, \ldots$. The energy of the entire system of all phonons is

$$U = \sum_k \left(n_k + \frac{1}{2}\right)\hbar \omega_k.$$

This result demonstrates the quantization of the energy of elastic waves on a line.

**Creation and Annihilation Operators**

It is helpful in advanced work to transform the phonon hamiltonian (16) into the form of a set of harmonic oscillators:

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

Here $a_k^+$, $a_k$ are harmonic oscillator operators, also called creation and destruction operators or boson operators. The transformation is derived below.

The boson creation operator $a^+$ which "creates a phonon" is defined by the property

$$a^+ |n\rangle = (n + 1)^{1/2} |n + 1\rangle,$$

when acting on a harmonic oscillator state of quantum number $n$, and the boson annihilation operator $a$ which "destroys a phonon" is defined by the property

$$a |n\rangle = n^{1/2} |n - 1\rangle.$
It follows that
\[ a^+ a |n\rangle = a^+ n^{1/2} |n - 1\rangle = n |n\rangle, \tag{25} \]
so that \(|n\rangle\) is an eigenstate of the operator \(a^+ a\) with the integral eigenvalue \(n\), called the quantum number or occupancy of the oscillator. When the phonon mode \(k\) is in the eigenstate labeled by \(n_k\), we may say that there are \(n_k\) phonons in the mode. The eigenvalues of (22) are \(U = \Sigma (n_k + \frac{1}{2})\hbar \omega_k\), in agreement with (21).

Because
\[ aa^+ |n\rangle = a(n + 1)^{1/2} |n + 1\rangle = (n + 1) |n\rangle, \tag{26} \]
the commutator of the boson wave operators \(a_k^+\) and \(a_k\) satisfies the relation
\[ [a, a^+] = aa^+ - a^+ a - 1. \tag{27} \]

We still have to prove that the hamiltonian (16) can be expressed as (19) in terms of the phonon operators \(a_k^+, a_k\). This can be done by the transformation
\[ a_k^+ = (2\hbar)^{-1/2}[(M \omega_k)^{1/2} Q_{-k} - i(M \omega_k)^{-1/2} P_k], \tag{28} \]
\[ a_k = (2\hbar)^{-1/2}[(M \omega_k)^{1/2} Q_k + i(M \omega_k)^{-1/2} P_{-k}]. \tag{29} \]

The inverse relations are
\[ Q_k = (\hbar/2M \omega_k)^{1/2}(a_k + a_{k}^+), \tag{30} \]
\[ P_k = i(\hbar M \omega_k/2)^{1/2}(a_k^+ - a_{-k}). \tag{31} \]

By (4), (5), and (29) the particle position operator becomes
\[ q_s = \sum_k (\hbar/2N M \omega_k)^{1/2}[a_k \exp(iks) + a_{k}^+ \exp(-iks)]. \tag{32} \]

This equation relates the particle displacement operator to the phonon creation and annihilation operators.

To obtain (29) from (28), we use the properties
\[ Q_{-k}^+ = Q_k; \quad P_k^+ = P_{-k} \tag{33} \]
which follow from (5) and (7) by use of the quantum mechanical requirement that \(q_s\) and \(p_s\) be hermitian operators:
\[ q_s = q_s^+; \quad p_s = p_s^+. \tag{34} \]

Then (28) follows from the transformations (4), (5), and (7). We verify that the commutation relation (33) is satisfied by the operators defined by (28) and (25):
\[ [a_k, a_{k}^+] = (2\hbar)^{-1}(M \omega_k [Q_k, Q_{-k}] - i[Q_k, P_k] + i[P_{-k}, Q_{-k}]
\[ + [P_{-k}, P_k]/M \omega_k]. \tag{35} \]
By use of \([Q_k, P_{k'}] = i\hbar \delta(k, k')\) from (10) we have
\[ [a_k, a^\dagger_{k'}] = \delta(k, k') . \] (36)

It remains to show that the versions of (16) and (22) of the phonon hamiltonian are identical. We note that \(\omega_k = \omega_{-k}\) from (15), and we form
\[ \hbar \omega_k (a^\dagger_k a_k + a_{-k} a_{-k}) = \frac{1}{2M} (P_k P_{-k} + P_{-k} P_k) + \frac{1}{2} M \omega_k^2 (Q_k Q_{-k} + Q_{-k} Q_k) . \]

This exhibits the equivalence of the two expressions (14) and (22) for \(H\). We identify \(\omega_k = (2C/M)^{1/2}(1 - \cos ka)^{1/2}\) in (15) with the classical frequency of the oscillator mode of wavevector \(k\).

APPENDIX D: FERMI-DIRAC DISTRIBUTION FUNCTION

The Fermi-Dirac distribution function\(^1\) may be derived in several steps by use of a modern approach to statistical mechanics. We outline the argument here. The notation is such that conventional entropy \(S\) is related to the fundamental entropy \(\sigma\) by \(S = k_B \sigma\), and the kelvin temperature \(T\) is related to the fundamental temperature \(\tau\) by \(\tau = k_B T\), where \(k_B\) is the Boltzmann constant with the value \(1.38066 \times 10^{-23}\) J K\(^{-1}\).

The leading quantities are the entropy, the temperature, the Boltzmann factor, the chemical potential, the Gibbs factor, and the distribution functions. The entropy measures the number of quantum states accessible to a system. A closed system might be in any of these quantum states and (we assume) with equal probability. The fundamental assumption is that quantum states are either accessible or inaccessible to the system, and the system is equally likely to be in any one accessible state as in any other accessible state. Given \(g\) accessible states, the entropy is defined as \(\sigma = \log g\). The entropy thus defined will be a function of the energy \(U\), the number of particles \(N\), and the volume \(V\) of the system.

When two systems, each of specified energy, are brought into thermal contact they may transfer energy; their total energy remains constant, but the constraints on their individual energies are lifted. A transfer of energy in one direction, or perhaps in the other, may increase the product \(g_1 g_2\) that measures the number of accessible states of the combined systems. What we call the fundamental assumption biases the outcome in favor of that allocation of the total energy that maximizes the number of accessible states: more is better, and more likely. This statement is the kernel of the law of increase of entropy.

\(^1\)This appendix follows closely the introduction to C. Kittel and H. Kroemer, *Thermal Physics*, 2nd ed., Freeman, 1990.
which is the general expression of the second law of thermodynamics.

We have brought two systems into thermal contact so that they may transfer energy. What is the most probable outcome of the encounter? One system will gain energy at the expense of the other, and meanwhile the total entropy of the two systems will increase. Eventually the entropy will reach a maximum for the given total energy. It is not difficult to show that the maximum is attained when the value of \((\partial \sigma/\partial U)_{N,V}\) for one system is equal to the value of the same quantity for the second system. This equality property for two systems in thermal contact is the property we expect of the temperature. Accordingly, the fundamental temperature \(\tau\) is defined by the relation

\[
\frac{1}{\tau} = \left(\frac{\partial \sigma}{\partial U}\right)_{N,V}.
\]  

(1)

The use of \(1/\tau\) assures that energy will flow from high \(\tau\) to low \(\tau\); no more complicated relation is needed.

Now consider a very simple example of the Boltzmann factor. Let a small system with only two states, one at energy 0 and one at energy \(\epsilon\), be placed in thermal contact with a large system that we call the reservoir. The total energy of the combined systems is \(U_0\); when the small system is in the state of energy 0, the reservoir has energy \(U_0\) and will have \(g(U_0)\) states accessible to it. When the small system is in the state of energy \(\epsilon\), the reservoir will have energy \(U_0 - \epsilon\) and will have \(g(U_0 - \epsilon)\) states accessible to it. By the fundamental assumption, the ratio of the probability of finding the small system with energy \(\epsilon\) to the probability of finding it with energy 0 is

\[
\frac{P(\epsilon)}{P(0)} = \frac{g(U_0 - \epsilon)}{g(U_0)} = \frac{\exp[\sigma(U_0 - \epsilon)]}{\exp[\sigma(U_0)]}.
\]  

(2)

The reservoir entropy \(\sigma\) may be expanded in a Taylor series:

\[
\sigma(U_0 - \epsilon) = \sigma(U_0) - \epsilon(\partial \sigma/\partial U)_{0} = \sigma(U_0) - \epsilon/\tau,
\]  

(3)

by the definition (1) of the temperature. Higher order terms in the expansion may be dropped. Cancellation of the term \(\exp[\sigma(U_0)]\), which occurs in the numerator and denominator of (2) after the substitution of (3), leaves us with

\[
P(\epsilon)/P(0) = \exp(-\epsilon/\tau).
\]  

(4)

This is Boltzmann's result. To show its use, we calculate the thermal average energy \(\langle \epsilon \rangle\) of the two-state system in thermal contact with a reservoir at temperature \(\tau\):

\[
\langle \epsilon \rangle = \sum_i \epsilon_i P(\epsilon_i) = 0 \cdot P(0) + \epsilon P(\epsilon) = \frac{\epsilon \exp(-\epsilon/\tau)}{1 + \exp(-\epsilon/\tau)},
\]  

(5)
where we have imposed the normalization condition on the sum of the probabilities:

\[ P(0) + P(\varepsilon) = 1 \]  

The argument can be generalized immediately to find the average energy of a harmonic oscillator at temperature \( \tau \), as in the Planck law.

The most important extension of the theory is to systems that can transfer particles as well as energy with the reservoir. For two systems in diffusive and thermal contact, the entropy will be a maximum with respect to the transfer of particles as well as to the transfer of energy. Not only must \((\partial \sigma / \partial U)_{U,V}\) be equal for the two systems, but \((\partial \sigma / \partial N)_{U,V}\) must also be equal, where \( N \) refers to the number of particles of a given species. The new equality condition is the occasion for the introduction\(^2\) of the chemical potential \( \mu \):

\[ -\frac{\mu}{\tau} = \left( \frac{\partial \sigma}{\partial N} \right)_{U,V} \]  

For two systems in thermal and diffusive contact, \( \tau_1 = \tau_2 \) and \( \mu_1 = \mu_2 \). The sign in (7) is chosen to ensure that the direction of particle flow is from high chemical potential to low chemical potential as equilibrium is approached.

The Gibbs factor is an extension of the Boltzmann factor (4) and allows us to treat systems that can transfer particles. The simplest example is a system with two states, one with 0 particles and 0 energy, and one with 1 particle and energy \( \varepsilon \). The system is in contact with a reservoir at temperature \( \tau \) and chemical potential \( \mu \). We extend (3) for the reservoir entropy:

\[ \sigma(U_0 - \varepsilon; N_0 - 1) = \sigma(U_0; N_0) - \varepsilon(\partial \sigma / \partial U) - 1 \cdot (\partial \sigma / \partial N_0) \]

\[ = \sigma(U_0; N_0) - \varepsilon/\tau + \mu/\tau . \]

By analogy with (4), we have the Gibbs factor

\[ P(1,\varepsilon)/P(0,0) = \exp[(\mu - \varepsilon)/\tau] , \]

for the ratio of the probability that the system is occupied by 1 particle at energy \( \varepsilon \) to the probability that the system is unoccupied, with energy 0. The result (9) after normalization is readily expressed as

\[ P(1,\varepsilon) = \frac{1}{\exp[(\varepsilon - \mu)/\tau] + 1} . \]

This is the Fermi-Dirac distribution function.

\(^2\)TP Chapter 5 has a careful treatment of the chemical potential.
APPENDIX E: DERIVATION OF THE $dk/dt$ EQUATION

The simple and rigorous derivation that follows is due to Kroemer. In quantum mechanics, for any operator $A$ we have

$$d\langle A \rangle/dt = (i\hbar)\langle [H, A] \rangle,$$  \hspace{1cm} (1)

where $H$ is the hamiltonian. See also C. L. Cook, American J. Phys. 55, 953 (1987).

We let $A$ be the lattice translation operator $T$ defined by

$$Tf(x) = f(x + a),$$  \hspace{1cm} (2)

where $a$ is a basis vector, here in one dimension. For a Bloch function

$$T\psi_k(x) = \exp(ika)\psi_k(x).$$  \hspace{1cm} (3)

This result is usually written for one band, but it holds even if $\psi_k$ is a linear combination of Bloch states from any number of bands, but having the identical wavevector $k$ in the reduced zone scheme.

The crystal hamiltonian $H_0$ commutes with the lattice translation operator $T$, so that $[H_0, T] = 0$. If we add a uniform external force $F$, then

$$H = H_0 - Fx,$$  \hspace{1cm} (4)

and

$$[H, T] = FaT.$$  \hspace{1cm} (5)

From (1) and (5),

$$d\langle T \rangle/dt = (i\hbar)(Fa)\langle T \rangle.$$  \hspace{1cm} (6)

From (6) we form

$$\langle T \rangle * d\langle T \rangle/dt = (iFa/\hbar)\langle T \rangle^2;$$

$$\langle T \rangle d\langle T^* \rangle/dt = -(iFa/\hbar)\langle T \rangle^2.$$

On addition,

$$d\langle T \rangle^2/dt = 0.$$  \hspace{1cm} (7)

This is the equation of a circle in the complex plane. The coordinate axes in the plane are the real and imaginary parts of the eigenvalue $\exp(ika)$. If $\langle T \rangle$ is initially on the unit circle, it will remain on the unit circle.

For $\psi$’s that satisfy periodic boundary conditions, $\langle T \rangle$ can lie on the unit circle only if $\psi_k$ is a single Bloch function or a superposition of Bloch functions from different bands, but with the same reduced $k$.

As $\langle T \rangle$ moves around the unit circle, the wavevector $k$ changes exactly at the same rate for the components of $\psi_k$ in all bands. With $\langle T \rangle = \exp(ika)$, we
have from (6) that
\[ i a \frac{dk}{dt} = i F a / \hbar \]  

or
\[ \frac{dk}{dt} = F / \hbar \]  

an exact result.

This does not mean that interband mixing (such as Zener tunneling) does not occur under the influence of applied electric fields. It just means that \( k \) evolves at a constant rate for every component of a wave packet. The result is easily extended to three dimensions.

**APPENDIX F: BOLTZMANN TRANSPORT EQUATION**

The classical theory of transport processes is based on the Boltzmann transport equation. We work in the six-dimensional space of Cartesian coordinates \( r \) and velocity \( v \). The classical distribution function \( f(r,v) \) is defined by the relation

\[ f(r,v)drdv = \text{number of particles in } drdv \]  

(1)

The Boltzmann equation is derived by the following argument. We consider the effect of a time displacement \( dt \) on the distribution function. The Liouville theorem of classical mechanics tells us that if we follow a volume element along a flowline the distribution is conserved:

\[ f(t + dt, r + dr, v + dv) = f(t, r, v) \]  

(2)

in the absence of collisions. With collisions

\[ f(t + dt, r + dr, v + dv) - f(t, r, v) = dt(\partial f / \partial t)_{\text{collisions}} \]  

(3)

Thus

\[ dt(\partial f / \partial t) + dr \cdot \text{grad}_r f + dv \cdot \text{grad}_v f = dt(\partial f / \partial t)_{\text{coll}} \]  

(4)

Let \( \alpha \) denote the acceleration \( dv/dt \); then

\[ \partial f / \partial t + v \cdot \text{grad}_r f + \alpha \cdot \text{grad}_v f = (\partial f / \partial t)_{\text{coll}} \]  

(5)

This is the Boltzmann transport equation.

In many problems the collision term \( (\partial f / \partial t)_{\text{coll}} \) may be treated by the introduction of a relaxation time \( \tau_c(r,v) \), defined by the equation

\[ (\partial f / \partial t)_{\text{coll}} = -(f - f_0) / \tau_c \]  

(6)
Here \( f_0 \) is the distribution function in thermal equilibrium. Do not confuse \( \tau_c \) for relaxation time with \( \tau \) for temperature. Suppose that a nonequilibrium distribution of velocities is set up by external forces which are suddenly removed. The decay of the distribution towards equilibrium is then obtained from (6) as

\[
\frac{\partial (f - f_0)}{\partial t} = -\frac{f - f_0}{\tau_c},
\]

if we note that \( \partial f_0 / \partial t = 0 \) by definition of the equilibrium distribution. This equation has the solution

\[
(f - f_0)_{t} = (f - f_0)_{t=0} \exp(-t/\tau_c).
\]

It is not excluded that \( \tau_c \) may be a function of \( \mathbf{r} \) and \( \mathbf{v} \).

We combine (1), (5), and (6) to obtain the Boltzmann transport equation in the relaxation time approximation:

\[
\frac{\partial f}{\partial t} + \alpha \cdot \nabla_v f + \mathbf{v} \cdot \nabla_r f = -\frac{f - f_0}{\tau_c}.
\]

In the steady state \( \partial f / \partial t = 0 \) by definition.

**Particle Diffusion**

Consider an isothermal system with a gradient of the particle concentration. The steady-state Boltzmann transport equation in the relaxation time approximation becomes

\[
v_x df / dx = -(f - f_0)/\tau_c,
\]

where the nonequilibrium distribution function \( f \) varies along the \( x \) direction. We may write (10) to first order as

\[
f_1 = f_0 - v_x \tau_c df_0 / dx,
\]

where we have replaced \( \partial f / \partial x \) by \( df_0 / dx \). We can iterate to obtain higher order solutions when desired. Thus the second order solution is

\[
f_2 = f_0 - v_x \tau_c df_1 / dx = f_0 - v_x \tau_c df_0 / dx + v_x^2 \tau_c^2 d^2 f_0 / dx^2.
\]

The iteration may be used in the treatment of nonlinear effects.

**Classical Distribution**

Let \( f_0 \) be the distribution function in the classical limit:

\[
f_0 = \exp((\mu - \epsilon)/\tau).
\]

We are at liberty to take whatever normalization for the distribution function is
most convenient because the transport equation is linear in \( f \) and \( f_0 \). We can take the normalization as in (13) rather than as in (1). Then
\[
df_0/dx = (df_0/d\mu)(d\mu/dx) = (f_0/\tau)(d\mu/dx) ,
\]
and the first order solution (11) for the nonequilibrium distribution becomes
\[
f = f_0 - (v_x \tau_c f_0/\tau)(d\mu/dx) .
\] (15)
The particle flux density in the \( x \) direction is
\[
J_n^x = \int v_x f D(\varepsilon) d\varepsilon ,
\] (16)
where \( D(\varepsilon) \) is the density of electron states per unit volume per unit energy range:
\[
D(\varepsilon) = \frac{1}{2\pi^2} \left( \frac{2M}{\hbar^2} \right)^{3/2} \varepsilon^{1/2} ,
\] (17)
as in (6.20). Thus
\[
J_n^x = \int v_x f_0 D(\varepsilon) d\varepsilon - (d\mu/dx) \int (v_x^2 \tau_c f_0/\tau) D(\varepsilon) d\varepsilon .
\] (18)
The first integral vanishes because \( v_x \) is an odd function and \( f_0 \) is an even function of \( v_x \). This confirms that the net particle flux vanishes for the equilibrium distribution \( f_0 \). The second integral will not vanish.

Before evaluating the second integral, we have an opportunity to make use of what we may know about the velocity dependence of the relaxation time \( \tau_c \). Only for the sake of example we assume that \( \tau_c \) is constant, independent of velocity; \( \tau_c \) may then be taken out of the integral:
\[
J_n^x = -(d\mu/dx)(\tau_c/\tau) \int v_x^2 f_0 D(\varepsilon) d\varepsilon .
\] (19)
The integral may be written as
\[
\frac{1}{3} \int v_x^2 f_0 D(\varepsilon) d\varepsilon = \frac{2}{3M} \int (\frac{1}{2} M v_x^2) f_0 D(\varepsilon) d\varepsilon = n\tau/M ,
\] (20)
because the integral is just the kinetic energy density \( \frac{3}{2} n\tau \) of the particles. Here \( \int f_0 D(\varepsilon) d\varepsilon = n \) is the concentration. The particle flux density is
\[
J_n^x = -(n\tau_c/M)(d\mu/dx) = -(\tau_c/\tau)(dn/dx) ,
\] (21)
because
\[
\mu = \tau \log n + \text{constant} .
\] (22)
The result (21) is of the form of the diffusion equation with the diffusivity
\[
D_n = \tau_c/\tau = \frac{1}{3} (v_x^2) \tau_c .
\] (23)

Another possible assumption about the relaxation time is that it is inversely proportional to the velocity, as in \( \tau_c = l/v \), where the mean free path \( l \) is constant. Instead of (19) we have
\[
J_n^x = -(d\mu/dx)(l/\tau) \int (v_x^2/v) f_0 D(\varepsilon) d\varepsilon ,
\] (24)
and now the integral may be written as
\[ \frac{1}{2} \int c \Delta D(\epsilon) d\epsilon = \frac{1}{3} n \bar{c}, \] (25)
where \( \bar{c} \) is the average speed. Thus
\[ J_n^x = -\frac{1}{2}(\bar{c} n/\tau)(d\mu/dx) = -\frac{1}{3} \bar{c} (dn/dx), \] (26)
and the diffusivity is
\[ D_n = \frac{1}{3} l \bar{c}. \] (27)

**Fermi-Dirac Distribution**

The distribution function is
\[ f_0 = \frac{1}{\exp[(\epsilon - \mu)/\tau] + 1}. \] (28)

To form \( df_0/dx \) as in (14) we need the derivative \( df_0/d\mu \). We argue below that
\[ df_0/d\mu = \delta(\epsilon - \mu), \] (29)

at low temperatures \( \tau \ll \mu \). Here \( \delta \) is the Dirac delta function, which has the property for a general function \( F(\epsilon) \) that
\[ \int_{-\infty}^{\infty} F(\epsilon) \delta(\epsilon - \mu) d\epsilon = F(\mu). \] (30)

Now consider the integral \( \int_{-\infty}^{\infty} F(\epsilon)(df_0/d\mu) d\epsilon \). At low temperatures \( df_0/d\mu \) is very large for \( \epsilon = \mu \) and is small elsewhere. Unless the function \( F(\epsilon) \) is very rapidly varying near \( \mu \) we may take \( F(\epsilon) \) outside the integral, with the value \( F(\mu) \):
\[ \int_{-\infty}^{\infty} F(\epsilon)(df_0/d\mu) d\epsilon = F(\mu) \int_{-\infty}^{\infty} (df_0/d\mu) d\epsilon = -F(\mu) \int_{-\infty}^{\infty} (df_0/d\epsilon) d\epsilon \] (31)
\[ = -F(\mu)[f_0(\epsilon)]^\infty_0 = F(\mu)f_0(0), \]
where we have used \( df_0/d\mu = -df_0/d\epsilon \). We have also used \( f_0 = 0 \) for \( \epsilon = \infty \). At low temperatures \( f(0) = 1 \); thus the right-hand side of (31) is just \( F(\mu) \), consistent with the delta function approximation. Thus
\[ df_0/dx = \delta(\epsilon - \mu) d\mu/dx. \] (32)

The particle flux density is, from (16),
\[ J_n^x = -(d\mu/dx)\tau_c \int v_x^2 \delta(\epsilon - \mu) D(\epsilon) d\epsilon, \] (33)
where \( \tau_c \) is the relaxation time at the surface \( \epsilon = \mu \) of the Fermi sphere. The integral has the value
\[ \frac{1}{3} v_F^2 (3n/2\epsilon_F) = n/m, \] (34)
by use of $D(\mu) = 3n/2\epsilon_F$ at absolute zero, where $\epsilon_F = \frac{1}{2}mv_F^2$ defines the velocity $v_F$ on the Fermi surface. Thus

$$J_\mu^x = -(n\tau_c/m)d\mu/dx .$$  \hspace{1cm} (35)

At absolute zero $\mu(0) = (\hbar^2/2m)(3\pi^2n)^{2/3}$, whence

$$d\mu/dx = \left\{\frac{3}{2}(\hbar^2/2m)(3\pi^2)^{2/3}n^{1/3}\right\}dn/dx = \frac{3}{2}(\epsilon_F/n)dn/dx ,$$  \hspace{1cm} (36)

so that (33) becomes

$$J_\mu^x = -(2\tau_c/3m)\epsilon_F \frac{dn}{dx} = -\frac{1}{3}v_F^2\tau_c \frac{dn}{dx} .$$  \hspace{1cm} (37)

The diffusivity is the coefficient of $dn/dx$:

$$D_n = \frac{1}{3}v_F^2\tau_c ,$$  \hspace{1cm} (38)

closely similar in form to the result (23) for the classical distribution of velocities. In (38) the relaxation time is to be taken at the Fermi energy.

We see we can solve transport problems where the Fermi-Dirac distribution applies, as in metals, as easily as where the classical approximation applies.

Electrical Conductivity

The isothermal electrical conductivity $\sigma$ follows from the result for the particle diffusivity when we multiply the particle flux density by the particle charge $q$ and replace the gradient $d\mu/dx$ of the chemical potential by the gradient $qd\phi/dx = -qE_x$ of the external potential, where $E_x$ is the $x$ component of the electric field intensity. The electric current density follows from (21):

$$J_q = (nq^2\tau_c/m)E ; \quad \sigma = nq^2\tau_c/m ,$$  \hspace{1cm} (39)

for a classical gas with relaxation time $\tau_c$. For the Fermi-Dirac distribution, from (35),

$$J_q = (nq^2\tau_c/m)E ; \quad \sigma = nq^2\tau_c/m .$$  \hspace{1cm} (40)

APPENDIX G: VECTOR POTENTIAL, FIELD MOMENTUM, AND GAUGE TRANSFORMATIONS

This section is included because it is hard to find the magnetic vector potential $A$ discussed thoroughly in one place, and we need the vector potential
in superconductivity. It may seem mysterious that the hamiltonian of a particle in a magnetic field has the form derived in (18) below:

\[ H = \frac{1}{2M} \left( p - \frac{Q}{c} A \right)^2 + Q\varphi, \]  

(1)

where \( Q \) is the charge; \( M \) is the mass; \( A \) is the vector potential, and \( \varphi \) is the electrostatic potential. This expression is valid in classical mechanics and in quantum mechanics. Because the kinetic energy of a particle is not changed by a static magnetic field, it is perhaps unexpected that the vector potential of the magnetic field enters the hamiltonian. As we shall see, the key is the observation that the momentum \( p \) is the sum of two parts, the kinetic momentum

\[ p_{\text{kin}} = Mv, \]

(2)

which is familiar to us, and the potential momentum or field momentum

\[ p_{\text{field}} = \frac{Q}{c} A. \]

(3)

The total momentum is

\[ p = p_{\text{kin}} + p_{\text{field}} = Mv + \frac{Q}{c} A, \]

(4)

and the kinetic energy is

\[ \frac{1}{2} Mv^2 = \frac{1}{2M} (Mv)^2 = \frac{1}{2M} \left( p - \frac{Q}{c} A \right)^2. \]

(5)

The vector potential\(^1\) is related to the magnetic field by

\[ B = \text{curl } A. \]

(6)

We assume that we work in nonmagnetic material so that \( H \) and \( B \) are treated as identical.

**Lagrangian Equations of Motion**

To find the hamiltonian, the prescription of classical mechanics is clear: we must first find the Lagrangian. The Lagrangian in generalized coordinates is

\[ L = \frac{1}{2} M\dot{q}^2 - Q\varphi(q) + \frac{Q}{c} \dot{q} \cdot A(q). \]

(7)

This is correct because it leads to the correct equation of motion of a charge in combined electric and magnetic fields, as we now show.

In Cartesian coordinates the Lagrange equation of motion is
\[
\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0 ,
\] (8)
and similarly for \( y \) and \( z \). From (7) we form
\[
\frac{\partial L}{\partial x} = -Q \frac{\partial \varphi}{\partial x} + \frac{Q}{c} \left( \dot{x} \frac{\partial A_x}{\partial x} + \dot{y} \frac{\partial A_y}{\partial x} + \dot{z} \frac{\partial A_z}{\partial x} \right) ;
\] (9)
\[
\frac{\partial L}{\partial \dot{x}} = M\ddot{x} + \frac{Q}{c} A_x ;
\] (10)
\[
\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} = M\ddot{x} + \frac{Q}{c} \frac{dA_x}{dt} = M\ddot{x} + \frac{Q}{c} \left( \frac{\partial A_x}{\partial t} + \dot{x} \frac{\partial A_x}{\partial x} + \dot{y} \frac{\partial A_x}{\partial y} + \dot{z} \frac{\partial A_x}{\partial z} \right) .
\] (11)

Thus (8) becomes
\[
M\ddot{x} + Q \frac{\partial \varphi}{\partial x} + \frac{Q}{c} \left[ \frac{\partial A_x}{\partial t} + \dot{y} \left( \frac{\partial A_x}{\partial y} - \frac{\partial A_y}{\partial x} \right) + \dot{z} \left( \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) \right] = 0 ,
\] (12)
or
\[
M \frac{d^2 x}{dt^2} = QE_x + \frac{Q}{c} [v \times B]_x ,
\] (13)
with
\[
E_x = -\frac{\partial \varphi}{\partial x} - \frac{1}{c} \frac{\partial A_x}{\partial t} ;
\] (14)
\[
B = \text{curl } A .
\] (15)

Equation (13) is the Lorentz force equation. This confirms that (7) is correct. We note in (14) that \( E \) has one contribution from the electrostatic potential \( \varphi \) and another from the time derivative of the magnetic vector potential \( A \).

**Derivation of the Hamiltonian**

The momentum \( p \) is defined in terms of the Lagrangian as
\[
p = \frac{\partial L}{\partial \dot{q}} = M\dot{q} + \frac{Q}{c} A ,
\] (16)
in agreement with (4). The hamiltonian \( H(p,q) \) is defined by
\[
H(p,q) = p \cdot \dot{q} - L ,
\] (17)
or
\[ H = M\dot{q}^2 + \frac{Q}{c} \dot{q} \cdot A - \frac{1}{2} M\dot{q}^2 + Q\varphi - \frac{Q}{c} \dot{q} \cdot A = \frac{1}{2M} \left( p - \frac{Q}{c} A \right)^2 + Q\varphi, \]  
(18)
as in (1).

**Field Momentum**

The momentum in the electromagnetic field that accompanies a particle moving in a magnetic field is given by the volume integral of the Poynting vector, so that
\[ p_{\text{field}} = \frac{1}{4\pi c} \int dV \, \mathbf{E} \times \mathbf{B}. \]  
(19)
We work in the nonrelativistic approximation with \( v \ll c \), where \( v \) is the velocity of the particle. At low values of \( v/c \) we consider \( \mathbf{B} \) to arise from an external source alone, but \( \mathbf{E} \) arises from the charge on the particle. For a charge \( Q \) at \( r' \),
\[ \mathbf{E} = -\nabla\varphi; \quad \nabla^2 \varphi = -4\pi Q \, \delta(r - r'). \]  
(20)
Thus
\[ p_f = -\frac{1}{4\pi c} \int dV \, \nabla \varphi \times \text{curl} \, \mathbf{A}. \]  
(21)
By a standard vector relation we have
\[ \int dV \, \nabla \varphi \times \text{curl} \, \mathbf{A} = -\int dV \left[ \mathbf{A} \times \text{curl} \, (\nabla \varphi) - \text{div} \, \nabla \varphi - (\nabla \varphi) \, \text{div} \, \mathbf{A} \right]. \]  
(22)
But \( \text{curl} \, (\nabla \varphi) = 0 \), and we can always choose the gauge such that \( \text{div} \, \mathbf{A} = 0 \). This is the transverse gauge.
Thus, we have
\[ p_f = -\frac{1}{4\pi c} \int dV \, \mathbf{A} \, \nabla^2 \varphi = \frac{1}{c} \int dV \, A Q \, \delta(r - r') = \frac{Q}{c} A. \]  
(23)
This is the interpretation of the field contribution to the total momentum \( p = Mv + QA/c \).

**GAUGE TRANSFORMATION**

Suppose \( H\psi = \epsilon\psi \), where
\[ H = \frac{1}{2M} \left( p - \frac{Q}{c} A \right)^2. \]  
(24)
Let us make a gauge transformation to \( A' \), where

\[
A' = A + \nabla \chi ,
\]

where \( \chi \) is a scalar. Now \( B = \operatorname{curl} A = \operatorname{curl} A' \), because \( \operatorname{curl} (\nabla \chi) = 0 \). The Schrödinger equation becomes

\[
\frac{1}{2M} \left( p - \frac{Q}{c} A' + \frac{Q}{c} \nabla \chi \right)^2 \psi = \epsilon \psi .
\]

What \( \psi' \) satisfies

\[
\frac{1}{2M} \left( p - \frac{Q}{c} A' \right)^2 \psi' = \epsilon \psi' ,
\]

with the same \( \epsilon \) as for \( \psi \). Equation (27) is equivalent to

\[
\frac{1}{2M} \left( p - \frac{Q}{c} A - \frac{Q}{c} \nabla \chi \right)^2 \psi' = \epsilon \psi' .
\]

We try

\[
\psi' = \exp(iQ\chi/\hbar c)\psi .
\]

Now

\[
p\psi' = \exp(iQ\chi/\hbar c)p\psi + \frac{Q}{c} (\nabla \chi)\exp(iQ\chi/\hbar c)\psi ,
\]

so that

\[
\left(p - \frac{Q}{c} \nabla \chi \right)\psi' = \exp(iQ\chi/\hbar c)p\psi
\]

and

\[
\frac{1}{2M} \left( p - \frac{Q}{c} A - \frac{Q}{c} \nabla \chi \right)^2 \psi' = \exp(iQ\chi/\hbar c)\frac{1}{2M} \left(p - \frac{Q}{c} A \right)^2 \psi
= \exp(iQ\chi/\hbar c)\epsilon \psi .
\]

Thus \( \psi' = \exp(iQ\chi/\hbar c)\psi \) satisfies the Schrödinger equation after the gauge transformation (25). The energy \( \epsilon \) is invariant under the transformation. The gauge transformation on \( A \) merely changes the local phase of the wavefunction.

We see that

\[
\psi'\psi' = \psi'\psi ,
\]

so that the charge density is invariant under a gauge transformation.

**Gauge in the London Equation**

Because of the equation of continuity in the flow of electric charge we require that in a superconductor
\[ \text{div } j = 0 , \]
so that the vector potential in the London equation \( j = -cA/4\pi\lambda_L^2 \) must satisfy
\[ \text{div } A = 0 . \] (32)

Further, there is no current flow through a vacuum/superconductor interface. The normal component of the current across the interface must vanish: \( j_n = 0 \), so that the vector potential in the London equation must satisfy
\[ A_n = 0 . \] (33)

The gauge of the vector potential in the London equation of superconductivity is to be chosen so that (32) and (33) are satisfied.

**APPENDIX II: COOPER PAIRS**

For a complete set of states of a two-electron system that satisfy periodic boundary conditions in a cube of unit volume, we take plane wave product functions
\[ \varphi(k_1,k_2;r_1,r_2) = \exp[i(k_1 \cdot r_1 + k_2 \cdot r_2)] . \] (1)

We assume that the electrons are of opposite spin.

We introduce center-of-mass and relative coordinates:
\[ R = \frac{1}{2}(r_1 + r_2) ; \quad r = r_1 - r_2 ; \] (2)
\[ K = k_1 + k_2 ; \quad k = \frac{1}{2}(k_1 - k_2) , \] (3)

so that
\[ k_1 \cdot r_1 + k_2 \cdot r_2 = K \cdot R + k \cdot r . \] (4)

Thus (1) becomes
\[ \varphi(K,k;R,r) = \exp(iK \cdot R) \exp(ik \cdot r) , \] (5)

and the kinetic energy of the two-electron system is
\[ E_K + E_k = (\hbar^2/m)(\frac{1}{4}K^2 + k^2) . \] (6)

We give special attention to the product functions for which the center-of-mass wavevector \( K = 0 \), so that \( k_1 = -k_2 \). With an interaction \( H_1 \) between the two electrons, we set up the eigenvalue problem in terms of the expansion
\[ \varphi(r) = \Sigma g_k \exp(ik \cdot r) . \] (7)
The Schrödinger equation is

\[(H_0 + H_1 - \epsilon)\chi(r) = 0 = \sum_k [(E_k - \epsilon)g_k + H_{1k}g_k] \exp(ik' \cdot r), \tag{8}\]

where \(H_1\) is the interaction energy of the two electrons. Here \(\epsilon\) is the eigenvalue.

We take the scalar product with \(\exp(ik' \cdot r)\) to obtain

\[(E_k - \epsilon)g_k + \sum_k g_k (k|H_1|k') = 0, \tag{9}\]

the secular equation of the problem.

Now transform the sum to an integral:

\[(E - \epsilon)g(E) + \int dE' g(E')H_1(E,E')N(E') = 0, \tag{10}\]

where \(N(E')\) is the number of two electron states with total momentum \(K = 0\) and with kinetic energy in \(dE'\) at \(E'\).

Now consider the matrix elements \(H_1(E,E') = (k|H_1|k')\). Studies of these by Bardeen suggest that they are important when the two electrons are confined to a thin energy shell near the Fermi surface—within a shell of thickness \(h\omega_D\) above \(E_F\), where \(\omega_D\) is the Debye phonon cutoff frequency. We assume that

\[H_1(E,E') = -V \tag{11}\]

for \(E,E'\) within the shell and zero otherwise. Here \(V\) is assumed to be positive.

Thus (10) becomes

\[(E - \epsilon)g(E) = V \int_{2\epsilon_f}^{2\epsilon_m} dE' g(E')N(E') = C, \tag{12}\]

with \(\epsilon_m = \epsilon_F + h\omega_D\). Here \(C\) is a constant, independent of \(E\).

From (12) we have

\[g(E) = \frac{C}{E - \epsilon} \tag{13}\]

and

\[1 = V \int_{2\epsilon_f}^{2\epsilon_m} dE' \frac{N(E')}{E' - \epsilon}. \tag{14}\]

With \(N(E')\) approximately constant and equal to \(N_F\) over the small energy range between \(2\epsilon_m\) and \(2\epsilon_F\), we take it out of the integral to obtain

\[1 = N_F V \int_{2\epsilon_f}^{2\epsilon_m} dE' \frac{1}{E' - \epsilon} = N_F V \log \frac{2\epsilon_m - \epsilon}{2\epsilon_F - \epsilon}. \tag{15}\]

Let the eigenvalue \(\epsilon\) of (15) be written as

\[\epsilon = 2\epsilon_F - \Delta, \tag{16}\]

which defines the binding energy \(\Delta\) of the electron pair, relative to two free
electrons at the Fermi surface. Then (15) becomes

\[ 1 = N_F V \log \frac{2\epsilon_m - 2\epsilon_F + \Delta}{\Delta} = N_F V \log \frac{2\hbar \omega_D + \Delta}{\Delta}, \]  
(17)

or

\[ \frac{1}{N_F V} = \log (1 + 2\hbar \omega_D/\Delta). \]  
(18)

This result for the binding energy of a Cooper pair may be written as

\[ \Delta = \frac{2\hbar \omega_D}{\exp(1/N_F V) - 1}. \]  
(19)

For \( V \) positive (attractive interaction) the energy of the system is lowered by excitation of a pair of electrons above the Fermi level. Therefore the Fermi sea is unstable in an important way. The binding energy (19) is closely related to the superconducting energy gap \( E_g \). The BCS calculations show that a high density of Cooper pairs may form in a metal.

**APPENDIX I: GINZBURG-LANDAU EQUATION**

We owe to Ginzburg and Landau an elegant theory of the phenomenology of the superconducting state and of the spatial variation of the order parameter in that state. An extension of the theory by Abrikosov describes the structure of the vortex state which is exploited technologically in superconducting magnets. The attractions of the GL theory are the natural introduction of the coherence length and of the wavefunction used in the theory of the Josephson effects in Chapter 12.

We introduce the **order parameter** \( \psi(r) \) with the property that

\[ \psi^*(r)\psi(r) = n_S(r), \]  
(1)

the local concentration of superconducting electrons. The mathematical formulation of the definition of the function \( \psi(r) \) will come out of the BCS theory. We first set up a form for the free energy density \( F_S(r) \) in a superconductor as a function of the order parameter. We assume that in the general vicinity of the transition temperature

\[ F_S(r) = F_N - \alpha |\psi|^2 + \frac{1}{2} \beta |\psi|^4 + (1/2m)|(-i\hbar \nabla - qA/c)\psi|^2 - \int B \cdot dB, \]  
(2)

with the phenomenological positive constants \( \alpha, \beta, \) and \( m, \) of which more will be said. Here:

1. \( F_N \) is the free energy density of the normal state.
2. $-\alpha|\psi|^2 + \frac{1}{2}\beta|\psi|^4$ is a typical Landau form (as in Chapter 13) for the expansion of the free energy in terms of an order parameter that vanishes at a second-order phase transition. This term may be viewed as $-\alpha n_s + \frac{1}{2}\beta n_s^2$ and by itself is a minimum with respect to $n_s$ when $n_s(T) = \alpha/\beta$.

3. The term $|\text{grad} \psi|^2$ represents an increase in energy caused by a spatial variation of the order parameter. It has the form of the kinetic energy in quantum mechanics.\(^1\) The kinetic momentum $-i\hbar \nabla$ is accompanied by the field momentum $-qA/c$ to ensure the gauge invariance of the free energy, as in Appendix C. Here $q = -2e$ for an electron pair.

4. The term $-\int M \cdot dB_\circ$, with the fictitious magnetization $M = (B - B_\circ)/4\pi$, represents the increase in the superconducting free energy caused by the expulsion of magnetic flux from the superconductor.

The separate terms in (2) will be illustrated by examples as we progress further. First let us derive the GL equation (6). We minimize the total free energy $\int dV F_s(r)$ with respect to variations in the function $\psi(r)$. We have

$$\delta F_s(r) = [-\alpha \psi + \beta|\psi|^2 \psi + (1/2m)(-i\hbar \nabla - qA/c)\psi \cdot (i\hbar \nabla - qA/c)]\delta \psi^* + \text{c.c.}$$

(3)

We integrate by parts to obtain

$$\int dV (\nabla \psi)(\nabla \delta \psi^*) = -\int dV (\nabla^2 \psi)\delta \psi^* ,$$

(4)

if $\delta \psi^*$ vanishes on the boundaries. It follows that

$$\delta \int dV F_s = \int dV \delta \psi^*[-\alpha \psi + \beta|\psi|^2 \psi + (1/2m)(-i\hbar \nabla - qA/c)^2 \psi] + \text{c.c.}$$

(5)

This integral is zero if the term in brackets is zero:

$$[(1/2m)(-i\hbar \nabla - qA/c)^2 - \alpha + \beta|\psi|^2]\psi = 0 .$$

(6)

This is the Ginzburg-Landau equation; it resembles a Schrödinger equation for $\psi$.

By minimizing (2) with respect to $\delta A$ we obtain a gauge-invariant expression for the supercurrent flux:

$$j_s(r) = -(iq\hbar/2m)(\psi^* \nabla \psi - \psi \nabla \psi^*) - (q^2/mc)\psi^* \psi A .$$

(7)

At a free surface of the specimen we must choose the gauge to satisfy the boundary condition that no current flows out of the superconductor into the vacuum: $\hat{n} \cdot j_s = 0$, where $\hat{n}$ is the surface normal.

\(^1\)A contribution of the form $|\nabla M|^2$, where $M$ is the magnetization, was introduced by Landau and Lifshitz to represent the exchange energy density in a ferromagnet; see QTS, p. 65.
**Coherence Length.** The intrinsic coherence length $\xi$ may be defined from (6). Let $\mathbf{A} = 0$ and suppose that $\beta|\psi|^2$ may be neglected in comparison with $\alpha$. In one dimension the GL equation (6) reduces to

$$-\frac{\hbar^2}{2m} \frac{d^2 \psi}{dx^2} = \alpha \psi \, .$$

(8)

This has a wavelike solution of the form $\exp(ix/\xi)$, where $\xi$ is defined by

$$\xi = (\hbar^2/2m\alpha)^{1/2} \, .$$

(9)

A more interesting special solution is obtained if we retain the nonlinear term $\beta|\psi|^2$ in (6). Let us look for a solution with $\psi = 0$ at $x = 0$ and with $\psi \to \psi_0$ as $x \to \infty$. This situation represents a boundary between normal and superconducting states. Such states can coexist if there is a magnetic field $H_c$ in the normal region. For the moment we neglect the penetration of the field into the superconducting region: we take the field penetration depth $\lambda \ll \xi$, which defines an extreme type I superconductor.

The solution of

$$-\frac{\hbar^2}{2m} \frac{d^2 \psi}{dx^2} - \alpha \psi + \beta|\psi|^2 \psi = 0 \, ,$$

(10)

subject to our boundary conditions, is

$$\psi(x) = (\alpha/\beta)^{1/2} \tanh(x/\sqrt{2} \xi) \, .$$

(11)

This may be verified by direct substitution. Deep inside the superconductor we have $\psi_0 = (\alpha/\beta)^{1/2}$, as follows from the minimization of the terms $-\alpha|\psi|^2 + \frac{1}{2} \beta|\psi|^4$ in the free energy. We see from (11) that $\xi$ marks the extent of the coherence of the superconducting wavefunction into the normal region.

We have seen that deep inside the superconductor the free energy is a minimum when $|\psi_0|^2 = \alpha/\beta$, so that

$$F_S = F_N - \frac{\alpha^2}{2} \beta = F_N - H_c^2/8\pi \, ,$$

(12)

by definition of the thermodynamic critical field $H_c$ as the stabilization free energy density of the superconducting state. It follows that the critical field is related to $\alpha$ and $\beta$ by

$$H_c = (4\pi\alpha^2/\beta)^{1/2} \, .$$

(13)

Consider the penetration depth of a weak magnetic field ($B \ll H_c$) into a superconductor. We assume that $|\psi|^2$ in the superconductor is equal to $|\psi_0|^2$, the value in the absence of a field. Then the equation for the supercurrent flux reduces to
\[ j_s(r) = -(e^2/mc)|\psi_0|^2 A, \]  
(14)

which is just the London equation \( j_s(r) = -(e/4\pi^2 \lambda^2) A \), with the penetration depth

\[ \lambda = \left( \frac{mc^2}{4\pi q^2 |\psi_0|^2} \right)^{1/2} = \left( \frac{mc^2\beta}{4\pi q^2 \alpha} \right)^{1/2}. \]  
(15)

The dimensionless ratio \( \kappa = \lambda/\xi \) of the two characteristic lengths is an important parameter in the theory of superconductivity. From (9) and (15) we find

\[ \kappa = \frac{mc}{q\hbar} \left( \frac{\beta}{2\pi} \right)^{1/2}. \]  
(16)

We now show that the value \( \kappa = 1/\sqrt{2} \) divides type I superconductors \( (\kappa < 1/\sqrt{2}) \) from type II superconductors \( (\kappa > 1/\sqrt{2}) \).

**Calculation of the Upper Critical Field.** Superconducting regions nucleate spontaneously within a normal conductor when the applied magnetic field is decreased below a value denoted by \( H_{c2} \). At the onset of superconductivity \( |\psi| \) is small and we linearize the GL equation (6) to obtain

\[ \frac{1}{2m} (-i\hbar \nabla - qA/c)^2 \psi = \alpha \psi. \]  
(17)

The magnetic field in a superconducting region at the onset of superconductivity is just the applied field, so that \( A = B(0,x,0) \) and (17) becomes

\[ -\frac{\hbar}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} \right) \psi + \frac{1}{2m} \left( i\hbar \frac{\partial}{\partial y} + \frac{qB}{c} x \right)^2 \psi = \alpha \psi. \]  
(18)

This is of the same form as the Schrödinger equation of a free particle in a magnetic field.

We look for a solution in the form \( \exp[i(k_y y + k_z z)] \phi(x) \) and find

\[ (1/2m)[-\hbar^2 d^2/dx^2 + \hbar^2 k_z^2 + (\hbar k_y - qBx/c)^2] \phi = \alpha \phi. \]  
(19)

this is the equation for an harmonic oscillator, if we set \( E = \alpha - (\hbar^2/2m)(k_y^2 + k_z^2) \) as the eigenvalue of

\[ (1/2m)[-\hbar^2 d^2/dx^2 + (q^2 B^2/c^2)X^2 - (2\hbar k_y qB/c)X] \phi = E \phi. \]  
(20)

The term linear in \( x \) can be transformed away by a shift of the origin from 0 to \( x_0 = \hbar k_y qB/2mc \), so that (20) becomes, with \( X = x - x_0 \),

\[ -\left[ \frac{\hbar^2}{2m} \frac{d^2}{dX^2} + \frac{1}{2m(qB/mc)^2} X^2 \right] \phi = (E + \hbar^2 k_y^2/2m) \phi. \]  
(21)
The largest value of the magnetic field $B$ for which solutions of (21) exist is given by the lowest eigenvalue, which is

$$\frac{1}{2} \hbar \omega = \hbar q B_{\text{max}}/2mc = \alpha - \hbar^2 k_z^2/2m,$$  \hspace{1cm} (22)

where $\omega$ is the oscillator frequency $qB/mc$. With $k_z$ set equal to zero,

$$B_{\text{max}} = H_c \equiv 2\alpha mc/q\hbar.$$  \hspace{1cm} (23)

This result may be expressed by (13) and (16) in terms of the thermodynamic critical field $H_c$ and the GL parameter $\kappa = \lambda/\xi$:

$$H_c \equiv H_{c2} = \frac{2\alpha mc}{q\hbar} \cdot \frac{H_c}{(4\pi \alpha^2/\beta)^{1/2}} = \sqrt{2} \frac{mc}{\hbar q} \sqrt{\frac{\beta}{2\pi}} H_c = \sqrt{2} \kappa H_c.$$  \hspace{1cm} (24)

When $\lambda/\xi > 1/\sqrt{2}$, a superconductor has $H_{c2} > H_c$ and is said to be of type II.

It is helpful to write $H_{c2}$ in terms of the flux quantum $\Phi_0 = 2\pi \hbar c/q$ and $\xi^2 = \hbar^2/2mc$:

$$H_{c2} = \frac{2mc\alpha}{q\hbar} \cdot \frac{q\Phi_0}{2\pi \hbar c} \cdot \frac{\hbar^2}{2mc\xi^2} = \frac{\Phi_0}{2\pi \xi^2}.$$  \hspace{1cm} (25)

This tells us that at the upper critical field the flux density $H_{c2}$ in the material is equal to one flux quantum per area $2\pi \xi^2$, consistent with a fluxoid lattice spacing of the order of $\xi$.

**APPENDIX J: ELECTRON-PHONON COLLISIONS**

Phonons distort the local crystal structure and hence distort the local band structure. This distortion is sensed by the conduction electrons. The important effects of the coupling of electrons with phonons are

- Electrons are scattered from one state $k$ to another state $k'$, leading to electrical resistivity.
• Phonons can be absorbed in the scattering event, leading to the attenuation of ultrasonic waves.
• An electron will carry with it a crystal distortion, and the effective mass of the electron is thereby increased.
• A crystal distortion associated with one electron can be sensed by a second electron, thereby causing the electron-electron interaction that enters the theory of superconductivity.

The deformation potential approximation is that the electron energy \( \epsilon(k,r) \) is coupled to the crystal dilation \( \Delta(r) \) or fractional volume change by

\[
\epsilon(k,r) = \epsilon_0(k) + C\Delta(r) ,
\]

where \( C \) is a constant. The approximation is useful for spherical band edges \( \epsilon_0(k) \) at long phonon wavelengths and low electron concentrations. The dilation may be expressed in terms of the phonon operators \( a_q, a_q^+ \) of Appendix C by

\[
\Delta(r) = i \sum_q \left( \hbar/2M\omega_q \right)^{1/2} |q| [a_q \exp(iq \cdot r) - a_q^+ \exp(-iq \cdot r)] ,
\]

as in QTS, p. 23. Here \( M \) is the mass of the crystal. The result (2) also follows from (C.32) on forming \( q_s - q_{s-1} \) in the limit \( k \ll 1 \).

In the Born approximation for the scattering we are concerned with the matrix elements of \( C\Delta(r) \) between the one-electron Bloch states \( |k\rangle \) and \( |k'\rangle \), with \( |k\rangle = \exp(ik \cdot r)u_k(r) \). In the wave field representation the matrix element is

\[
H' = \int d^3x \psi^+(r)C\Delta(r)\psi(r) = \sum_{k'k} c_{k'}^+c_k |k'\rangle \langle C\Delta|k\rangle
\]

\[
= iC \sum_{k'k} c_{k'}^+c_k \sum_q \left( \hbar/2M\omega_q \right)^{1/2} |q| (a_q \int d^3x u_{k'}^*u_k e^{i(k-k'+q) \cdot r} - a_q^+ \int d^3x u_{k'}^*u_k e^{i(k-k'-q) \cdot r} ) ,
\]

where

\[
\psi(r) = \sum_k c_k \varphi_k(r) = \sum_k c_k \exp(ik \cdot r)u_k(r) ,
\]

where \( c_k^+, c_k \) are the fermion creation and annihilation operators. The product \( u_{k'}^*(r)u_k(r) \) involves the periodic parts of the Bloch functions and is itself periodic in the lattice; thus the integral in (3) vanishes unless

\[
k - k' \pm q = \begin{cases} 0, & \text{vector in the reciprocal lattice.} \\ \end{cases}
\]
In semiconductors at low temperatures only the possibility zero (i.e. processes) may be allowed energetically.

Let us limit ourselves to N processes, and for convenience we approximate \( \int d^3x \ u_k \cdot u_k \) by unity. Then the deformation potential perturbation is

\[
H' = iC \sum_{kq} \left( \frac{\hbar}{2M\omega_q} \right)^{1/2} |q| (a_q c_{k+q}^+ c_k - a_q^+ c_{k-q}^+ c_k) .
\]  

(5)

**Relaxation Time.** In the presence of the electron-phonon interaction the wavevector \( k \) is not a constant of the motion for the electron alone, but the sum of the wavevectors of the electron and virtual phonon is conserved. Suppose an electron is initially in the state \( |k\rangle \); how long will it stay in the same state?

We calculate first the probability \( w \) per unit time that the electron in \( k \) will emit a phonon \( q \). If \( n_q \) is the initial population of the phonon state,

\[
 w(k - q; n_q + 1|k; n_q) = (2\pi/\hbar) |\langle k - q; n_q + 1|H'|k; n_q \rangle|^2 \delta(\epsilon_k - \hbar\omega_q - \epsilon_{k-q}) ,
\]  

(6)

by time-dependent perturbation theory. Here

\[
|\langle k - q; n_q + 1|H'|k; n_q \rangle|^2 = (C^2 \hbar q/2Mc_s)(n_q + 1) .
\]  

(7)

The total collision rate \( W \) of an electron in the state \( |k\rangle \) with a phonon system at absolute zero is, with \( n_q = 0 \),

\[
W = \frac{C^2}{4\pi \rho c_s} \int_{-1}^{1} d(\cos \theta_q) \int_{0}^{q_m} dq \ q^3 \delta(\epsilon_k - \epsilon_{k-q} - \hbar\omega_q) ,
\]  

(8)

where \( \rho \) is the mass density.

The argument of the delta function is

\[
\frac{\hbar^2}{2m^*} (2k \cdot q - q^2) - \hbar c_s q = \frac{\hbar^2}{2m^*} (2k \cdot q - q^2 - q_c) ,
\]  

(9)

where \( q_c = 2\hbar m^* c_s \), with \( c_s \) the velocity of sound. The minimum value of \( k \) for which the argument can be zero is \( k_{\text{min}} = \frac{1}{2} (q + q_c) \), which for \( q = 0 \) reduces to \( k_{\text{min}} = \frac{1}{2} q_c = m^* c_s / \hbar \). For this value of \( k \) the electron group velocity \( v_g = k_{\text{min}} / m^* \) is equal to the velocity of sound. Thus the threshold for the emission of phonons by electrons in a crystal is that the electron group velocity should exceed the acoustic velocity. This requirement resembles the Cerenkov threshold for the emission of photons in crystals by fast electrons. The electron energy at the threshold is \( \frac{1}{2} m^* c_s^2 \sim 10^{-27} \cdot 10^{11} \sim 10^{-16} \text{ erg} \sim 1 \text{ K} \). An electron of energy below this threshold will not be slowed down in a perfect crystal at absolute zero, even by higher order electron-phonon interactions, at least in the harmonic approximation for the phonons.
For $k > q_c$ we may neglect the $q q_c$ term in (9). The integrals in (8) become
\[
\int_{-1}^{1} d\mu \int dq \frac{1}{2m^*/\hbar^2} q^3 (2m^*/\hbar^2 q) \delta (2k\mu - q) = (8m^*/\hbar^2) \int_{0}^{1} d\mu \frac{k^2 \mu^2}{1} = 8m^*/3\hbar^2 \ ,
\]
and the phonon emission rate is
\[
W(\text{emission}) = \frac{2C^2m^*k^2}{3\pi\hbar c\hbar^2} ,
\]
directly proportional to the electron energy $\epsilon_k$. The loss of the component of wavevector parallel to the original direction of the electron when a phonon is emitted at an angle $\theta$ to $k$ is given by $q \cos \theta$. The fractional rate of loss of $k_z$ is given by the transition rate integral with the extra factor $(q/k) \cos \theta$ in the integrand. Instead of (10), we have
\[
(2m^*/\hbar^2 k) \int_{0}^{1} d\mu \ 8k^3 \mu^4 = 16m^*/5\hbar^2 \ ,
\]
so that the fractional rate of decrease of $k_z$ is
\[
W(k_z) = 4C^2m^*/5\pi\hbar c\hbar^2 .
\]
This quantity enters into the electrical resistivity.

The above results apply to absolute zero. At a temperature $k_B T > \hbar c k$ the integrated phonon emission rate is
\[
W(\text{emission}) = \frac{C^2m^*k k_B T}{\pi c_s^2 \rho \hbar^3} .
\]
For electrons in thermal equilibrium at not too low temperatures the required inequality is easily satisfied for the rms value of $k$. If we take $C = 10^{-12}$ erg, $m^* = 10^{-27}$ g, $k = 10^7$ cm$^{-1}$, $c_s = 3 \times 10^5$ cm s$^{-1}$, $\rho = 5$ g cm$^{-3}$, then $W = 10^{12}$ s$^{-1}$. At absolute zero (13) gives $W = 5 \times 10^{10}$ s$^{-1}$ with these same parameters.
Subject Index

Absolute thermoelectric power, 228
Acceptor ionization energies, 225
Acceptor states, 224
Acoustical phonon branch, 104
Acronyms, 328
Activation energy, 544
Adiabatic demagnetization, 431
Alfvén waves, 302
Alkali halide crystals, table, 73
Alloys, 611
  strength, 600
  transition metal, 624
Alnico V, 476
Amorphous solids, 521
  ferromagnets, 529
  semiconductors, 530
Anharmonic interactions, 129
Anisotropy energy, 470, 471
Annihilation operators, 641
Antiferroelectric crystals, 404, 406
Antiferromagnetic crystals, 462
  magnons, 466
  resonance, 506, 508
Anti-Stokes line, 323
Atomic form factor, 45
Atomic radii, 77
  table, 78
Band gap, 175, 199
Band structure, GaAs, 218
  germanium, 215
Barium titanate, 397
Basis, 4, 5
Binding, crystalline, 54
Biomagnetism, 478
Bloch equations, 492
Bloch frequency, 230
Bloch function, 179, 246
Bloch oscillator, 229
Bloch T^3 law, 456
Bloch theorem, 179, 185
Bloch wall, 472
Bohr magneton, 421
Boltzmann transport equation, 647
Boson operators, 641
Boundary conditions, periodic, 120
Bragg law, 29
Bravais lattice, 10, 556
Brillouin function, 422
Brillouin scattering, 306
Brillouin zone, 37, 217, 266
  first, 39, 100, 176
  volume, 51
Bulk moduli, table, 59
Bulk modulus, 86
Bulk modulus, electron gas, 169
  Fermi gas, 303
Burgers vector, 592
Cauchy integral, 309
Causality, 329
Cell, primitive, 6
  unit, 6

Wigner-Seitz, 8, 38
Centipoise, 527
Central equation, 186
Cesium chloride structure, 186
Chalcogenide glasses, 531
Charge density waves, 301
Chemical potential, 146, 153, 154, 169
Classical distribution, 648
Clogston relation, 349
Clausius-Mossotti relation, 390
Coercive force, small particle, 472
Coercivity, 470, 475
Coherence length, 352, 660
  intrinsic, 353
Cohesive energy, 57, 66, 80
  Fermi gas, 170
  sodium metal, 251
  square well potential, 266
Cohesive energies, table, 57
Collisionless electron gas, 311
Color centers, 548
Compressibility, 86
Compressibilities, table, 59
Concentration, table, 24
Conduction band edge, 199
Conduction electron
  ferromagnetism, 438
Conduction electrons, susceptibility, 433
Conductivity, electrical, 156, 158, 221
  impurity, 222
  ionic, 298
  thermal, 131, 166
Conductivity sum rule, 330
<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conductor, perfect</td>
<td>339</td>
</tr>
<tr>
<td>Configurational heat capacity</td>
<td>630</td>
</tr>
<tr>
<td>Contact hyperfine interaction</td>
<td>497</td>
</tr>
<tr>
<td>Continuum wave equation</td>
<td>112</td>
</tr>
<tr>
<td>Cooper pairs</td>
<td>356</td>
</tr>
<tr>
<td>Covalent bond</td>
<td>72</td>
</tr>
<tr>
<td>Covalent crystals</td>
<td>72</td>
</tr>
<tr>
<td>Creation operators</td>
<td>641</td>
</tr>
<tr>
<td>Creep</td>
<td>603</td>
</tr>
<tr>
<td>Critical currents</td>
<td>372</td>
</tr>
<tr>
<td>Critical field</td>
<td>339</td>
</tr>
<tr>
<td>Critical thin films</td>
<td>375</td>
</tr>
<tr>
<td>Critical magnetic field</td>
<td>372</td>
</tr>
<tr>
<td>Critical point exponents</td>
<td>445</td>
</tr>
<tr>
<td>Critical points</td>
<td>312</td>
</tr>
<tr>
<td>Critical shear stress</td>
<td>588</td>
</tr>
<tr>
<td>Critical temperature</td>
<td>399</td>
</tr>
<tr>
<td>Crystal field</td>
<td>426</td>
</tr>
<tr>
<td>Crystal growth</td>
<td>25</td>
</tr>
<tr>
<td>Crystal momentum</td>
<td>109</td>
</tr>
<tr>
<td>Crystal structure, elements</td>
<td>23</td>
</tr>
<tr>
<td>Crystalline binding</td>
<td>54</td>
</tr>
<tr>
<td>Crystallography, surface</td>
<td>556</td>
</tr>
<tr>
<td>Cubic lattices</td>
<td>12</td>
</tr>
<tr>
<td>Cubic zinc sulfide structure</td>
<td>20</td>
</tr>
<tr>
<td>Curie constant</td>
<td>423</td>
</tr>
<tr>
<td>Curie-Weiss law</td>
<td>444</td>
</tr>
<tr>
<td>Cyclotron frequency</td>
<td>164</td>
</tr>
<tr>
<td>Cyclotron resonance</td>
<td>213</td>
</tr>
<tr>
<td>Spherical energy surface</td>
<td>231</td>
</tr>
<tr>
<td>Dangling bonds</td>
<td>554</td>
</tr>
<tr>
<td>Davydov splitting</td>
<td>331</td>
</tr>
<tr>
<td>Deformation</td>
<td>80</td>
</tr>
<tr>
<td>De Haas-van Alphen effect</td>
<td>257</td>
</tr>
<tr>
<td>Debye model, density of states, 122</td>
<td></td>
</tr>
<tr>
<td>Debye period</td>
<td>266</td>
</tr>
<tr>
<td>Direct gap semiconductor</td>
<td>213</td>
</tr>
<tr>
<td>Dislocation densities</td>
<td>598</td>
</tr>
<tr>
<td>Dislocation multiplication</td>
<td>600</td>
</tr>
<tr>
<td>Dislocations</td>
<td>589</td>
</tr>
<tr>
<td>Dispersion relation, phonons, 100, 104</td>
<td></td>
</tr>
<tr>
<td>Direct photon absorption</td>
<td>201</td>
</tr>
<tr>
<td>Diffusion</td>
<td>495</td>
</tr>
<tr>
<td>Diffusive</td>
<td>544</td>
</tr>
<tr>
<td>Dilution</td>
<td>82</td>
</tr>
<tr>
<td>Dilatation</td>
<td>82</td>
</tr>
<tr>
<td>Direct electron scattering</td>
<td>522</td>
</tr>
<tr>
<td>Debye T³ law</td>
<td>124</td>
</tr>
<tr>
<td>Debye temperature</td>
<td>122</td>
</tr>
<tr>
<td>Debye-Waller factor</td>
<td>632</td>
</tr>
<tr>
<td>Defects, paramagnetic</td>
<td>499</td>
</tr>
<tr>
<td>Deficit semiconductors</td>
<td>222</td>
</tr>
<tr>
<td>Degeneracy</td>
<td>145</td>
</tr>
<tr>
<td>Degenerate semiconductor</td>
<td>286</td>
</tr>
<tr>
<td>Demagnetization, isentropic</td>
<td>431</td>
</tr>
<tr>
<td>Demagnetization factors</td>
<td>504</td>
</tr>
<tr>
<td>Density, table</td>
<td>24</td>
</tr>
<tr>
<td>Density of modes (see density of states)</td>
<td></td>
</tr>
<tr>
<td>Density of states, 128, 149</td>
<td></td>
</tr>
<tr>
<td>Debye model, 120</td>
<td></td>
</tr>
<tr>
<td>Einstein model, 124</td>
<td></td>
</tr>
<tr>
<td>General result</td>
<td>127</td>
</tr>
<tr>
<td>One dimension</td>
<td>118</td>
</tr>
<tr>
<td>Singularity</td>
<td>139</td>
</tr>
<tr>
<td>Three dimensions</td>
<td>119</td>
</tr>
<tr>
<td>Depolarization factors</td>
<td>384</td>
</tr>
<tr>
<td>Depolarization field</td>
<td>384</td>
</tr>
<tr>
<td>Diamagnetism</td>
<td>417</td>
</tr>
<tr>
<td>Diamond structure, 19, 51, 195</td>
<td></td>
</tr>
<tr>
<td>Dielectric constant</td>
<td>388</td>
</tr>
<tr>
<td>Dielectric function, 307</td>
<td></td>
</tr>
<tr>
<td>Electron gas</td>
<td>271</td>
</tr>
<tr>
<td>Lindhard</td>
<td>282</td>
</tr>
<tr>
<td>Thomas-Fermi</td>
<td>282</td>
</tr>
<tr>
<td>Dielectrics</td>
<td>381</td>
</tr>
<tr>
<td>Dielectric susceptibility</td>
<td>385</td>
</tr>
<tr>
<td>Diffraction, Josephson junction, 369</td>
<td></td>
</tr>
<tr>
<td>Diffraction conditions, wave, 34</td>
<td></td>
</tr>
<tr>
<td>Diffusion</td>
<td>495</td>
</tr>
<tr>
<td>Diffusive</td>
<td>544</td>
</tr>
<tr>
<td>Dilatation</td>
<td>82</td>
</tr>
<tr>
<td>Dilatation</td>
<td>82</td>
</tr>
<tr>
<td>Direct gap semiconductor</td>
<td>213</td>
</tr>
<tr>
<td>Direct photon absorption</td>
<td>201</td>
</tr>
<tr>
<td>Diffusion</td>
<td>495</td>
</tr>
<tr>
<td>Diffusive</td>
<td>544</td>
</tr>
<tr>
<td>Dilatation</td>
<td>82</td>
</tr>
<tr>
<td>Dilatation</td>
<td>82</td>
</tr>
<tr>
<td>Direct electron scattering</td>
<td>522</td>
</tr>
<tr>
<td>Debye T³ law</td>
<td>124</td>
</tr>
<tr>
<td>Debye temperature</td>
<td>122</td>
</tr>
<tr>
<td>Table</td>
<td>126</td>
</tr>
<tr>
<td>Debye-Waller factor</td>
<td>632</td>
</tr>
<tr>
<td>Planck, 101</td>
<td></td>
</tr>
<tr>
<td>Dog's bone orbit</td>
<td>263</td>
</tr>
<tr>
<td>Domains, origin of</td>
<td>474</td>
</tr>
<tr>
<td>Closure, 474</td>
<td></td>
</tr>
<tr>
<td>Ferroelectric</td>
<td>406</td>
</tr>
<tr>
<td>Magnetic bubble</td>
<td>481</td>
</tr>
<tr>
<td>Donor ionization energies</td>
<td>224</td>
</tr>
<tr>
<td>Donor states</td>
<td>222</td>
</tr>
<tr>
<td>Doping</td>
<td>222</td>
</tr>
<tr>
<td>Dulong and Petit value</td>
<td>127</td>
</tr>
<tr>
<td>Edge dislocation</td>
<td>589</td>
</tr>
<tr>
<td>Effective mass, 209, 210, 212</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>211</td>
</tr>
<tr>
<td>Semiconductors, table, 212, 214</td>
<td></td>
</tr>
<tr>
<td>Tensor, 210</td>
<td></td>
</tr>
<tr>
<td>Thermal</td>
<td>156</td>
</tr>
<tr>
<td>Einstein equation</td>
<td>561</td>
</tr>
<tr>
<td>Einstein model, density of states</td>
<td></td>
</tr>
<tr>
<td>Elastic compliance</td>
<td>83</td>
</tr>
<tr>
<td>Elastic limit</td>
<td></td>
</tr>
<tr>
<td>Elastic stiffness</td>
<td>83, 91</td>
</tr>
<tr>
<td>Elastic strain</td>
<td>80</td>
</tr>
<tr>
<td>Elastic waves</td>
<td>87</td>
</tr>
<tr>
<td>Elastic wave quantization</td>
<td></td>
</tr>
<tr>
<td>Electric field, local, 386</td>
<td></td>
</tr>
<tr>
<td>Macroscopic</td>
<td>382</td>
</tr>
<tr>
<td>Electric quadruple moment</td>
<td>503</td>
</tr>
<tr>
<td>Electrical conductivity</td>
<td>156</td>
</tr>
<tr>
<td>Table</td>
<td>160</td>
</tr>
<tr>
<td>Electrical resistivity</td>
<td>159</td>
</tr>
<tr>
<td>Table</td>
<td>160</td>
</tr>
<tr>
<td>Electromechanical coupling factor</td>
<td>410</td>
</tr>
<tr>
<td>Electron affinities</td>
<td>68</td>
</tr>
<tr>
<td>Electron compound</td>
<td>615, 616</td>
</tr>
<tr>
<td>Electron density distribution</td>
<td>67</td>
</tr>
<tr>
<td>Electron heat capacity, table, 257</td>
<td></td>
</tr>
<tr>
<td>Electron orbits</td>
<td>242</td>
</tr>
<tr>
<td>Electron spin resonance</td>
<td>485</td>
</tr>
<tr>
<td>Electron work functions</td>
<td>560</td>
</tr>
<tr>
<td>Table</td>
<td>562</td>
</tr>
<tr>
<td>Electron-electron collisions, 294</td>
<td></td>
</tr>
<tr>
<td>Electron-electron interaction, 294</td>
<td></td>
</tr>
</tbody>
</table>
Electron-hole drops, 321
Electron-lattice interaction, 297
Electron-phonon collisions, 662
Electronic polarizabilities, table, 391
Electronic structure, surface, 560
Electrostatic screening, 280
Elementary excitations, 98
Empirical pseudopotential method, 253
Empty core model, 253
Empty lattice approximation, 188
ENDOR, 500
Energy band calculation, 244
Energy bands, 173, 175
Energy gap, 173, 175
superconductors, 344
table, semiconductors, 201
Energy levels, one dimension, 144
Energy loss, fast particles, 326
Energy loss spectra, electrons, 279
Entropy, superconductors, 342
Entropy of mixing, 620
Equation of motion, hole, 208
Equations of motion, electron, 203
ESCA, 324
Eutectics, 622
Ewald construction, 37 sphere, 559
sums, 634
Ewald-Kornfeld method, 637
Exchange energy, 446
field, 443
frequency resonance, 516
integral, 443
interaction, 74
narrowing, 510
Excitons, 312
Exclusion principle, Pauli, 64
Exotic superconductors, 156
Extended zone scheme, 238
Extinction coefficient, 307
Extremal orbits, 260
Fact, atomic form, 43, 45
structure, 44
Fermi energy, 145
gas, 141, 146
level, 144
liquid, 294
surface, copper, 262
surface, gold, 262
surface parameters, table, 150
surfaces, 238
Fermi-Dirac distribution, 146, 643
Fermions, heavy, 156, 337
Ferroelasticity, 410
Ferroelectricity criterion, 412
crystals, 381, 393
domains, 385
linear array, 414
Ferromagnetism, crystals, 443, 449
amorphous, 529
conduction electron, 438
domains, 468
order, 458
resonance, 503
Fiber optics, 535
Fick's law, 544
Fine structure constant, 566
First Brillouin zone, 39, 100, 176
First-order transition, 402
Flux quantization, 356
Fluxoid, 358, 363
Fourier analysis, 30, 42, 183
Fractional quantized Hall effect (FQHE), 570
Frank-Read source, 600
Free-energy, superconductors, 343
Frenkel defect, 542, 551
Frenkel exciton, 314, 316
Friedel oscillations, 283
Fullerenes, 373
Fused quartz, 524
Gap, direct, 202
Gap, indirect, 202
Gap plasmons, 304
Gauge transformation, 654
Geomagnetism, 478
Ginzburg-Landau equation, 658
Ginzburg-Landau parameter, 354, 661
GLAC theory, 360
Glass, 527
chalcogenide, 531
transition temperature, 527
Grain boundary, small-angle, 595
Group velocity, 102
Gruneisen constant, 129
Gyromagnetic ratio, 420
Gyroscopic equation, 490
Hagen-Rubens relation, 335
Hall coefficient, 164
table, 167
Hall effect, 164, 373
two carrier types, 231
Hall resistance, 165
Hall resistivity, 564
Hardness, 606
He\(^3\), liquid, 170
Heat capacity, configurational, 630
electron gas, 151
glasses, 533
metals, 155
phonon, 117
superconductors, 342
Heavy fermions, 156, 337
Heisenberg model, 446
Helicon waves, 303
Heterojunction, 576
Heterostructures, 574
Hexagonal close-packed structure, 17, 25
Hexagonal lattice, 51
High temperature superconductors, 371
Hole, equation of motion, 206, 208
Hole orbits, 242
Hole trapping, 298
Hooke's law, 83
HTS, 371
Hume-Rothery rules, 424
Hund rules, 424
Hydrogen bonds, 76
Hyperfine constant, 498
Hyperfine effects, ESR in metals, 515
Hyperfine splitting, 497
Hysteresis, 475
Ilmenite, 396
F centers, 499, 548, 552
Impurity conductivity, 222
Impurity orbits, 230
Index system, 12
Indirect gap, 202
Indirect photon absorption, 201
Inelastic scattering, phonons, 110
Inert gas crystals, table, 60
Injection laser, 577
Insulators, 193
Interband transitions, 312
Interface plasmons, 302
Interfacial polarization, 412
Intrinsic carrier
concentration, 200, 216
Intrinsic coherence length, 353
Intrinsic mobility, 220
Intrinsic semiconductors, 199
Inverse spinel, 460
 Ionic bond, 66
 Ionic character, table, 76
 Ionic conductivity, 298, 542
 Ionic crystals, 66
 Ionic radii, 79
 table, 78
 Ionization, thermal, 226
 Ionization energies, acceptor, 225
donor, 224
table, 61
Iron garnets, 461
Iron group, 425
Isentropic demagnetization, 431
Isotope effect,
superconductors, 346, 347
Isotope effect, thermal
conduction, 138
Jahn-Teller effect, 221, 429
trapping, 551
Josephson tunneling, 367, 369, 377
Jump frequency, 546
 Junction, superconducting, 364
Junctions, p-n, 570
k·p perturbation theory, 267
 Kelvin relation, 228
Knight shift, 500, 502
Kohn anomaly, 113
Kondo effect, 628
Kramers-Heisenberg
dispersion, 309
Kramers-Kronig relations, 308, 310
Kronig-Penney model, 180, 195
 reciprocal space, 187
LA modes, 104
Lagrangian equations, 652
Landau gauge, 570
Landau level, 268, 278, 566
Landau theory, Fermi liquid, 294
Landau theory, phase
transition, 400
Landau-Ginzburg equations, 354
Langevin diamagnetism, 417
Langevin result, 419
Lanthanide group, 425
Larmor frequency, 418
Larmor theorem, 418
Laser, 510, 577
 injection, 578
ruby, 512
semiconductor, 577
Lattice, 4
Bravais, 10
Lattice constants,
equilibrium, 65
Lattice frequencies, table, 292
Lattice momentum, 205
Lattice sums, dipole arrays,
637
Lattice translation operator,
647
Lattice types, 8, 10
Lattice vacancies, 541
Lattices, cubic, 12
Laue equation, 36, 581, 631
Law of mass action, 219
LCIO approximation, 245
LEED, 557, 558
Lennard-Jones potential, 63, 64
Lenz's law, 417
Lindhard dielectric function,
282
Line width, 494
Liquid He₃, 170
Liquidus, 622
LO modes, 104
Local electric field, 386
London equation, 349, 655
London gauge, 656
London penetration depth,
353, 373, 377
Longitudinal plasma
oscillations, 276
Longitudinal relaxation time,
490
Long-range order, 620
Lorentz field, 388
Lorenz number, 168
Low-angle grain boundary,
595
LST relation, 291, 294
Luminescence, 315
Lyddane-Sachs-Teller relation,
291
Madelung constant, 69
Madelung energy, 68
Magnetic breakdown, 268
Magnetic bubble domains,
480
Magnetic force microscopy,
478
Magnetic monopole, 358
Magnetic resonance, 487
Magnetic resonance
tomography, 487
Magnetic susceptibility, 417
Magnetite, 461
Magnetization, 417
saturation, 446
Magnetocconductivity, tensor,
170, 271
Magnetocrystalline energy,
471
Magnetoeleastic coupling, 479
Magnetogyratic ratio, 420
Magneton numbers, 425
Magnetoplasma frequency, 303
Magnetoresistance, 564
two carrier types, 231
Magnetotaxis, 478
Magnons, 450
 antiferromagnetic, 466
dispersion relation, 481
Maser action, 510
Matthiessen's rule, 161
Maxwell equations, 382
Mean field approximation,
443
enching, orbital angular momentum, 426
E, 564

ial distribution function, 523
Van effect, 322
scattering, 306
dom network, 526
dom stacking, 22
earth ions, 423
high attenuation, 536
real lattice points, 38
prosthetic lattice vectors, 33
combination radiation, 321
construction, 554
ification, 570
ured zone scheme, 195, 235
ectance, 293, 307, 308
ction, normal incidence, 330
ctivity coefficient, 307
active index, 307
nation, 556
rect, 492
bach, 492
aman, 492
ation time, 490
itudinal, 490
an-lattice, 490
ance, 470
stance, surface, 171
stance per square, 171
itivity, electrical, 158, 159
itivity ratio, 162
ence, electron gas, 304
ence function, 309
uration, 516
ED, 560
son-Dushman equation, 562
Y theory, 628
ing coordinate system, 515
 laser, 512
raction magnetization, 446
raction polarization, 412
vation, rf, 516
ning tunneling microscopy, 20
Schottky barrier, 573
Schottky defect, 541
Schottky vacancies, 551
Screened Coulomb potential, 282
Screening, electrostatic, 280
Screw dislocation, 590
Second-order transition, 402
Self-diffusion, 547
Self-trapping, 221
Semiconductor, degenerate, 286
Semiconductor crystals, 174, 199
Semiconductor lasers, 577
Semiconductors, deficit, 222
Semimetals, 174, 203, 228
Shear constant, 94
Shear strength, single crystals, 587
Shear stress, critical, 587
Short-range order, 621
Shubnikov-de Haas effect, 583
Singularity, Van Hove, 129
Singularity, density of states, 139
Slater-Pauling plot, 627
Slip, 588
Sodium chloride structure, 15
Sodium metal, 142
Soft modes, 400, 413
Solar cells, 573
Solids, 622
Solitons, 473
Solubility gap, 622
Spectroscopic splitting factor, 429
Spheres, metal, 171
Spinel, 459
Spin-lattice interaction, 491
Spin-lattice relaxation time, 490
Spin wave (see also magnon), 450
quantization, 454
resonance, 506
Square lattice,
Stability criteria, 95
Stabilization free energy, 347
Stacking fault, 589
STM, 20
Stokes line, 323
Strain component, 81
Stress component, 4
Structural phase transitions, 393
Structure factor, 42, 635
bce lattice, 44
diamond, 51
fcc lattice, 45
Substrate, 555
Superconductivity, table, 336
exotic, 156, 337
type I, 340
type II, 340, 360
Superlattices, 618, 630
Superparamagnetism, 477
Surface crystallography, 556
electronic structure, 560
ets, 558
plasmons, 279, 302
resistance, 171
states, 562
subbands, 562
transport, 564
Susceptibility, dielectric, 385
magnetic, 417
Symmetry operation, 4
TA modes, 104
Taxis, 478
Temperature, Debye, 122
Temperature dependence, 633
reflection lines, 633
Tetrahedral angle, 25
Thermal conductivity, 131, 166
glasses, 534
isotope effect, 138
metals, 166
table, 126
Thermal dilation, 139
Thermal effective mass, 156
Thermal excitation of magnons, 455
Thermal expansion, 130
Thermal ionization, 226
Thermal resistivity, phonon gas, 133
Thermionic emission, 561
Thermodynamics, superconducting transition, 347
Thermoelectric effects, 227
Thomas-Fermi approximation, 280
Thomas-Fermi dielectric function, 282
Three-level maser, 512.
Tight binding method, 245
TO modes, 104
Tomography, magnetic
resonance, 342, 487
Transistor, MOS, 564
Transition, displacive, 394
first-order, 402
metal-insulator, 284
order-disorder, 394
second-order, 402
Transition metal alloys, 624
Transition temperature, glass, 527
Transitions, displacive, 398
Translation operation, 4
Translation vector, 4
Transparency, alkali metals, 274
Transport, surface, 564
Transport equation, Boltzmann,
Transverse optical modes, plasma, 274
Transverse relaxation time, 492
Triplet excited states, 437
Tunneling, 364
Josephson, 366
Tunneling microscope, 580
Twining, 589
Two-fluid model, 372, 376
Two-level system, 440
Type I superconductors, 340
Type II superconductors, 340, 360
Ultraviolet transmission
limits, 275
Umklapp processes, 135, 162
Unit cell, 6
UPS, 324
Upper critical field, 661
Valence band edge, 199
Van der Waals interaction, 54, 56, 304
Van Hove singularities, 129
Van Vleck paramagnetism, 430
Vector potential, 651
Vickers hardness number, 606
Viscosity, 527
Vitreous silica, 524
Vortex state, 340, 361, 376
Wannier functions, 267
Wannier ladder, 230
Wave equation, continuum, 112
periodic lattice, 184
periodic potential, 183
Wavevectors, energy gap, 196
Weiss field, 443
Whiskers, 604
Wiedemann-Franz law, 166
Wigner-Seitz boundary
condition, 250, 251
Wigner-Seitz cell, 8, 38, 251
Wigner-Seitz method, 244, 248
Work function, 560
Work-hardening, 603
XPS, 324
Young's modulus, 94
Yttrium iron garnet, 505
Zener breakdown, 231
Zener tunneling, 230
Zero-point lattice
displacement, 139
Zero-point motion, 55, 62
Zero-field splitting, 510
ZnS structure, 20

**Table of SI Prefixes**

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Prefix</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-15}$</td>
<td>femto</td>
<td>f</td>
</tr>
<tr>
<td>$10^{-12}$</td>
<td>pico</td>
<td>p</td>
</tr>
<tr>
<td>$10^{-9}$</td>
<td>nano</td>
<td>n</td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>micro</td>
<td>μ</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>milli</td>
<td>m</td>
</tr>
<tr>
<td>$10^{3}$</td>
<td>kilo</td>
<td>k</td>
</tr>
<tr>
<td>$10^{6}$</td>
<td>mega</td>
<td>M</td>
</tr>
<tr>
<td>$10^{9}$</td>
<td>giga</td>
<td>G</td>
</tr>
<tr>
<td>$10^{12}$</td>
<td>tera</td>
<td>T</td>
</tr>
</tbody>
</table>
# Table of Values

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>Value</th>
<th>CGS</th>
<th>SI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity of light</td>
<td>(c)</td>
<td>2.997925</td>
<td>(10^{10}) cm s(^{-1})</td>
<td>(10^8) m s(^{-1})</td>
</tr>
<tr>
<td>Proton charge</td>
<td>(e)</td>
<td>1.60219</td>
<td>–</td>
<td>(10^{-19}) C</td>
</tr>
<tr>
<td>Planck's constant</td>
<td>(h)</td>
<td>6.62620</td>
<td>(10^{-27}) erg s</td>
<td>(10^{-34}) J s</td>
</tr>
<tr>
<td></td>
<td>(\hbar = h/2\pi)</td>
<td>1.05459</td>
<td>(10^{-27}) erg s</td>
<td>(10^{-34}) J s</td>
</tr>
<tr>
<td>Avogadro's number</td>
<td>(N)</td>
<td>(6.02217 \times 10^{23}) mol(^{-1})</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Atomic mass unit</td>
<td>(\text{amu})</td>
<td>1.66053</td>
<td>(10^{-24}) g</td>
<td>(10^{-27}) kg</td>
</tr>
<tr>
<td>Electron rest mass</td>
<td>(m_e)</td>
<td>9.10956</td>
<td>(10^{-28}) g</td>
<td>(10^{-31}) kg</td>
</tr>
<tr>
<td>Proton rest mass</td>
<td>(M_p)</td>
<td>1.67261</td>
<td>(10^{-24}) g</td>
<td>(10^{-27}) kg</td>
</tr>
<tr>
<td>Proton mass/electron mass</td>
<td>(M_p/m_e)</td>
<td>1836.1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Reciprocal fine structure</td>
<td>(1/\alpha)</td>
<td>137.036</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>constant (\hbar c/e^2)</td>
<td></td>
<td></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Electron radius (e^2/me^2)</td>
<td>(r_e)</td>
<td>2.81794</td>
<td>(10^{-13}) cm</td>
<td>(10^{-15}) m</td>
</tr>
<tr>
<td>Electron Compton</td>
<td></td>
<td></td>
<td>(10^{-11}) cm</td>
<td>(10^{-13}) m</td>
</tr>
<tr>
<td>wavelength (\hbar /mc)</td>
<td>(\lambda_e)</td>
<td>3.86159</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Bohr radius (\hbar^2/me^2)</td>
<td>(r_0)</td>
<td>5.29177</td>
<td>(10^{-9}) cm</td>
<td>(10^{-11}) m</td>
</tr>
<tr>
<td>Bohr magneton (\hbar^2/2mc)</td>
<td>(\mu_B)</td>
<td>9.27410</td>
<td>(10^{-21}) erg G(^{-1})</td>
<td>(10^{-24}) J T(^{-1})</td>
</tr>
<tr>
<td>Rydberg constant (me^4/2\hbar^2)</td>
<td>(R_{\infty}) or Ry</td>
<td>2.17991</td>
<td>(10^{-11}) erg</td>
<td>(10^{-16}) J</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13.6508 eV</td>
<td></td>
</tr>
<tr>
<td>1 electron volt</td>
<td>(\text{eV})</td>
<td>1.60219</td>
<td>(10^{-12}) erg</td>
<td>(10^{-19}) J</td>
</tr>
<tr>
<td></td>
<td>(\text{eV}/h)</td>
<td>(2.41797 \times 10^{14}) Hz</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(\text{eV}/hc)</td>
<td>8.06546</td>
<td>(10^3) cm(^{-1})</td>
<td>(10^5) m(^{-1})</td>
</tr>
<tr>
<td></td>
<td>(\text{eV}/k_B)</td>
<td>(1.16048 \times 10^4) K</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Boltzmann constant</td>
<td>(k_B)</td>
<td>1.38062</td>
<td>(10^{-16}) erg K(^{-1})</td>
<td>(10^{-23}) J K</td>
</tr>
<tr>
<td>Permittivity of free space</td>
<td>(\varepsilon_0)</td>
<td>–</td>
<td>1</td>
<td>(10^7/4\pi c^2)</td>
</tr>
<tr>
<td>Permeability of free space</td>
<td>(\mu_0)</td>
<td>–</td>
<td>1</td>
<td>(4\pi \times 10^{-7})</td>
</tr>
</tbody>
</table>